Spring 5-15-2017

Essays in Entrepreneurship, Innovation and Labor

Andres Gonzalo Hincapie Norena
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Economics

Dissertation Examination Committee:
Barton Hamilton, Chair
George-Levi Gayle
Limor Golan
Robert Miller
Carl Sanders

Essays in Entrepreneurship, Innovation and Labor

by

Andrés Gonzalo Hincapié Noreña

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2017
St. Louis, Missouri

Table of Contents

List of Figures . . . v
List of Tables . . . vii
Acknowledgements . . . ix
Abstract . . . x
1 Introduction . . . 1
2 Entrepreneurship over the Life Cycle . . . 7
2.1 Introduction . . . 7
2.2 Data . . . 13
2.2.1 Incorporated and Unincorporated Entrepreneurs . . . 15
2.2.2 Preliminary Analysis of the Data . . . 18
2.3 Model . . . 22
2.3.1 Occupations and Individual Characteristics . . . 23
2.3.2 Preferences . . . 24
2.3.3 Income and Learning . . . 25
2.3.4 Optimal Choices . . . 27
2.4 Estimation . . . 32
2.4.1 First Stage: Income Parameters and Learning Structure . . . 34
2.4.2 Second Stage: Utility Parameters . . . 35
2.4.3 Identification . . . 36
2.5 Parameter Estimates and Economic Forces at Play . . . 37
2.5.1 Income Parameters and the Distribution of Ability . . . 38
2.5.2 Utility Parameters . . . 45
2.5.3 Model Fit . . . 51
2.5.4 Decomposition Exercises . . . 51
2.6 Policy Counterfactuals . . . 59
2.6.1 Subsidies . . . 60
2.6.2 Entrepreneurship Education . . . 62
2.7 Conclusion . . . 65

3 Innovation and Diffusion of Medical Treatment . . . 68
3.1 Introduction . . . 68
3.2 Data . . . 75
3.2.1 Summary Statistics . . . 76
3.2.2 Consumer Demand . . . 78
3.2.3 Market-Level Innovation . . . 83
3.3 Model . . . 87
3.3.1 Summary and Timing within a Period . . . 87
3.3.2 Supply . . . 88
3.3.3 Demand . . . 94
3.4 Estimation . . . 104
3.4.1 Overview . . . 105
3.4.2 Moment Condition . . . 106
3.5 Parameter Estimates and Choice Dynamics . . . 110
3.5.1 The Distributions of Innovations and New Products . . . 110
3.5.2 Utility Parameter Estimates . . . 113
3.5.3 Transitions and Outcomes . . . 115
3.5.4 Simulated Choice Dynamics and Model Fit . . . 119
3.6 Alternative Choice and Technology Paths . . . 122
3.6.1 The Distribution of Technology Paths . . . 122
3.6.2 Demand Pull: How Consumer Choices Affect Innovation . . . 128
3.7 Demand Externalities and a Constrained Planner . . . 130
3.8 Conclusion . . . 135
Bibliography . . . 138
Appendix . . . 145
A Appendix to Chapter 2 . . . 145
A.1 Data Appendix . . . 145
A.1.1 PSID Data . . . 145
A.1.2 Bond Price Data . . . 155
A.2 Model Appendix . . . 156
A.2.1 Claim about belief variance. . . . 156
A.2.2 Proof of proposition 2.1 . . . 157
A.2.3 Proof of Proposition 2.2 . . . 159
A.3 Estimation Appendix . . . 161
A.3.1 First Stage Detailed . . . 161
A.3.2 Second Stage Detailed . . . 164
A.4 Results Appendix . . . 167

A.4.1 Solving the Model . . . 167
A.4.2 Model Fit . . . 168
A.4.3 Certainty Equivalent . . . 173
A.4.4 Monetary Value of Entry Costs . . . 174
A.4.5 Alternative Regimes . . . 175
B Appendix to Chapter 3 . . . 179
B.1 Data Appendix . . . 179
B.1.1 Main Variables . . . 179
B.1.2 Products and Product Components . . . 181
B.2 Estimation Appendix . . . 191
B.2.1 k-means Clustering Algorithm . . . 191
B.2.2 Product Characteristics . . . 192
B.2.3 GMM Estimation Algorithm . . . 192
B.2.4 Estimator . . . 201
B.2.5 CCP Estimation and Fit . . . 204
B.2.6 Proof of Proposition 3.1 . . . 209
B.3 Additional Tables and Figures . . . 210

List of Figures

2.1 Occupational Choice: Age Profile 20
2.2 Probability of Switching into Entrepreneurial Occupations 21
2.3 Probability of Switching by Occupation-Specific Experience 22
2.4 Returns to Own-Occupation Experience 40
2.5 Returns to Cross-Occupation Experience 42
2.6 Implied Correlation Between Abilities 44
2.7 Prior Variance Eliminated After 5 Years of Occupation-Specific Experience . 45
2.8 Certainty Equivalent 46
2.9 Monetary Equivalent of Entry Costs 49
2.10 Monetary Equivalent of non-pecuniary Benefits 50
2.11 Simulated versus Observed Choice Rates 52
2.12 Decomposition: Entry 55
2.13 Decomposition: Timing 57
2.14 Decomposition: Ability 58
2.15 Decomposition: PVI 59

3.1 Death rate in the sample. 80
3.2 Health and side effects summary trends over time. 81
3.3 Consumer demand over time. 82
3.4 Diffusion of Products Over Time 84
3.5 Treatment Evolution 86
3.6 Timing 89
3.7 The Distribution of Innovations, F_ν. 111
3.8 Distribution of Number of New Products 113
3.9 Health Effect on Outcomes 116
3.10 Goodness of Fit Figures 121
3.11 Distribution of Technology Paths: Health 123
3.12 Distribution of Technology Paths: No Ailments 125
3.13 Distribution of Technology Paths: Survival and Product Consumption 126

3.14 Distribution of Technology Paths: Entry and Exit 127
3.15 Alternative Choice Regimes: Health and No Ailments 129
3.16 Alternative Choice Regimes: Survival, Consumption, Entry, and Exit 130
3.17 Constrained Planner: Random Assignment to Clinical Trials 134

A.1 Proportion of Individuals Switching between t and $t + 1$ Over Years in the Labor Market (Collar) 149
A.2 Average Wealth 153
A.3 Model Fit by Age 170
A.4 Switching into Entrepreneurial Occupations (Correlated Learning) 172
A.5 Switching into Entrepreneurial Occupations (Uncorrelated Learning) 172

B.1 CCPs Goodness of fit: Stayers 207
B.2 CCPs Goodness of fit: Trial 207
B.3 CCPs Goodness of fit: No Product 208

List of Tables

2.1 Summary Statistics: Individuals . . . 15
2.2 Summary Statistics: Individual-Years . . . 16
2.3 Transition Matrix . . . 18
2.4 First Entry . . . 19
2.5 Occupation Spells . . . 19
2.6 Income Parameters . . . 39
2.7 Population Ability Covariance Matrices . . . 43
2.8 Idiosyncratic Variance . . . 43
2.9 Utility Parameters . . . 47
2.10 First Entry: Observed and Simulated . . . 53
2.11 Young Incorporated Entrepreneurship Subsidy . . . 61
2.12 Young Incorporated Entrepreneurship Education . . . 64

3.1 Summary Statistics: Subjects. Visit 14-47 (1990-2007) . . . 77
3.2 Summary Statistics: Subjects-Visits. Visits 14-47 (1990-2007) . . . 79
3.3 Moments of the Distribution of Innovations, F_ν . . . 112
3.4 Distribution of Number of New Products, F_N . . . 112
3.5 Utility Parameters, y_{it} . . . 114
3.6 Gross Income, m_{it} . . . 118
3.7 Tobit Model for Out-of-pocket Payments, o_{it} . . . 119
3.8 Logit Model for Labor Supply, l_{it} . . . 120
3.9 Logit model for Death, $1 - S_{it+1}$ . . . 120
3.10 Constrained Planner: Assignment Rules . . . 132

A.1 Parameters of the Wealth Profile Equation . . . 153
A.2 Full time and Part Time observations (1968-1996) . . . 154
A.3 Transition Patterns: Observed and Simulated . . . 169
A.4 Spells: Observed and Simulated . . . 169
A.5 Income: Observed and Simulated . . . 171

A.6 Average Income for Counterfactual: No Learning-by-doing 176
A.7 Comparison of Counterfactual Regimes 178

B.1 Trial Components 182
B.2 Chemical Formulae of Product Components 186
B.3 Combination Components 187
B.4 Products in the Market, Entry and Exit 188
B.5 Health Characteristics of Treatments 210
B.6 No Ailments Characteristics of Treatments 216

Acknowledgments

I thank my advisers Barton Hamilton, George-Levi Gayle, Limor Golan, and Robert Miller for their help and support throughout these years. I also thank the faculty of the Department of Economics for the knowledge they shared with me and the faculty of the Olin Business School from whom I received precious advice in countless occasions.

For financial support during part of this research, I thank the Center for Research in Economics and Strategy (CRES) at the Olin Business School. Additionally, I thank the Venezuelan Association in Missouri and Grupo Atlántico Dance Company for financial support during part of this research through their Hispanic Academic and Artistic-Academic scholarships.

I also want to thank my family for their continuous support during my years as a PhD student. My mum Margarita, my dad Eduardo, my aunt Rosita, and my sister Gisela, who helped me find peace of mind in the hardest moments. I also thank my dear friend Lauren Wilmore for her invaluable friendship during these years. Finally, I thank my lovely partner Briana Harrison, who has kept me grounded for a number of years and who I hope will continue doing so for years to come. ¡Gracias!

Andrés Hincapié

Washington University in St. Louis

May 2017

ABSTRACT OF THE DISSERTATION

Essays in Entrepreneurship, Innovation and Labor

by

Andrés Gonzalo Hincapié Noreña

Doctor of Philosophy in Economics

Washington University in St. Louis, 2017

Professor Barton Hamilton, Chair

I explore the role of experimentation on the career choices of individuals deciding whether to be paid employees or entrepreneurs, and on the decisions of consumers deciding what medical treatment to buy. In the labor market, experimentation entails the accumulation of information that allows individuals to improve their occupational choices. In the product market, experimentation entails discovering the quality of new products and it may have an effect on the evolution of technology: less experimentation by individuals may slow down the process of innovation.

I start in Chapter 2 with the observation that most individuals do not start a business and, if they do, tend to do so well into their thirties. While policies encouraging young, would-be entrepreneurs are popular, little is known about whether they are effective. Using data from the Panel Study of Income Dynamics, I estimate a dynamic Roy model with imperfect information about ability to evaluate the relative importance of various economic determinants of entrepreneurial participation. Risk-averse, forward-looking individuals sequentially select entrepreneurial and paid-employment occupations based on their returns to experience, information value, non-pecuniary benefits, and entry costs. Results show that the main barriers faced by young entrepreneurs are entry costs and information frictions. I consider two policy counterfactuals: a subsidy targeting entry costs and entrepreneurship

education targeting information frictions. I extend previous literature providing a mapping from the information quality of entrepreneurship education into career choices and long-term outcomes. A subsidy for young entrepreneurs increases participation but has small long-term effects. Entrepreneurship education can have sizable effects on participation and present value of income flows, even for low information quality. Nevertheless, the value of any particular entrepreneurship education program will depend on its cost and its information quality.

In Chapter 3 we develop and estimate a dynamic structural model of demand for a product line whose spectrum of characteristics evolves over time because innovation is endogenous to consumer demand. To achieve this goal, we provide a new approach to the econometric challenge of estimating the process of technological change where innovation under uncertainty includes both frequent and incremental modifications along with sporadic major breakthroughs. Quality in our model is a multidimensional object: new products that are superior in some dimensions might be inferior in others. For example, new medicines more effective in combating disease than existing products sometimes have harsher side effects. In our model, consumer choices determine both the speed and the direction of product innovation. Demand externalities arise because the aggregate choices of atomistic individuals drive innovation. We apply our framework to analyze consumer choice and the realized path of innovations over a long time horizon in a maturing product market: HIV drugs. In this market, we observe the introduction of hundreds of new products, marking mostly modest, but sometimes major innovations over existing technologies. Our estimates are obtained through simulations of alternative hypothetical worlds that might have arisen if the innovations had taken different paths to the ones we observe. We use our estimates to assess the effects of policies that internalize the externalities affecting innovation and consumer welfare by modifying consumer choices. We find that experimentation in clinical trials is one of the mechanisms through which the externality operates.

Chapter 1

Introduction

In the labor market, as well as in the market for consumption goods, individuals face uncertainty. New workers are uncertain about how good they would be as managers or as entrepreneurs. Consumers are uncertain about the characteristics of new products. This lack of information drives the decisions of individuals who choose safer alternatives in an attempt to avoid uncertainty. In the labor market, this means that individuals may engage less in entrepreneurship. In the market for consumption goods, it means that individuals may stick to older products even as new ones become available.

In environments with imperfect information, experimentation is characterized as the process of gathering information that can be used to reduce uncertainty and adjust future choices. In entrepreneurship, experimentation can be undertaken by investors trying to reduce uncertainty about the potential demand for a specific product or it can be undertaken by entrepreneurs themselves trying to reveal whether their entrepreneurial potential is high enough to make them successful. In product markets with innovation, experimentation by consumers can determine the speed at which new products are adopted in the market thereby also affecting the frequency with which new products are introduced. An example is the market for medical treatments, where experimentation in clinical trials is fundamental to unveil the safety and efficacy of treatments before they are made available in the market.

My dissertation explores the role of experimentation as a mechanism driving choices in the labor market and in the market for medical treatments. In Chapter 2, experimentation is part of an individual's career and it partially determines why individuals choose between more risky occupations (entrepreneurship) and less risky occupations (paid employment). In Chapter 3 experimentation by potential consumers of medical treatments shapes the process of product innovation.

Chapter 2 introduces two stylized facts: most individuals do not attempt entrepreneurship during their careers and if they do, tend to do so when they are older and have accumulated some paid-employment experience. I aim to explain these facts asking what determines whether somebody attempts entrepreneurship during their life cycle and what determines the gap in first entry ages between paid employment and self-employment.

The idea of experimentation in a labor market setting has been discussed by Jovanovic (1979) and Miller (1984). In their models, individuals have incomplete information about their ability and must discover how productive they are in every occupation. Recent papers by Sanders (2010), James (2011), and Antonovics and Golan (2012) have expanded on those ideas. In particular, James (2011) introduces an estimation method that captures the extent to which learning about ability can happen across paid employment occupations.[1]

Experimentation to gather information is also a mechanism explaining participation in entrepreneurship and it has consequences for the estimation of entrepreneurship returns. Manso (2014) discusses the "experimentation" bias when estimating the life-time returns of entrepreneurship out of a cross-section. This bias is caused by the fact that the value of experimenting and finding out whether or not an idea is to be successful is not accounted for in the cross-section, which will then underestimate the returns to entrepreneurship. Dillon and Stanton (2016) extend on the ideas of Manso (2014) and study the option value of en-

[1] In the multi-armed bandit literature in statistics this phenomenon is called correlated learning.

trepreneurship using a dynamic occupational choice model in which individuals must discover their entrepreneurial ability by experimenting with entrepreneurship. Because individuals only learn about entrepreneurial ability from attempting entrepreneurship, their framework does not capture James (2011) suggestion that learning may happen across occupations.

Experimentation in entrepreneurship has also been discussed in Kerr et al. (2014) who focus on the costs and constraints to experimentation by investors trying to reduce uncertainty about potential consumer demand. Those constraints may take the form of inflexibilities in the labor market, sunk technology costs, or inefficient failure policies. Gottlieb et al. (2016) use a Canadian reform that extended job-protected maternity leave to show that reluctance to experiment increases if failure is penalized. In their setting, the penalty for new mothers who engage in entrepreneurship and fail, before the extension was granted, is the absence of the option to return to the old job.

Chapter 2 extends the literature by evaluating the role of the information friction on explaining entrepreneurship, relative to the roles of other forces suggested in the literature such as learning by doing, risk aversion, entry costs associated with wealth, and non pecuniary benefits. In particular, Chapter 2 is one of the first attempts at quantifying the role of risk aversion on preventing individuals from trying entrepreneurship in a dynamic setting. Additionally, Chapter 2 extends our knowledge of cross-occupation learning between paid employment and self-employment.

Various mechanisms determining occupational choices are analyzed using the structural dynamic model of occupational choice in Chapter 2. In the model, individuals' careers unravel as they experiment trying to find their position in the distribution of ability. The model is estimated with data from the Panel Study of Income Dynamics and the estimated structure is used to quantify the roles of various mechanism. Using the framework developed, Chapter 2 also provides an evaluation of short-term and long-term effects of policies that attempt to promote young entrepreneurship by targeting the main barriers they face. These barriers are obtained as the main mechanisms explaining the gap in first entry ages between paid employment and self-employment.

In order to analyze the various mechanisms determining occupational choices, Chapter 2 provides a structural dynamic model of occupational choice in which individuals' careers unravel as they experiment trying to find their position in the distribution of ability. The model is estimated with data from the Panel Study of Income Dynamics and the estimated structure is used to quantify the role of each mechanism. Using the framework developed, Chapter 2 also provides an evaluation of short-term and long-term effects of policies that attempt to promote young entrepreneurs by targeting the main barriers they face. These barriers are obtained as the main mechanisms explaining the gap in first entry ages between paid employment and self-employment.

Chapter 3 explores the interaction between aggregate consumer choices and the process of innovation in the market for medical treatments. We ask whether innovation and diffusion depend upon consumers' dynamic optimization or whether these processes are independent of consumer demand. Furthermore, we ask whether there is a demand externality through experimentation in the innovation and diffusion processes and we analyze the consequences of the externality.[2]

Dynamic demand under uncertainty, a fundamental component of the study in Chapter 3, is usually linked to the idea of *experience* goods—goods whose characteristics are difficult to assess before a purchase is made and consumption is exercised. Empirical studies of dynamic demand of experience goods include Erdem and Keane (1996) and Crawford and Shum (2005). Erdem and Keane (1996) use data on detergents to model the decisions of consumers and how they depend on previous consumption experience. In their framework consumer, experimentation influences future choices because it represents a process of accumulation

[2]This chapter is part of a joint project with Barton Hamilton (Olin Business School), Robert Miller (Tepper School of Business), and Nicholas Papageorge (Johns Hopkins University).

of information about product characteristics. Highlighting the cost of experimentation, one of their findings is that risk aversion discourages consumers from trying brands they have not yet collected information about. In a paper that focuses on a market closer to the one we explore in Chapter 3, Crawford and Shum (2005) analyze demand under uncertainty using data of anti-ulcer prescriptions. Their model allows for consumers to experiment with products that have multiples characteristics (symptomatic and curative effects), and incorporates the effects of these characteristics on consumers' state variables that determine patient behavior in a dynamic way.

In terms of product markets with innovation, papers by Goettler and Gordon (2011) and Gowrisankaran and Rysman (2012) have made important contributions studying dynamic demand for durables. Using data from the personal computer microprocessor industry, Goettler and Gordon (2011) explore the role of market structure on the process of innovation. They built and estimate a structural model in which firms compete in an oligopoly over time, and innovation is endogenous to their investment decisions. They find that the presence of a second firm can slow innovation. However, they also find that consumer surplus would fall in a monopolistic market because the losses from increases in price outweigh the gains from higher innovation. Gowrisankaran and Rysman (2012) focus on the dynamic demand for new durables in the digital camcorder industry. Their model allows for consumer heterogeneity and a changing number of models. In specifying the expectations of consumers over the characteristics of future products, Gowrisankaran and Rysman (2012) make a simplifying assumption: "consumers expect that the evolution of the value of purchase will follow a simple one dimensional Markov process." The authors make this major assumption to gain tractability when computing expectations over future choice sets, a problem that any model of dynamic demand with unanticipated innovation is bound to face.

In Chapter 3 we introduce an empirical framework to capture how aggregate consumer demand and experimentation affect innovation in markets for medical treatments along multiple dimensions of quality. Using our framework we estimate a distribution that describes the size and direction of innovations, and embed this distribution into a structural model of dynamic demand. In the model, forward-looking consumers make choices after forming expectations about the characteristics of future products. In particular, consumers can experiment with new products that are not yet available in the market by participating in a clinical trial. Externalities arise because aggregate consumer behavior changes the path of innovation, thereby changing dynamic payoffs. We depart from existing work on dynamic demand in how we model consumer expectations over the path of innovation. Most papers take the existing set of products as given or exogenous to the model and focus on demand responses to new products. In contrast, we explicitly model how consumers form expectations about future innovations, allowing them to take into account that aggregate market shares can shift the direction of innovation. We apply this framework to data from the Multicenter AIDS Cohort Study, analyzing consumer choices and the realized path of innovations in the market for HIV treatments.

Chapter 2

Entrepreneurship over the Life Cycle

2.1 Introduction

Entrepreneurship has long been considered an engine of innovation, growth, and development. These connections, already suggested by Schumpeter (1911) and more recently described by Baumol (2002), have found empirical support in the literature (Guiso et al., 2004; Chatterji et al., 2013). In particular, young entrepreneurship is routinely celebrated by the popular press, as seen in the *30 under 30* collection by *Forbes* magazine, and by policy makers, who highlight the economic potential gains from fostering young entrepreneurs (OECD, 2013). However, most people do not start a business during their careers and, if they do, tend to do so well into their thirties.[1] Although there are many policies attempting to stimulate entrepreneurship among individuals entering the labor market, it is not clear to what extent they induce young people to start businesses, what types of entrepreneurs they attract (e.g., high or low ability), and what the long term consequences of these policies are.[2]

[1] According to data from the Panel Study of Income Dynamics.

[2] Examples of such initiatives include the Thiel Fellowship (`http://thielfellowship.org`), the National Young Entrepreneurship Challenge (`https://www.nfte.com`), the Veale Young Entrepreneurship Forum (`http://vealeentrepreneurs.org`), and the Small Business Administration's Learning Center (`https://www.sba.gov/tools/sba-learning-center/training/young-entrepreneurs`).

In this chapter, I develop a life-cycle framework analyzing entrepreneurial entry and exit based on the Roy (1951) model of occupational choice, and focusing on the self-employment decisions of young labor market entrants. In the model, risk-averse individuals sequentially choose salaried and entrepreneurial occupations based on their returns to experience, information value, uncertainty, non-pecuniary benefits, and entry costs. In particular, individuals decide to become entrepreneurs based on their beliefs about their own abilities. This chapter adds to the literature by quantifying the relative importance of economic forces that determine whether and when people become entrepreneurs. Three elements are particularly novel to the empirical literature on entrepreneurship: first, an analysis of the timing of the choice; second, an assessment of the relative importance of risk aversion in a dynamic setting; and finally, an evaluation of the role of cross-occupation learning between salaried and entrepreneurial occupations.[3] In this framework, cross-occupation learning refers to the ability to transfer skills learned in one occupation to another, as well as the ability to learn about one's entrepreneurial ability from paid-employment success.[4]

A number of economic forces have been suggested in the literature to explain why individuals, both young and old, attempt entrepreneurship. Learning-by-doing, commonly characterized as experience accumulation that increases productivity, is one such force (Lazear, 2005; Lafontaine and Shaw, 2016).[5] Learning-by-doing predicts that individuals are more inclined to try entrepreneurship if they do not have to climb the productivity ladder first. Learning about one's entrepreneurial ability also determines entry.[6] If individuals are uncertain about their entrepreneurial ability, but the performance of their business helps them

[3]Munk (2015) approaches the timing of the entrepreneurial choice in a reduced-form setting revisiting the question of whether self-employment pays.

[4]For instance, cross-occupation learning by doing is behind the "rank" mechanism in Liang et al. (2014), which suggests that individuals in certain types of occupations are more likely to develop entrepreneurial skills.

[5]A well known example in the occupational choice literature is Keane and Wolpin (1997).

[6]Models of occupational choice with incomplete information include Jovanovic (1979), Sanders (2010), and Antonovics and Golan (2012).

learn about it, they attempt entrepreneurship as long as their prior variance is high—they will want to learn whether they are in the "right" part of the distribution.[7] The option value of entrepreneurship, which intersects both learning-by-doing and learning about ability, also affects entry. Individuals attempt entrepreneurship because they can always switch back to paid employment if they discover that entrepreneurship is not the best option for them (Manso, 2014; Dillon and Stanton, 2016). However, reluctance to experiment increases if failure is penalized (Gottlieb et al., 2016).

Risk aversion, credit constraints, and non-pecuniary benefits also affect the entry margin. Risk aversion pushes individuals away from entrepreneurship because it is a more uncertain occupation (Iyigun and Owen, 1998). Credit constraints in starting a business or in reaching optimal scale prevent less affluent individuals from trying their luck as entrepreneurs (Evans and Jovanovic, 1989; Hurst and Lusardi, 2004; Buera, 2009). Finally, non-pecuniary motivations such as "being one's own boss" and "wanting flexibility over schedule" could also provide incentives for entry (Hamilton, 2000; Hurst and Pugsley, 2011, 2015).[8]

More importantly, some of the forces explaining entry have predictions for whether entrepreneurs will be young. For instance, learning-by-doing implies that individuals who want to become highly productive entrepreneurs should start at an early age. Learning about ability gives the same prediction. High ability variance in entrepreneurship encourages individuals to seek to discover their place in the distribution as early as possible (Miller, 1984). Credit constraints preclude young individuals with weaker credit histories and less disposable wealth from entering. Risk aversion has a less clear effect on the timing of entry. Overall, risk aversion will prevent entry at any stage in an individual's career. However, if learning about ability reduces uncertainty over time (e.g. if learning is Bayesian), then the effect of risk aversion on entry can be attenuated as individuals acquire more experience.

[7]The high variation observed in entrepreneurial income suggests this could be the case.

[8]Other determinants explored in the literature include peer-effects (Nanda and Sørensen, 2010) and personality traits (Hamilton et al., 2016; Humphries, 2016).

If success in paid employment correlates with higher entrepreneurial ability, favorable outcomes in paid employment may be associated with switching into self-employment. However, if paid-employment outcomes are uninformative of one's entrepreneurial ability, successful workers would be more willing to stay in paid employment not only because they seem better at it, but also because there is less uncertainty about the fact that they are.

There are several econometric challenges for estimation of the framework developed here. Given that individuals select occupations based on their beliefs, selection bias cannot be accounted for by using first-differences estimators of occupation-specific income equations and panel data (Gibbons et al., 2005).[9] Therefore, this chapter accounts for selection bias by using the likelihood function implied by the model. However, this econometric decision can come at the cost of multidimensional integration over unobserved ability vectors. To get around the computational burden, estimation of the parameters of the model follows a two-stage procedure using a combination of an Expectation-Maximization (EM) algorithm and a conditional choice probabilities (ccp) estimator (Hotz and Miller, 1993; Arcidiacono and Miller, 2011; James, 2011). The EM algorithm in the first stage bypasses the need for multidimensional integration. The ccp estimator in the second stage allows for a flexible treatment of the large state space of the problem, which includes continuous beliefs and experience for each occupation. The ccp estimator delivers such flexibility because the structural parameters can be estimated without solving the dynamic optimization problem at every candidate parameter vector in the search algorithm.

In the empirical analysis, this chapter uses data from the Panel Study of Income Dynamics (PSID). The sample is restricted to white and black men between the years 1968 and 1996. Moreover, following suggestive evidence in Levine and Rubinstein (Forthcoming), entrepreneurship is disaggregated by incorporation status. Entrepreneurs with promising

[9] Since individual's beliefs change over time as they acquire more information, their ability "does not translate into a fixed effect" in an income equation (Gibbons et al., 2005). Cross-section data are even more problematic as they provide no historic information to model the process of belief formation.

abilities benefit more from the incorporated organizational form because it encourages entrepreneurial risk taking by offering limited liability and by facilitating fund raising through the issuance of stock. Less promising entrepreneurs, on the other hand, benefit more from the less complex unincorporated form that offers lower administrative costs and regulatory burden. Results using the PSID suggest that incorporated entrepreneurs are more similar to white collar workers than they are to unincorporated entrepreneurs.

The framework is used to quantify the importance of the economic forces at play by comparing the baseline model against counterfactual regimes that shut down each economic force. Results indicate that learning-by-doing and entry costs have the largest effects preventing individuals from attempting entrepreneurship. The role of learning-by-doing is evaluated by turning the profile of returns to experience into a flat average return. Given that the profile of returns to experience in entrepreneurship is steep, the flat average in the counterfactual is high, making individuals more willing to experiment. Risk aversion and information frictions also play important roles. For instance, shutting down risk aversion increases the percentage of individuals who attempt incorporated entrepreneurship by 40%, and eliminating information frictions increases this number by 35%. Eliminating cross-occupation learning reduces the proportion of individuals attempting entrepreneurship by 10%. Although these effects seem small, in the long term they are stronger: eliminating cross-occupation learning decreases the present value of income (PVI) of incorporated entrepreneurs by about a quarter.

Results also show that the two main barriers to young entrepreneurship are entry costs and information frictions. In the model, entry costs capture barriers to entrepreneurship not explicitly modeled, such as credit constraints.[10] In order to make the link between entry cost and credit constraints, these costs are interacted with age and a permanent wealth component (estimated as a fixed effect outside of the model). Estimates show that younger individuals as well as individuals with lower permanent wealth face higher barriers to entry. Flattening

[10]Unfortunately, the PSID lacks wealth data for most years during the period studied.

the profile of entry costs, allowing individuals of all ages to face the same average entry cost, closes the gap in average first-entry age between white collar work and entrepreneurship by about 70%. Eliminating information frictions, providing full information about ability, induces individuals to enter entrepreneurship earlier, closing the first-entry age gap by 20%.

Focusing on the main barriers to young entrepreneurship, the chapter undertakes two policy counterfactuals: a subsidy that targets entry costs and entrepreneurship education that targets information frictions.[11] Previous literature has shown that entrepreneurship education can shift beliefs (von Graevenitz et al., 2010; Oosterbeek et al., 2010). The chapter takes this result as given and extends the literature by providing a mapping from movements in beliefs, generated by entrepreneurial education of a given quality, into career choices and long-term outcomes. Results suggest that a blanket subsidy for young incorporated entrepreneurship increases participation and has a small positive effect on the average PVI net of the subsidy.[12] Additionally, incorporated entrepreneurship education can generate sizable increases in young incorporated entrepreneurship and PVI even at low levels of information quality. Nevertheless, caution must be taken when reading these results. The information quality of any specific policy may be different, and its cost may well exceed the additional income it generates.

The rest of the chapter is organized as follows. Section 2.2 presents the data and describes the main regularities motivating the research question and modeling choices. Section 2.3 describes the characteristics of the model and its implications. Section 2.4 describes the estimation method. Section 2.5 discusses the estimated parameters and presents the decomposition exercise. Section 2.6 introduces the policy counterfactuals. Section **??** concludes the chapter.

[11]Young entrepreneurs are defined as those who attempt entrepreneurship during the first five years of their labor market careers.

[12]Assuming that the incorporated self-employed are closer than their unincorporated peers to what is commonly thought of as "the entrepreneur," these counterfactuals focus on incorporated entrepreneurship.

2.2 Data

The Panel Study of Income Dynamics (PSID) is used in the empirical application.[13] There are three main reasons that motivate this choice. First, many individuals in the PSID are observed at the onset of their labor market careers and yearly from that point forward. This allows for the construction of measures of accumulated occupation-specific experience and occupational income at any point in a respondent's career. Second, whenever an individual declares himself to be self-employed, the survey's questions on self-employment allow for disaggregation into incorporated and unincorporated self-employment. Multiple differences between these self-employment options have been previously identified by Levine and Rubinstein (Forthcoming). Finally, although not available for most periods, the PSID collects data on wealth that will prove useful in the analysis of entry costs.

The sample is restricted to white and black men between years the 1968 and 1996. It contains survey information on occupation, self-employment status, business ownership, incorporation status, labor income, business income, working hours, completed education, age, race, marital status, and wealth.[14] Individuals' labor market careers are set to start at the beginning of their potential experience.[15] Both types of occupations, salaried and entrepreneurial, are disaggregated to exploit differences in returns to experience as well as differences in terms of the information they provide. For paid employment, the three-digit occupation code is used to generate two categories: blue collar and white collar.[16] This aggregation has been used previously in occupational choice models studying paid employ-

[13]The study started in 1968 with a representative sample of about 18,000 individuals in 5,000 families in the United States. Information about these individuals and their descendants was collected yearly until 1996, after which the study became biennial.

[14]Biennial data collected after 1996 was not used, since it would require making assumptions about occupational choices and income in years where no data were collected.

[15]Potential experience starts at the end of their completed education. All individuals are assumed to start school at age 6. Appendix A.1 contains a detailed description of how each variable is constructed.

[16]The PSID provides three-digit occupation codes from the 1970 Census of Population, constructed using the Alphabetical Index of Industries and Occupations issued June 1971 by the U.S. Department of Commerce and the U.S. Census Bureau.

ment (Keane and Wolpin, 1997). For self-employment, which is interchangeably referred to as entrepreneurship in this study, the incorporation status questions in the survey are used to create two categories: unincorporated and incorporated.

The income measure for paid employment corresponds to the individual's reported annual labor income. For self-employment, measuring income is less transparent. Since incorporated individuals are not asked about their business income in the survey, their reported labor income is used as their income measure.[17] For unincorporated individuals, who are not sheltered from the losses of their ventures through limited liability, the measure corresponds to the sum of the quantity reported as labor income plus the quantity reported as business income. Income measures are converted to hourly rates by dividing annual income figures by reported annual working hours.

The PSID includes a measure of wealth for selected years starting in 1984. It is constructed as the sum of six types of assets (farm business, checking or savings accounts, real estate other than main home, stocks, vehicles, and other assets) net of debt value plus value of home equity. The survey does not include data on wealth at every period. Therefore, in the current analysis a measure of permanent wealth will be considered instead of the separate wealth observations. The measure of permanent wealth, denoted ω_i, is obtained as the constant plus the fixed effect of a regression of wealth on a second degree polynomial on age. In estimation, only individuals with at least three wealth data points are considered. The variable ω_i is meant to capture long-run differences in access to resources.

After dropping observations of individuals who lack data on relevant variables, the analytic sample contains 1,506 individuals and 21,334 individual-year observations. Table 2.1 shows that about one-fifth of the sample is African American and 42% of the individuals have college education or more.[18] The average entry age to the labor market is about 22

[17]This measure corresponds to what Hamilton (2000) terms "the draw" or the difference between net profit and retained earnings. The characteristics of corporations as separate legal entities from the business owners justify the use of this measure.

[18]Notably, for the period of study, the proportion of individuals with college education in the sample is

Table 2.1: Summary Statistics: Individuals

Individuals	1506
Black	0.22
College or more	0.42
Age at entry	21.88
	[1.96]
Permanent Wealth	400
	[674]

Notes: Permanent wealth is in thousands of dollars of 2000. Standard deviation is in square brackets.

years. Finally, the average permanent wealth is $400,000 (measured in year 2000 USD) with a large standard deviation of $674,000.[19] In addition, consistent with the higher complexity of the incorporated organizational structure, Table 2.2 shows that incorporated individuals tend to be more educated than their unincorporated counterparts. Finally, self-employed individuals, especially incorporated entrepreneurs, are more likely to be married than paid employees.

2.2.1 Incorporated and Unincorporated Entrepreneurs

Using the individual's incorporated status, this chapter distinguishes between two types of self-employment. This disaggregation follows Levine and Rubinstein (Forthcoming), who introduce the differences between the organizational forms of unincorporated and incorporated businesses and the differences between the individuals they attract. They highlight how the organizational form of corporations facilitates growth and risk-taking behavior. Individuals seeking to establish businesses with high potential for development tend to be more attracted

higher than what it was for the U.S. adult population, 22% (Ryan and Siebens, 2012). The disparity arises when the selection criteria requires that individuals must be observed from the beginning of their labor market careers and widens more once observations without enough wealth data points are dropped. The lower the education the earlier they would enter the job market and the less likely the PSID is to observe them from the beginning of their careers.

[19] As mentioned above, the measure of individual permanent wealth corresponds to the constant plus the individual fixed effect of a regression of wealth on an age profile. Hence, the level of this measure depends on the shape of the polynomial implemented. The estimated value of the constant is about $418,000 (see Appendix A.1).

Table 2.2: Summary Statistics: Individual-Years

	All	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
Observations	21334	8902	9957	1403	602	470
%	100.00	41.73	46.67	6.58	2.82	2.20
Marital Status	0.76	0.74	0.77	0.79	0.86	0.50
High School	0.28	0.50	0.10	0.22	0.13	0.52
Some College	0.28	0.35	0.22	0.29	0.24	0.23
College	0.21	0.10	0.30	0.21	0.29	0.08
Some Grad	0.23	0.05	0.39	0.27	0.34	0.17
Age	31.04	28.92	32.21	33.93	36.94	30.45
	[7.27]	[6.65]	[7.18]	[7.35]	[7.00]	[8.06]
Wrkhrs	2147	2096	2234	2329	2703	
	[693]	[617]	[559]	[819]	[724]	
Hr Labor Income	18.71	14.16	21.24	21.30	37.91	
	[16.12]	[7.97]	[14.26]	[23.08]	[51.69]	
Residual Hr Labor Income		[6.99]	[12.32]	[20.41]	[44.29]	

Notes: White collar occupations are: professional, technical, and kindred workers; managers and administrators, except farm related; sales workers; clerical and kindred workers. Blue collar occupations are: craftsmen and kindred workers; operatives, except transport (including armed forces); transport equipment operatives; laborers, except farm related; service workers, except private household. Farm related occupations and military personnel dropped. Individuals are classified as unemployed if they reported to be not working or working for less than 2.5% the total amount of available hours in a year. Monetary quantities are in real dollars of 2000. Standard deviation is in square brackets. Residual income computed from occupation-specific OLS regressions on race, education and second degree polynomials on occupation-specific experience. One unit of hourly income represents 10$ per hour.

to this organizational form.

There are three distinctive characteristics of incorporated businesses. First, they are separate legal entities from their owners. This allows the corporation to own property, carry on business after the death of its owners, incur liabilities, and sue or be sued. Importantly, it also means that corporations can operate isolated from sudden shocks in an owner's personal finances. Second, corporations have limited liability against creditors. In other words, creditors seeking debt repayment can go after a shareholder's assets only to the extent of her investment in the business. This is precisely one of the reasons that motivates investors and venture capitalists to invest. Instead, unincorporated businesses owners have their personal assets exposed to the losses of their business. Third, corporations can issue shares of stock. This makes it easier for them to raise money in order to develop the business. It also makes transferability of ownership simpler than for sole proprietors or general partners.

But the advantages of incorporating a business come at the costs of more complex administrative activities, higher administrative costs, and potentially higher taxes.[20] Consequently, non-employer self-employed individuals and other small business owners will find incorporation unattractive (e.g. individual construction contractors, car repair shop owners). They are less likely to develop the business much further, so they have less incentive to incorporate. In contrast, incorporated self-employed individuals seek to take advantage of the organizational structure of the corporation to grow and develop their businesses.

The organizational forms of these two types of businesses also suggest that the abilities required from individuals trying to sort into them and the skills developed while working in each of them may be different. This observation, which further motivates the separation made in this chapter, finds empirical support in the data. Table 2.3 shows that entrepreneurs moving into paid employment differ in terms of the occupations they switch to. Unincorporated entrepreneurs are just as likely to switch to either blue collar or white collar work, while incorporated entrepreneurs tend to transition to white collar work. Further differences between these two types of self-employment will be introduced in the preliminary analysis of the data below. Notably, it will be shown that successful white-collar workers, characterized by higher residual income, are more likely to switch into incorporated entrepreneurial activities.[21] In general, the data suggest that incorporated entrepreneurs look more like white collar workers than unincorporated entrepreneurs.[22]

[20]These activities include holding annual meetings, recording meeting minutes, and keeping up at all times corporate documents such as the register of directors, the share register and the transfer register. Additionally, corporations are taxed and their shareholders are also taxed on their dividends.

[21]Residual income in occupation-specific regressions after controlling for demographics and a quadratic in occupation-specific experience.

[22]In the PSID, self-employed individuals, as well as paid employees, report occupation and industry. In computations not shown here, incorporated individuals are about 20 percent points more likely to report working in white collar occupations than unincorporated individuals. On the other hand, unincorporated entrepreneurs are about 20 percent points more likely to report belonging to the construction and repair industry than incorporated entrepreneurs.

Table 2.3: Transition Matrix

	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
Blue Collar	0.87	0.09	0.02	0.00	0.02
White Collar	0.07	0.89	0.02	0.01	0.01
Unincorporated	0.10	0.10	0.74	0.04	0.01
Incorporated	0.03	0.14	0.07	0.76	0.01
Unemployed	0.37	0.16	0.04	0.00	0.43

Notes: Matrix entry i, j represents the proportion of people in occupation in row i who move into occupation in column j between t and $t + 1$.

2.2.2 Preliminary Analysis of the Data

Several stylized facts in the data motivate this research and the dynamic model of occupational choice with learning used for analysis. Most individuals do not attempt entrepreneurship during their careers and they are even less likely to start their careers as entrepreneurs—individuals entering entrepreneurship tend to be older and have accumulated some paid employment experience prior to entry. Besides, those who attempt entrepreneurship tend to transition out of it faster than those who enter paid employment. In terms of hourly income, entrepreneurial occupations display higher variation than paid employment occupations, even after controlling for observables. Finally, successful white collar workers, characterized by higher residual income, are more likely to switch into incorporated entrepreneurial activities than their less successful peers.

Entrepreneurship is less common in an individual's career than paid employment. Table 2.4 shows that the proportion of individuals who attempt entrepreneurial occupations is less than half the proportion of individuals who attempt salaried occupations.[23] Furthermore, most of the 4,294 occupational spells in the sample occur in paid employment occupations and they are more than 60 percent longer than spells in entrepreneurial occupations (see Table 2.5). This is consistent with the transition patterns in Table 2.3, as salaried occupations tend to be more absorbing than entrepreneurial occupations.

[23] In separate calculations, the percentage of individuals who try at least one type of entrepreneurship by age 50 is about 34%. Virtually everybody in the sample tries paid employment

Table 2.4: First Entry

	Blue Collar	White Collar	Unincorporated	Incorporated
Ever	0.65	0.87	0.28	0.15
At First Entry				
Age	23.16	25.60	32.23	35.50
exp_{bc}	-	2.81	3.88	2.42
exp_{wc}	1.30	-	5.13	8.44
exp_{eu}	0.11	0.14	-	1.38
exp_{ei}	0.02	0.04	0.52	-

Notes: Statistics computed using individuals that are observed from the beginning of their careers until at least age 40. This leaves 486 unique individuals. No observations are used beyond age 50.

Table 2.5: Occupation Spells

	All	Blue Collar	White Collar	Unincorporated	Incorporated	Unemployed
Total	4294	1707	1652	453	194	288
Percent		39.75	38.47	10.55	4.52	6.71
Duration	4.97	5.21	6.03	3.10	3.10	1.63
First		52.06	42.56	2.19	0.27	2.92

Notes: **Duration** is the average duration of spells in years. **First** is the percentage of first spells that belong to a particular occupation.

Table 2.4 introduces the puzzle of young entrepreneurship. It shows that those who attempt entrepreneurial occupations tend to do so later in their careers, after accumulating more than 8 years of paid-employment experience on average. This translates into a gap of more than 7 years in average entry age between entrepreneurial occupations and salaried occupations. This gap runs opposite to the prediction of a parsimonious model of uncorrelated learning about ability (Miller, 1984) and does not seem suggestive of individuals having to climb the productivity ladder early on to become highly productive entrepreneurs. Notably, although few individuals start their careers as entrepreneurs (see Table 2.5), Figure 2.1 shows that participation increases as individuals age.

In terms of hourly income, entrepreneurial occupations display higher variation across observations than paid employment occupations. This is shown in Table 2.2, where individual-year observations are summarized by occupation. In particular, the variance of hourly income in incorporated entrepreneurship is more than three times as large as the variance in white

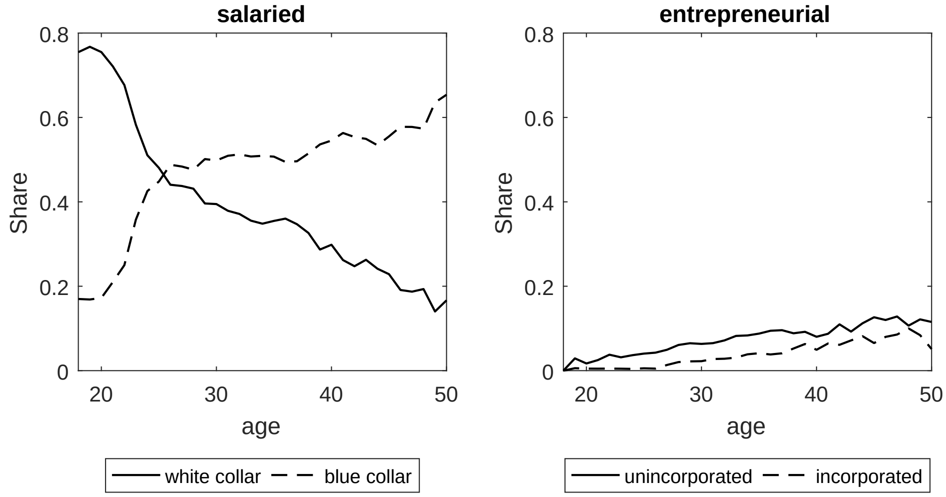


Figure 2.1: Occupational Choice: Age Profile

Notes: Participation rates into each occupation by age.

collar work. This difference remains even after controlling for demographics and occupation-specific experience. Interestingly, even though unincorporated entrepreneurship has a higher hourly income variance than white collar work, they share a similar average hourly income. Incorporated entrepreneurship, however, has a mean hourly income that is 75% higher.

Finally, successful white-collar workers, characterized by higher residual income, are more likely to transition into incorporated entrepreneurship than their less successful counterparts. This is shown in Figure 2.2. On the x-axis is the quintile of average residual income at t. On the y-axis is the probability of switching into the two types of entrepreneurship at $t+1$. Higher residual income in either white collar or blue collar work is generally associated with a smaller probability of switching into unincorporated entrepreneurship. This is consistent with uncorrelated learning, where unexplained success is only informative of ability in the current occupation. However, successful white-collar workers, as measured by higher residual income, are more likely to switch into incorporated entrepreneurship than their less successful peers. This is consistent with correlated learning about ability between white collar work and incorporated entrepreneurship, and is also consistent with the similarities between incorporated entrepreneurs and white collar workers mentioned in the previous section.

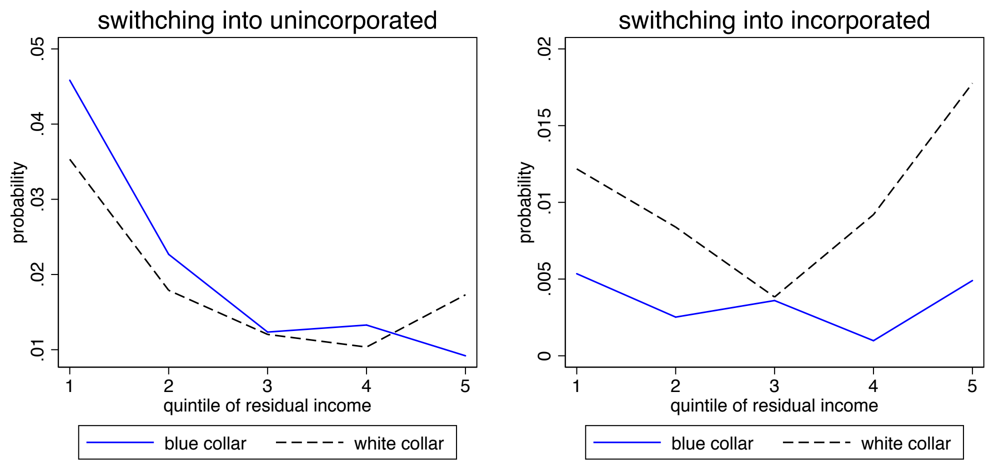


Figure 2.2: Probability of Switching into Entrepreneurial Occupations

Notes: Probability of switching into entrepreneurial occupations in $t+1$ by decile of residual income in t. Residual income is computed from occupation-specific regressions of hourly income on occupation-specific experience, general experience squared, race, education and marital status.

These facts are interpreted through the lens of a dynamic model of occupational choice with accumulation of experience (learning-by-doing) and learning about ability. Learning about ability is often modeled using Bayesian decision makers that draw information from their labor market outcomes and combine it with potentially heterogeneous priors in order to update their beliefs about their aptitude at various firms or occupations (Jovanovic, 1979; Miller, 1984). A prediction of these models, that the separation probability of a worker is decreasing in her tenure on the job, is found in Figure 2.3 at the occupation level.[24] Notably, the probability of switching from entrepreneurship decreases at a much higher rate than the probability of switching from salaried occupations during the first five years of accumulated occupation-specific experience.

The stylized facts also motivate the introduction of cross-occupation learning. Cross-

[24]This relation partially results from workers being less uncertain of their competence, which ameliorates the effect of new information.

occupation learning-by-doing aims to capture the nature of the mix of salaried experience that individuals acquire before entering each entrepreneurial occupation (see Table 2.4). Moreover, cross-occupation learning about ability (correlated learning) aims to capture the relationship between unexplained success in white collar work and entry into incorporated entrepreneurship. .

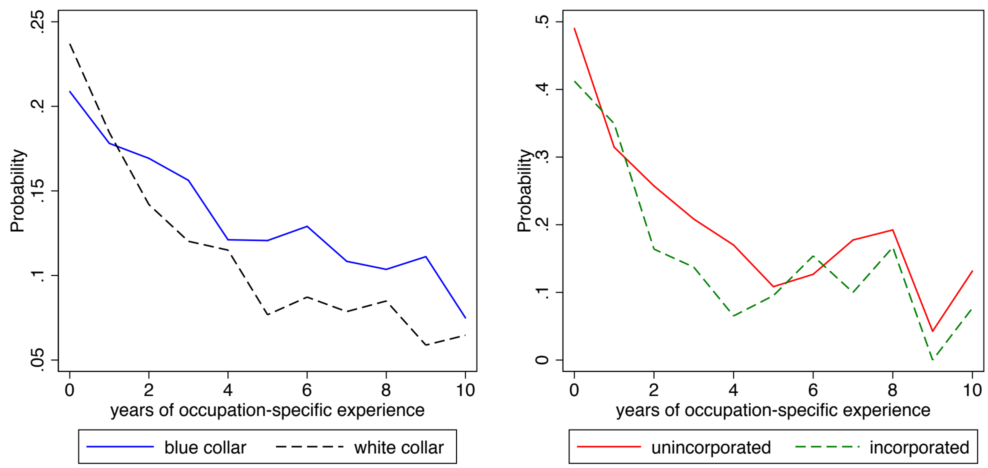


Figure 2.3: Probability of Switching by Occupation-Specific Experience

Notes: Probability of switching from occupation k next period conditional on years of experience in occupation k. Figure considers only individuals who are observed for at least 10 years in the sample.

2.3 Model

In the model, forward-looking, risk-averse individuals face dynamic incentives that reflect two processes: accumulation of experience (learning-by-doing) and accumulation of information (learning about unobserved ability). The model captures the transferability of acquired skills as well as spillovers of information. For instance, a financial manager who decides to become an entrepreneur later in his career may transfer the managerial skills he has acquired into his business. Additionally, his success or failure as a manager may reveal his entrepreneurial ability as well.

Individuals enter the labor market immediately after finishing their education. Their productivity in each occupation is determined by their experience, their unobserved ability, and idiosyncratic shocks that prevent them from learning their ability immediately after the first period. Therefore, they receive noisy income signals that they use to update beliefs about their own ability. Individuals maximize their expected utility using their updated beliefs to compute expectations. They have differential preferences for occupations and are able to smooth consumption over time.

2.3.1 Occupations and Individual Characteristics

Immediately after finishing his education, individual i of age t_{i0} enters the labor market.[25] He then decides which occupation to join and how much to consume. If he decides to work, he can be a white collar or blue collar paid employee, or he can be unincorporated or incorporated self-employed. Denote $d_{kit} \in \{0, 1\}$ as an indicator of whether or not he chooses alternative k at age t.

Every individual has a vector of observable characteristics, h_{it}, that partly determines his productivity in each occupation. h_{it} includes his race (white or black), his education level (high school or less, some college, college, and more than college), as well as observable characteristics that change over time (age, marital status, and accumulated experience). He has perfect foresight over his marital status. His experience is collected in a 4-dimensional vector, x_{it}, that contains his accumulated experience in every occupation. He starts his career with no experience in any occupation, and from that point forward the kth component of his experience vector evolves as a function of his choices as follows

$$x_{kit+1} = x_{kit} + d_{kit} \tag{2.1}$$

[25]The empirical analysis uses data for males only. Therefore, masculine pronouns are used to describe the model.

2.3.2 Preferences

The individual is infinitely-lived and discounts next period's utility by the factor β. He works until age T and begins his retirement at age $T+1$.[26] After reaching retirement age, he only decides how to smooth his remaining savings. In order to capture the effects of more uncertain entrepreneurial outcomes on occupational choices, while retaining tractability, the individual's flow utility is characterized by a CARA function of consumption, c_{it}, with absolute risk aversion parameter ρ. His lifetime utility at period $t \leq T$ is given by

$$-\sum_{s=t}^{T}\sum_{k=0}^{4}\beta^{s-t}d_{kit}\alpha_{kit}(h_{it})\exp\{-\rho c_{it}-\varepsilon_{kit}\}-\sum_{s=T+1}^{\infty}\beta^{s-t}\exp\{-\rho c_{it}\} \tag{2.2}$$

The marginal contribution of consumption to his utility is occupation-specific and is determined by the non-pecuniary cost of each occupation, α_{kit}, which is allowed to vary by education level. Given an education level, the non-pecuniary cost is a function of his vector of observables

$$\alpha_{kit}(h_{it})=\exp\{\alpha_{k0}+\alpha_{k1}black_i+\alpha_{k2}married_{it}+1\{x_{kit}=0\}(\alpha_{k3}+\alpha_{k4}t+\alpha_{k5}\omega_i+\alpha_{k6}t\omega_i)\} \tag{2.3}$$

Notably, α_{kit} includes a first-time entry cost which is a function of permanent wealth (ω_i) and age. This is a reduced-form way of capturing barriers to entry that are not explicitly modeled. For entrepreneurship, the age profile of the entry costs is meant to capture the difficulties that young individuals with weaker credit histories and less savings may face. Additionally, the introduction of the permanent wealth measure is meant to capture whether more affluent individuals, in a long-term sense, face smaller entry costs to entrepreneurship. For identification reasons, the non-pecuniary cost of not working, α_{0it}, is normalized to one. Finally, ε_{kit} is the kth component of a 5-dimensional vector of choice-specific taste shocks

[26] Retirement age is set at $T+1=51$ for data availability reasons.

that he observes before choosing an alternative. The taste shocks are unobserved to the econometrician and are assumed to be drawn from a Type-I Extreme Value distribution, independent across individuals, periods, and options.

2.3.3 Income and Learning

Individual i of education level s starts his labor market career endowed with a vector of occupation-specific abilities $\mathcal{M}_i = \{\mu_{1i}, \ldots, \mu_{4i}\}$ drawn from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Delta}_s$.[27] His ability partially determines his productivity in each occupation. His hourly income at the beginning of period $t+1$, from choosing occupation k at period t, is given by

$$y_{kit+1} = f_k(h_{it}; \theta_k) + \mu_{ki} + \eta_{kit+1} \tag{2.4}$$

His hourly income is the sum of an idiosyncratic productivity shock η_{kit+1}, his unobserved occupation-specific ability μ_{ki}, and a function of his observable characteristics $f_k(\cdot)$, which is characterized by a vector of parameters θ_k.[28] Productivity shocks η_{kit+1} are distributed $N(0, \sigma^2_{\eta_k})$ and are independent over time and across individuals and occupations.

The individual does not observe μ_{ki} and η_{kit+1} separately, which prevents him from learning his ability immediately. Instead, he observes their sum, $\zeta_{kit+1} = y_{kit+1} - f_k(h_{it}; \theta_k)$ after choosing occupation k.[29] He follows Bayes' Rule and uses the information he has acquired, i.e. his residual income signal ζ_{kit+1}, to form beliefs $\mathbb{B}_{it+1}$ about his ability. At the beginning of his career, he believes he is no different from any of his peers, and his prior beliefs correspond to the population distribution of ability for people with his education level. Therefore,

[27] The distribution is set around zero because it is not possible to identify the education level effect in the income equation as well as the mean of the distribution, as it depends on education.

[28] In particular, $f_k(\cdot)$ captures the returns to experience using step-functions of the experience vector x_{it}.

[29] In this sense, he is paid his actual productivity as opposed to his expected productivity. This assumption, although less compelling for paid employment, is a natural assumption for entrepreneurial income, which is not contracted upon.

by virtue of the joint normality of the distribution of ability, his initial beliefs can be characterized by the mean and variance of the population distribution: $\mathbb{B}_{it_{i0}} = \langle \mathbf{0}, \mathbf{\Delta}_s \rangle$. The normality of the prior and the idiosyncratic shocks yields a posterior which is also multivariate normal. Therefore, his beliefs at any period follow a normal distribution and can be characterized by a mean vector $\mathbb{E}_{it}$ and a covariance matrix $\mathbb{V}_{it}$.

Updating rules for similar problems have been previously obtained in the literature (DeGroot, 1970; James, 2011). Define the 4-dimensional vector ζ_{it} with characteristic component $\zeta_{\{k\}it}$ and the 4×4 diagonal matrix Σ_{it} with characteristic component $\Sigma_{\{k,k\}it}$ as follows

$$\zeta_{\{k\}it} = \begin{cases} \zeta_{kit} \text{ if } d_{kit-1} = 1 \\ 0 \text{ otherwise} \end{cases} \qquad \Sigma_{\{k,k\}it} = \begin{cases} 1/\sigma^2_{\eta_k} \text{ if } d_{kit-1} = 1 \\ 0 \text{ otherwise} \end{cases} \tag{2.5}$$

After receiving his income signal from last period's work, he updates his beliefs to

$$\mathbb{E}_{it} = \left[\mathbb{V}^{-1}_{it-1} + \Sigma_{it}\right]^{-1} \left[\mathbb{V}^{-1}_{it-1}\mathbb{E}_{it-1} + \Sigma_{it}\zeta_{it}\right] \tag{2.6}$$

$$\mathbb{V}_{it} = \left[\mathbb{V}^{-1}_{it-1} + \Sigma_{it}\right]^{-1} \tag{2.7}$$

These updating rules reflect how his beliefs change as a function of his experience and information received. Equation (2.6) implies that the effect of a very noisy signal on the prior mean is minor.[30] Moreover, equation (2.6) determines the extent to which learning about ability can happen across occupations. For instance, the direction and magnitude of the adjustment in beliefs of a white collar worker regarding his entrepreneurial ability is determined by the off-diagonal terms in the variance matrix $\mathbb{V}_{it}$. The larger these covariances are, the larger the adjustment will be.[31] Equation (2.7) implies that the prior variance at t is

[30] A noisy signal is characterized by high idiosyncratic variance $\sigma^2_{\eta_k}$.

[31] The marginal effect of a signal from occupation k on the next period's prior mean of occupation k' equals $(1/\eta_k)\mathbb{V}_{\{k,k'\}it}$.

a deterministic map of the vector of accumulated experience x_{it} and the covariance matrix of the ability distribution (see Appendix A.2). Hence, conditional on x_{it}, the order in which the individual samples occupations prior to t is irrelevant to determining the posterior variance. More importantly, provided experience is already included in the individual's state, equation (2.7) implies that $\mathbb{V}_{it}$ is redundant information.

2.3.4 Optimal Choices

At the beginning of any period before retirement, the individual receives his income from last period's occupation and observes his vector of taste shifters ε_{it}. Using his income signal, he updates his beliefs. Given that he can smooth consumption over time, he simultaneously chooses his consumption and asset portfolio, as well as whether to work and which occupation to join.[32] The set up for the consumption smoothing problem follows Margiotta and Miller (2000) and Gayle et al. (2015) whose main result will hold here: the indirect utility from optimal consumption has closed form, and the occupational choice will be independent of disposable wealth. This result will substantially facilitate estimation, and its adoption is also data-driven given the lack of wealth data in the PSID for most years in the period studied. Therefore, credit constraints are not explicitly modeled, and occupational choices will depend on relative differences in non-pecuniary benefits, expected flow payoffs given experience and beliefs, and continuation values from future experience and beliefs induced by each alternative.

Consumption

The set up of the consumption smoothing problem aims to relax the hand-to-mouth assumption that would force individuals to absorb the entire variation in income every period. The

[32] Upon retirement, he simply decides on his consumption and assets portfolio in order to smooth his remaining wealth.

individual has access to a contingent-claims market for consumption goods to smooth his consumption using his wealth. However, income is assumed uninsurable to capture "unobservable insurance risk or unobserved levels of effort in labor supply" (Green, 1987).

Let λ_τ denote the derivative of the price measure for claims to consumption at date τ.[33] If he decides to work in occupation k, individual i supplies $\bar{L}_k$ hours inelastically.[34] Dropping the occupation indicator, working at age t yields him annual income $\bar{L}y_{it+1}$ at the beginning of $t+1$. Hence, the law of motion for his disposable wealth is:

$$E_t[\lambda_{t+1}\xi_{it+1}|d_{it}, h_{it}, \mathbb{E}_{it}] + \lambda_t c_{it} \leq \lambda_t \xi_{it} + E_t[\lambda_{t+1}\bar{L}y_{it+1}|d_{it}, h_{it}, \mathbb{E}_{it}]$$

where the expectation is conditional on his choice at period t, d_{it}. His budget constraint reflects his financial resources, which are allocated to current period consumption and next period savings.

Similar to results in Margiotta and Miller (2000), the assumptions on the market for consumption claims and the CARA nature of the flow utility yield an expression for the value function that can be separated into two factors: an indirect utility function for wealth and an index that captures the value of human capital and information. This expression, presented below, satisfies a portfolio separation property (Altuğ and Labadie, 1994): consumers will hold only a few securities. The first one is a bond b_τ that, contingent on the history through calendar date τ, pays a unit of consumption from period τ in perpetuity in date-τ prices. The second one is a security a_τ that pays the random quantity $(\ln \lambda_s - s \ln \beta)$ of consumption

[33] The commodity space for consumption goods is formed by consumption units at date 0 and claims to consumption at calendar date τ contingent on how history unfolds. λ_τ denotes the derivative of the price measure for claims to consumption at date τ, Λ_τ. Therefore, the price of a unit of consumption to be delivered with certainty at date τ in terms of date 0 consumption units is $E[\lambda_\tau]$.

[34] $\bar{L}_k$ is specified as as the average number of hours worked by individuals in occupation k in the sample.

units from period τ in perpetuity, in date-τ prices. The prices of these assets are given by

$$b_\tau \equiv E_\tau \left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau}\right] \qquad a_\tau \equiv E_\tau \left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau}(\ln \lambda_s - s \ln \beta)\right] \tag{2.8}$$

Individuals in the model can accurately forecast the price of both assets. The state space of the individual's dynamic problem is then formed by his vector of observable characteristics, his beliefs, his wealth, and the prices of these assets. Let $\tau(t)$ be the calendar date when the individual is of age t and let ξ_{it} denote his disposable wealth at t. Following Margiotta and Miller (2000), and dropping the index i for simplicity, the value function solving his savings problem at retirement age $T+1$, in present value terms is

$$V_{T+1}(h_{T+1}, \mathbb{E}_{T+1}, \xi_{T+1}, a_{\tau(T+1)}, b_{\tau(T+1)}) = -\lambda_{\tau(T+1)} b_{\tau(T+1)} \exp\left(\frac{-(\rho\xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right) \tag{2.9}$$

His occupational ability and his experience become irrelevant once he retires.[35] In fact, as he receives no retirement flow income, his present value only depends on his remaining wealth and the price of the assets $a_{\tau(T+1)}$ and $b_{\tau(T+1)}$. At any age before retirement, he chooses an occupation in addition to his consumption and asset portfolio. Proposition 2.1 provides the ex-ante value function of an individual at any age before retirement.[36]

Proposition 2.1. *At any age t before retirement, $t \leq T$, the value function of an individual who has not yet observed his taste shocks, ε_t, can be written as*

$$V_t(h_t, \mathbb{E}_t, \xi_t, a_{\tau(t)}, b_{\tau(t)}) = -\lambda_{\tau(t)} b_{\tau(t)} \exp\left(\frac{-(\rho\xi_t + a_{\tau(t)})}{b_{\tau(t)}}\right) A_t(h_t, \mathbb{E}_t) \tag{2.10}$$

where $A_t(h_t, \mathbb{E}_t)$ is defined recursively as

[35] More realistic models would have ability as well as accumulated human capital generating an income stream after retirement. This chapter abstracts from such considerations, but acknowledging that retirement considerations would play a stronger role in such models as various occupational paths summarized by their experience vector would generate alternative retirement income flows.

[36] The ex-ante value function is defined as the value function before knowing the realization of the vector of taste shocks, ε_{it}

$$A_t(h_t, \mathbb{E}_t) = \sum_{k=0}^{4} p_{kt}(h_t, \mathbb{E}_t)\alpha_{kt}(h_t)^{1/b_{\tau(t)}} E_\varepsilon[e^{-\varepsilon^*_{kt}/b_{\tau(t)}}] E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})v_{kt+1}|\mathbb{E}_t, h_t]^{1-1/b_{\tau(t)}} \quad (2.11)$$

with $A_{T+1}(h_{T+1}, \mathbb{E}_{T+1}) \equiv 1$ *and* $v_{kt+1} \equiv \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right)$.

The probability of choosing k *at age* t *conditional on characteristics and beliefs is denoted* $p_{kt}(h_t, \mathbb{E}_t)$. ε^*_{kt} *is the value of the taste shock* ε_{kt} *conditional on alternative* k *being chosen at* t. *The deterministic transition from* h_t *into* h_{t+1} *is denoted* $\bar{H}_{kt+1}(h_t)$, *and the stochastic transition from yesterday's beliefs into today's is denoted* $\mathbb{E}_{kt+1}$; *both are conditional on choosing* k *at* t.

Proof: See Appendix A.2.

The ex-ante value function in (2.10) can be separated into two factors: the indirect utility from wealth and an index that represents the value of accumulated experience and information. The index $A_t(h_t, \mathbb{E}_t)$ is a strictly positive average of expected outcomes weighted by the conditional choice probability of each alternative. The function v_{kt+1} in the recursive formulation of $A_t(h_t, \mathbb{E}_t)$ is a utility measure of income from occupation k received at the beginning of period $t+1$, adjusted for consumption smoothing. Higher values of the prior mean or higher values of human capital are associated with lower values of the index via the term v_{kt+1}. Additionally, the value of human capital and beliefs in occupation k decreases with the size of the non-pecuniary costs, α_{kt}.

Occupation

Using equations (2.10) and (2.11), and applying logs to transform the problem, it can be shown that at any age t before retirement the individual chooses an occupation to solve

$$\max_k \sum_{k=0}^{4} d_{kt}\{\varepsilon_{kt} - \ln \alpha_{kt}(h_t) - (b_{\tau(t)} - 1)\ln E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})v_{kt+1}|\mathbb{E}_t, h_t]\} \quad (2.12)$$

His occupational choice is independent of his current level of wealth. This is a consequence of the multiplicative separability of the ex-ante value function obtained in Proposition 2.1.[37] Trade-offs between occupations are characterized by differences in non-pecuniary utility components, ε_{kt} and $\alpha_{kt}(h_t)$, and differences in the expected utility from income, v_{kt+1}, scaled by the index capturing the value of human capital accumulation and beliefs evolution, $A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})$. Since the taste shocks are assumed to be distributed Type-I Extreme Value, the expression in equation (2.12) becomes a standard logit. Following Hotz and Miller (1993), Proposition 2.2 uses the recursive nature of the index $A_{t+1}(h_t, \mathbb{E}_t)$ and yields a representation of the logarithm of the odds ratio in terms of future choice probabilities and utility parameters.

Proposition 2.2. *For any choice $k > 0$, the logarithm of the likelihood ratio between choosing occupation k and choosing not to work is given by*

$$\ln\left(\frac{p_{kt}(h_t, \mathbb{E}_t)}{p_{0t}(h_t, \mathbb{E}_t)}\right) = -\ln \alpha_{kt}(h_t) - (b_{\tau(t)} - 1)\ln E_t\left[v_{kt+1}\prod_{s=1}^{T-t}\left(\frac{p_{0t+s}(h_{kt}^{(s)}, \mathbb{E}_{kt}^{(s)})}{p_{0t+s}(h_{0t}^{(s)}, \mathbb{E}_{0t}^{(s)})}\right)^{\phi_t(s)}\middle| \mathbb{E}_t, h_t\right] \quad (2.13)$$

where

$$\phi_t(s) = \frac{1}{b_{\tau(t)+s}}\prod_{r=1}^{s-1}(1 - 1/b_{\tau(t)+r}) \quad (2.14)$$

and where $h_{kt}^{(s)}$ and $\mathbb{E}_{kt}^{(s)}$ indicate the value of the state variables at future age $t+s$, conditional

[37] In fact, relaxing the model to allow the use of savings as collateral would break this separability property as low levels of wealth will not allow the individual to obtain the optimal scale of his business, rendering the occupational choice dependent on wealth.

on the decision path described by making $d = 1$ for all $d \in \{d_{kt}, d_{0t+1}, d_{0t+2}, \ldots, d_{0T}\}$.

Proof: See Appendix A.2.

Equation (2.13) shows that the logarithm of the likelihood ratio between working in any occupation $k > 0$ and the decision not to work is a function of the trade-offs described in equation (2.12).[38] Higher non-pecuniary costs and lower expected utility from compensation make alternative k less likely. Moreover, if choosing alternative k makes the individual less likely to work in the future, thereby reducing the value of his human capital or his information, then alternative k is also less likely to be chosen today.

2.4 Estimation

Estimation of the parameters of the model is done in two stages using a combination of an Expectation-Maximization (EM) algorithm and a conditional choice probabilities (ccp) estimator (Hotz and Miller, 1993; Arcidiacono and Miller, 2011; James, 2011). The EM algorithm in the first stage permits fast estimation of income and ability parameters, bypassing the need for multidimensional integration over unobserved ability vectors. The ccp estimator in the second stage follows naturally from the expression derived in Proposition 2.2. It allows for a flexible treatment of the large state space of the problem, which includes continuous beliefs and experience. It yields consistent estimates of the utility parameters without having to solve the dynamic optimization problem at every candidate parameter vector in the search algorithm. The flexibility provided by the ccp estimator facilitates the disaggregation of occupations used in this chapter.

[38]The derivation of equation (2.13) relies on the assumption that the decision not to work changes neither the beliefs nor the vector of accumulated human capital. In general, an expression similar to that in Proposition 2.2 can be obtained in terms of any path of future choices. Such alternative representation, however, will not be as tractable because expectations over functions of future income signals will be necessary.

Individuals in the model select on beliefs about their ability rather than selecting on their actual ability. In other words, conditional on the history of income signals up to t, mapped into beliefs $\mathbb{E}_{it}$, choices at t are independent of ability. Let Λ be the collection of parameters of the utility function, let Θ be the collection of income parameters, including the variance parameters of the productivity shocks, σ_{η_k}, and let $\boldsymbol{\Delta}_s$ be the covariance matrix of the population ability distribution conditional on education level s. Therefore, the likelihood of the data—hourly income and choices—for a person with education level s can be written as:

$$\mathcal{L}_i = \prod_{t=t_{i0}}^{T_i} \prod_{k=0}^{4} \Pr\left[d_{kit} = 1 | h_{it}, \mathbb{E}_{it}; \Lambda, \Theta\right]^{d_{kit}} \times \int_{\tilde{\mathcal{M}}} \left\{ \prod_{t=t_{i0}}^{T_i} \prod_{j=1}^{4} \Pr\left[y_{jit+1} | h_{it}, \tilde{\mu}_j; \Theta\right]^{d_{jit}} \right\} dF\left(\tilde{\mathcal{M}}; \boldsymbol{\Delta}_s\right) \tag{2.15}$$

Equation (2.15) reflects two characteristics of the model. First, individuals are heterogeneous in their unobserved ability, $\mathcal{M}_i$. Second, rather than their unobserved ability, a function of the history of their income signals—their belief $\mathbb{E}_{it}$—shapes their occupational decisions. This has the convenient effect of taking the choices part of the likelihood out of the multi-dimensional integral.[39] The log likelihood of (2.15) is then additively separable:

$$\ln \mathcal{L}_i = \ln \mathcal{L}_i^d + \ln \mathcal{L}_i^y$$

The first stage of the estimation procedure utilizes the income term of the log likelihood to obtain estimates of Θ and $\boldsymbol{\Delta}_s$. These estimates are used in the second stage to estimate Λ. The scale of Θ, $\boldsymbol{\Delta}_s$, and ρ depends on the units in which income and consumption are measured. Hourly income is expressed in \$10 units and consumption in \$1,000 units.

[39] No measurement error is assumed in the hourly income data, which allows for people's beliefs to be backed out using income data. Allowing for measurement error would render the two-stage procedure non-viable as integration over the error terms would be necessary over the entire expression in equation (2.15).

Therefore, converting hourly income into annual income for occupation k in the model entails dividing $\bar{L}_k y_{kt+1}$ by 100. The standard errors provided in the chapter are uncorrected for the two-stage estimation.[40]

2.4.1 First Stage: Income Parameters and Learning Structure

The first stage uses an Expectation-Maximization (EM) algorithm, an iterative method that yields maximum likelihood estimates when a portion of the data is unobserved. In the model, the unobserved part of the data is the individual's ability, $\mathcal{M}_i$. In order to implement the EM algorithm, assume $\mathcal{M}_i$ is observed for all i. Hence, the income term of the log likelihood for individual i becomes

$$\ln \mathcal{L}_i^y(\mathcal{M}_i) = \sum_{t=t_{i0}}^{T_i} \sum_{j=0}^{4} d_{jit} \ln \Pr\left[y_{jit+1} | h_{it}, \mu_j; \Theta\right]$$

Starting from a guess of parameters $\langle \Theta^0, \Delta^0 \rangle$, implementation of the EM algorithm entails iteration over the following two steps to obtain maximum likelihood estimates:

1. *Expectation Step*: compute the expected value of $\ln \mathcal{L}_i^y(\mathcal{M}_i)$, conditional on the data actually observed and the parameters at the mth iteration

$$E_m[\ln \mathcal{L}_i^y(\mathcal{M}_i)|\cdot]$$

2. *Maximization Step*: find the new iterated value of the vector of parameters by maximizing the expression obtained in the expectation step:

$$\langle \Theta^{m+1}, \Delta^{m+1} \rangle = \max_{\langle \Theta, \Delta \rangle} \sum_i E_m[\ln \mathcal{L}_i^y(\mathcal{M}_i)|\cdot]$$

[40] For the parameters in the first stage, which uses an EM algorithm, computation of the standard errors follows the NDS procedure described in Jamshidian and Jennrich (2000).

For use in the second stage, consistent estimates of individual beliefs are obtained. This computation uses the point estimates for Θ and $\mathbf{\Delta}_s$, the history of signals received by every individual, Bayes' Rule, and the rational expectations assumption regarding the individual's prior. The first stage of the estimation algorithm is detailed in Appendix A.3.

2.4.2 Second Stage: Utility Parameters

In order to estimate Λ, the second stage follows Hotz and Miller (1993) and takes advantage of the expression derived in Proposition 2.2 mapping future choice probabilities and utility parameters into current choice probabilities. The Type-I Extreme Value assumption regarding the distribution of preference shocks implies that the choice probabilities can be written as

$$p_{kit}(h_{it}, \mathbb{E}_{it}) = \frac{\exp(V_k(h_{it}, \mathbb{E}_{it}))}{1 + \sum_{k'>0} \exp(V_{k'}(h_{it}, \mathbb{E}_{it}))} \tag{2.16}$$

where $V_0 = 0$ and for any $k > 0$

$$V_k(h_{it}, \mathbb{E}_{it}) = -\ln \alpha_{kit}(h_{it}) - (b_{\tau(t)} - 1) \ln E_t \left[v_{kit+1} \prod_{s=1}^{T-t} \left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{0it}^{(s)})} \right)^{\phi_t(s)} \middle| \mathbb{E}_{it}, h_{it} \right] \tag{2.17}$$

An iterative algorithm is implemented that maximizes the log likelihood of the data while searching over the space of parameters and ccps. This procedure is akin to the swapping of the nested fixed point algorithm described in Aguirregabiria and Mira (2002). The procedure is initialized with flexible parametric versions of the future conditional choice probabilities estimated from the data.[41] It entails the following two steps:

1. *Maximization Step*: plug the estimated ccps in equation (2.16) and maximize the log likelihood of the observed choices. Notably, forming $V_k(h_t, \mathbb{E}_t)$ requires knowledge of the bond prices $b_{\tau(t)}$. Following Gayle and Miller (2009), obtain bond prices using

[41] When computing the ccps, beliefs, estimated in the first stage, are also treated as data.

series from the Federal Reserve Economic Data (see Appendix A.1).

2. *CCP Step*: use the estimated parameters at the current iteration to solve the model backwards and obtain new ccps implied by the model.

The parameter vector that yields the minimum log likelihood is chosen. The second stage of the estimation method is further explained in Appendix A.3.

2.4.3 Identification

The sources of variation that identify the parameters of the model are observed income and occupational choices over time, as well as variation in observed experience and demographics. There are three sets of parameters to be identified: returns to experience and education, parameters of the Bayesian learning structure including the distribution of abilities, and parameters of the utility function. Identification of each set of parameters is discussed below.

The main challenge for causal estimation of the returns to occupation-specific experience in the income equation is that individuals select on beliefs (Gibbons et al., 2005). This is accounted for in estimation using the likelihood induced by the model.[42] Provided that selection is accounted for, income variation for different levels of occupation-specific experience identifies the returns to experience. This is the learning-by-doing piece of the model. The returns to cross-occupation experience are identified from the variation in income of switchers. Notably, since education indicators are introduced linearly in the income equations, the mean of the distribution of ability is not identified.[43]

Identification of the latent distribution of ability under normality assumptions using income data and occupation choices has been shown in Heckman and Honoré (1990). In-

[42] Instrumental variables have also been used to account for selection on beliefs (see Altonji and Williams (2005), Dillon and Stanton (2016)).

[43] The distribution of ability is also conditional on education and its mean enters linearly in the income equation.

dividuals with different education levels are likely to have different distributions of ability. Therefore, given that education is not endogenized, the distribution of ability is made education-specific. The parameters of the covariance matrix of the ability distribution are then identified from variation in residual income. In particular, the off-diagonal terms of the variance matrix are identified from the covariation in residuals of switchers. Finally, the variance of the idiosyncratic shocks, which is not made education-specific, is identified from the excess variation in residual income per occupation.[44]

Following results in Arcidiacono and Miller (2015), given that the distribution of the choice-specific taste shocks, the subjective discount factor, and the transition function of beliefs—estimated in the first stage—are known, the flow payoffs are identified up to the normalization that the flow payoffs from unemployment are zero in each state and time period.[45] Hence, the functional form assumptions regarding the utility function provide over-identifying restrictions. Notably, separate identification of the risk aversion parameter from prior beliefs may be difficult. This is because similar choice patterns would be generated by overconfident low-variance priors and high risk aversion, or by under confident high-variance priors with low risk aversion. The panel dimension of the data helps separate the variation in choices due to risk aversion. Over time and regardless of priors, Bayesian learning implies that individuals' beliefs will get arbitrarily close to their true ability, and the remaining idiosyncratic variation would help identify the risk aversion parameter.

2.5 Parameter Estimates and Economic Forces at Play

In this section, the estimates of the structural model are discussed, and the economic forces at play in the decision to become an entrepreneur are evaluated. Especial attention is paid

[44]These points can be seen more clearly in the updating rules of the EM algorithm in Appendix A.3.

[45]Identification in models of dynamic discrete choice is also discussed in Magnac and Thesmar (2002) and Rust (1994).

to the role that these forces play in the timing of the entrepreneurial choice. Section 2.5.1 presents estimated parameters of the income equation and the distribution of ability. Section 2.5.2 discusses utility parameter estimates and Section 2.5.3 discusses model fit. Finally, the role of the forces at play is quantified using a decomposition exercise in Section 2.5.4.

2.5.1 Income Parameters and the Distribution of Ability

Two key economic forces in the model are learning-by-doing and learning about ability. In this section the estimated results for these processes are described. Beginning with learning-by-doing, Table 2.6 presents the estimated returns to experience as well as other parameters of the income equation in (2.4). The returns to experience are specified using step-functions that capture the marginal increase in productivity for an extra year of experience.[46]

The estimated increments in productivity from Table 2.6 are illustrated in Figure 2.4. As an example, one year of blue collar experience adds .81 dollars to the hourly income from blue collar work. A second year of blue collar experience adds another .74 dollars and so forth. Figure 2.4 shows that the returns to blue collar experience are the flattest while the returns to incorporated experience are the steepest. Individuals trying to reach the high productivity levels available in incorporated entrepreneurship should start climbing the ladder when they are young. .

Learning-by-doing can also happen across-occupations. That is, experience accumulated in one occupation may have non-zero returns in another. Figure 2.5 illustrates the estimated returns to cross-occupation experience. As an example, one year of white collar experience adds 1.33 dollars to the hourly income from blue collar work, and any extra year of white collar experience up to five does not add anything. The most striking result coming out of

[46]The choice of representing the returns profile using a step-function is data driven. It is difficult to obtain smooth profiles for entrepreneurial experience given that the number of individuals with high numbers of entrepreneurial experience is low. The location of the steps was chosen from preliminary OLS regressions. No steps beyond the 10th year of experience were significant in the OLS exercise, so it is assumed that individuals reach the top of the productivity ladder by the 10th year in the occupation.

Table 2.6: Income Parameters

	Blue Collar		White Collar		Unincorporated		Incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
Constant	0.927**	0.008	0.847**	0.015	1.304**	0.059	0.730**	0.244
Black	-0.204**	0.008	-0.122**	0.014	-0.180**	0.052	-0.097	0.278
Some College	0.222**	0.009	0.129**	0.014	-0.095**	0.041	0.597**	0.237
College	0.323**	0.017	0.456**	0.018	0.537**	0.083	0.411*	0.216
More than College	0.251**	0.021	0.624**	0.020	0.934**	0.080	1.842**	0.366
Married	0.053**	0.005	0.213**	0.008	0.019	0.039	0.718**	0.114
$1\{exp_{bc} \geq 1\}$	0.081**	0.006	-0.077**	0.011	-0.233**	0.053		
$1\{exp_{bc} \geq 2\}$	0.074**	0.007					-0.268	0.179
$1\{exp_{bc} \geq 3\}$	0.041**	0.007						
$1\{exp_{bc} \geq 4\}$	0.056**	0.008						
$1\{exp_{bc} \geq 5\}$	0.067**	0.008					-0.508	0.319
$1\{exp_{bc} \geq 6\}$	0.142**	0.007						
$1\{exp_{bc} \geq 7\}$			0.119**	0.015				
$1\{exp_{wc} \geq 1\}$	0.133**	0.006	0.126**	0.012				
$1\{exp_{wc} \geq 2\}$			0.127**	0.013	0.019	0.063	-0.765**	0.182
$1\{exp_{wc} \geq 3\}$			0.192**	0.012	0.044	0.072	0.824**	0.228
$1\{exp_{wc} \geq 5\}$			0.140**	0.011				
$1\{exp_{wc} \geq 6\}$	0.209**	0.013			0.480**	0.061		
$1\{exp_{wc} \geq 7\}$			0.178**	0.014				
$1\{exp_{wc} \geq 8\}$			0.410**	0.013			0.627**	0.124
$1\{exp_{eu} \geq 1\}$			-0.177**	0.016	0.219**	0.033		
$1\{exp_{eu} \geq 3\}$							0.229	0.241
$1\{exp_{eu} \geq 4\}$			0.743**	0.032	0.407**	0.034		
$1\{exp_{eu} \geq 7\}$					0.280**	0.038		
$1\{exp_{ei} \geq 1\}$					-0.451**	0.060	0.433**	0.086
$1\{exp_{ei} \geq 2\}$			1.004**	0.028				
$1\{exp_{ei} \geq 3\}$					1.867**	0.109		
$1\{exp_{ei} \geq 5\}$							1.078**	0.085
$1\{exp_{e} \geq 1\}$	-0.022*	0.011						
$1\{exp_{e} \geq 2\}$	-0.156**	0.017						
$1\{exp_{e} \geq 5\}$	0.365**	0.026						
Obs	8902		9957		1403		602	

Notes: Hourly income measured in \$10s. *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. Estimated parameters of equation (2.4). Returns to experience are estimated as step functions. As an example, $1\{exp_{eu} \geq 3\}$ indicates that the individual has three years or more of unincorporated experience. In blue collar work, experience from both entrepreneurial occupations is pooled: $exp_e = exp_{eu} + exp_{ei}$. Steps functions were chosen to avoid out of sample return estimates especially on entrepreneurial occupations. The steps were chosen using statistical significance in a preliminary OLS.

Figure 2.5 is that, while expertise in entrepreneurial activities always increases productivity in paid employment, low levels of entrepreneurial experience reduce it. This finding is similar to some of the results in Jovanovic and Nyarko (1996), where switching technologies can

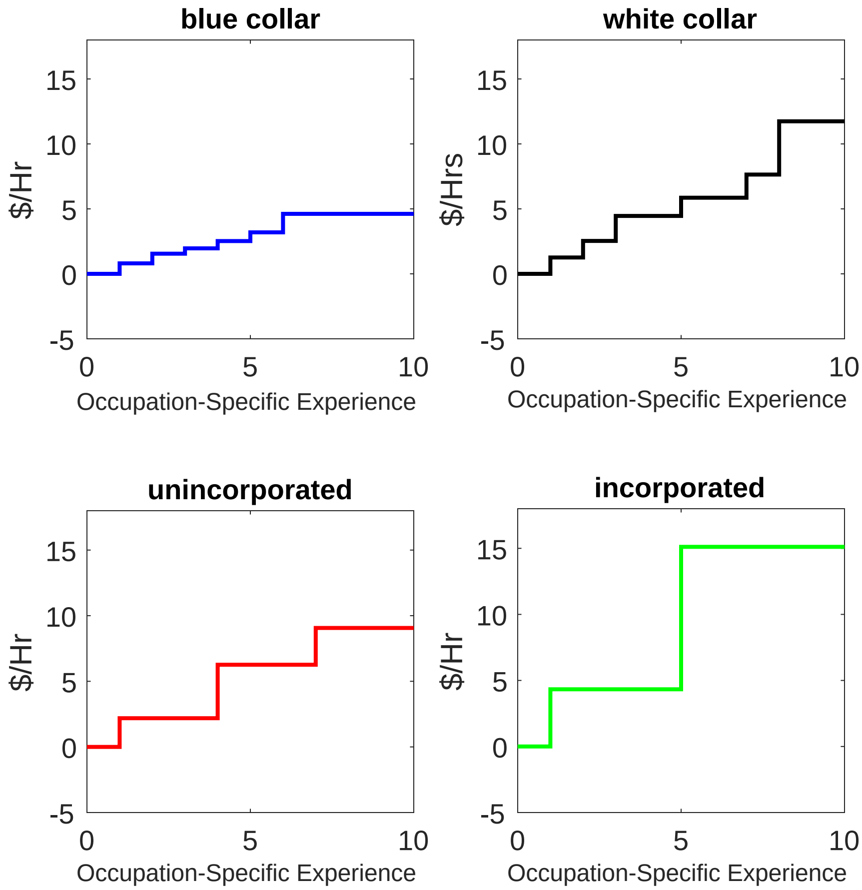


Figure 2.4: Returns to Own-Occupation Experience

Notes: Returns implied by the marginal step estimates in Table 2.6.

reduce productivity by reducing expertise.[47] These results suggest that, unless mastered, skills learned in entrepreneurship may harm paid-employment productivity.[48]

Another form of learning allowed in the model is learning about ability. In the model, individuals start their careers with a common prior that corresponds to the population distribution of ability for their education level. Over time, their success or failure in each occupation allows them to update their beliefs. The estimated covariance matrices for the

[47]Manso (2014) finds that entrepreneurial experience generates a premium for salaried workers whenever it is more than two years.

[48]Some skills commonly considered entrepreneurial are the ability to sell, innovative thinking, initiative, and self-reliance.

ability distribution, which constitute the initial priors, are presented in Table 2.7.[49] Recall that the ability vector is measured in \$10s/hr. Therefore, an individual with more than college education and ability that is one standard deviation above the population mean would have an incorporated ability of $10 \times \sqrt{10.88} \approx \$33/\text{hr}$. The same individual, would have white collar ability of $10 \times \sqrt{0.87} \approx \$9/\text{hr}$. This comparison highlights the most important results from Table 2.7: there is higher variation in entrepreneurial ability than in paid-employment ability. In addition, more education tends to be associated with higher variation in ability. In light of the model, learning about ability implies that individuals should enter entrepreneurship as early as possible in order to find out their position in the high-variance ability distribution. .

Learning about ability can also happen across-occupations because occupational abilities are correlated. This can be seen in the off-diagonal terms of the covariance matrix in Table 2.7. To better understand this relationship, the correlations between abilities are computed and presented in Figure 2.6. Individuals with high ability in white collar work tend to have high ability in entrepreneurial activities. Moreover, consistent with the discussion in Section 2.2, the correlation between white collar ability and incorporated ability is higher than the correlation between both entrepreneurial abilities at any education level. These results are also consistent with the differences between entrepreneurial occupations presented in Levine and Rubinstein (Forthcoming).

Estimates indicate that there is an incentive for young individuals to attempt entrepreneurship in order to learn whether they are high-ability. Alternatively, they can use white collar success as an indicator of their entrepreneurial ability. However, these learning possibilities depend on how noisy the signals are in each occupation. As Table 2.8 shows, the high idiosyncratic variation of entrepreneurial occupations threatens their informational value. .

[49] As mentioned above, the mean of the ability distribution is normalized to zero because it is not separately identified from the linear returns to education in equation (2.4) (see Table 2.6).

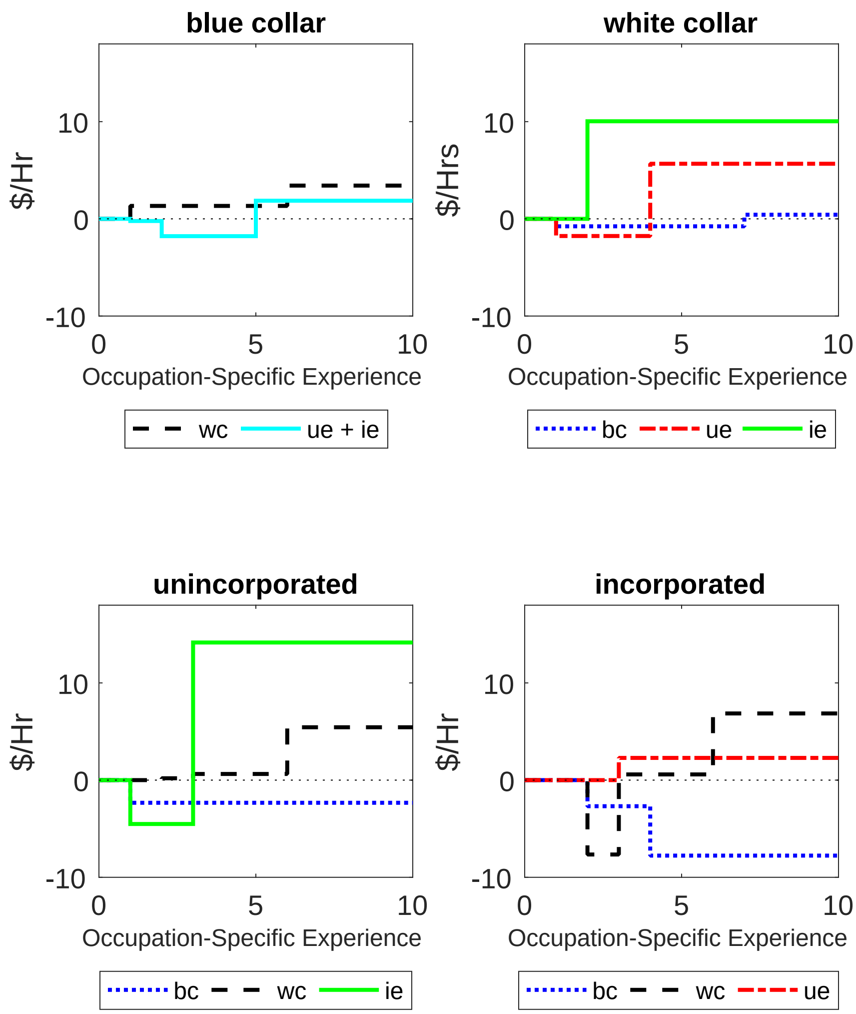


Figure 2.5: Returns to Cross-Occupation Experience

Notes: Returns implied by the marginal step estimates in Table 2.6. Occupations are: white collar (wc), blue collar (bc), unincorporated entrepreneurship (ue), and incorporated entrepreneurship (ie). In blue collar work, experience from both entrepreneurial occupations is pooled.

To get a sense of how fast own- and cross-occupation learning about ability can actually happen, Figure 2.7 presents the percent of prior uncertainty about entrepreneurial ability that is eliminated after working for 5 years in each occupation. Consider the right panel and individuals with more than college education—the rightmost set of 4 bars in the figure. On the one hand, own-occupation learning indicates that initial uncertainty about incorporated ability is reduced by almost 90 percent, after 5 years of incorporated experience. On the

Table 2.7: Population Ability Covariance Matrices

High School

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.15**	0.003						
white collar	0.13**	0.007	0.11**	0.008				
unincorporated	0.20**	0.020	0.17**	0.029	0.27**	0.040		
incorporated	0.04	0.479	0.03	0.313	0.05	0.920	0.04	0.293

Some College

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.23**	0.006						
white collar	0.26**	0.009	0.32**	0.009				
unincorporated	0.16**	0.032	0.14**	0.032	0.28**	0.033		
incorporated	0.45**	0.128	0.77**	0.071	0.35*	0.179	4.60**	0.395

College

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.45**	0.021						
white collar	0.33**	0.026	0.57**	0.016				
unincorporated	0.29**	0.099	0.38**	0.059	3.52**	0.249		
incorporated	0.71**	0.079	0.85**	0.038	-0.11	0.528	1.66**	0.179

More than College

	blue collar		white collar		unincorporated		incorporated	
	coeff	se	coeff	se	coeff	se	coeff	se
blue collar	0.37**	0.020						
white collar	0.22**	0.035	0.87**	0.022				
unincorporated	-0.41**	0.110	0.66**	0.075	3.03**	0.175		
incorporated	-0.29	0.685	1.82**	0.149	2.35**	0.706	10.88**	0.665

Notes: *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. Covariance matrix of the joint distribution of unobserved ability conditional on education, denoted $\mathbf{\Delta}_s$.

Table 2.8: Idiosyncratic Variance

blue collar		white collar		unincorporated		incorporated	
coeff	se	coeff	se	coeff	se	coeff	se
0.30**	0.001	0.96**	0.004	2.47**	0.03	8.00**	0.134

Notes: *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. Idiosyncratic hourly income variance in every occupation, denoted $\sigma^2_{\eta_k}$.

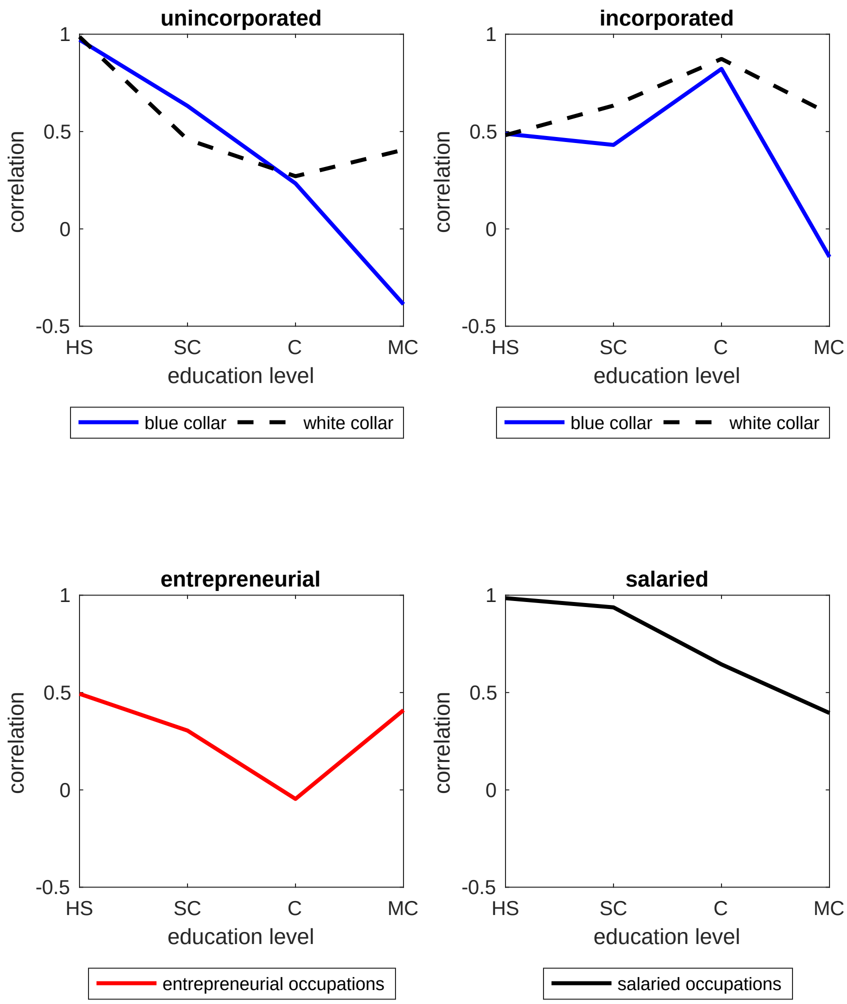


Figure 2.6: Implied Correlation Between Abilities

Notes: Correlation between abilities per education level implied by the estimates in Table 2.7. Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). Top figures show the implied correlations of unobserved ability between salaried occupations and each of the entrepreneurial occupations. Bottom left figure shows the implied correlation of abilities between unincorporated and incorporated entrepreneurship. Bottom right figure shows the implied correlation of abilities between blue and white collar.

other, cross-occupation learning indicates that initial uncertainty about incorporated ability is reduced by about 30 percent, after 5 years of white collar experience. Surprisingly, Figure 2.7 shows that for college educated individuals, paid employment can be a better source for learning about incorporated ability than incorporated entrepreneurship itself. Correlated learning about ability provides incentives for young college educated individuals to attempt entrepreneurship later in life. .

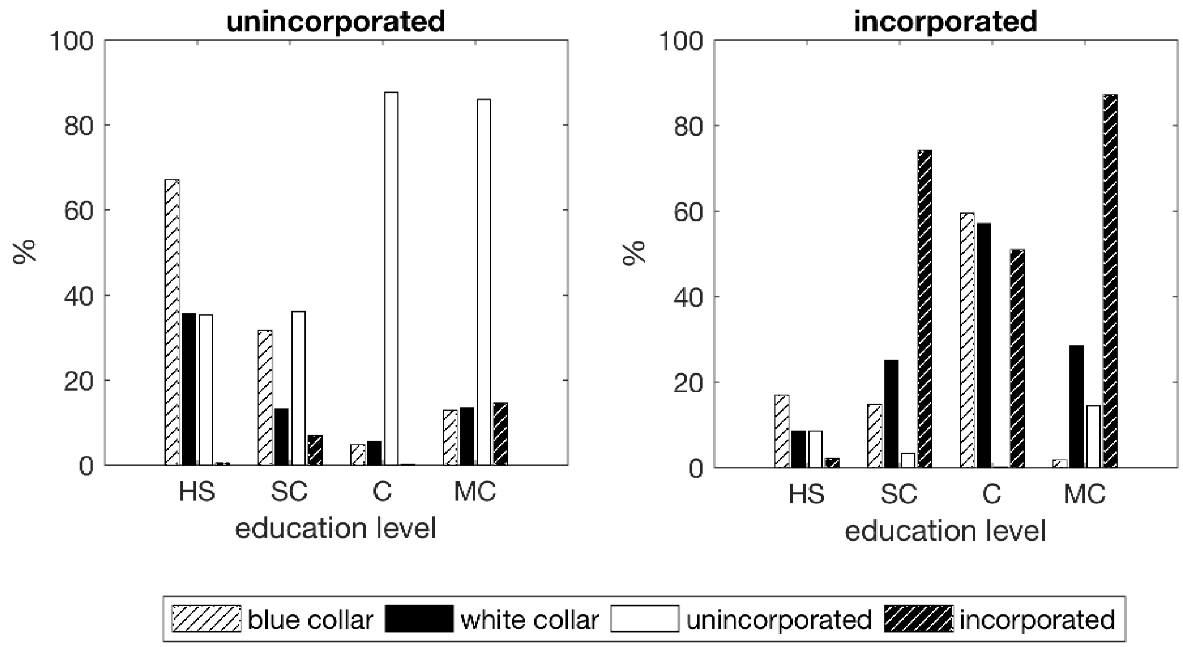


Figure 2.7: Prior Variance Eliminated After 5 Years of Occupation-Specific Experience

Notes: Figure shows how much prior variance remains in occupation k after accumulating five years of experience in occupation k' and zero years in any other occupation. On the y-axis of the panel devoted to occupation k is the percent quantity $(1 - \mathbb{V}_{\{k,k\},5}(k')/\mathbb{V}_{\{k,k\},0}) * 100$. The numerator is the belief variance of occupation k after accumulating five years of experience in occupation k' and zero in any other occupation. The denominator is the prior variance of occupation k. Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). Quantities are obtained using the estimates in Table 2.7.

2.5.2 Utility Parameters

Estimates of the utility function parameters are presented in Table 2.9. First, individuals are risk-averse—the point estimate for ρ is greater than zero. To better assess the importance of risk aversion, Figure 2.8 depicts static and dynamic measures of certainty equivalence. On the y-axis of both panels is the proportion of expected annual income that would be necessary to make individuals indifferent between accepting the income variation in the occupation and receiving the certainty equivalent for sure.[50] The static certainty equivalent,

[50]By construction, provided individuals are risk-averse, this measure is always bounded above by 1 in the static case. The measure is further explained in Appendix A.4.

on the left panel in Figure 2.8, highlights the role of risk aversion as it does not account for the future value of experience and beliefs. Not surprisingly, incorporated entrepreneurship has the lowest static certainty equivalent as it offers the highest income variation. However, dynamic considerations about future human capital and beliefs mitigate the effects of risk aversion (right panel of Figure 2.8).

. Besides risk aversion, first-time entry costs are a barrier to young entrepreneurship.

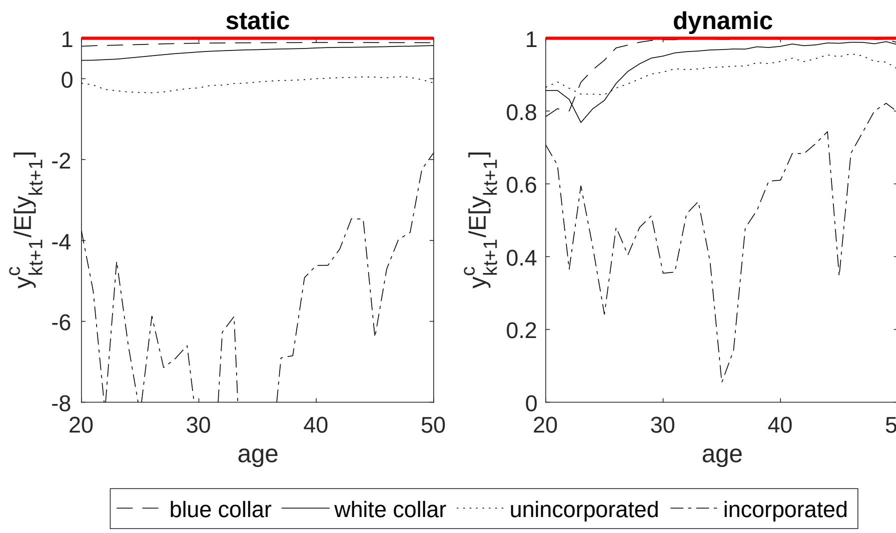


Figure 2.8: Certainty Equivalent

Notes: On the y-axis of both figures is the average of a scale free measure of risk aversion computed across individuals of a given age with positive expected income, $E[y_{kt+1}|h_t, \mathbb{E}_t] > 0$. The measure can be understood as the proportion of expected annual income that would be necessary to let a static/dynamic individual indifferent between taking the gamble of going into the occupation and receiving the certainty equivalent for sure without entering the occupation. In the numerator is the certainty equivalent, y^c_{kt+1}, static and dynamic, obtained in equations (A.29) and (A.30) in Appendix A.4. In the denominator is the expected hourly income conditional on beliefs, $E[y_{kt+1}|h_t, \mathbb{E}_t]$.

Table 2.9 suggests that entry costs for entrepreneurship are higher for young individuals as well as individuals with low permanent wealth. Figure 2.9 introduces monetary equivalents for entry cost estimates.[51] On the x-axis of each panel is the education level and on the y-axis is the monetary equivalent in thousands of dollars (year, 2000). Entry costs are more responsive to age than to permanent wealth. On the one hand, the monetary equivalent of

[51]In the specification of the model, entry costs and non-pecuniary benefits can be treated as the indirect utility representation of terms in the budget constraints (see Appendix A.4).

Table 2.9: Utility Parameters

ρ	coeff	se
	0.040**	0.0005

		blue collar		white collar		unincorporated		incorporated	
High School	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-1.804**	0.042	-0.939**	0.048	-0.549**	0.061	-1.459**	0.104
	black	0.738**	0.041	1.173**	0.047	0.895**	0.061	1.724**	0.135
	married	-0.684**	0.040	-0.364**	0.047	-0.544**	0.060	-0.051	0.112
Entry Cost	*constant*	-4.723**	0.338	2.975**	0.143	5.298**	0.154	12.645**	0.251
	$age/10$	2.837**	0.151	-0.132*	0.053	-0.536**	0.054	-2.090**	0.084
	$\omega_i/10^3$	2.681**	0.705	-1.529**	0.377	0.465*	0.231	-0.258	0.196
	$(age/10)\cdot(\omega_i/10^3)$	-1.108**	0.287	0.421**	0.143	-0.234**	0.080	0.083	0.067
Some College	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-1.986**	0.052	-1.920**	0.053	-1.717**	0.066	-1.361**	0.090
	black	0.466**	0.058	0.857**	0.061	1.069**	0.081	1.248**	0.109
	married	-0.933**	0.058	-0.784**	0.060	-0.686**	0.071	-0.874**	0.099
Entry Cost	*constant*	-3.838**	0.213	3.807**	0.171	4.974**	0.148	11.559**	0.200
	$age/10$	2.061**	0.087	-0.109	0.067	-0.331**	0.050	-1.649**	0.060
	$\omega_i/10^3$	-0.515	0.413	-2.989**	0.456	0.024	0.077	-0.526**	0.069
	$(age/10)\cdot(\omega_i/10^3)$	0.236	0.168	1.085**	0.188	-0.029	0.023	0.158**	0.017
College	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-2.667**	0.101	-3.596**	0.100	-2.987**	0.113	-4.394**	0.142
	black	0.216*	0.104	1.465**	0.104	1.755**	0.169	16.547	319.588
	married	-0.208	0.107	0.381**	0.106	0.067	0.119	1.002**	0.146
Entry Cost	*constant*	-3.117**	0.390	1.199**	0.291	5.810**	0.215	7.482**	0.288
	$age/10$	1.887**	0.149	0.771**	0.110	-0.534**	0.068	-0.700**	0.085
	$\omega_i/10^3$	-1.243	0.957	-1.912**	0.729	-3.583**	0.341	-2.451**	0.297
	$(age/10)\cdot(\omega_i/10^3)$	0.800*	0.372	0.375	0.287	1.106**	0.127	0.610**	0.102
More than College	α	coeff	se	coeff	se	coeff	se	coeff	se
Non Peuniary	*constant*	-1.038**	0.066	-1.902**	0.063	-1.287**	0.077	-0.457**	0.109
	black	-0.637**	0.116	-0.298**	0.114	0.423*	0.166	-0.809**	0.144
	married	-0.436**	0.076	-0.336**	0.072	-0.740**	0.085	-0.966**	0.117
Entry Cost	*constant*	-5.888**	0.442	-3.414**	0.479	5.186**	0.238	9.253**	0.301
	$age/10$	2.689**	0.166	2.481**	0.187	-0.200**	0.072	-0.875**	0.086
	$\omega_i/10^3$	1.951*	0.809	1.712	0.897	-0.531	0.284	-3.499**	0.241
	$(age/10)\cdot(\omega_i/10^3)$	-0.459	0.301	-0.921**	0.357	0.070	0.088	0.897**	0.074

Notes: *, ** indicate statistical significance at the .10 and .05 levels, respectively. Standard errors have not been corrected for 2-stage estimation yet. ω_i is defined as the individual's permanent wealth in Section 2.2 and it is measured in thousands of dollars of 2000. Estimated parameters of equations (2.2) and (2.3).

entry costs into incorporated entrepreneurship decreases from about \$200,000, for individuals age 20, to less than \$150,000, for individuals age 40. On the other, entrepreneurial entry costs decrease \$20,000 or less when permanent wealth goes from the 10th percentile to the 90th. The negative sign of the relation between entry costs and permanent wealth suggests, in a reduced form sense, that individuals are able to ease their barriers to entrepreneurship using their life-time potential. However, the steeper profile of entry costs associated with age captures barriers to entry not explicitly modeled, such as tighter credit constraints for young individuals with weaker credit histories or less capital.[52] .

Non-pecuniary motivations such as "being one's own boss" and "wanting flexibility over schedule" (Hamilton, 2000; Hurst and Pugsley, 2011, 2015) have also been suggested in the literature. This association would be even stronger for risk-averse individuals trying to avoid the higher variation of entrepreneurial outcomes. However, the dynamic treatment of the entrepreneurial choice, as well as the integration of its information value, suggests a more nuanced story. Estimates of the non-pecuniary benefits not associated with first entry are presented in Table 2.9 and are converted to their monetary equivalent in Figure 2.10. Overall dominance of entrepreneurial non-pecuniary benefits does not emerge once dynamic considerations are introduced.[53] On the one hand, entrepreneurial activities are always ranked below blue collar work for low educated individuals. In monetary terms, this difference is equivalent to at least \$20,000 per year. On the other, entrepreneurial occupations become more attractive in non-pecuniary benefits for individuals with college or more. In particular, the non-pecuniary benefits of incorporated entrepreneurship for the college educated are higher that those from any other occupation at any education

[52]One more pattern emerges from the figure. The dispersion in entry costs across-occupations decreases with age as entrepreneurial entry costs decline, but salaried entry costs increase. This phenomenon may result from older individuals finding it harder to start careers in paid employment due to difficulties in obtaining or regaining skills at old ages.

[53]These results are in line with those in Dillon and Stanton (2016), who account for dynamics highlighting the option value of the entrepreneurial choice.

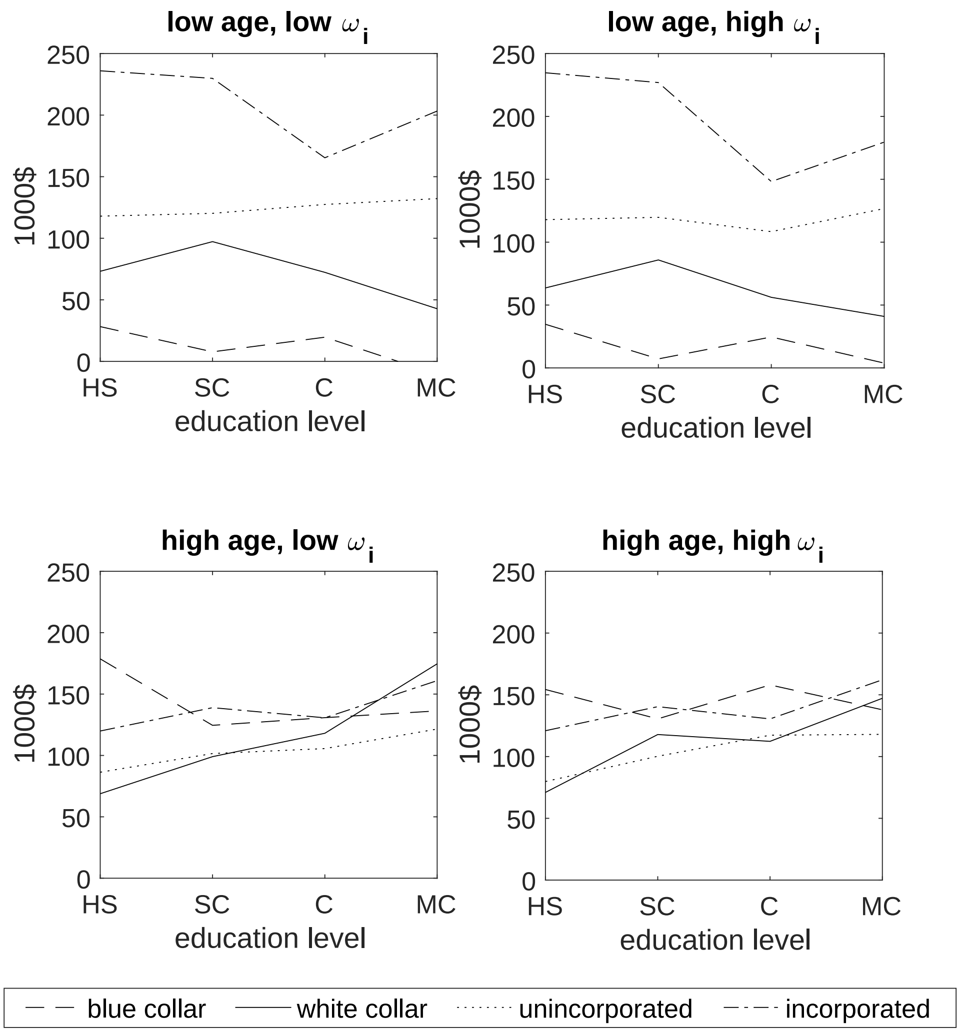


Figure 2.9: Monetary Equivalent of Entry Costs

Notes: On the x-axis of each panel is the education level. Education levels are: high school (HS), some college (SC), college (C), and more than college (MC). On the y-axis is the monetary equivalent of entry costs, obtained using the estimates in Table 2.9 and equation (A.31) in Appendix A.4. The top two panels correspond to individuals 20 years old, while the bottom two correspond to individuals 40 years old. The left panels correspond to individuals in the 10th percentile of permanent wealth ω_i, while the right panels correspond to individuals in the 90th percentile.

level. Hence, it is possible that the importance of non-pecuniary benefits explaining the entrepreneurial choice has an education gradient not explored in previous literature.

.

Finally, even though the treatment of the entrepreneurial choice in this chapter attempts

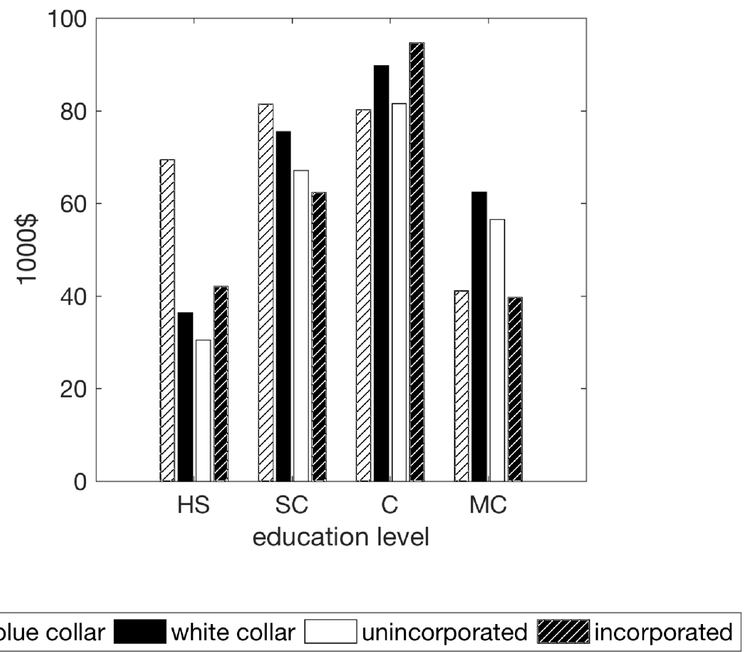


Figure 2.10: Monetary Equivalent of non-pecuniary Benefits

Notes: On the x-axis are levels education: high school (HS), some college (SC), college (C), and more than college (MC). Occupations are: white collar (wc), blue collar (bc), unincorporated entrepreneurship (ue), and incorporated entrepreneurship (ie). On the y-axis is the monetary equivalent of the non-pecuniary benefits not related to entry. They are obtained using the estimates in Table 2.9 and a similar derivation as in equation (A.31) in Appendix A.4. The benefits are computed for a white, married man.

to capture many of its economic determinants, there are at least two caveats of this analysis. First, credit constraints are not modeled explicitly. Data limitations as well as model tractability kept this project away from incorporating the role of savings in entrepreneurship. Credit constraints remain as a common explanation for the lack of higher entrepreneurial participation (Evans and Jovanovic, 1989; Hurst and Lusardi, 2004; Buera, 2009). In reduced form, the age profile of entry costs in the model tries to get at this issue. Consistent with the credit constraints hypothesis, results show young individuals face higher entry costs to entrepreneurship. Second, the decision of hours worked is not modeled. Hence, the non-pecuniary benefits from working in a given occupation are net of the disutility from working. Given that entrepreneurs work more hours, it is possible that accounting for the disutility from working would yield non-pecuniary benefits that are higher for entrepreneurial occupations, at least for the highly educated.

2.5.3 Model Fit

Goodness of fit is assessed by solving the model and comparing simulated data against the sample.[54] Although white collar and blue collar participation rates are predicted less precisely, Figure 2.11 shows that the model successfully generates the incorporated and unincorporated participation rates over the life cycle.[55] Additionally, first-entry statistics presented in Table 2.10 show that the model replicates the absence of young entrepreneurs. The model captures well the proportion of individuals who attempt entrepreneurial occupations by age 40, and it captures reasonably well the average age at first entry into all occupations. More interestingly, the model captures the nature of the experience obtained before first entry. Consistent with the similarities between white collar work and incorporated entrepreneurship highlighted in Section 2.2, simulated individuals attempting entrepreneurship tend to have more prior experience. In particular, first-time unincorporated entrepreneurs tend to have more blue collar experience, whereas the opposite is true for first-time incorporated entrepreneurs. Further measures of model fit in terms of transitions, spells, and realized income are presented in Appendix A.4. .

2.5.4 Decomposition Exercises

In this section, the effects of the economic forces at play are quantified by comparing simulated data from the estimated model (baseline) against simulated data from a number of counterfactual regimes that disable parts of the structure. The quantification is done in four dimensions: entry, timing, ability, and present value of income. Given that incorporated

[54] Appendix A.4 illustrates how the model is solved.

[55] In a separate exercise, not shown here for space considerations, choice rates are simulated at any age t taking the state observed in the data as given. This exercise shows a much better fit for salaried occupations suggesting that the model has a harder time capturing transitions.

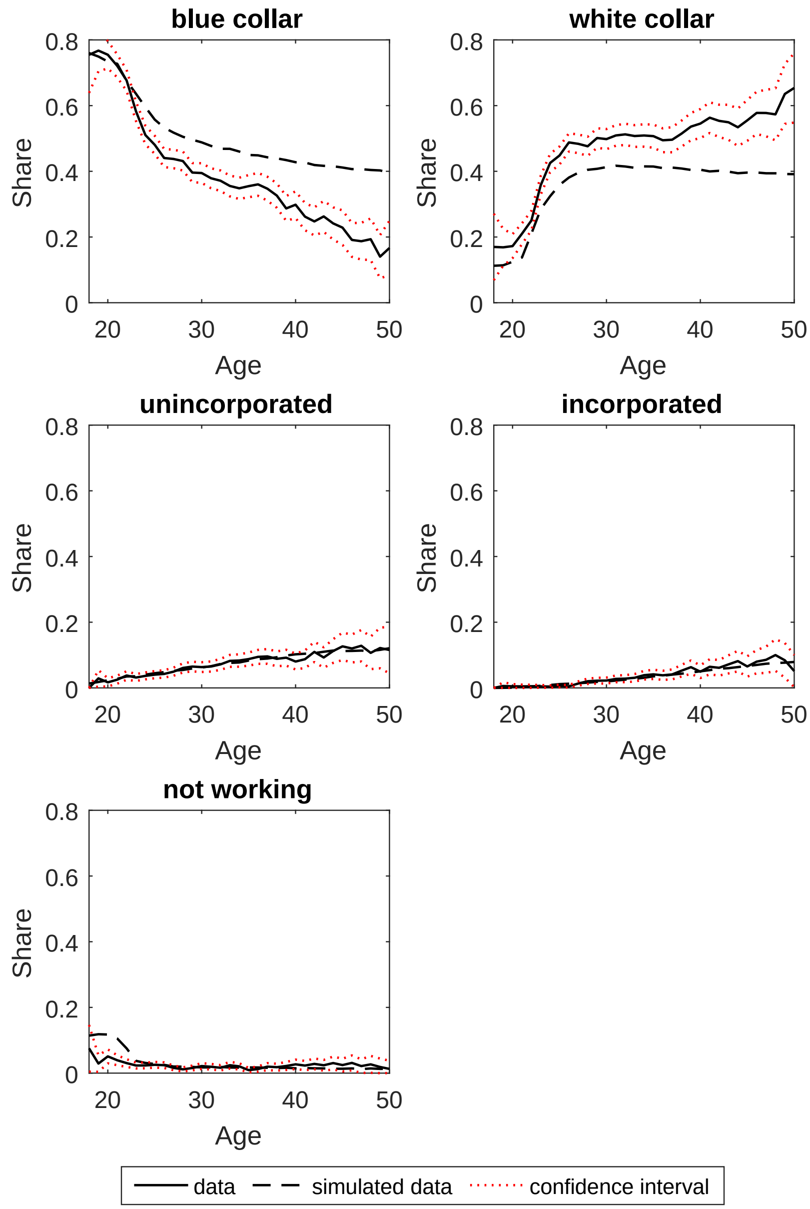


Figure 2.11: Simulated versus Observed Choice Rates

Notes: Actual and simulated choices by age.

entrepreneurs seem most comparable to what is commonly considered as "the entrepreneur" (see Section 2.2 as well as Levine and Rubinstein (Forthcoming)), the discussion will include both types but will center on incorporated entrepreneurs. Comparisons reveal that the two main barriers to young entrepreneurship are entry costs and lack of information.

Table 2.10: First Entry: Observed and Simulated

Data

	blue collar	white collar	unincorporated	incorporated
Tried by age 40	0.65	0.83	0.23	0.11
At first entry				
Age	22.84	24.81	29.57	32.82
exp_{bc}	-	1.99	3.48	2.05
exp_{wc}	1.08	-	3.37	6.58
exp_{eu}	0.03	0.15	-	0.96
exp_{ei}	0.01	0.02	0.22	-

Model

	blue collar	white collar	unincorporated	incorporated
Tried by age 40	0.77	0.73	0.23	0.09
At first entry				
Age	22.23	25.84	29.98	32.40
exp_{bc}	-	3.32	4.51	3.73
exp_{wc}	0.58	-	3.17	5.53
exp_{eu}	0.08	0.25	-	0.48
exp_{ei}	0.00	0.04	0.11	-

Notes: Statistics computed using individuals that are observed from the beginning of their careers until at least age 40. Only data when individuals are 40 years old or below are used.

Moreover, information frictions have a large long-term effect: fully-informed incorporated entrepreneurs have a present value of income (PVI) that is about 50% higher than in the baseline. Additionally, the decomposition points to a considerable long-term effect from using paid-employment outcomes to predict incorporated entrepreneurial success.

The counterfactual regimes used for decomposition are:[56]

- *No leaning-by-doing*: productivity does not increase with own or cross-occupation experience. Instead, occupational skill is constant and pays an average return.[57]

- *No learning about ability*: individuals know their ability vector $\mathcal{M}_i$ but the initial level

[56]The counterfactual regimes are further explained in Appendix A.4. The initial state is fixed across counterfactual regimes. Extended results from these comparisons are also presented in Appendix A.4.

[57]Let $R_k(x)$ be the return to experience in occupation k for somebody who has worked x years in occupation k and zero years in any other occupation (see Figure 2.4). The fixed hourly return to observed skill for

of uncertainty remains unchanged.[58]

- *No X (cross-occupation) leaning-by-doing*: productivity in one occupation is invariant to experience in another. The cross-occupation returns to experience (see Figure 2.5) are set to zero.

- *No X (correlated) learning about ability*: individuals believe that their success in one occupation is uninformative of their ability in another. Their initial prior variance is $\mathbf{\Delta}_s$ diagonalized.

- *No uncertainty*: individuals know their ability vector $\mathcal{M}_i$ and there is no idiosyncratic variation around their hourly income.

- *Uniform entry cost*: entry cost does not vary with age. Instead, individuals pay the cost faced by a 35-year old individual with their same education level.

Why are there not more entrepreneurs? Having to climb the productivity ladder has the strongest effect discouraging incorporated entry. Close behind are the effects of uncertainty (in ability and idiosyncratic variation) and entry costs. Figure 2.12 displays the ratio of the share of individuals who attempt entrepreneurship during their careers in each of the counterfactuals relative to the baseline. It shows that shutting down learning-by-doing has the strongest effect on incorporated entry: if people did not have to learn-by-doing their way up through the steep incorporated productivity ladder (see Figure 2.4), they would be almost twice as likely to attempt incorporated entrepreneurship. In isolation, shutting down learning about ability (i.e. sorting on ability) only increases the rate of incorporated

individuals in occupation k under this counterfactual regime is

$$\bar{y}_k = \sum_{x=0}^{20} R_k(x)$$

[58]In terms of equation (2.4), this amounts to changing the value of the idiosyncratic income variance in occupation k from just σ_{η_k} to $\sigma_{\eta_k} + \mathbf{\Delta}_{s,\{k,k\}}$.

entrepreneurship by about 35% over the baseline.[59] However, once all uncertainty is eliminated, the effect on entry is similar to the effect from flattening the entry cost: individuals would be 75% more likely to attempt incorporated entrepreneurship.[60] Shutting down either type of cross-occupation learning decreases incorporated participation. This latter result makes sense because white collar workers, who are set to gain the most from switching, cannot increase their productivity or improve their beliefs using their experience and paid-employment success. .

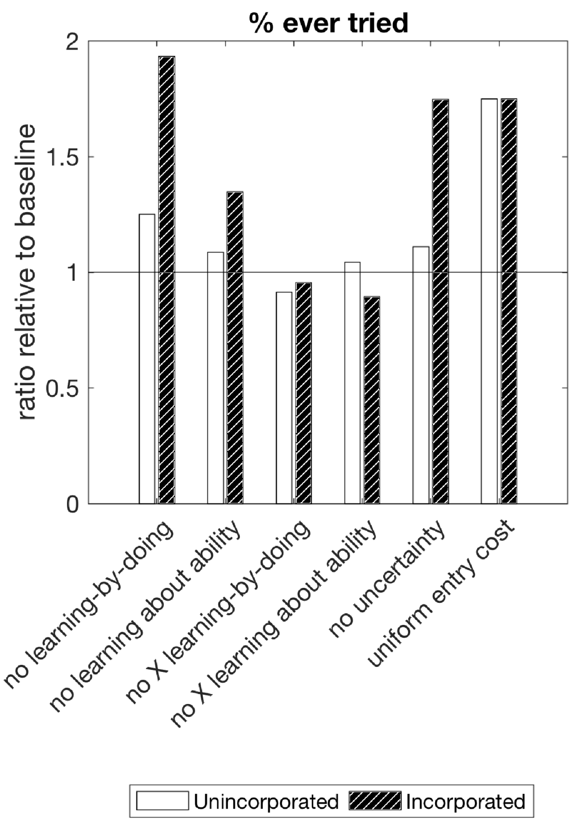


Figure 2.12: Decomposition: Entry

What about young entrepreneurship? The two main barriers to young entrepreneurs

[59]Arcidiacono et al. (2016) study the role of information frictions in a correlated learning framework with schooling and labor market participation decisions.

[60]Given the results in Table 2.9, flattening the entry costs with respect to age amounts to young people facing a lower entry cost than in the baseline.

are entry costs and lack of information. Table 2.4, introduced in Section 2.2, showed the gap in first-entry ages between entrepreneurial occupations and salaried occupations. Figure 2.13 shows the percentage of the gap, relative to white collar, that is closed under each of the counterfactual regimes. Flattening the entry costs closes 70% of the gap, and sorting on ability closes 20%. If individuals knew their ability, more of them would enter early. Risk aversion and correlated learning do indeed induce young individuals to bypass some of the risk of attempting entrepreneurship by acquiring some paid-employment experience first. Figure 2.13 shows that eliminating all uncertainty reduces the gap by an extra 5% on top of the reduction attained from providing full information about ability. Interestingly, eliminating correlated learning on its own widens the gap by about 3%. In the uncorrelated learning case, this happens because young individuals who avoid starting their careers as entrepreneurs must first find out that they really are not good at paid employment in order to switch into entrepreneurship. .

However, individual participation is not all that matters. The ability of those attempting entrepreneurship can have strong effects for the economy as a whole. Figure 2.14 displays the ratio of ability at first entry in the counterfactual regime relative to the baseline. Not surprisingly, the ability of fully informed individuals entering incorporated entrepreneurship for the first time is higher—about three times as large as in the baseline. Perhaps more surprising is the fact that the ability of fully informed individuals entering unincorporated entrepreneurship is twelve times as large as in the baseline. This suggests that under full information the returns from unincorporated entrepreneurship are relatively less attractive. Hence, individuals choosing to enter must be of very high unincorporated ability. Shutting down learning-by-doing and flattening entry costs reduce incorporated ability at first entry relative to the baseline because they ease the threshold to enter. Interestingly, shutting down correlated learning not only reduces ability at first entry but makes it negative. The reason for this is the mismatch between prior variance and population variance. In the

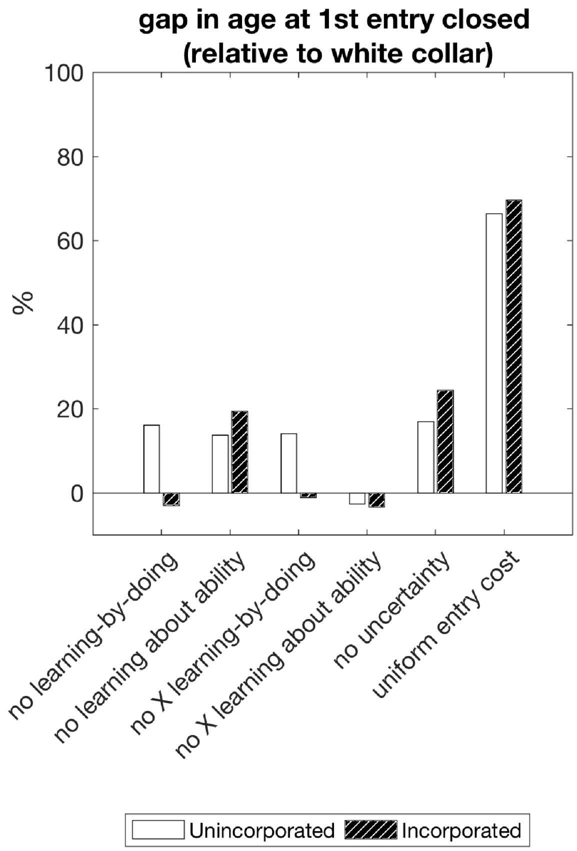


Figure 2.13: Decomposition: Timing

counterfactual regime, although individuals only learn about ability if they attempt the occupation, the true ability distributions are still characterized by the covariance matrices in Table 2.7. Provided abilities are mostly positively correlated (see Figure 2.6), this means that those who switch to entrepreneurship are likely to be low ability entrepreneurs. After all, they switch after discovering their low paid-employment ability. .

Finally, the economic forces in the model have long-term effects on the outcomes of entrepreneurs. Figure 2.15 shows the present value of the entrepreneurs' realized stream of income relative to the baseline. Results indicate that the effect of the information frictions is large: fully informed incorporated entrepreneurs have a present value of income (PVI) that is about 50% higher than in the baseline. Flattening entry costs also increases the PVI of the

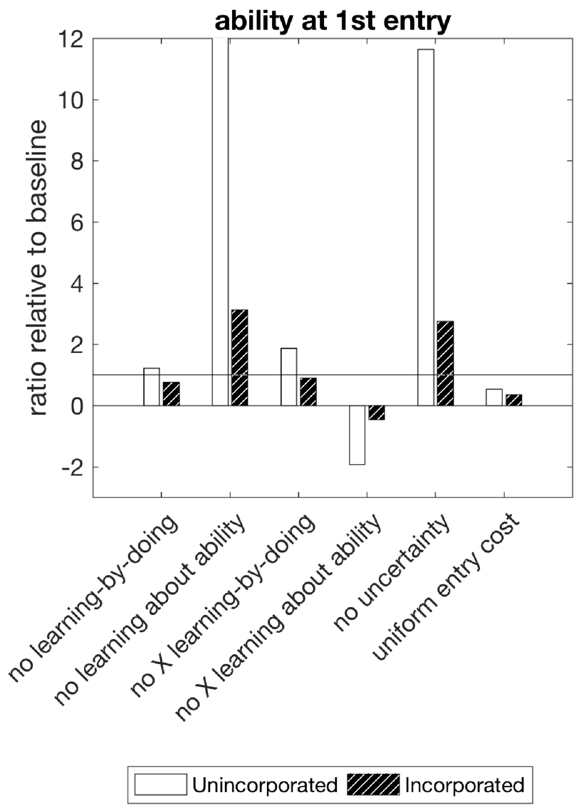


Figure 2.14: Decomposition: Ability

incorporated, although only by about 10%, because successful young entrepreneurs will enjoy the returns for longer. Shutting down learning-by-doing or shutting down cross-occupation learning-by-doing both reduce the PVI by about 20%. Since the former is an extension of the latter, what really decreases the incorporated PVI is the lack of transferability of skills learned. Notably, shutting down correlated learning decreases the incorporated PVI by about 25%. The long-term effect of using paid-employment outcomes to predict incorporated entrepreneurial success is certainly not negligible.[61] .

[61]Simulations in Appendix A.4 show that a model with uncorrelated learning does not reproduce the trends observed in Figure 2.2. The baseline model with correlated learning does reproduce the trends.

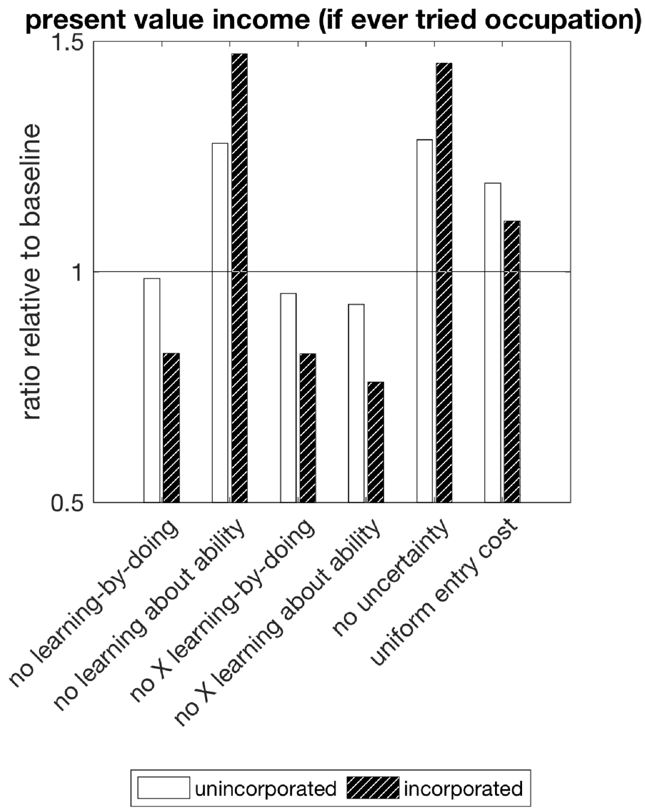


Figure 2.15: Decomposition: PVI

2.6 Policy Counterfactuals

According to the decomposition exercises in Section 2.5, the two main barriers to young entrepreneurship are entry costs and information frictions. In this section, policies focusing on incorporated entrepreneurship target these barriers. This section extends the literature by providing a mapping from entrepreneurial education that shifts beliefs, into career choices and long-term outcomes. Results suggest that a blanket subsidy increases entrepreneurship but has a small long-term effect as measured by the PVI. Additionally, results show that entrepreneurial education that provides information can have a sizable effect on participation and present value of income flows, even for low information quality.

2.6.1 Subsidies

The previous section showed that entry costs are a strong barrier to young entrepreneurship. Consequently, the intervention considered here is a subsidy for young incorporated entrepreneurs. The policy consists of giving either $25,000 or $50,000 to any individual who decides to start his career as a young incorporated entrepreneur. This is, the subsidy is granted only if the individual becomes an entrepreneur immediately after finishing his education.[62] Table 2.11 summarizes the effects of the intervention in terms of young incorporated entrepreneurs and in terms of the overall pool of incorporated entrepreneurs (long-term effects).

The results reflect the effects of lowering the threshold of entry when information is imperfect. There is more participation but less entrepreneurial quality. The $50,000 subsidy more than doubles young incorporated entrepreneurship as measured by the number of individuals who attempt entrepreneurship during their first five years in the labor market. However, the average ability at first entry decreases by about 60%.

Regardless of this adverse effect on the quality of young incorporated entrepreneurs, the subsidy may be justified. For instance, the subsidy may attract high-ability individuals who were not entering due to high entry costs associated to their age or lower permanent wealth. However, as measured by the 95th percentile of the ability of young entrepreneurs, this is not the case. The subsidy could also have long-term effects. As a consequence of the $50,000 subsidy, the number of people who attempt incorporated entrepreneurship in their careers increases by 50%. Besides, the average net present value of income of all individuals in the economy (those who experiment with entrepreneurship and those who do not) increases by about 3%. These results suggest that the marginal individuals induced to experiment by the subsidy turn out to be more productive as entrepreneurs than what they would have been

[62] One such example of interventions providing funds for young entrepreneurs is the Thiel Fellowship (`http://thielfellowship.org`). However, this intervention relies on a tournament that may reveal information about the participant's quality.

Table 2.11: Young Incorporated Entrepreneurship Subsidy

	Subsidy in $1000s		
	0	25	50
Young Entrepreneurs			
Tried in first 5 years	0.02	0.03	0.05
Mean belief ($ per hour) at 1st entry	3.7	2.5	1.2
Mean ability ($ per hour) at 1st entry	5.0	3.6	2.1
Bias (belief-abiliity)	-1.3	-1.1	-0.9
95th pctile ability ($ per hour) at 1st entry	46.5	44.0	37.8
All entrepreneurs			
Tried	0.15	0.17	0.22
Participation rate at age 40	0.04	0.05	0.06
PVI net of subsidy ($1000s)	757	770	770
Mean belief ($ per hour) at 1st entry	6.4	6.2	5.2
Mean ability ($ per hour) at 1st entry	5.5	5.4	4.6
Bias (belief-abiliity)	0.9	0.8	0.6
All individuals			
PVI net of subsidy ($1000s)	508	513	526

Notes: Subsidy given only to individuals who decide to attempt incorporated entrepreneurship immediately after finishing their education. Young entrepreneurs are those who tried incorporated entrepreneurship for the first time within their first five years in the labor market. *Rows*: Summary statistics are computed separately for young incorporated entrepreneurs and for all incorporated entrepreneurs. PVI stands for the present value of income. *Columns*: The no-subsidy column corresponds to the baseline model.

as paid employees.

Looking at the average entrepreneurial ability of new entrants, the results here seem consistent with those in Hamilton et al. (2016) and underscore the arguments presented in Shane (2009) against blindly subsidizing entrepreneurship. However, once the long-term effects are evaluated, it appears that policies that relax entry costs and attract marginal entrepreneurs may be effective. Although small, there are gains in terms of PVI from young successful entrepreneurs attracted early on by the subsidy.

2.6.2 Entrepreneurship Education

Many policies trying to foster young entrepreneurship focus on entrepreneurship education.[63] To the extent that these policies help reveal entrepreneurial potential, the emphasis on entrepreneurial education is consistent with the decomposition exercises in Section 2.5 showing that information frictions are a main barrier to young entrepreneurs. The empirical literature has provided evidence that entrepreneurship education programs can shift individuals' elicited beliefs and intentions (von Graevenitz et al., 2010; Oosterbeek et al., 2010; Souitaris et al., 2007; Peterman and Kennedy, 2003)), but the value of these policies will critically depend on the quality of the information they provide. Results here extend the literature providing a mapping from entrepreneurship education of a given quality, through shifts in beliefs, into career choices and long-term outcomes. In other words, the chapter provides a dynamic framework to assess the value of any given policy that provides information.[64]

The counterfactuals operate as follows: all individuals draw noisy information—regarding their ability as incorporated entrepreneurs—from their outcomes in an entrepreneurship education program. Individuals use this information to update their beliefs before beginning their careers. This policy effectively induces initial heterogeneity in entrepreneurial beliefs that will depend on ability as well as on luck. Additionally, because abilities are correlated, the policy will induce heterogeneity in beliefs across all occupations. In the language of the model, the entrepreneurship education program yields every individual a signal about his incorporated ability ($\mu_{4,i} \in \mathcal{M}_i$) given by

$$\zeta_i^p = \mu_{4,i} + \nu_i \tag{2.18}$$

[63]Examples of such programs are the BizCamps or the Regional Young Entrepreneurship Challenge by the Network for Teaching Entrepreneurship and the Junior Achievement Young Enterprise Student Mini-Company (SMC) program.

[64]The chapter remains silent as to what the information quality of entrepreneurship education is. Evaluation of the information quality of a specific policy could be attained from observing the information signal generated through the program (for instance, rankings, scores, simulated profits) to randomly selected individuals.

where ν_i are iid $N(0, \sigma_\nu^2)$. Individuals use the information contained in ζ_i^p to update their beliefs before entering the labor market. It is assumed that no entrepreneurship education program can provide better information than actually becoming an entrepreneur in the job market for one period. In other words, the noise variance from this intervention is bounded below by the idiosyncratic variance ($\sigma_{\eta 4}^2$) estimated in Section 2.5 (see Table 2.8). Therefore, the noise variance from the entrepreneurship education program can be written as

$$\sigma_\nu^2 = s \cdot \sigma_{\eta 4}^2, \quad \text{with } s \geq 1 \tag{2.19}$$

Four counterfactuals are considered here that differ in terms of how noisy their signals are: $s \in \{1, 2, 5, 10\}$. For instance, when s equals 10 the quality of information from the entrepreneurship education program is 10% the quality of information from actually becoming an incorporated entrepreneur for one period. Results from these interventions are presented in Table 2.12.

For the range of information quality considered ($s \in \{1, 2, 5, 10\}$), results suggest that the lower the quality, the higher the percentage of young entrepreneurs and the bias in their beliefs. Table 2.12 also shows a decline in average ability of young entrepreneurs for information quality below 50% ($s = 2$). This reflects the number of young entrepreneurs who are attracted by lucky signals in programs with lower information quality. Providing noisy information magnifies the role of overestimation of ability in fostering experimentation. Young incorporated entrepreneurs go from having an hourly-income negative bias of -\$1.3/hr in the baseline to a positive bias of \$81/hr from entrepreneurial education providing 10% information quality ($s = 10$). These results are consistent with previous literature suggesting that overconfidence influences entrepreneurial entry (Roll, 1986; Camerer and Lovallo, 1999). However, in the framework of this chapter, overestimation of ability is not a different psychological trait of entrepreneurs or the result of differential analysis of the information

Table 2.12: Young Incorporated Entrepreneurship Education

	Noise Variance Scale, s				
	inf	10	5	2	1
Young Entrepreneurs					
Tried in first 5 years	0.02	0.12	0.11	0.10	0.08
Mean belief ($ per hour) at 1st entry	3.7	82.4	75.1	61.3	48.4
Mean ability ($ per hour) at 1st entry	5.0	1.3	2.1	5.5	10.7
Bias (belief-abiliity)	-1.3	81.1	73.0	55.8	37.7
95th pctile ability ($ per hour) at 1st entry	46.5	47.0	47.5	49.9	59.1
Overall					
Tried	0.15	0.26	0.26	0.25	0.24
Participation rate at age 40	0.04	0.12	0.12	0.12	0.11
PVI ($1000s)	757	944	941	961	983
Mean belief ($ per hour) at 1st entry	6.4	45.9	41.8	33.8	26.5
Mean ability ($ per hour) at 1st entry	5.5	2.0	2.3	4.0	6.6
Bias (belief-abiliity)	0.9	43.9	39.5	29.8	19.9
All individuals					
PVI ($1000s)	508	573	575	578	581

Notes: Individual-specific signal about incorporated ability given to everybody immediately after finishing their education. Interventions are characterized by the noise variance of their signals (σ_ν), expressed in terms of the noise variance of trying incorporated entrepreneurship in reality (σ_{η_4} in Table 2.8): $\sigma_\nu = s \cdot \sigma_{\eta_4}$. Young entrepreneurs are those who tried incorporated entrepreneurship for the first time within their first five years in the labor market. *Rows*: Summary statistics are computed separately for young incorporated entrepreneurs and for all incorporated entrepreneurs. PVI stands for the present value of income. *Columns*: The *inf* column corresponds to the baseline model where no signal is received. This column is identical to the no-subsidy column in Table 2.11.

received (March and Shapira, 1987; Busenitz, 1999). Instead, overestimation at first entry emerges endogenously from uninformed rational individuals who are fortunate to receive large positive signals.

Entrepreneurial education can also provide long-term gains even when the quality of information is as low as 10%. Table 2.12 shows that the share of incorporated entrepreneurship at age 40 triples, the percentage of individuals who attempt incorporated entrepreneurship increases by about 70%, and the PVI of incorporated entrepreneurs increases by 25%. Additionally, entrepreneurial education could benefit all individuals, not only those who eventually become entrepreneurs. A first approximation to the benefits from entrepreneurial education to the average individual is given by the difference in average PVI relative to the baseline. According to the last row in Table 2.12, this number varies from \$65,000 when $s = 10$ to \$73,000 when $s = 1$.[65]

The results in this section extend the literature providing a mapping from movements in beliefs, generated by entrepreneurial education of a given quality, into career choices and long-term outcomes. Nevertheless, caution must be taken when reading these results. The information quality of any specific policy may be different, and its cost may well exceed the additional income it generates. At the limit, as s goes to infinity, entrepreneurship education loses all its information value.[66]

2.7 Conclusion

On the basis of their potential economic benefits, entrepreneurship, and in particular young entrepreneurship, is the target of many policy interventions. These interventions are a response to the fact that most individuals do not start a business and, if they do, tend to do so well into their thirties after they have accumulated some salaried experience. While interventions aiming to foster young entrepreneurship spread, their effectiveness is still an open question.

This chapter explores the reasons why individuals attempt entrepreneurship in their careers as well as the reasons explaining the gap in first-entry ages between paid employment and self-employment. The chapter extends the literature by quantifying the relative im-

[65] To reflect the differences in entrepreneurial potential at every education level, these quantities can also be computed by education level. Results imply that the difference in expected PVI from entrepreneurial education of 10% of quality goes from \$700 for high schoolers to about \$200,000 for individuals with more than college education.

[66] A more general characterization of the information value of entrepreneurship education programs could be to write

$$\zeta_i^p = \kappa \cdot \mu_{4,i} + \nu_i$$

for $\kappa \in [0, 1]$. An uninformative entrepreneurial education program could also be one in which entrepreneurial ability is only weakly related to outcomes, i.e. where κ approaches zero.

portance of various determinants of entrepreneurial choice studied separately in previous research, namely, accumulation of experience (learning-by-doing), accumulation of information (learning about ability), risk aversion, and entry costs. In addition to the quantification, three elements are particularly novel to the empirical literature on entrepreneurship: an analysis of the gap in first-entry ages, an assessment of the relative importance of risk aversion in a dynamic setting, and an evaluation of the role of cross-occupation learning between paid employment and entrepreneurship.

Using the structure of the model to quantify effects, a decomposition exercise indicates that learning-by-doing and entry costs have the largest effects preventing individuals from attempting entrepreneurship. Risk aversion and information frictions also play important roles. For instance, shutting down risk aversion increases the percentage of individuals who attempt incorporated entrepreneurship by 40%, and eliminating information frictions increases this number by 35%. Eliminating cross-occupation learning reduces by 10% the percentage of individuals attempting entrepreneurship. Although these effects may seem small, in the long term the effects become stronger: eliminating cross-occupation learning decreases the present value of income of incorporated entrepreneurs by about 25%.

Results also indicate that the main determinants of the gap in first entry ages are entry costs and information frictions. The chapter evaluates the effects of policies that target these barriers. Previous literature has shown that entrepreneurship education can shift beliefs (von Graevenitz et al., 2010; Oosterbeek et al., 2010). This chapter extends the literature by providing a mapping from the information quality of entrepreneurship education into career choices and outcomes. Results show that a blanket subsidy increases young entrepreneurship by lowering the threshold of entry but has a limited long-term effect. Additionally, results indicate that entrepreneurship education can have a large effect on the rate of young entrepreneurship, and on the present value of income for all individuals. Nevertheless, the information content of any specific policy must be assessed separately. It is possible for some policies to be very uninformative, compromising the benefits from implementing them.

Finally, the motivation for many entrepreneurship policies goes beyond a simple desire to attract more entrepreneurs. Many policies seeking to foster entrepreneurship are motivated by the jobs entrepreneurs create. However, beyond attracting new entrants, entrepreneurship policies may or may not affect the decision to hire employees (Fairlie and Miranda, Forthcoming). Therefore, more work is needed to evaluate these policies, taking into account not only the number of new entrepreneurs but also their quality and their propensity to generate jobs. The framework introduced here is the first step towards that goal. Future research could account for the effects of entrepreneurship policies on job creation by extending the model and acquiring data on the number of employees hired.

Chapter 3

Innovation and Diffusion of Medical Treatment

3.1 Introduction

In many product markets, innovation can lead to substantial quality changes from one point in time to the next. Research on product innovation tends to emphasize demand responses and consumer surplus. Going back to Hicks (1932) economists have recognized that market demand not only responds to, but also drives innovation. Sometimes knows as "demand pull", the idea is that firms respond to consumer preferences by shifting resources towards the development of products that meet potential demand (Schmookler, 1966; Scherer, 1982). Demand pull implies a possible externality if the market fails to price the impact of consumer choices on innovation (Jovanovic and MacDonald, 1994; Waldfogel, 2003; Finkelstein, 2004; Goettler and Gordon, 2011). A potential implication is that innovation does not progress as quickly as it would if the externality were priced.

Several features of the market for pharmaceuticals make it an interesting context to study demand pull. First, medical products have two dimensions of quality (efficacy and side effects) and consumer preferences are heterogeneous. Thus, it is not generally meaningful to see one product as strictly better than another, and many differentiated products can coexist in a given market. Moreover, innovations can be better on one dimension and worse along another, a leading example being new medicines that are more effective at curing illness, but have harsher side effects. Second, in medical markets product quality is often uncertain, especially when products are new. Experimentation is therefore common among consumers and helps to drive both the speed and the direction of future innovation (Bolton and Harris, 1999; Dranove et al., 2014). Though experimentation occurs in many markets, a unique feature of the market for pharmaceuticals is that patients often resort to trying possibly dangerous new medical compounds (for example through participation in clinical trials) only when they are sick and lack access to better options. It is therefore the most desperate among consumers who drive medical innovation, which benefits healthier patients along with generations of potential future patients.[1] An implication for medical markets is that correctly pricing the externality may improve both efficiency and equity. The reasoning is that incentivizing experimentation among all consumers (rather than relying solely upon the sickest) not only speeds innovation as in other markets, but also distributes the burden of innovating more evenly across patient groups.

In this paper, we introduce an empirical framework to capture how innovation along multiple dimensions of quality is endogenous to aggregate consumer demand. Our framework centers on estimating a multi-dimensional distribution of innovations, which is then embedded into a structural model of dynamic demand. In the model, forward-looking consumers make choices after forming expectations over potential future innovations. Optimal consumer choices are then aggregated into market shares, which help to drive both the speed and the direction of innovation by determining how new products are drawn from the distribution of

[1] This point is linked to the model of endogenous growth in Romer (1986) where producer innovations may generate profits for potential future producers.

innovations. In this sense, the model takes explicit account of demand externalities, which arise since the aggregate behavior of atomistic consumers affects dynamic payoffs through its impact on innovation. We match our model to data on the realized path of innovations, product quality and consumer choices over a long time horizon in a maturing product market: HIV drugs.[2]

We highlight three features of our framework, which depart from earlier literature. First, we allow consumers to experiment with products in several ways. Consumers form expectations over the qualities of the products available to them and become fully aware of their qualities only after they have used a product at least once. Alternatively, consumers may experiment with new technologies that are not yet on the market. In software, this is known as beta-testing; in medicine, this is done through participation in clinical trials. Experimental products may be superior to products already on the market, but they may also be of dangerously low quality. Second, consumers are neither fully aware of how the product market will evolve, nor are they fully unaware, in which case technology change amounts to unexpected regime change. Rather, we use the full history of product introductions to estimate a stochastic process of product innovation and assume that agents use this process to form expectations about future innovations. Third, atomistic consumers in our model are fully aware that aggregate demand ultimately drives the path of innovation. In forming expectations, each consumer takes account of this.

We apply our framework to the market for HIV drugs. HIV is a medical condition that reduces the ability of the immune system to fight off routine infections (a condition known as AIDS).[3] It reached epidemic proportions in several countries, including the U.S., starting in 1984. In developed countries, where access to medication is widespread and subsidized, HIV has reached a point where the condition is manageable and side effects of medications

[2]HIV stands for human immunodeficiency virus.
[3]AIDS stands for acquired immunodeficiency syndrome.

are fairly mild. However, this was not always the case. In the early years of the epidemic, available treatments were not only largely ineffective, but also had uncomfortable, painful and even deadly side effects. Each year brought innovations that were incremental at best. Indeed, as we will show, some new products were worse since they were more toxic without being more effective. In the mid-nineties, a new set of treatments (collectively known as HAART) was introduced, which effectively transformed HIV from a virtual death sentence to a chronic condition.[4] Within two years, the introduction of HAART reduced mortality rates by over 80% among HIV+ men (Bhaskaran et al., 2008). HAART therefore marked a clear departure from existing products in the market for HIV treatments. However, HAART involved drugs that were highly toxic, leading to side effects that were often intolerable and drove some people to avoid using them. In other words, HAART comprised treatments that were better on some dimensions, but worse on others. Thereafter, a series of new drugs were introduced, which were effective and had fewer side effects.

We use data on HIV+ men's treatment decisions and health outcomes over approximately 20 years. The benefit of observing a long panel in the market for HIV drugs is that we can watch the path of innovation unfold. Since, we observe the same individuals over time, the evolving market allows us to identify both the stochastic process of innovation and consumer preferences. We exploit observed consumer decisions over time and the realized path of innovation to better understand how expectations were formed. *Ex post*, we observe that the realized path of technological innovation occurred in fits and spurts. The path includes fairly incremental changes to drug qualities along with massive innovations that drastically altered the lives of consumers with HIV. An example of the latter is the introduction of HAART, which constitutes a key source of variation in product characteristics that we exploit to help

[4] HAART stands for highly active anti-retroviral treatment. There is no vaccine or cure for HIV or AIDS, but HAART is the current standard treatment. In general, 1996 is marked as the year when two crucial clinical guidelines that comprise HAART came to be commonly acknowledged. First, protease inhibitors (made widely available towards the end of 1995) would be an effective HIV treatment. Second, several anti-retroviral drugs taken simultaneously could indefinitely delay the onset of AIDS.

identify our model. However, in our framework, large and drastic changes in the product market, such as HAART, are due to less likely draws from the same underlying distribution that generates more likely, smaller, incremental improvements.

We contribute to three separate literatures. The first studies how consumer behavior affects innovation. Schmookler (1966) formalized the idea, calling it "demand pull".[5] Building on this idea, several papers have demonstrated that market size affects the speed of innovation. In a seminal contribution, Goettler and Gordon (2011) show that market structure also drives innovation. They find that in the market for computer processors, the presence of a second firm can slow innovation (since firms do not expect to capture all profits), but that consumer surplus falls in the absence of a competing firm due to monopolistic prices. In another contribution, Finkelstein (2004) shows that policies promoting vaccine use accelerate the development of vaccines. Also in the medical context, Dranove et al. (2014) identify a "social value" of pharmaceutical innovation, showing that Medicare Part D spurred the development of some drugs. A common idea in this literature is that if consumer behavior drives innovation, which benefits other consumers, it follows that a demand externality arises. Waldfogel (2003) uses the term "preference externalities" to describe the mechanism through which market shares can influence products, thus benefitting consumers with similar tastes. He also highlights the individuals with different tastes benefit less.[6] More closely related to us, Bolton and Harris (1999) argue that a free-riding problem emerges if experimenting accelerates innovation. In our context, if clinical trials provide social benefits by spurring innovation, individually rational consumers may choose to participate less than is socially optimal.

[5]Theoretical models of demand-driven innovation include Jovanovic and Rob (1987) and Miller (1988). Models of diffusion of products include Bass (1969), Jovanovic and MacDonald (1994).

[6]Demand externalities have been discussed in a variety of scenarios, including sorting into neighborhoods (Bayer and McMillan, 2012) and the emergence of food deserts (Allcott et al., 2015). In the context of obesity, Bhattacharya and Packalen (2012) provide evidence that individual efforts to prevent obesity can shrink the market size for obesity treatments, which slows technological progress. If so, individuals may over-invest in preventative care compared to the social optimum.

A second literature we contribute to studies dynamic demand under uncertainty. Following Petrin (2004), each product in our model is a bundle of characteristics.[7] Moreover, in our framework, characteristics can have dynamic impacts on consumers (Gowrisankaran and Rysman, 2012). Literature on product choice has considered the idea that consumers are unaware of product characteristics or match value. Erdem and Keane (1996) study the value of experimentation with new products to learn about their qualities. Learning has been incorporated into dynamic models of pharmaceutical demand.[8] Examples are Crawford and Shum (2005) and Chan and Hamilton (2006), where the latter paper explicitly incorporates consumer distaste for side effects. We incorporate learning and uncertainty into our model in several ways. First, and similar to existing work, we model consumers as learning about existing market products that they have never used. Second, consumers can experiment with new products that are not yet widely available by participating in a clinical trial. Third, we depart from existing work on dynamic demand in how we model consumer expectations over the path of innovation. Most papers take the existing set of products as given or exogenous to the model and focus on demand responses to new products. In contrast, we explicitly model how consumers form expectations about future innovations, and allow them take into account that aggregate market shares can shift the direction of innovation.

Methodologically, we build on Hotz and Miller (1993) and Hotz et al. (1994) in using forward simulation to incorporate how individuals form expectations about future innovations.[9] In our context, the choice set that individuals face is non-stationary. We handle this problem by re-defining the current state of technology using a stationary distribution of innovations and a non-stationary reference point or centroid for innovation that emerges endogenously from consumer demand. This is similar to what Goettler and Gordon (2011) do

[7]Studies pioneering the 'characteristics approach' include Stigler (1945), Lancaster (1966) and Rosen (1974).

[8]Empirical models of learning and experimentation also include Miller (1984) and Hincapié (2016).

[9]We also build on Altuğ and Miller (1998) in providing an empirical dynamic model with aggregate shocks.

in their framework when studying microprocessor speed. However, there are some important differences to our setting and thus to our modeling choices. In their setting, product quality is one-dimensional and the innovation distribution is effectively binary (either improving by a fixed amount or not). They also assume consumer homogeneity, which means that the choice set in their context is also effectively limited to upgrading to the best technology or staying with the current one.

While the Goettler and Gordon (2011) model is well-equipped to analyze the market for microprocessors, in our case, we need to account for demand externalities where product quality is multi-dimensional. This means that new product qualities can move in many different directions on a two-dimensional plane. Also, as we show, the empirical distribution of innovations for HIV drugs is not well-approximated as movements with a fixed distance. Finally, we must account for a larger choice set since multiple dimensions of product quality coupled with consumer preference heterogeneity imply that many products can co-exist in a single market. In light of these features of our setting, when computing lifetime utility associated with each choice, we use forward simulation to capture how consumers make decisions after forming expectations about potential future innovations.[10]

The remainder of this paper is organized as follows. Section 3.2 describes the data set we use. In Section 3.3, we specify the structural model and in Section 3.4 we discuss estimation. In Section 3.5, we present parameter estimates and describe model implications for the distribution of innovations. In Section 3.6, we study counterfactual technology paths and the link between consumer choice and innovation. In Section 3.7, we examine the choice externality and consumer welfare. Section 3.8 concludes.

[10]It is important to point out that, unlike Goettler and Gordon (2011), we do not explicitly model firm interaction or dynamic decisions. Therefore, we are unable to conduct policy analysis related to market structure using our framework. An interesting extension of the current paper would be to merge the two approaches by integrating firm decision-making into a model where products have multiple qualities.

3.2 Data

In this section we introduce the data set used in this paper and describe some of the key empirical patterns we use to identify structural parameters. We use the public data set from the Multi-Center AIDS cohort Study (MACS). The MACS is an ongoing longitudinal investigation (beginning in 1984) of HIV infection in men who have sex with men (MSM) conducted at four sites: Baltimore, Chicago, Pittsburgh and Los Angeles.[11] At each semi-annual visit, survey data are collected on HIV+ men's treatment decisions, out-of-pocket treatment expenditures, physical ailments, which can reflect drug side effects, along with sociodemographic information, such as labor supply, income, race, and education.

In addition, blood tests are administered at each visit to objectively measure health status. Our main objective measure of immune system health is *CD4 count*, defined as the number of white blood cells per cubic millimeter of blood. Absent HIV infection, a normal range is between 500 and 1500. For HIV+ individuals, a count below 500 indicates that the immune system has begun to deteriorate due to HIV, but can still fight off infections such that the individual is not symptomatic. When CD4 count drops below about 300, a patient is said to suffer from AIDS.[12] AIDS means that the immune system becomes unable to fight off routine infections and survival probability drops.

---

[11] Data in this manuscript were collected by the Multi-Center AIDS Cohort Study (MACS) with centers (Principal Investigators) at The Johns Hopkins Bloomberg School of Public Health (Joseph B. Margolick, Lisa P. Jacobson), Howard Brown Health Center, Feinberg School of Medicine, Northwestern University, and Cook County Bureau of Health Services (John P. Phair, Steven M. Wolinsky), University of California, Los Angeles (Roger Detels), and University of Pittsburgh (Charles R. Rinaldo). The MACS is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute. UO1-AI-35042, 5-MO1-RR-00052 (GCRC), UO1-AI-35043, UO1-AI-35039, UO1-AI-35040, UO1-AI-35041. Website located at `http://www.statepi.jhsph.edu/macs/macs.html`.

[12] AIDS stands for acquired immunodeficiency syndrome. The CD4 cutoff below which AIDS occurs varies between 200 and 350.

3.2.1 Summary Statistics

The full MACS data set contains information on 6,972 subjects at 49 possible semi-annual visits for a total of 111,271 observations in the form of subject-visit dyads. We limit our attention to HIV+ individuals, leaving us with 47,753 observations. Due to lack of data on gross income and out-of-pocket treatment costs at earlier visits, we drop observations prior to visit 14 (roughly, late 1990) and for robustness in the reporting of survival we also drop observations after visit 47 (about 2008). These sample period restrictions leave us with 29,523 observations and 2,420 individuals. Next, we drop observations where data are missing on at least one of the variables used in subsequent analysis (though we conduct various robustness checks to insure that our results are not driven by these exclusions). After these exclusions, the remaining analytic sample consists of 1,719 unique individuals and 16,851 observations.

Summary statistics by individual are reported in Table 3.1. The first column presents statistics for the analytic sample.[13] 68% of sample subjects are white, 22% are black and about 9% are hispanic. Race variation in our sample is important since previous research has emphasized difficulties in recruiting blacks into clinical trials, which may reflect different costs associated with treatments or variation in expected health outcomes (Harris et al., 1996). About 86% of the sample received some secondary education or more and nearly a quarter (23%) attended graduate school. Consistent with previous research studying medication choice using the MACS data set, there is evidence of substantial variation in labor supply (Papageorge, 2016). 74% of the sample is observed working at least once and 68% of the sample is observed not working at least once.

Underscoring the seriousness of HIV infection, about 40% of the HIV+ subjects we observe at least once over the sample period die prior to the end of the sample period. However, product market innovation led to drastic changes for HIV+ men. The most striking

[13]For comparison, the third column reports statistics for a larger sample of 2,420 individuals, where we have not dropped observations due to missing data on any particular variable.

Table 3.1: Summary Statistics: Subjects. Visit 14-47 (1990-2007)

	Restricted Sample	
Subjects	1719	
	mean	std dev
Black	0.22	
Hispanic	0.09	
White	0.68	
High School	0.14	
Some College	0.29	
College	0.34	
More than College	0.23	
Died	0.40	
Died Conditional	0.20	
Ever Take Market Product	0.83	
Ever Take Trial Product	0.24	
Ever Work	0.74	
Ever Not Work	0.68	
Age in 1991	36.04	8.72

Notes: Standard deviation in square brackets. Data for unique individuals. *Ever Market Product* stands for ever consumed a market product during the period from visit 14 to visit 47. Similar definition holds for *Ever Trial Product. Died Conditional* is the proportion of individuals who died conditional on surviving until year 1995.

example is the introduction of HAART in the mid-1990s, which was much more effective at improving underlying health compared to the treatments that preceded it. Conditional on surviving until the invention of HAART, 20% of subjects are observed dying. This understates the impact of HAART since the sample under study is an aging cohort, i.e., observed survival rates are much higher even when the cohort is older after HAART becomes available. Further, according to Table 3.1, about 83% of subjects are observed using a market product at least once. Moreover, nearly a quarter (24%) opt for early access by participating in a clinical trial at least once during the sample period, suggesting that patients are willing to try experimental products where quality is uncertain.

3.2.2 Consumer Demand

In this section, we study consumer demand in the maturing market for HIV drugs. We emphasize two key patterns in the data. First, consumers are willing to use drugs with side effects when drugs are also effective. Otherwise, they often avoid drugs altogether. Second, consumers participate in clinical trials when they are very sick and when existing technologies are of low quality. Once good technology comes available, willingness to experiment plunges. Together, these patterns in the data support two ideas that underlie our theoretical model developed in Section 3.3. First, product quality in the medical context is multi-dimensional. Second, experimentation is a rational choice to gain access to unavailable and possibly superior technology.

In conducting our preliminary analysis of consumer demand, we pay close attention to comparisons of behavior before and after the introduction of HAART. Since HAART marked a large innovation on earlier treatments, it induced strong and observable consumer responses that help to identify consumer preferences over medications. Summary statistics for subject-visit dyads are found in Table 3.2 for the full analytic sample (column [1]) and then separately for the pre and the post-HAART eras (columns [2] and [3], respectively). We split the sample by HAART era to illustrate substantial changes to choices and outcomes after HAART was introduced.

Perhaps the most striking example of the impact of HAART on consumers is through its effect on survival. In Figure 3.1, we plot the probability of dying between periods t and $t + 1$ conditional on survival until t. Death rates are much higher prior to HAART introduction and despite a multitude of new treatments coming available. After HAART, death rates plunge, and continue to fall until 2007, as smaller innovations occur that make drugs incrementally more effective and less toxic. HAART introduction also affected immune system health, as measured by CD4 count. According to Table 3.2, average CD4 count among

Table 3.2: Summary Statistics: Subjects-Visits. Visits 14-47 (1990-2007)

	Analytic Sample	Pre Haart	Post Haart
Obs	16851	6972	9879
Ailments	0.43	0.45	0.41
Market Product	0.65	0.49	0.76
Trial Product	0.07	0.09	0.05
Work	0.63	0.70	0.58
Age	44.48	40.89	47.01
	[8.03]	[6.99]	[7.75]
CD4	475	407	524
	[297]	[298]	[287]
Gross Income	17567	19036	16531
	[8787]	[8733]	[8677]
Out-of-pocket Pay	266	179	327
	[706]	[598]	[767]

Notes: Standard deviation in square brackets. Income and Out-of-pocket are semestral and measured in real dollars of 2000. Pre HAART era corresponds to visit <= 24 or (roughly) year <= 1995.

HIV+ men in our sample is 407 in the pre-HAART era, rising to 524 in the post-HAART era. In Figure 3.2(a), we plot average CD4 count over time for people on market drugs and no treatment for HIV. Over time, health for people taking no drugs remains fairly constant while health for individuals in a market drug rises.[14]

Given the impact of HAART on health, it is important to understand why many consumers did not use it. In Figure 3.3(a), we plot the proportion of HIV+ consumers using an HIV treatment. Notice that treatment consumption is about 50% in 1990 and actually falls prior to HAART introduction. This reflects that products available on the market are of fairly low quality. Still, if quality were uni-dimensional, even a low quality drug would be better than no drug at all. Moreover, even after HAART is invented, though there is a considerable rise in market product usage, there is a substantial proportion of HIV+ men not using treatment.

Treatment costs are one possible explanation. In Table 3.2, we see that treatment costs

[14]Notice that average age rises and labor supply and income decline after HAART, consistent with the fact that we observe an aging cohort, which is more likely to retire and report lower gross income over time.

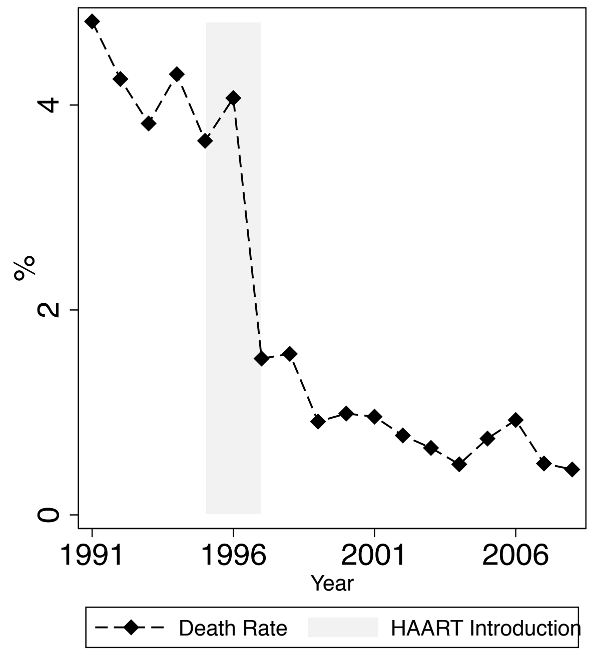
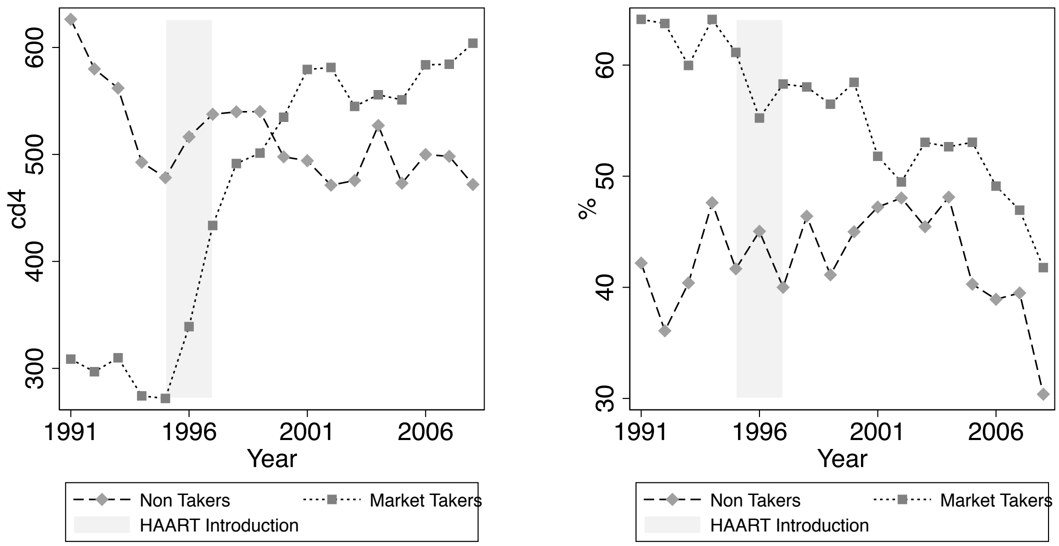


Figure 3.1: Death rate in the sample.

Notes: More than 1500 surveyed individuals died for AIDS-related causes during our analysis period.

rise after HAART introduction, from about \$179 to \$327 for six months of treatment. In other words, even in the post-HAART era, costs are fairly low given that individuals earning on average are about \$37,000 per year. It is worth mentioning, moreover, that non-users of market drugs pay non-zero costs for drugs, perhaps spending more money on medication to fight opportunistic infections. In other words, the incremental out-of-pocket cost of effective HIV treatments does not appear sufficient to explain why some people avoid HIV treatments.

Another possibility is that drug quality is multi-dimensional in which case demand reflects a distaste for another feature of HIV drugs. Given data on physical ailments, we explore the possibility that consumer demand reveals a distaste for side effects. Interestingly, after HAART introduction, the proportion of individuals reporting physical ailments declines only slightly (45% to 41%). The small change reflects the net effect of two countervailing dynamics (Papageorge, 2016). HAART improved health on average, which lowered reported ailments



(a) (b)

Figure 3.2: Health and side effects summary trends over time.

Notes: Panel 3.2(a) shows the mean CD4 over time by consumption status. Panel 3.2(b) contains mean ailments over time by consumption status.

attributable to symptoms of HIV. However, HAART also led to side effects among users, thereby increasing reports of ailments. The increase in side effects also reflects how use of HIV treatment rose with the introduction of HAART, from 45% to 76%. We also plot physical ailments over time in Figure 3.2(b). For non-users of HIV medications, ailments remain fairly steady. For users of HIV medications, ailments drop prior to HAART introduction and then rise after HAART, which is consistent with HAART being a highly effective drug with side effects. However, after 2001, ailments decline for individuals using HIV drugs. This reflects later improvements to medications, which lowered their side effects.

Further evidence in support of the idea two dimensions of quality is market consumption by CD4 count, plotted in Figure 3.3(a). Sicker people are far more willing to take low effective medications despite side effects in the years before HAART. After HAART, notice a striking convergence in the proportion of men using medications, driven largely by healthy individuals going onto medication. Thus, the rise in consumption of HIV treatments after

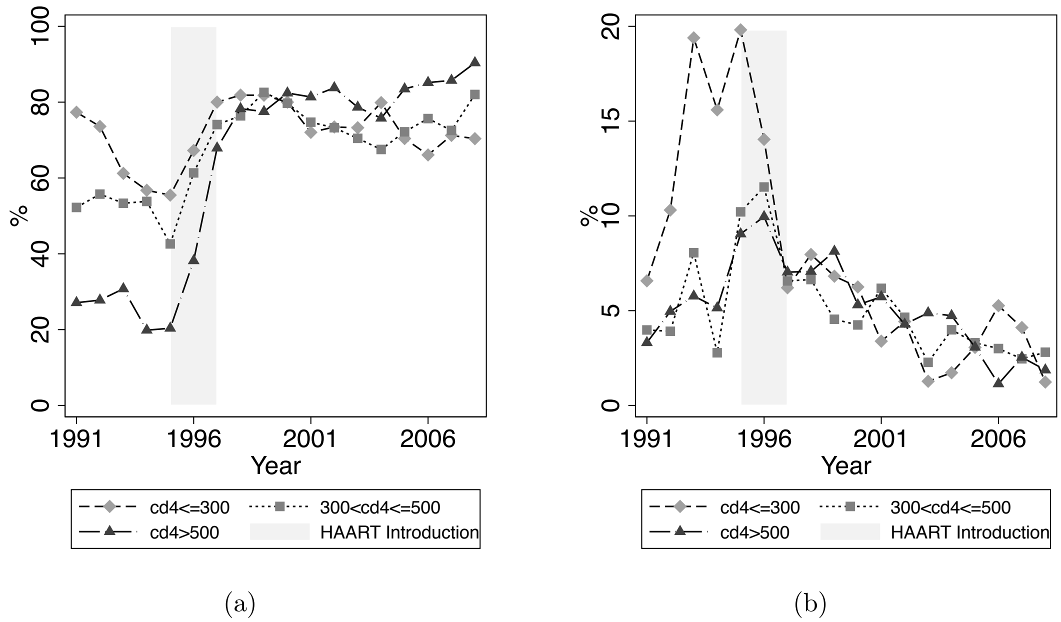


Figure 3.3: Consumer demand over time.

Notes: Panel 3.3(a) shows treatment consumption over time by health status. Panel 3.3(b) shows trial participation over time by health status.

HAART was introduced suggests that patients are more likely to use drugs despite side effects if the utility cost of suffering ailments is offset by expected improvements to health. HAART was more effective than earlier drugs, which encouraged people to use it despite its side effects. This would explain the rapid rise in use of HIV treatments after HAART is introduced since individuals would be more willing to use drugs with side effects as long as drugs are effective at improving underlying health.

Another option for individuals in the product market we study is to join a clinical trial to gain early access to new products. Studying how individuals experiment with new drugs by joining a clinical trial further highlights how consumers respond to innovations in the market for HIV drugs. Trial participation over time and by health status is plotted in Figure 3.3(b). The figure reveals several dynamics. First, early trial participation is driven largely by individuals with low CD4 counts. This suggests that, as individuals become ill, they also become more willing to experiment with new products of uncertain qualities. Second, in the years just prior to HAART introduction, the drugs that comprise HAART, including protease inhibitors, marked a substantial improvement over drugs available on the market. In those years, trial participation gave individuals early access to much better products. This relates to the idea of *beta testing* in markets where some consumers are willing to experiment with new products with high potential quality.

After HAART, notice that trial participation plunges after HAART is introduced as a market option. After HAART, there is a marked convergence by health status in the proportion of patients in trials. This means that once effective drugs are available, it is no longer possible to explain trial participation as an option for people who are very sick and therefore willing to face uncertainty in exchange for early access to a high-quality product. The reason is that individuals no longer need to participate in a clinical trial (and face therefore more uncertainty) to access good drugs. Together, these dynamics suggest that experimentation is a rational choice to gain access to new technology, especially in a maturing market where existing technology is not particularly good and patients are desperate for something better.

3.2.3 Market-Level Innovation

In the previous section, we studied how consumer responses to innovation shed light on consumer preferences. The patterns we have described until now are consistent with the idea that patients value their health, but are also concerned with side effects. Moreover, side effects seem to play a larger role in demand after survival is more or less assured. However, our preliminary evidence also suggests that preferences are not lexicographic. Patients seem willing to use toxic (or experimental) medication if the alternative is a large rise in the probability of dying, but patients will also forgo treatments with harsh side effects if drugs are not effective and the survival gains are limited.

In this section, we consider market-level innovation. To start, we illustrate innovation

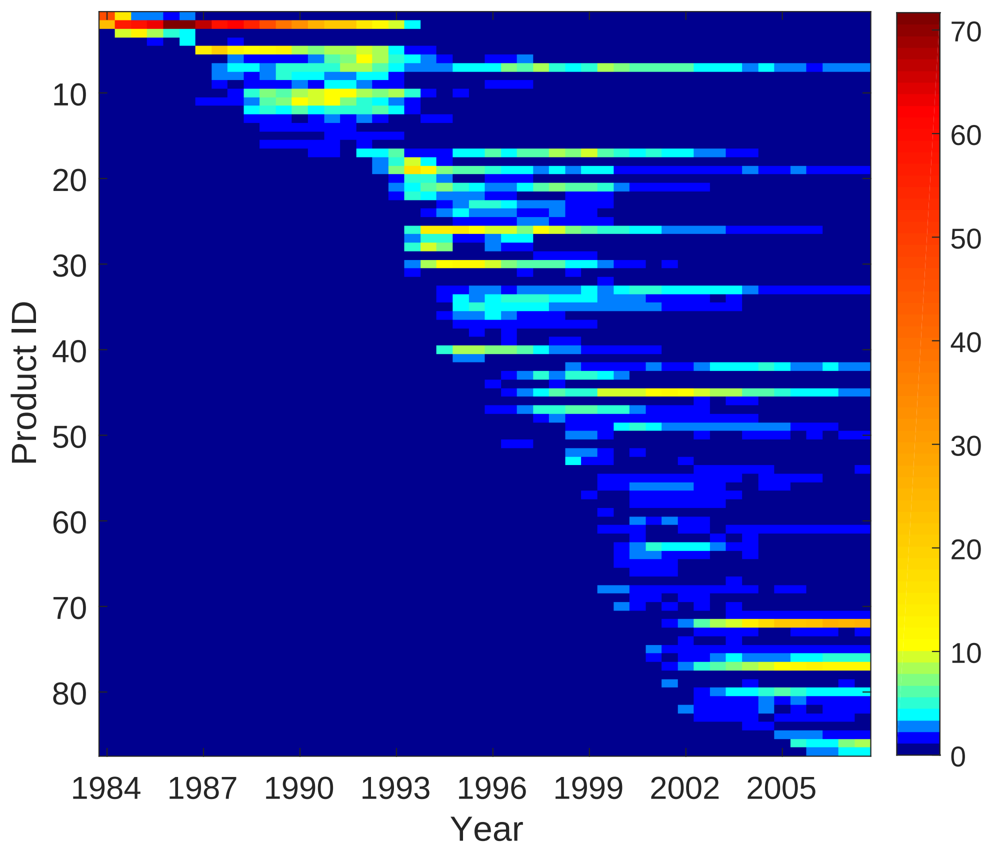


Figure 3.4: Diffusion of Products Over Time

Notes: HIV treatments from 1984 to 2008. Each id—or row—represents a product. Color indicates the share of the market that the product captures. Shares are conditional on consuming a product. Early on there are few products with high shares, as time passes new products strip market share from incumbents and less popular products exit.

and diffusion of new products over time in the market for HIV treatment using a "heat map" displayed in Figure 3.4. For the approximately 90 drugs that were most used, we compute market share over our sample period.[15] Dark blue corresponds to no market share and warmer colors mean higher market shares. In the early years of the epidemic, there are only a few drugs with high market share. Over time, many new drugs emerge, most with lower market share. The heat map captures two important patterns in the data. First, many new

[15]Appendix B.1 contains additional information on individual drugs and treatment combinations. Table B.1 discusses which drugs or combinations are taken in clinical trials. Table B.2 lists the chemical compositions of each drug. Table B.3 shows how drugs are combined into treatments. Table B.4 discusses "core treatments", which are the main sets of treatments we observe, including the individual drugs they are composed of, whether or not they count as HAART and their entry and exit visits.

drugs were invented over time. In other words, the market for HIV drugs was active over our sample period. Second, most old drugs eventually exit and are replaced by new drugs, which means that new drugs marked improvements upon older ones. A striking shift occurs in the mid-1990's, after which point most earlier drugs exit, replaced by new drugs. This corresponds to HAART introduction, when protease inhibitors (PI's) were introduced and became a standard part of HIV treatment. After HAART, moreover, many drugs became obsolete.[16]

Finally, we provide suggestive evidence that the observed innovation path can be seen as a response to consumer preferences. In other words, we search for evidence of the idea that demand drives the direction of innovation. In Figure 3.5, we plot drug qualities (effectiveness and side effects) for different periods of time. The figure illustrates the path of technology over time. Notice how there is a large innovation in the direction of improved health in the mid-1990's. This improvement is the introduction of HAART. Moreover, notice that there is a rightward shift in later years as innovations reduce side effects without offering much of an improvement in efficacy. This rightward shift is important as it corresponds to changes on the relative importance of one dimension of taste over another.

We argued previously that consumer demand patterns, as the market for HIV drugs matured, seem to show a preference for drugs with fewer side effects, especially when survival is less of a concern. The path of innovation seems to have followed shifts in market demand after HAART was introduced. Therefore, preliminary empirical patterns provide support for the idea that innovation responded to consumer demand.

[16] An exception is AZT, which remained a standard component of HAART.

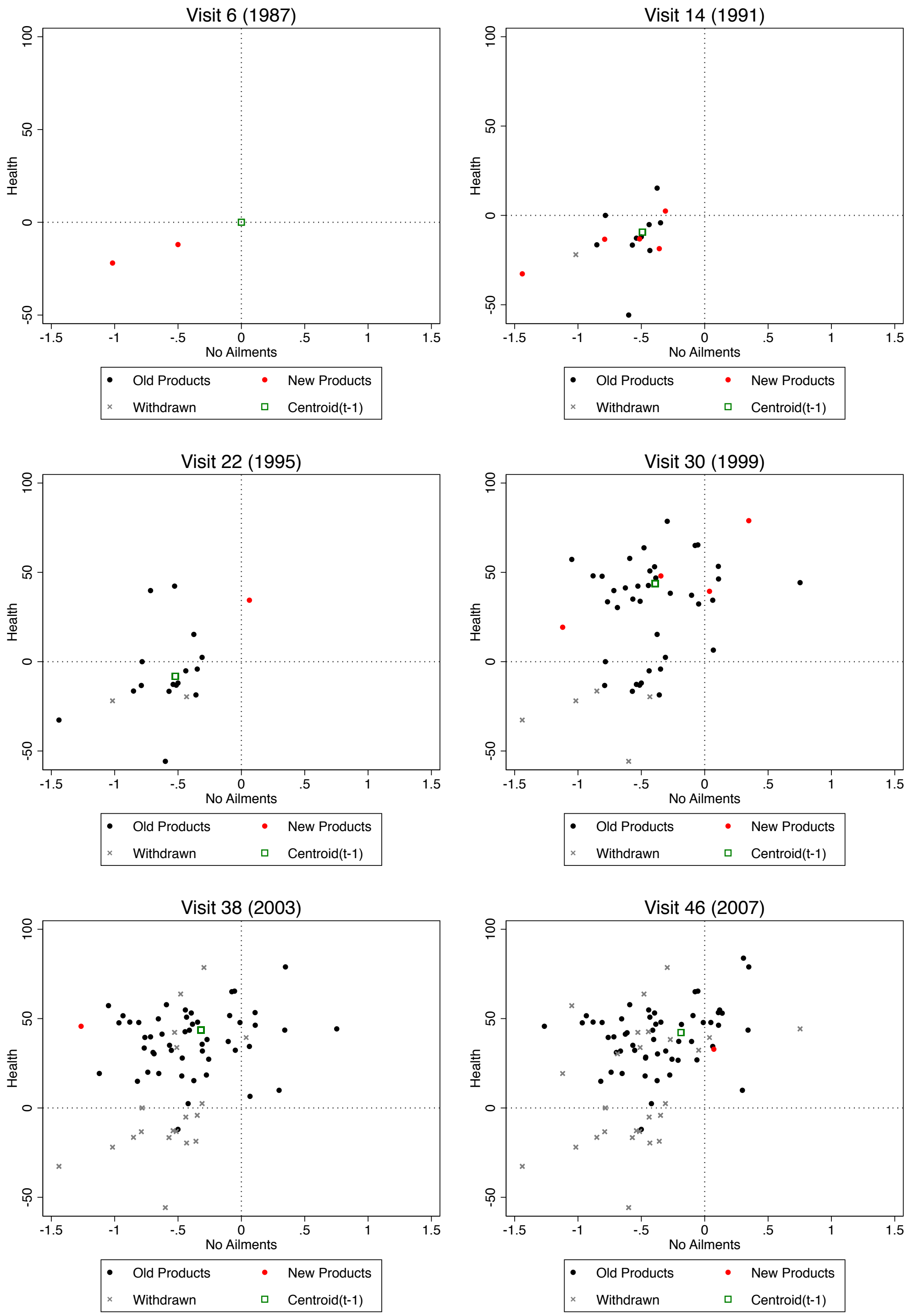


Figure 3.5: Treatment Evolution

Notes: Figure shows snapshots of the evolution of the state of the product market at the different stages. Products are two-dimensional. On the x-axis is a measure of a treatments ability to not cause side effects. On the y-axis is a measure of its contribution to underlying health. Dimensions are measured in different scales. Incumbent products are shown in black. New products are shown in red. Withdrawn products are shown as x. The green square is a measure of the prevalent technology in the previous period.

3.3 Model

This section describes a model of innovation in the market for HIV treatments. Products are multidimensional: they can improve health and increase lifespan, but have potential side effects, which affect survival and labor market outcomes. New products are developed in clinical trials and both the entry of new products and the exit of incumbent products are determined by an endogenous stochastic processes. Consistent with the nature of our data, we are more detailed in our treatment of the demand side. Individuals maximize lifetime utility by choosing an HIV medical treatment. Consumers can choose a product that is available on the market, opt for no treatment at all or experiment with a new treatment by participating in a clinical trial. All treatments on the market cost the same to the consumer. In making decisions, consumers face several sources of uncertainty caused by individual or aggregate choices. First, they are uncertain about current-period outcomes, including their income and side effects. Second, consumers are uncertain about the evolution of other individual-specific state variables, notably health and survival. Finally, they face uncertainty over the evolution of the product market since new treatments may enter the market and some incumbent treatments may drop out.

In describing the model, we begin with a summary of the timing within a period (Section 3.3.1). Second, we discuss the supply of treatment, including entry and exit of new products from the market (Section 3.3.2). Third, we specify consumer demand for treatment, including choice sets, utility and individual state-to-state transitions (Section 3.3.3).

3.3.1 Summary and Timing within a Period

The timing of the model within a period proceeds as follows, where we begin with the aggregate state. In period t, the aggregate state is denoted Ξ_t and consists of current and previous individual product characteristics $\mathbb{P}_t$, previous market shares $\mathbb{S}_t$ and the distribution of current-period consumer characteristics $\mathcal{F}_t$. Together, these factors determine the average state of technology ω_t along with entry (the quantity and qualities of new products) and exit of products that are withdrawn from the market prior to the start of the following period. Current period market shares along with the qualities of new products and the distribution of consumer characteristics constitute the one-period-ahead aggregate state.

Upon entering period t, the consumer observes his state vector $\mathcal{Z}_{it}$, along with his choice set $\mathcal{C}_t$. His state vector consists of individual-level components $z_{it}^{\mathcal{I}}$ (e.g., health) along with aggregate market components $z_{it}^{\mathcal{M}}$ that the consumer uses to forecast the future. When choosing among products, he takes account of how each choice affects current outcomes (e.g., side effects, income) and future states (e.g., health, labor participation). He also forecasts the characteristics of future products, which may affect the relative payoffs to his current choice. The consumer's choice of treatment maximizes his expected discounted lifetime utility.[17] Once a consumer makes a choice, outcome variables are realized and he receives his flow utility y_{jit}. Thereafter, their state variables update and the consumer enters the next period.

3.3.2 Supply

We specify a reduced-form model of supply that we use to capture the evolution of product characteristics. We do not model firm behavior, strategic pricing and R&D decisions.[18] Entry and exit occur at the end of the period immediately before the next period begins. We start by describing the aggregate state at the beginning of period t, followed by entry and exit.

[17]Given the large number of individual products in the market, consumers are assumed to have limited information about individual product characteristics. As will be explained below, products are clustered into categories (using a k-means algorithm) and the consumer knows the weighted average of qualities in each cluster W_t, where the weights are the probability of being assigned a specific treatment.

[18]Modeling the supply side in this way obviously limits the sorts of counterfactuals we can perform. For example, our model would be ill-equipped to evaluate policies affecting, for example, market structure.

Supply

$\longrightarrow$

AGG. STATE	EXPERIMENTS	ENTRY	EXIT	UPDATE
$\Xi_t = \{\mathbb{P}_t, \mathbb{S}_t, \mathcal{F}_t\}$		$\theta = \omega_{t-1} + \nu_{rt}$		
$\mathbb{P}_t = \{\mathcal{P}_t, \mathcal{P}_{t-1}, \mathcal{P}_{t-2}\}$		$\nu_{rt} \sim F_\nu$	$Products$	$\Xi_{t+1} = \{\mathbb{P}_{t+1}, \mathbb{S}_{t+1}, \mathcal{F}_{t+1}\}$
$\mathbb{S}_t = \{s_{t-1}, s_{t-2}\}$	$TrialShare_t$	$New_{t+1} \sim F_N$	$Withdrawn_t$	$\mathbb{P}_{t+1} = \{\mathcal{P}_{t+1}, \mathcal{P}_t, \mathcal{P}_{t-1}\}$
$\omega_t = f_1^S(\mathbb{P}_t, \mathbb{S}_t)$	$= f_2^S(\mathbb{P}_t, \mathcal{F}_t)$	$Q_t = f_3^S(\mathbb{P}_t, \mathbb{S}_t)$	$= f_4^S(\mathbb{P}_t, \mathbb{S}_t, \mathcal{F}_t\}$	$\mathbb{S}_{t+1} = \{s_t, s_{t-1}\}$
$\mathcal{Z}_{it} \equiv \langle z_{it}, \varepsilon_{it}\rangle$	$W_t = f_1^D(\mathcal{P}_t)$	$y_{jit} + \epsilon_{jit}$	$\mathcal{C}_{t+1}$	$\mathcal{Z}_{i,t+1} \equiv \langle z_{i,t+1}, \varepsilon_{i,t+1}\rangle$
$z_{it} \equiv \langle z^{\mathcal{I}}_{it}, z^{\mathcal{M}}_t\rangle$	choose $j \in \mathcal{C}_{it}$		W_{t+1}	$z_{i,t+1} \equiv \left\langle z^{\mathcal{I}}_{i,t+1}, z^{\mathcal{M}}_{t+1}\right\rangle$
STATE	CHOICES	UTILITY	FORECAST	TRANSITIONS

$\longrightarrow$

Demand

Figure 3.6: Timing

The Aggregate State

The aggregate state is denoted Ξ_t and summarizes market-level quantities at period t. The state contains current and previous product qualities and market shares along with the distribution of consumer characteristics in the market, each described below.

Product Characteristics ($\mathbb{P}_t$): The aggregate state contains a set $\mathbb{P}_t$ of characteristics of current and previous products up to two periods into the past:

$$\mathbb{P}_t = \{\mathcal{P}_t, \mathcal{P}_{t-1}, \mathcal{P}_{t-2}\}$$

where $\mathcal{P}_t$ denotes the characteristics of products available at t.

Market Shares ($\mathbb{S}_t$): The aggregate state also contains a set of previous shares going back in time two periods:

$$\mathbb{S}_t = \{s_{t-1}, s_{t-2}\}$$

where s_{t-1} denotes a set containing the shares of products available at $t-1$.

Consumer Characteristics ($\mathcal{F}_t$): Finally, the current distribution of consumer characteristics $\mathcal{F}_t$ is also contained in the aggregate state. The initial distribution of consumer characteristics in the test market is denoted $\mathcal{F}_0$.[19] Thus, the aggregate state is given by:

$$\Xi_t = \{\mathbb{P}_t, \mathbb{S}_t, \mathcal{F}_t\}$$

Entry

In each period, entry of new products occurs according to a reference point for innovation or *centroid*, denoted ω_{t-1}, a distribution of characteristics of new products $F_{\theta|\omega_{t-1}}$ and a distribution of number of new products F_N.

Centroid (ω_t): At any period t, the centroid for innovation is a weighted average among users of products available last period, given by:

$$\begin{aligned} \omega_t &= f_1^S(\mathbb{S}_t, \mathbb{P}_t) \\ &= \sum_{k \in \mathcal{P}_{t-1}} s_{kt-1}\theta_k \end{aligned} \tag{3.1}$$

where θ_k are the characteristics of product k and s_{kt-1} is the ratio of individuals who consume treatment k relative to the number of individuals who consume a treatment. If nobody uses a treatment the base for innovation remains the same, i.e. $\omega_t = \omega_{t-1}$.

[19]The taste market is the market based on which innovation is undertaken.

Characteristics of New Products ($F_{\theta|\omega_t}$): Every new product r introduced at t, characterized by θ_{rt}, is an innovation around the centroid of the previous period:

$$\theta_{rt} = \omega_{t-1} + \nu_{rt} \tag{3.2}$$

where ν_{rt} represents a vector of disturbances (innovations) around the current technology, drawn from F_ν, a stationary distribution of innovations. F_ν embodies the outcomes of R&D efforts made by firms and the government. We do not specify a parametric form for F_ν. As will be explained in Section **??**, F_ν is a non-parametric distribution estimated off of the history of observed innovations around the centroid. Clearly, F_ν and ω_{t-1} determine $F_{\theta|\omega_{t-1}}$.

Number of New Products (F_N): The number of new products at any period is equivalent to the number of draws to be taken from $F_{\theta|\omega_{t-1}}$. In the data, we observe that the number of new products introduced in the market varies across time. Moreover, the number of products seems to be related to the size of previous discoveries as well as to the share of individuals joining a trial. We capture these facts in our specification of F_N. At the end of any period, a number New_{t+1} of new products enters the market. This number follows a negative binomial process that permits dispersion in the mean:

$$\begin{aligned} New_{t+1}|_{\mu^*} &\sim Poisson\left(\mu_t^*\right) \\ \mu_{t+1}^* &\sim Gamma\left(1/\alpha^N, \alpha^N \mu_t\right) \\ \mu_{t+1} &= \exp(\beta_0^N + \beta_1^N Q_t + \beta_2^N TrialsShare_t) \\ \ln \alpha^N &= \alpha_0^N + \alpha_1^N Q_t \end{aligned} \tag{3.3}$$

The binomial model is conditioned on two covariates. First, the quality of previous innovations, denoted Q_t, captures the relatively higher number of new products that follow the appearance of better innovations. Second, the share of individuals who consume a trial product, endogenously given by the characteristics of the test market as $TrialsShare_t = f_2^S(\mathbb{P}_t, \mathcal{F}_t)$, captures the fact that more experiments can be conducted if more consumers participate in clinical trials.

The quality of previous innovations measures the distance between the previous period's new products and the previous period's centroid. The relative change is computed for each of the two dimensions of product characteristics (health, h, and lack of ailments, x) and is scaled by the maximum change observed over the sample period.[20] It is computed as follows:

$$\begin{aligned} Q_t &= f_3^S(\mathbb{S}_t, \mathbb{P}_t) \\ &= \sum_{r \in \{h,x\}} \frac{\max_{\theta^r \text{ new at } t} \left\{\theta^r - \omega_{t-1}^r\right\}}{\max_{\theta^r \text{new at } \tau, \forall \tau} \left\{\theta^r - \omega_{\tau-1}^r\right\}} \end{aligned} \tag{3.5}$$

The specifications of ω_t, $F_{\theta|\omega_{t-1}}$, and F_N render the path of innovation endogenous. Individual choices, summarized by market shares, affect the centroid in equation (3.1). By affecting ω_t, market shares affect the characteristics of every new product θ in equation (3.2). Intuitively, treatments that keep patients alive and those associated with fewer ailments will capture larger shares of the market and firms will innovate on drugs with larger market shares. Additionally, individuals' choices affect the path of innovation through their effect on the distribution of number of new products.

Exit

Incumbent drugs may exit the market. Exit happens at two different levels: *exit for switchers* and *overall exit*. *Exit for switchers* happens when the product is no longer available for

[20] Note that in order to compute Q_{t-1} we need the scaling quantities given by

$$\max_{\theta^r \text{ new at } \tau, \forall \tau} \left\{\theta^r - \omega_{\tau-1}^r\right\} \tag{3.4}$$

for $r \in \{h, x\}$ which are estimated consistently by their data counterparts.

individuals who have yet to use it, but is still available for those who were consuming the product in the prior period. *Overall exit* happens when the product is no longer available to any consumer. Exit happens according to the following rules that aim to reconcile empirical observations and theory—where expected shares must be positive due to model assumptions on the taste shocks of consumers.[21]

1. If the ratio of people switching to product k relative to the number of people switching falls bellow $\tilde{\sigma}_1$ during three consecutive periods, the product is withdrawn from the market. $\tilde{\sigma}_1$ is chosen as the minimum conditional share observed in the data and the number 3 is chosen to smooth the market spells of products.

2. If the ratio of people consuming product k, either by staying or switching, relative to the number of people consuming a market product falls below $\tilde{\sigma}_2$ during two consecutive periods, the product is withdrawn from the market. $\tilde{\sigma}_2$ is chosen as the minimum conditional share observed in the data and the number 2 is chosen to smooth the market spells of products.

The exit criteria can be written in terms of the aggregate state of the problem as

$$ProductsWithdrawn_{t+1} = f_4^S(\mathbb{S}_t, \mathbb{P}_t, \mathcal{F}_t)$$

The Evolution of the Aggregate State

Given the current aggregate state Ξ_t and the exogenous distribution of innovations, aggregate choices induce a new distribution of consumer characteristics $\mathcal{F}_{t+1}$. Through the entry and exit mechanisms, a new set of available products comes available and is denoted $\mathcal{P}_{t+1}$, which can be used to form $\mathbb{P}_{t+1}$. Finally, consumer choices can be summarized into market shares $\mathcal{S}_t$, which can be used to form $\mathbb{S}_{t+1}$. Thus, we have all the components of one-period-ahead

[21]These shocks will be discussed in the demand portion of the model.

aggregate state Ξ_{t+1}. We now turn to consumer demand.

3.3.3 Demand

The individual chooses medical treatment to maximize expected discounted lifetime utility. In making decisions, he observes his current state which includes individual-specific variables, such as health, along with market-level variables, such as the current state of medical technology. Individuals use market-level variables to form expectations over the future path of innovation. In specifying the individual's problem, we discuss state variables, the choice set, flow utility and stochastic processes governing outcomes and state-to-state transition probabilities. Next, we specify the value function.

State Variables

The state for individual i at period t is denoted $\mathcal{Z}_{it}$, where

$$\mathcal{Z}_{it} \equiv \langle z_{it}, \varepsilon_{it} \rangle \tag{3.6}$$

z_{it} is a set of state variables that is further sub-divided into a set of individual-specific variables, denoted $z_{it}^{\mathcal{I}}$, and a set of aggregate variables denoted $z_t^{\mathcal{M}}$:

$$z_{it} \equiv \left\langle z_{it}^{\mathcal{I}}, z_t^{\mathcal{M}} \right\rangle \tag{3.7}$$

The individual-specific state variables, $z_{it}^{\mathcal{I}}$, are

$$
\begin{aligned}
b_i &: \text{ a set of race indicators} \\
edu_i &: \text{ a set of education indicators} \\
h_{it-1} \in \mathbb{R}_+ &: \text{ health at the start of period } t \\
a_{it-1} \in \{25, 25.5, \ldots\} &: \text{ age at the start of period } t \\
l_{it} \in \{0, 1\} &: \text{ working during period } t \\
q_{it-1} = \left\{q_{it-1}^x, q_{it-1}^h\right\} \in \mathbb{R}^2 &: \text{ characteristics of product consumed last period} \\
\eta_i &: \text{ person-specific income characteristic}
\end{aligned}
$$

The individual can be either white, black or Hispanic. He belongs to one of four mutually exclusive educational categories: high school, some college, college, and more than college. His health, measured by CD4 count, is a continuous positive number.[22] His age is measured in half-year increments, corresponding to the frequency of MACS data collection. l_{it} indicates whether he will work during period t.

Each HIV treatment has two characteristics: its effectiveness at raising CD4 count, which we denote θ^h, and its propensity to cause side effects, denoted θ^x. We collect these into a vector denoted $\theta \in \mathbb{R}^2$. If the individual consumed a market product in the prior period, the characteristics of that product, denoted q_{it-1}, are part of his current state space. Finally, all elements of $z_{it}^{\mathcal{I}}$ are observed to the econometrician except η_i, which is an exogenous person-specific characteristic that affects the income process and is described below. Individuals also observe a vector of choice-specific additive utility disturbances ε_{it}, which are unobserved to the econometrician and assumed independent across time, individuals and choices. Besides individual-specific variables, z_{it} contains aggregate level components, collected in $z_t^{\mathcal{M}}$, which individuals use to forecast the evolution of the market. $z_t^{\mathcal{M}}$ will be described below when we

[22] CD4 ranges from 0 to 2915 in our analytic sample with a median of 448. Healthy CD4 counts are those above 500 units per mm^3 and typically range between 500 and 1,500.

describe the consumer's information.

Choices

At each period t the individual chooses whether or not to use medication. If he opts for medication, he may choose the same product he consumed in the last period or he may choose from the set of other treatments that are currently available on the market. Alternatively, he may choose a trial treatment. The individual faces uncertainty about the quality of both market and trial treatments.

We begin with uncertainty over market treatments. The individual learns about the quality of a product immediately after using it. Hence, if he chooses the same market treatment he consumed in the prior period, he faces no uncertainty regarding its characteristics.[23] Alternatively, if he decides to try a different market drug, his alternative is to choose one among several groups or *clusters* of drugs with similar qualities. The agent is then randomly assigned a drug within the cluster he selected.

Formally, at every period t there is a set of market products $\mathcal{P}_t$ clustered in several groups collected in $\mathcal{G}_t$. $\mathcal{G}_t$ denotes both the collection of clusters available at t and the cardinality of the collection. When individual i decides to consume a market treatment that is different from the one he consumed in the prior period, he must choose from a cluster $g_t \in \mathcal{G}_t$. By selecting group g_t he chooses a gamble among all products in group g_t. The distribution of products within the group is given by weights that are a function of the treatment characteristics and the number of products in the group. The estimation of these weights is explained below. The moments of the within cluster distribution are generated by the products in the cluster and their weights. However, for tractability we assume that agents do not observe the cluster components and instead they only observe the first two

[23] As discussed above, his state space includes the characteristics of the drug consumed in the prior period q_{it-1}.

moments of the within cluster distributions and that these are sufficient to describe the distribution. These moments form the set of characteristics of groups of products available at time t, $W_t = f_1^D(\mathcal{P}_t)$.

Our clustering process is a device to make the model tractable and estimation feasible by reducing the state space significantly while still allowing individuals to choose among different options in the market.[24] In order to avoid issues emerging from differences in scales, when using our clustering algorithm, we assume that clustering occurs with respect to scaled product characteristics denoted $\tilde{\theta} \in [-1,1] \times [-1,1]$.[25] Then we obtain product groups at t by solving a k-means algorithm that approximates the solution of the following objective function[26]

$$\min_{1\{k \in g\}_{k \in \mathcal{P}_t} | \mathcal{G}_t} \quad \sum_{g=1}^{\mathcal{G}_t} \sum_{k \in \mathcal{P}_t} 1\{k \in g\} \left\| \tilde{\theta}_k - \tilde{\theta}_k^c \right\|^2$$
$$s.t. \sum_{g \in \mathcal{G}_t} 1\{k \in g\} = 1 \text{ for all } k \in \mathcal{P}_t \tag{3.8}$$

where the centroid of cluster k, $\tilde{\theta}_k^c$, is defined as

$$\tilde{\theta}_k^c = \frac{\sum_{k \in \mathcal{P}_t} 1\{k \in g\} \tilde{\theta}_k}{\sum_{k \in \mathcal{P}_t} 1\{k \in g\}} \tag{3.9}$$

The algorithm is explained in detail in Appendix B.2. At any given period we set the maximum value of $\mathcal{G}_t$ at $\mathcal{G}^{\max}$ so that the individual knows how many groups will be available every period. $\mathcal{G}^{\max}$ is chosen so that there is a non negligible number of consumers choosing each group in the data. We set $\mathcal{G}^{\max} = 3$.

We do not model the variation of within cluster assignment endogenously. Instead,

[24]This approach is close to reality if individuals only observe product labels and do not know their characteristics beyond the fact that groups of product labels are associated to a certain mean and variance of characteristics.

[25]The transformation is explained in Appendix B.2

[26]See Duda and Hart (1973) and Andrew W. Moore's *K-means and Hierarchical Clustering* tutorial at http://www.cs.cmu.edu/∼awm/tutorials.html.

we develop the concept of within cluster weights as functions of products's characteristics. Weights are estimated in the following fashion:

1. We compute a nonlinear regression of within cluster shares on treatment characteristics:

$$s_{k|g_t} = \exp\left(X^w_{k,t}\beta^w\right) + \epsilon^w_{k|g_t} \tag{3.10}$$

   where $X^w_{k,t}$ includes a constant term, the ranking (within its cluster) of the characteristics of the product, the number of members in the cluster, whether the product is new, and several interactions.

2. We obtain predicted within cluster shares $\hat{s}_{k|g_t}$ and compute the weight of product k in cluster g_t as

$$\tilde{s}_{k|g_t} = \frac{\hat{s}_{k|g_t}}{\sum_{r\in g_t} \hat{s}_{r|g_t}} \tag{3.11}$$

If the individual chooses neither to try a cluster nor to stay in his previous treatment, he may instead join a clinical trial to get an experimental treatment. Trial product characteristics are unknown. However, he knows that innovation occurs in trials around the centroid ω_t.[27] Therefore, he knows that the product characteristic of trial treatments are distributed according to $F_{\theta|\omega_t}$, which was introduced as the distribution of new product characteristics in the supply section above. A key difference between consuming group g_t and the trial treatment is that once the individual chooses a group and a treatment is assigned to him, he has the chance of choosing that treatment with certainty the next period.

Having described each option, we now formally specify the choice set. Let d_{jit} be the choice indicator that takes the value of one if agent i in period t chooses medical treatment

[27] One way to think about this point is that consumers entering a trial see ω_t as the quality of a placebo drug administered in a trial. This would make sense in the context of HIV since new drugs were not tested against no drug at all, but were instead tested against "current best practices" (see for example Ickovics and Meisler (1997)).

j in the choice set $\mathcal{C}_{it}$. The choice set is time-specific because the characteristics of available products evolve as new products enter the market and incumbent products exit. The choice set is also individual specific since individuals who chose a market treatment in the prior period may choose that treatment again. If the individual did not choose a market treatment in the prior period his choice set is:

$$\mathcal{C}_{it} = \begin{cases} 0 & \text{No Treatment} \\ 1 & \text{Cluster } g_t = 1 \\ 2 & \text{Cluster } g_t = 2 \\ \vdots & \vdots \\ \mathcal{G}^{max} & \text{Cluster } g_t = \mathcal{G}^{max} \\ \mathcal{G}^{max} + 1 & \text{Trial} \end{cases} \tag{3.12}$$

If the individual chose a market treatment in the prior period his choice set $\mathcal{C}_{it}$ is augmented by one alternative to include the possibility of consuming his previous period treatment with certainty about its characteristics.

Utility

Next, we specify the flow utility function to capture how the individual's product choices are driven by the effects of each treatment choice on health, ailments, income, out-of-pocket payments, and non pecuniary benefits. For choice $j \in \mathcal{C}_{it}$ and state z_{it}, the utility at period t for individual i is a function of his health, ailments, and net income given by

$$y_{jit} + \varepsilon_{jit} = \alpha_{jit}\,(z_{it}) + \alpha_m(m_{jit} - o_{jit}) + \alpha_x x_{jit} + \alpha_{xp} x_{jit}(1 - d_{0it}) + \varepsilon_{jit} \tag{3.13}$$

where m_{jit} is gross income, o_{jit} are out-of-pocket payments, x_{jit} is an indicator for whether the individual does not suffer ailments, d_{0it} is the indicator of whether he chooses not to consume a treatment, and ε_{jit} are unobserved choice-specific taste shocks. The interaction of the no-ailments indicator and the treatment choice indicator is used to capture a distaste for side effects, which are ailments arising from treatment consumption.

In equation (3.13), $\alpha_{jit}(z_{it})$ are choice-specific preference parameters that depend on observables. They are defined as

$$\alpha_{jit}(z_{it}) \equiv \alpha'_{jb}b_i + \alpha_{ja}a_{it-1} + \alpha_{jh}h_{it-1} \tag{3.14}$$

We assume that consumer preferences over clusters are fully captured by cluster characteristics. Therefore, we assume parameters α'_{jb}, α_{ja}, and α_{jh} to be constant across clusters. This is the characteristics approach commonly used in structural models of demand which explains consumer choices as a function of product qualities. In contrast, participating in a clinical trial may offer differential benefits related to the psychological costs (or benefits) from being part of an experiment. We also allow the choice of remaining in the same product to have differential non pecuniary benefits in order to capture factors, such as switching costs and a preference for certainty, which explain why consumers may continue using a product they have used before even as better products enter the market. Finally, for identification purposes we normalize the non pecuniary benefits from not consuming a treatment to zero (Magnac and Thesmar, 2002; Arcidiacono and Miller, 2015).

Outcomes and Transitions

In this section, we specify the stochastic processes governing state variables in z_{it} as well as the outcome variables: income, out-of-pocket payments, ailments, and survival.

Income: Gross income is a function of today's state, z_{it}, and ailments, x_{jit}. It is given by

$$m_{jit} = X^m_{jit}\Gamma^m + \eta_i + \epsilon^m_{it} \tag{3.15}$$

where $X^m_{jit} = [1, h_{it-1}, \ldots, h^7_{it-1}, a_{it-1}, a^2_{it-1}, b_i, edu_i, l_{it}, x_{jit}]$. Gross income does not include product cost, which is accounted for in the payments equation below. Equation (3.15) is estimated using random effects and individual-specific income characteristics are estimated consistently as

$$\hat{\eta}_i = \sum_t \sum_j d_{jit} \left(m_{jit} - X^m_{jit} \hat{\Gamma}^m \right)$$

Individuals observe the income iid shocks ϵ^m_{it} before making their choice.

Payments: Out-of-pocket payments are censored at zero. They are given by the following tobit specification

$$o_{jit} = o\left(X^o_{jit}, \epsilon^o_{it}; \Gamma^o \right) \tag{3.16}$$

where $X^o_{jit} = [1, h_{it-1}, \ldots, h^6_{it-1}, a_{it-1}, a^2_{it-1}, b_i, edu_i, \{d_{jit}\}^5_{j=0}, l_{it}, x_{jit}]$ and ϵ^o_{it} is the error term in the underlying equation. Since we do not directly observe prices, and in order to simplify the problem, we assume a constant cost of participating in a trial as well as a constant cost of consuming a market product.[28]

Labor Supply: We do not model labor supply explicitly as a choice as it is not the main purpose of this paper. However, labor supply may be affected by treatment choices, e.g., through health status and physical aliments. Moreover, labor supply also affects income and therefore utility. To capture this, we treat labor supply as a state variable that individuals know at the beginning of the period before making their treatment decision. Individuals draw their labor market participation from the distribution characterized by

$$\Pr[l_{it} = 1 | X^l_{it}] = \frac{1}{1 + \exp(X^l_{it} \Gamma^l)} \tag{3.17}$$

where $X^l_{it} = [1, l_{it-1}, h_{it-1}, \ldots, h^4_{it-1}, a_{it-1}, a^2_{it-1}, b_i, edu_i]$.

Physical Ailments: First, define a mapping from the choice to the characteristics of the

[28]End-users customarily pay a standardized deductible that is a fraction of the brochure price of the drug paid by the insurance company. Median out-of-pocket drug costs are about \$300 every six months for a regime of drugs that would cost the insurance company between \$5,000 and \$15,000.

treatment

$$\theta(d_{jit}) = \{\theta^x(d_{jit}), \theta^h(d_{jit})\} \tag{3.18}$$

where $\theta(d_{jit}) = q_{it-1}$ if the individual consumes his prior-period market treatment. $\theta(d_{jit})$ is a stochastic variable is the individual chooses a cluster or if he joins a trial.

A production function transforms drug characteristics and health into aliments. Let x_{jit} be an indicator that takes the value of 1 if the individual does not suffer ailments in t after choosing alternative $j \in \mathcal{C}_{it}$. The probability of not having physical ailments for individual i is modeled as:

$$\Pr\left[x_{jit} = 1|\cdot\right] = \frac{\exp\left(\sum_{m=0}^{5} \gamma_m^x h_{it-1}^m + \theta^x\left(d_{jit}\right)\right)}{1 + \exp\left(\cdot\right)} \tag{3.19}$$

<u>Health</u>: CD4 count is our objective measure of health. Like ailments, health at the beginning of period $t+1$ is a function of previous health and drug characteristics. The production function of health is specified as:

$$h_{jit} = \sum_{m=0}^{5} \gamma_m^h h_{it-1}^m + \theta^h\left(d_{jit}\right) + \epsilon_{it}^h \tag{3.20}$$

The distribution of the health disturbance is estimated non-parametrically using the residuals of the health production function. We assume that $\mathbb{E}[\epsilon_{it}^h|\cdot] = 0$, where the expectation is conditional on the vector of regressors of the health production function.[29]

<u>Survival</u>: At the end of any period t individuals may survive into the next, denoted by $S_{it+1} = 1$, with the following probability

$$D_{it+1}\left(z_{it+1}\right) \equiv \Pr[S_{it+1} = 1|z_{it+1}] = \frac{1}{1 + \exp(X_{it}^d \Gamma^d)} \tag{3.21}$$

where $X_{it}^d = [1, h_{jit}, \ldots, h_{jit}^5, a_{it}, a_{it}^2, b_i, edu_i, x_{jit}]$.

[29] Here, it is important to point out that each individual drug in our sample has a set of characteristics that are observed by the econometrician. The agent only observes cluster attributes. However, these are constructed from individual-level data.

Consumer Information and Aggregate State Forecasts

Consumers must form expectations not only about the evolution of their individual-level state variables, but also about the characteristics of future choice sets. The following example underscores the importance of this type of forecast in explaining consumer choices: a relatively healthy consumer may avoid choosing from a group of effective drugs with strong side effects in the current period if he expects effective drugs with fewer side effects to emerge soon; in contrast, a sick consumer may not want to avoid medication despite side effects if he fears that he may not survive until better drugs are introduced.

We assume that consumers have rational expectations, but they do not observe the entire aggregate state Ξ_t. Instead, they observe a reduced aggregate state $z_t^{\mathcal{M}}$, which is a mapping from Ξ_t, and integrate over what they do not observe—these unobserved objects may be past, present and future. The aggregate portion of the individual's state is given by

$$z_t^{\mathcal{M}} \equiv \langle \omega_t, W_t, \mathcal{F}_t \rangle \tag{3.22}$$

The individual observes the centroid for innovation ω_t, described in Section 3.3.2, which determines the expected characteristics of trial products. His information set also contains the characteristics W_t of the clusters of products he observes, described in Section 3.3.3. Finally, he observes the current distribution of consumer characteristics $\mathcal{F}_t$.

The Value Function

We define the value function conditional on choice $j \in \mathcal{C}_{it}$, net of taste shocks, for individual i at time t as follows:

$$v_{jit}(z_{it}) = \mathbb{E}\left[y_{jit} + \beta \left[D_{it+1}\left(z_{it+1}\right) \max_{c \in \mathcal{C}_{it+1}} \left\{ v_{cit+1}(z_{it+1}) + \varepsilon_{cit+1} \right\} \right] \middle| \, z_{it}, j \right] \tag{3.23}$$

Expectations are taken over product characteristics affecting the flow utility and the evolution of both observed and unobserved state variables. Expectations over the evolution of unobserved state variables are independent conditional on the current set of state variables. Therefore, we can rewrite equation (3.23) as

$$v_{jit}(z_{it}) = \mathbb{E}_y[y_{jit}|z_{it}] + \beta \mathbb{E}_z \left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_\epsilon \left[\max_{c \in \mathcal{C}_{it+1}} \left\{ v_{cit+1}(z_{it+1}) + \varepsilon_{cit+1} \right\} \right] \middle| z_{it}, j \right] \quad (3.24)$$

The first expectations operator, $\mathbb{E}_y$, denotes expectations over outcomes that affect flow utility, including income and physical ailments. The second operator, $\mathbb{E}_z$, denotes expectations over the evolution of observed state variables z_{it}, including health and the mapping from the aggregate state that he observes, $z_t^{\mathcal{M}}$. The third operator, $\mathbb{E}_\epsilon$, denotes expectations taken over the joint distribution of unobserved choice-specific taste shifters.

3.4 Estimation

In this section, we describe how we estimate parameters of the model specified in Section 3.3 using a GMM estimator. A more extensive treatment of the estimation procedure, including a more detailed algorithm, is found in Appendix B.2. In Section 3.4.1, we provide an overview of the estimation procedure, summarizing the algorithm. The first eight steps can be seen as a "first stage", where quantities are computed that do not change as utility parameters change. These quantities are computed a single time and are then used to construct moments used in GMM estimation of utility parameters, described in the final step of the algorithm. In Section 3.4.2, we provide further details on the GMM estimator, describing the theoretical moment conditions and their sample analogs. We also emphasize unique features of our forward simulation procedure.

3.4.1 Overview

Our estimation procedure can be summarized in the following steps.

1. *Products as treatments.* Our estimation starts with the definition of products. We define a product as a combination of single-product components. Examples of products are AZT or the combination of AZT+3TC+Saquinavir.

2. *Outcome equations.* We estimate processes for income, out-of-pocket payment, labor supply and survival. Heath and no-ailments equations will be estimated in the next step (see equations (3.15), (3.16), (3.17), and (3.21)).

3. *Product characteristics.* Given products defined in step 1, we estimate product characteristics (see equations (3.19) and (3.20)).[30]

4. *Clusters.* Using the estimated product characteristics in step 3, we use a k-means algorithm to obtain clusters of products for every period (see equation (3.8)). Then, using the characteristics of the products in each cluster, we obtain within clusters weights as predictions from non-linear regressions of within cluster shares on covariates (see equations (3.10) and (3.11)). Finally, using the within cluster weights we compute cluster characteristics—mean and variance matrix.

5. *Centroid.* Using product characteristics from step 3, we back out innovation centroids for every period (see equation (3.1)).

6. *Distribution of innovations.* Every new product is modeled as a draw around the centroid (see equation (3.2)). Hence, for every new product at a given period we

[30] As described in the model, individuals make treatment decisions based on cluster attributes, which are probabilities of health and side effects associated with each cluster governed by equations (3.19) and (3.20). However, their realized outcome will depend on the individual drug they are randomly assigned to from the cluster. Thus, we must also estimate drug-specific characteristics, which are version of equations (3.19) and (3.20), but for each treatment rather than for each cluster. The treatment specific equations are equations (B.2) and (B.3) in Appendix **??**).

compute the realized innovation around the centroid as the residual from subtracting the centroid (step 5) from the product characteristic (step 3). Using the realized innovations we non-parametrically estimate the stationary distribution of innovations, F_v.

7. *Distribution of number of draws.* Using data regarding the amount of new products per period we estimate the distribution of number of new products (see equations (3.3) and (3.5)).

8. *Conditional choice probabilities (CCPs).* Using cluster characteristic from step 4, centroids from step 5 and other aggregate and individual-specific state variables we estimate flexible parametric CCPs (see Appendix B.2).

9. *Structural utility parameters.* We follow Hotz et al. (1994) and use forward simulation to generate choice and technology paths as well as future individual states that will serve as inputs to the simulated continuation value. In our forward simulation we use the estimated CCPs (step 8), the distribution of number of draws (step 7) and the distribution of innovations (step 6) as well other estimated processes (step 2 through step 5). Finally, we implement a GMM estimator using a moment condition which is a function of the forward simulated data, CCPs, and utility parameters.

3.4.2 Moment Condition

We use GMM to estimate utility parameters. Our moment conditions appeal to well-known results following from our assumption that the taste shocks ε_{jit} are iid Extreme Value Type I distributed (see for example Hotz and Miller (1993)). The moment conditions rely on differences between alternative representations of the difference in conditional value functions $v_{jit}(z_{it}) - v_{oit}(z_{it})$. Let $J = 6$ be the maximum possible cardinality of the individual's choice set, let $p_{jit}\left(z_{it}\right)$ be the probability that individual i chooses option j at time t conditional on his state z_{it}, and let $w(z_{it})$ be a vector of instruments orthogonal to the difference between alternative representations. We can form the following moment conditions:

$$\mathbb{E}\left\{ w(z_{it}) \otimes \begin{bmatrix} \ln\left(\frac{p_{oit}(z_{it})}{p_{1it}(z_{it})}\right) + v_{1it}(z_{it}) - v_{oit}(z_{it}) \\ \vdots \\ \ln\left(\frac{p_{oit}(z_{it})}{p_{J-1it}(z_{it})}\right) + v_{J-1it}(z_{it}) - v_{oit}(z_{it}) \end{bmatrix} \right\} = 0. \tag{3.25}$$

The first representation of the difference in conditional value functions is the log odds ratio formed with current-period conditional choice probabilities. The second representation relies on the results in Proposition 3.1, which yields the conditional value function as a mapping of future conditional choice probabilities and utility parameters.

Proposition 3.1. *Let $V(z_{it}, \varepsilon_{it})$ be the value function for individual i at period t who has a state given by z_{it} and ε_{it}. Define $P_j^{o(s-1)}$ as the probability of surviving until period $t+s-1$ conditional on the state at t, decision j at t, and optimal behavior, denoted d_i^o, up to some period $T^* > t$.*[31] *Define $\psi_{kit}(z_{it}) \equiv \mathbb{E}_\varepsilon[\varepsilon_{kit} | d_{it}^o = k, z_{it}]$ as the expected value of the kth taste shock conditional on alternative k being optimal. Finally, let γ be the Euler constant. Then, the conditional value function can be written as*

$$v_{jit}(z_{it}) = \mathbb{E}\left[y_{jit} | z_{it}\right] + \sum_{s=1}^{T^*} \beta^s P_j^{o(s-1)}(z_{it}) \times$$

$$\mathbb{E}_z\left[D_{it+s}(z_{it+s}) \sum_{k \in C_{t+s}} p_{kit+s}(z_{it+s}) \left[y_{kit+s}(z_{it+s}) + \psi_{kit+s}(z_{it+s})\right] \middle| z_{it}, d_{jit} = 1, S_{it+s-1} = 1, d_i^o \right]$$

$$+\beta^{T^*+1} P_j^{o(T^*)}(z_{it}) \mathbb{E}_z\left[D_{it+T^*+1}(z_{it+T^*+1}) V(z_{it+T^*+1}, \varepsilon_{it+T^*+1}) | z_{it}, j, S_{it+T^*} = 1, d_i^o\right] \tag{3.26}$$

and

$$\psi_{kit}(z_{it}) = \gamma - \ln(p_{kit}(z_{it})) \tag{3.27}$$

[31] Since any individual present at t has evidently survived until t, $P_j^{o(0)}(\cdot) \equiv 1$. Recall that S_{it+s} is the survival indicator and D_{it+s} is the one-period-ahead probability of survival, defined in Section 3.3.3

Proof: see Appendix B.2

Sample Analog and Forward Simulation

To form sample analogs of the moments in equation (3.25), we first substitute the theoretical log odds ratio using the estimated CCPs. Second, we use Proposition 3.1 to obtain differences in conditional value functions using forward simulation (Hotz et al., 1994). In our forward simulation procedure, for every individual i at time t facing choice set $\mathcal{C}_{it}$, we fix choice j and use the estimated stochastic processes governing outcomes and transition to simulate his state variables at $t+1$. We then use the estimated parameters of the CCPs to simulate $t+1$ choices conditional on the new simulated state. We continue the same process until T^*, whose value is set high enough so that the product $\beta^{T^*+1} P_j^{o(T^*)}(z_{it})$ approaches zero, eliminating further differences in conditional value functions beyond T^*.

Forward simulation is used in a variety of settings to compute conditional value functions. There is one feature of our estimation procedure that distinguishes our approach from prior literature. As was mentioned in Section 3.3, individuals are not only aware of the stochastic processes governing their individual state transitions, but also they are aware of the stochastic process that links aggregate behavior and innovation. Using the information contained in their state, consumers form expectations about future choice sets. This contrasts with setups where agents are either fully aware of future technologies or, alternatively, where they are fully unaware, in which case an innovation like HAART amounts to a regime change. Our forward simulation procedure explicitly incorporates how aggregate behavior affects individual expectations over future innovations.

For each observation (individual i at period t), we first construct an artificial technological path by simulating aggregate behavior forward. In other words, we forward simulate the choices of all individuals in the sample at period t, and collect the technological path generated by their choices. Then, because individuals are atomistic, for each observation we can generate several sequences of future choices and payoffs, taking as given its observation-specific artificial technological path, to reduce individual simulation error.[32] There are two features from our forward simulation that we underscore. First, our conditional choice probabilities, insofar as they condition on $z_t^{\mathcal{M}}$, capture the way in which individuals compute expectations. This guarantees that our method to obtain individuals' expectations by simulating future paths of the aggregate state Ξ_t generated from the simulated choices of individuals with limited information is accurate. Clearly, because we simulate the aggregate future path of Ξ_t, we can also obtain the simulated future path of $z_t^{\mathcal{M}}$. Second, we simulate separate artificial technological paths for every observation. This serves two purposes. It maintains the assumption, needed for consistency of the estimator, that the sample draws from the moment conditions—the contribution from each observation—are independent from each other. Additionally, simulating a separate artificial technological path for every observation prevents simulation errors in the technological paths from propagating across all observations.

The Distribution of Innovations

The forward simulation of future choice sets relies on the distribution of innovations. In our framework, the characteristics of new products entering the market today determined by last period's centroid and a draw from the distribution of innovations, F_ν. This distribution, which we assume is stationary, provides the location of new products relative to the current centroid. To estimate the distribution of innovations we use all periods in the MACS data with relevant information on treatment consumed, health, and ailments (1986 to 2008). We then define centroids for innovation, ω_t, given by equation (3.1). For each new product at t, characterized by θ, we compute a realized innovation vector as

$$\nu^\theta = \theta - \omega_{t-1}$$

[32]We generate 10 sequences per observation.

We do not impose that innovation vectors cannot be strictly negative. In other words, inferior products with lower quality in both dimensions (health and ailments) may be introduced if the non parametric distribution allocates mass to the the south-east quadrant of its domain. This is not at odds with what we observe in the data, and theoretical reasons why this may happen have been provided in the literature (Miller, 1988). Over the time span we observe there are 76 realizations from the innovations distribution which we use to obtain a nonparametric empirical distribution for ν.

3.5 Parameter Estimates and Choice Dynamics

In this section, we discuss estimates of the structural model. We organize our discussion around the key factors driving choices. We introduce the distribution of innovations in Section 3.5.1. We discuss the utility function in Section 3.5.2 and describe outcomes and transitions in Section 3.5.3. Finally, in Section 3.5.4 we use the model to simulate choices over time, which allows us to assess model fit.

3.5.1 The Distributions of Innovations and New Products

In our model, every new product is an innovation about the centroid. How far new products land from the centroid is stochastically given by the distribution of innovations, F_ν, that we estimate non parametrically. Figure 3.7 shows that F_ν is bimodal and it does not appear to be well approximated by a standard parametric distribution. While one of the modes is located approximately at the status quo point $(0, 0)$, a second mode is located north of the first one along the health axis. Since the probability distribution is not always decreasing as we move away from the centroid, the process on innovation is rather jumpy.

As shown in Table 3.3, the distribution of innovations has a positive mean in terms of health quality, but a mean in terms of no-ailments quality that is not statistically different

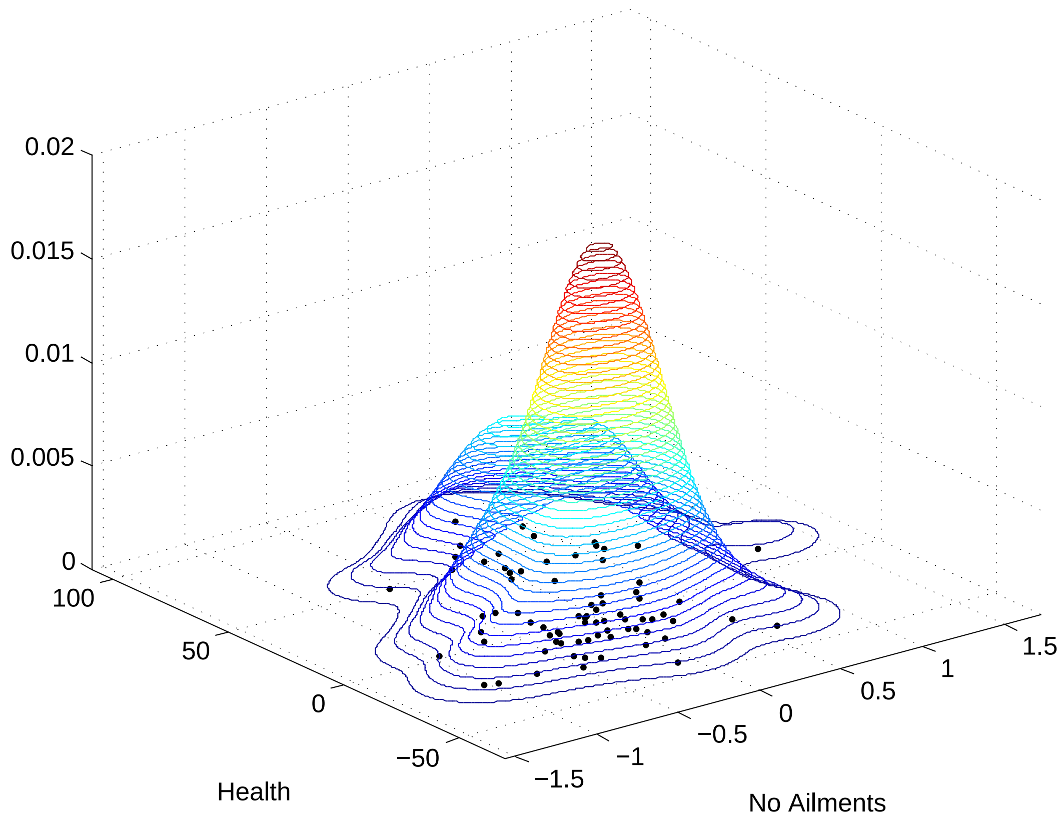


Figure 3.7: The Distribution of Innovations, F_ν.

Notes: Distribution estimated non parametrically from the realized innovation vectors of the form $\nu_{\theta_{tr}} = \theta_{tr} - \omega_{t-1}$, where θ_{tr} are the characteristics of the new product r at t and ω_{t-1} is the centroid for innovation at the time the product was in trials.

from zero. In other words, new products are on average more efficacious but no better in terms of side effects. These results suggests that, if consumers were to choose products at random, on average the quality of products would improve over time in terms of health quality but would remain largely unchanged in terms of no-ailments quality. However, in our model the centroid is a mapping from consumer demand that does not happen at random. Since the centroid anchors the distribution of innovations, the characteristics of future products are more likely to be close to the characteristics of products with larger market shares, which shapes the path of innovation. In this sense, innovation is endogenous to consumer choices.

Table 3.3: Moments of the Distribution of Innovations, F_ν

	mean	covariance matrix		Ho: $mean = 0$ (p-value)
$Health$	8.10	747.57	4.05	0.01
$NoAilments$	-0.02	4.05	0.16	0.62

Notes: Last column tests separately whether the each component of the mean is different than zero.

Table 3.4: Distribution of Number of New Products, F_N

	coef.	se
μ		
$MaxChange_{t-1}$	0.432	0.246
$TrialsShare_{t-1}$	6.177	2.462
$\ln\alpha$		
$Constant$	-0.206	0.451
$MaxChange_{t-1}$	-1.019	0.626

Notes: Model specified in (3.3). The variable Q_{t-1} measures the distance between the previous period's new products and the previous period's centroid. It captures the relatively higher number of new products that follow the appearance of better innovations. The variable $TrialsShare_{t-1}$ is the share of individuals going into a trial the previous period. According to the model in (3.3), $E[New_t] = \mu$ and $Var[New_t] = \mu(1 + \alpha\mu)$.

The process of innovation also depends on the magnitude of previous innovations and on participation in clinical trials. Estimates for the distribution of number of new products shown in Table 3.4 show that large positive innovations are likely to be followed by the appearance of a multitude of products, which is consistent with firms vying for market share following a breakthrough. The magnitude of previous innovations also reduces the dispersion around the number of new products that enter. The share of consumers opting for the trial product in the prior period also increases the likelihood of more products entering the market. The reason is that, as more consumers select trial products, firms have more room for experimentations which provides them with valuable information about the viability of new treatments that they can now introduce into the market more rapidly. The fit of our distribution of new products is shown in Figure 3.8. It shows that the empirical distribution is not far from the average (over time periods) of the predicted probabilities generated by

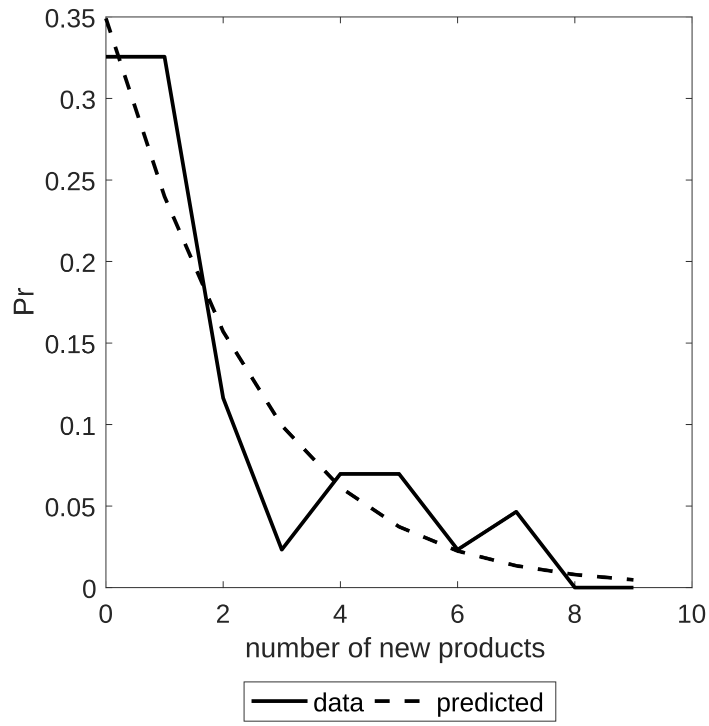


Figure 3.8: Distribution of Number of New Products

Notes: Model specified in (3.3). Figure shows the empirical distribution of new products and the average over time of the predicted probabilities using the estimated parameters in Table 3.4.

the model.

3.5.2 Utility Parameter Estimates

Utility parameters are reported in Table 3.5. The utility that individuals obtain from each choice (drawing from a cluster, joining a trial, staying on their current treatment) depends on their socio-demographic and health characteristics. These interactions help to explain heterogeneity in choices across groups that are not attributable to variation in continuation payoffs or variation in current net income. In interpreting parameter estimates, note that the non pecuniary utility from no treatment is normalized to zero across groups. Therefore, parameter estimates govern flow non pecuniary utility for different groups relative to what they gain from not taking a treatment.

According to parameter estimates, clusters and trials lead to a utility cost and, generally, these penalties are higher for non-white patients. Black men face a particularly high penalty

Table 3.5: Utility Parameters, y_{it}

parameter	variable	coef.	se
α_{4w}	$1\{cluster\} \cdot white$	-1.385	0.206
α_{4b}	$1\{cluster\} \cdot black$	-1.868	0.210
α_{4l}	$1\{cluster\} \cdot hispanic$	-1.075	0.835
α_{4a}	$1\{cluster\} \cdot a_{it-1}$	0.003	0.005
α_{4h}	$1\{cluster\} \cdot h_{it-1}/10^3$	-3.385	0.134
α_{5w}	$1\{trial\} \cdot white$	-2.678	0.168
α_{5b}	$1\{trial\} \cdot black$	-3.755	0.170
α_{5l}	$1\{trial\} \cdot hispanic$	-2.902	0.354
α_{5a}	$1\{trial\} \cdot a_{it-1}$	0.051	0.003
α_{5h}	$1\{trial\} \cdot h_{it-1}/10^3$	-1.702	0.082
α_{6w}	$1\{stay\} \cdot white$	0.525	0.157
α_{6b}	$1\{stay\} \cdot black$	0.396	0.159
α_{6l}	$1\{stay\} \cdot hispanic$	0.480	0.674
α_{6a}	$1\{stay\} \cdot a_{it-1}$	0.019	0.003
α_{6h}	$1\{stay\} \cdot h_{it-1}/10^3$	1.048	0.101
α_x	x_{it}	0.522	0.292
α_{xp}	$x_{it} \cdot 1\{product\}$	-3.575	0.226
α_m	$m_{it} - o_{it}$	0.141	0.023

Notes: Estimation of equation (3.13). Discount factor $\beta = .8$. $1\{cluster\}$ indicates whether the individual chose one of the three clusters of products available. $1\{product\}$ indicates whether the individual consumes a product in t, $1\{product\} = 1\{cluster\} + 1\{stay\} + 1\{trial\}$.

of trial participation, a finding that is consistent with a broad literature investigating historical reasons why blacks are reluctant to enter trials to use experimental drugs (Harris et al., 1996). Moreover, healthier individuals have a lower utility of choices where they face uncertainty, including clusters or trials. Interestingly, healthier individuals gain utility from using drugs they have used before. These results suggest that healthier individuals dislike uncertainty about drugs and, perhaps, switching costs relative to their less healthy counterparts. We also find that the utility costs of treatment relative to no treatment are stronger for younger individuals. This is perhaps reflective of age-dependent tolerance for medication, especially if older individuals have grown accustomed to using medications for other health problems.

Finally, individuals dislike ailments regardless of which product they are using. This utility parameter is key as it helps explain why individuals eschew medications that have high dynamic payoffs in the form of better future health. This finding is consistent with individuals who consider their quality of life in a multidimensional manner. Similar results have been found in Chan and Hamilton (2006) and Papageorge (2016), who showed that even in the context of a deadly infection (HIV), individual treatment choices reflect a distaste for side effects. Finally, our results show that individuals gain positive utility from income, which reflects consumption utility and is expected.

3.5.3 Transitions and Outcomes

Next, we discuss the processes describing how state variables produce outcomes or transition to other states. Individuals judge the quality of any product in terms of its ability to raise their future CD4 count and its ability not to generate ailments. We estimate the processes for health and ailments, which includes estimating the characteristics of products, using equations (3.19) and (3.20). To conserve on space, we present the coefficient estimates in Tables B.5 and B.6 in Appendix B.3 (see Column 5 in both tables). The top panels of Figure 3.9 show how current health transforms into future health and lack of ailments. While the slight concavity of the production function for health could be well approximated by a linear function, the production function for ailments is very non linear. The figures suggest that changes is health below a CD4 count of 250 units generate much larger movements in the log odds ratio of getting ailments than changes in health above that threshold.

Health also exhibits strongly non-linear relationships with other outcomes. This result helps to explain differences in optimal choice for individuals with somewhat similar health profiles (as measured by CD4 count). Figure 3.9, plots the estimated relationship between health and several outcomes: income, out-of-pocket payments, labor supply and survival. According to the figure, income increases steeply with CD4 count for very sick individuals but the effect of health flattens substantially for individuals with CD4 counts above 250.

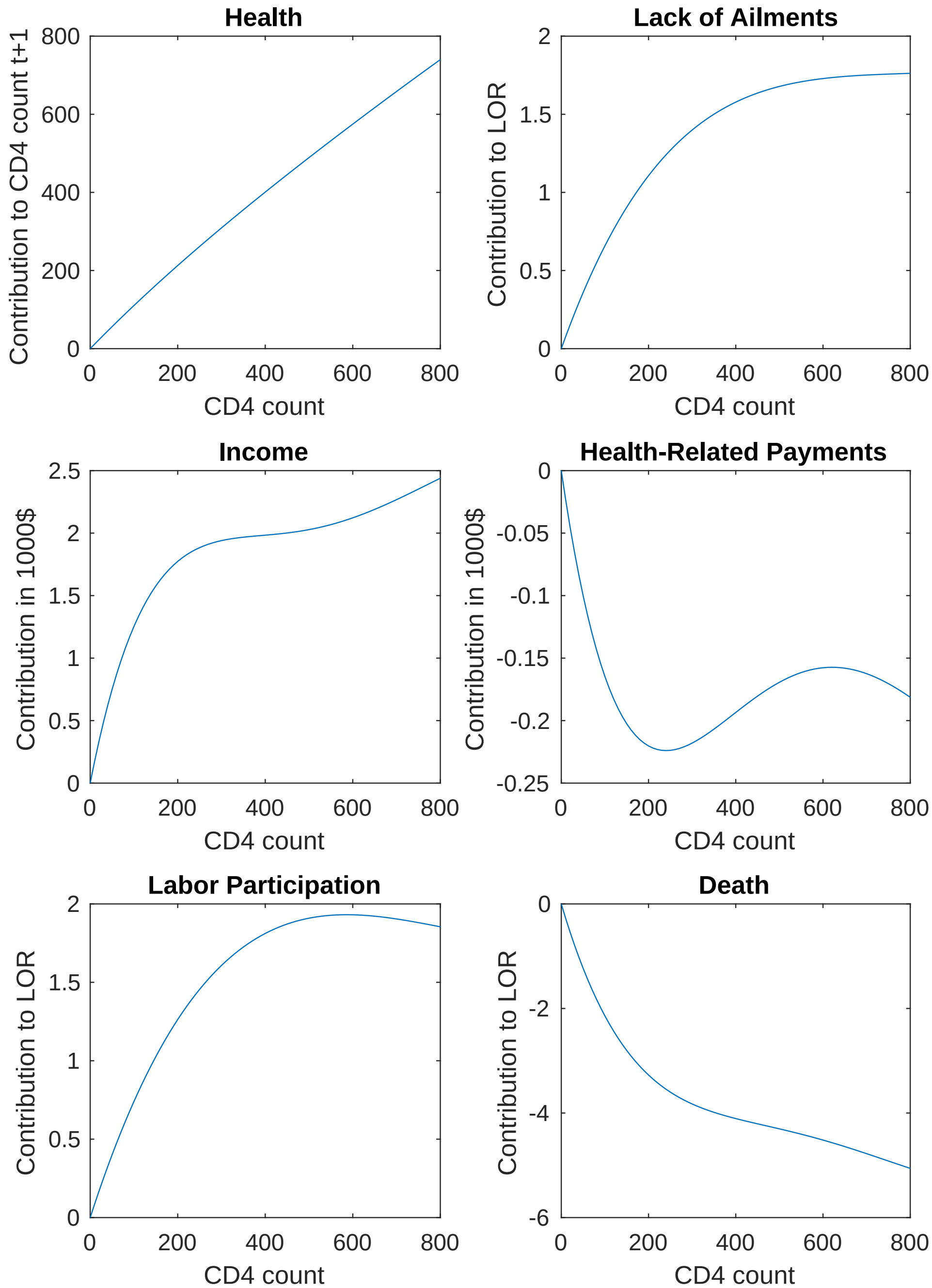


Figure 3.9: Health Effect on Outcomes

Notes: CD4 Count measured in hundreds of cells per microliter. LOR stands for log odds ratio. Semestral income measured in thousands of dollars of 2000.

The health profile of out-of-pocket payments in Figure 3.9 is the mirror image of the health profile for income with deeper decreases in payments as health increases for the sickest. This makes sense as health expenditures due to opportunistic infections, for example, would be expected to decline precipitously as a result of small health increases at low health levels. Similarly, the odds of working increases with health until a CD4 count of about 350 units and then it flattens. Finally, the effect of health increases on survival are more dramatic the more sick individuals are. Even though the positive impact of health on survival remains at higher health levels, this relationship diminishes considerably after a CD4 count of about 250 units.

In general, the health profiles in Figure 3.9 tell a very consistent story about CD4 count and HIV infection. The effect of marginal health increases on outcomes is much stronger for individuals with low CD4 counts and it seems to flatten after individuals surpass well-known cutoffs below which AIDS occurs. This is consistent with the idea that low CD4 counts have little discernible impact on symptoms or survival unless the AIDS threshold is reached. Below that threshold, further reductions have large effects on outcomes since the body's immune system becomes increasingly compromised and is therefore unable to fight off routine infections. These results underscore the importance of modeling the relationship between health and outcomes in a non linear fashion in the context of HIV.

We also estimate other sources of variation in outcomes. Table 3.6 presents our results for the income equation. Individuals who do not suffer ailments have higher income as their productivity is likely to be higher. Income is concave in age and it increases with labor participation and education. Minorities have lower income. At any period individuals may incur out-of-pocket costs related to their treatment consumption decision. According to Table 3.7, conditional on having out-of-pocket expenditures, these payments increase with age. Minorities spend less and more educated people spend more. Similarly, individuals that suffer ailments spend more, perhaps because they are managing other health conditions.

Table 3.6: Gross Income, m_{it}

variable	coef.	se
h_{it-1}	0.018	0.004
$h^2_{it-1}/10^3$	-0.064	0.019
$h^3_{it-1}/10^7$	1.138	0.381
$h^4_{it-1}/10^{10}$	-1.030	0.381
$h^5_{it-1}/10^{14}$	4.854	1.950
$h^6_{it-1}/10^{18}$	-11.270	4.850
$h^7_{it-1}/10^{20}$	0.101	0.046
a_{it-1}	0.482	0.114
a^2_{it-1}	-0.006	0.001
black	-5.534	0.366
hispanic	-4.167	0.570
some college	2.497	0.442
college	5.812	0.457
more than college	8.203	0.500
l_{it}	5.738	0.220
x_{it}	0.207	0.084
constant	-2.095	2.620

Notes: Estimation of equation (3.15). Random effects regression of gross-income on covariates. m_{it} is measured in thousands of real dollars of 2000. Health is given by the CD4 count measured in hundreds of cells per microliter.

Even with heavy subsidization in the HIV treatments market, individuals wanting to consume a product must pay part of the cost and this is reflected in higher expected payments. Labor market participation increases expected payments, which may reflect different pricing schemes for public versus private insurance.

Labor participation is stochastic in our model and it is revealed to individuals at the beginning of the period. Estimates in Table 3.8 show that the log odds ratio of working versus not working increases with age until about age 40 and then decreases. The odds of working increase with education and they increases substantially if the individual had worked the previous period. At the end of every period individuals face the possibility of death. Estimates in Table 3.9 imply that the log odds ratio of death versus survival decreases with age until about age 35 and then increases. The likelihood of death is smaller for black individuals and for individuals who are not suffering ailments.

Table 3.7: Tobit Model for Out-of-pocket Payments, o_{it}

variable	coef.	se
h_{it-1}	-0.002	0.000
$h^2_{it-1}/10^3$	0.009	0.002
$h^3_{it-1}/10^7$	-0.133	0.029
$h^4_{it-1}/10^{10}$	0.090	0.022
$h^5_{it-1}/10^{14}$	-0.266	0.071
$h^6_{it-1}/10^{18}$	0.279	0.083
a_{it-1}	0.037	0.007
a^2_{it-1}	0.000	0.000
black	-0.240	0.021
hispanic	-0.119	0.025
some college	0.169	0.026
college	0.318	0.033
more than college	0.336	0.030
market product	0.429	0.026
trial product	0.313	0.043
l_{it}	0.105	0.016
x_{it}	-0.122	0.017
constant	-1.459	0.182
σ^o	0.862	0.066

Notes: Estimation of equation (3.16). $market\ product = \sum_{k=1}^{4} d_{kit}$. o_{it} is measured on thousands of real dollars of 2000. Health is given by the CD4 count measured in hundreds of cells per microliter.

3.5.4 Simulated Choice Dynamics and Model Fit

In Figure 3.10, we plot observed treatment choices over time along with those generated by the model.[33] In general, we are able to capture basic trends, including the rise in treatment usage as drugs improve through innovation. We also capture trials participation dynamics fairly well, but we have a harder time reproducing the spike in participation shortly before HAART introduction. The reason for this may be that, although our model accounts for changes in the demand for trials, there was also a shift in the supply of trials as a number of new drugs were tested that would eventually comprise HAART. Hence, the spike in partici-

[33]The fit of our parametric ccps is discussed in Appendix B.2

Table 3.8: Logit Model for Labor Supply, l_{it}

variable	coef.	se
h_{it-1}	0.009	0.001
$h_{it-1}^2/10^3$	-0.013	0.002
$h_{it-1}^3/10^7$	0.075	0.023
$h_{it-1}^4/10^{10}$	-0.013	0.007
a_{it-1}	0.102	0.032
a_{it-1}^2	-0.001	0.000
black	-0.168	0.073
hispanic	-0.040	0.125
some college	0.312	0.105
college	0.537	0.103
more than college	0.613	0.108
l_{it-1}	4.458	0.056
constant	-5.914	0.742

Notes: Estimation of equation (3.17). Health is given by the CD4 count measured in hundreds of cells per microliter.

Table 3.9: Logit model for Death, $1 - S_{it+1}$

variable	coef.	se
h_{it-1}	-0.028	0.003
$h_{it-1}^2/10^3$	0.079	0.015
$h_{it-1}^3/10^7$	-1.104	0.292
$h_{it-1}^4/10^{10}$	0.704	0.220
$h_{it-1}^5/10^{14}$	-1.610	0.561
a_{it-1}	-0.116	0.058
a_{it-1}^2	0.002	0.001
black	-0.509	0.199
hispanic	0.034	0.235
some college	0.060	0.185
college	-0.353	0.185
more than college	-0.512	0.207
x_{it}	-1.140	0.159
constant	1.682	1.358

Notes: Estimation of equation (3.21). Health is given by the CD4 count measured in hundreds of cells per microliter.

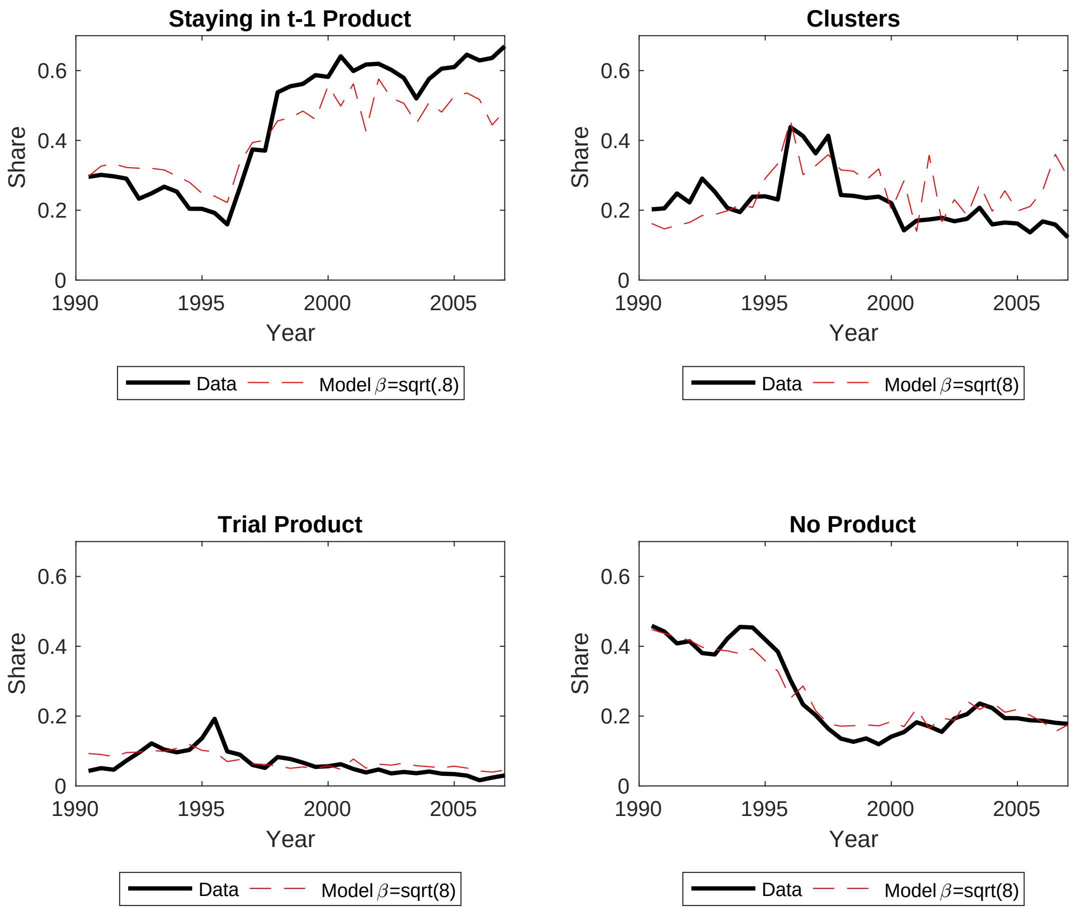


Figure 3.10: Goodness of Fit Figures

Notes: Simulated and empirical choice rates over time.

pation would not be fully captured by our model as it focuses on patient demand.[34] Beyond this spike, however, our model can capture the main contours of choice dynamics.

[34] In a companion paper, we model supply of trials more explicitly and demonstrate the the increase in trials prior to the introduction of HAART in part explains the observed spike in the likelihood of participation. In the current framework, we could model supply shifts in a reduced-form manner as a temporary decrease in the utility cost of joining a trial, which would reflect the ease of finding a trial in which to participate. We abstract from supply here, however, since the focus of our model is on demand shifts and innovation.

3.6 Alternative Choice and Technology Paths

Conditional on an initial state, our estimated model can generate a distribution of technological paths. We start this section by illustrating this feature of our model and by assessing the likelihood of the observed technological path (Section 3.6.1). Next, we we go on to discuss how different kinds of choice dynamics would influence the distribution. We demonstrate that alternative market shares could speed or slow the development of technologies that potentially increase social welfare (Section 3.6.2). This naturally leads to a discussion of policies that could raise consumer welfare, which we examine in Section 3.7

3.6.1 The Distribution of Technology Paths

Imbedded in our estimation procedure is the simulation of different innovation paths that are generated by the same distribution of innovations that yields the realized path. This means that conditional on an initial aggregate state, we can contrast the realized technological path against the full distribution of paths. In particular, we take the 1990 distribution of state variables (before HAART introduction) as given and simulate forward 1000 paths of technology, choices and state variables for 18 years. We repeat the procedure using the 1997 distribution of state variables, once HAART has been introduced. We plot the mean of the simulated distributions across time, and contrast it against the realized trends in the data.

First, we consider the path of aggregate health and the path of the health component of the centroid. Results are plotted in Figure 3.11, where the green line is the realized path, the black line is the mean simulated path and the dotted lines are bands of one standard deviation. Considering the plots on the left, where the simulation begins in 1990, it is clear that HAART introduction was a tail event. The observed path of innovations follows the simulated paths quite well until 1996. Thereafter, the health centroid, which summarizes effectiveness of market drugs weighted by their share, and the aggregate health

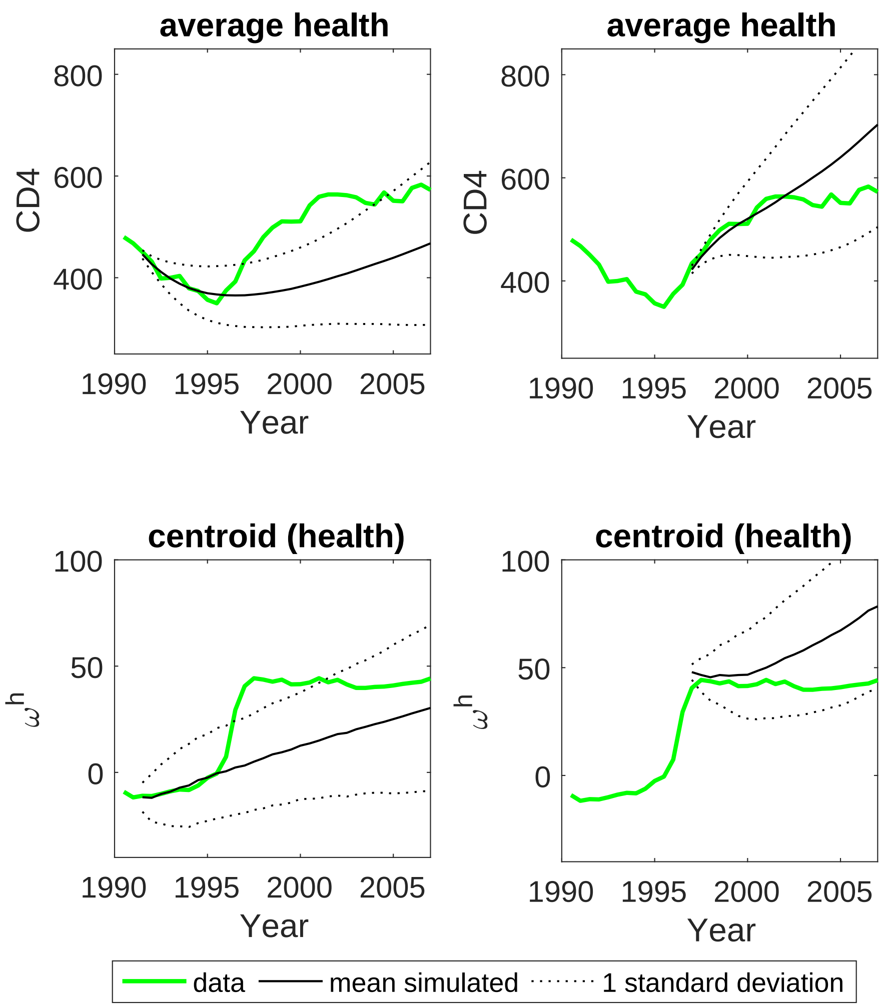


Figure 3.11: Distribution of Technology Paths: Health

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

of consumers in the market are far above what would have been expected. Between the years 1996 and 2000, the realized path of the centroid is outside of one-standard deviation band. Interestingly, the expected centroid approaches the realized centroid as time goes on. This means that the gradual progress of technology was expected to improve drug effectiveness until eventually something nearly as good as HAART would have come along. However, the timing was different because products with efficacy similar to HAART treatments were expected to appear far later than it actually did. Looking at the right side of Figure 3.11, where the simulation begins in 1997, notice that the realized path underperforms the average simulated path. This means that technology, measured by effectiveness, was expected to improve more than it did from the perspective of 1997.

In Figure, 3.12, we perform a similar analysis for the other product quality: lack of side effects. Here, realized product quality measured by the centroid, as well as aggregate ailments, seems to have underperformed what would have been expected from the distribution of innovations. In fact, one of the disappointments with regard to early versions of HAART treatments is that their side effects were quite harsh, which led many HIV+ men to avoid using them despite their effectiveness (Papageorge, 2016). In Figure 3.13 we consider the paths of survival and consumption. The results match those on health: HAART introduction was a tail event, which increased survival and product consumption. Finally, in Figure 3.14, we compare simulated paths with the realized path in terms of product entry and exit. The realized path of product entry is often outside the one standard deviation band around the mean. The realized path of product exit is within the bands but is often below the average path. In other words, the entry path with several high-entry periods seems unlikely given our distribution. Given our results in Figure 3.8, we argue that our under-estimation of entry does not mean that our model fits data poorly. Rather, our model is successful at treating breakthroughs (and subsequent entry of products) as tail events.

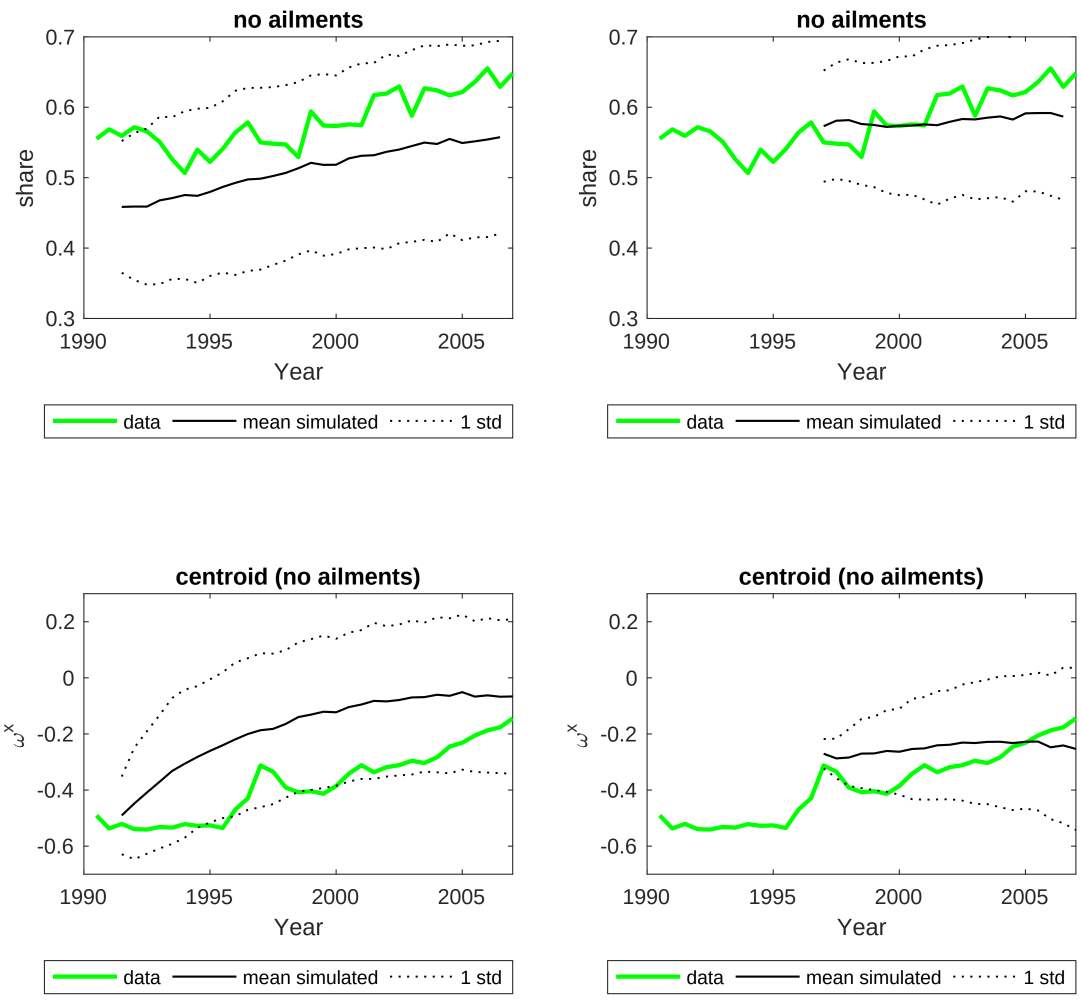


Figure 3.12: Distribution of Technology Paths: No Ailments

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

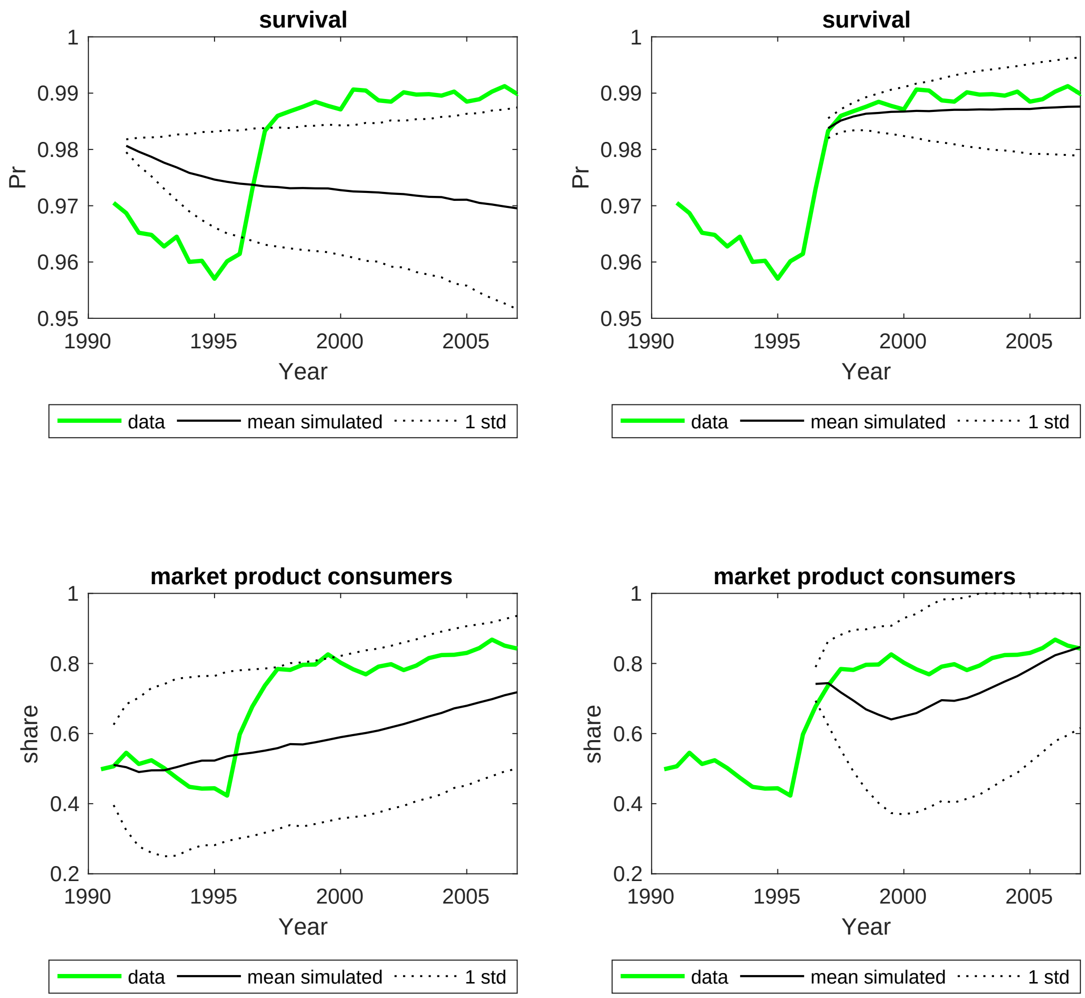


Figure 3.13: Distribution of Technology Paths: Survival and Product Consumption

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

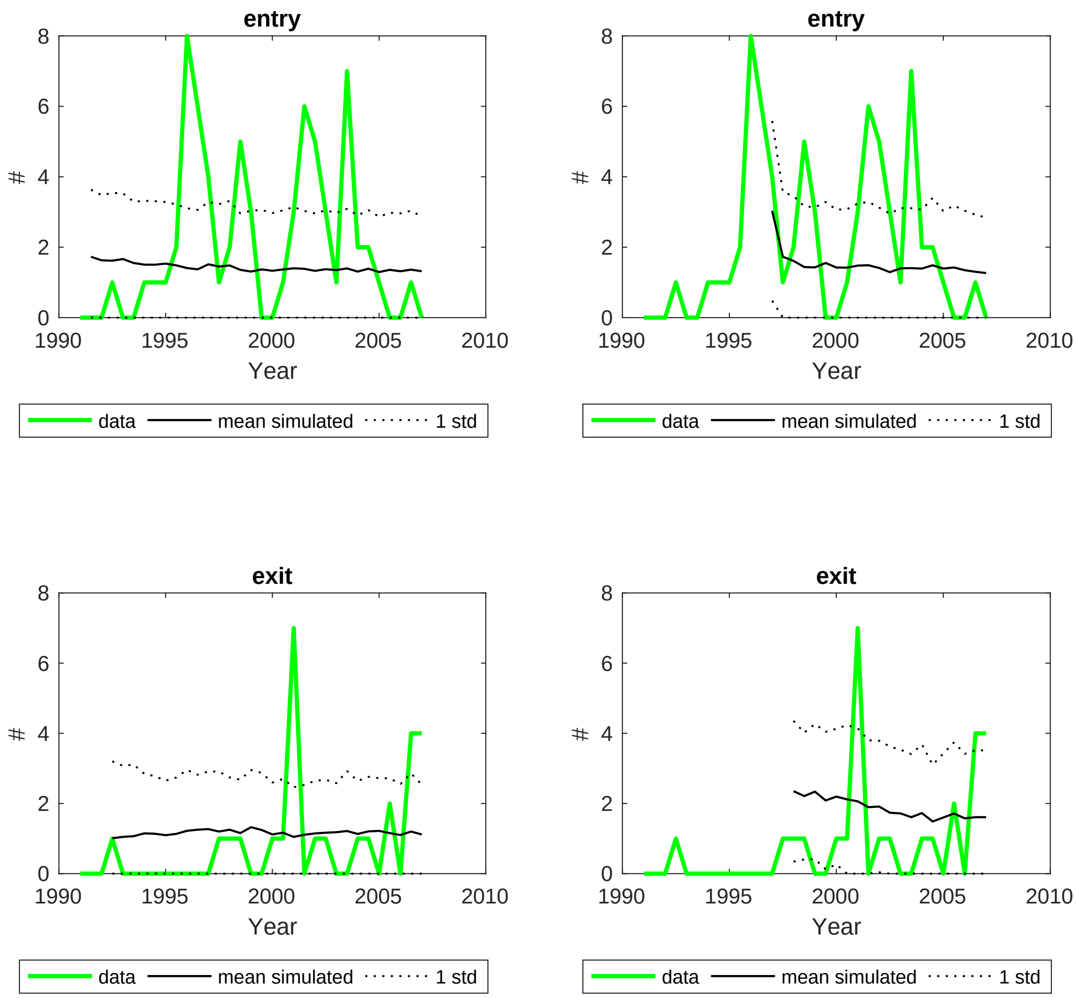


Figure 3.14: Distribution of Technology Paths: Entry and Exit

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

3.6.2 Demand Pull: How Consumer Choices Affect Innovation

Findings from this section demonstrate that a policy that changes consumer choices will affect the path of innovation. We consider the evolution of technology and aggregate outcomes under two choice regimes taking as initial condition the state of the world in 1991. In the first, just like in our model, individuals are dynamic optimizers. The distribution of technological paths under this regime was obtained in Section 3.6.1. In the second regime consumers choose options at random, which neutralizes the dependence of the technological path on the preferences and characteristics of consumers that would otherwise determine their optimal choices.

In Figure 3.15, we consider the mean—over 1000 simulated paths—of average consumer health and ailments under each regime (left-hand-side plots) and the mean health and ailments components of the centroid (right-hand-side plots). In Figure 3.16, we consider the mean—over 1000 simulated paths—of survival, share of consumers, entry and exit. Results in Figure 3.15 show that the random choice regime outperforms dynamic optimal choice in all but physical ailments technology. This occurs because the random choice regime yields higher product entry as a consequence of higher experimentation due to the randomization of choices (see Figure 3.16). As we showed in Table 3.3, new products are on average better in health terms due mostly to the second mode of the distribution of innovations. Since efficacy moves faster in the random choice regime, consumer ailments improve through the health channel instead of being a consequence of improvements in the ailments characteristics of products (see equation (3.19)). Figure 3.16 also shows the dynamic optimal regime generate lower survival rates. This result underscores how individuals value their quality of life beyond solely caring about health, and it also shows how individuals' preferences tilt the path of technology, in this case towards fewer side effects early on. In general, these results suggest that dynamic payoffs could rise overtime through technology improvements under

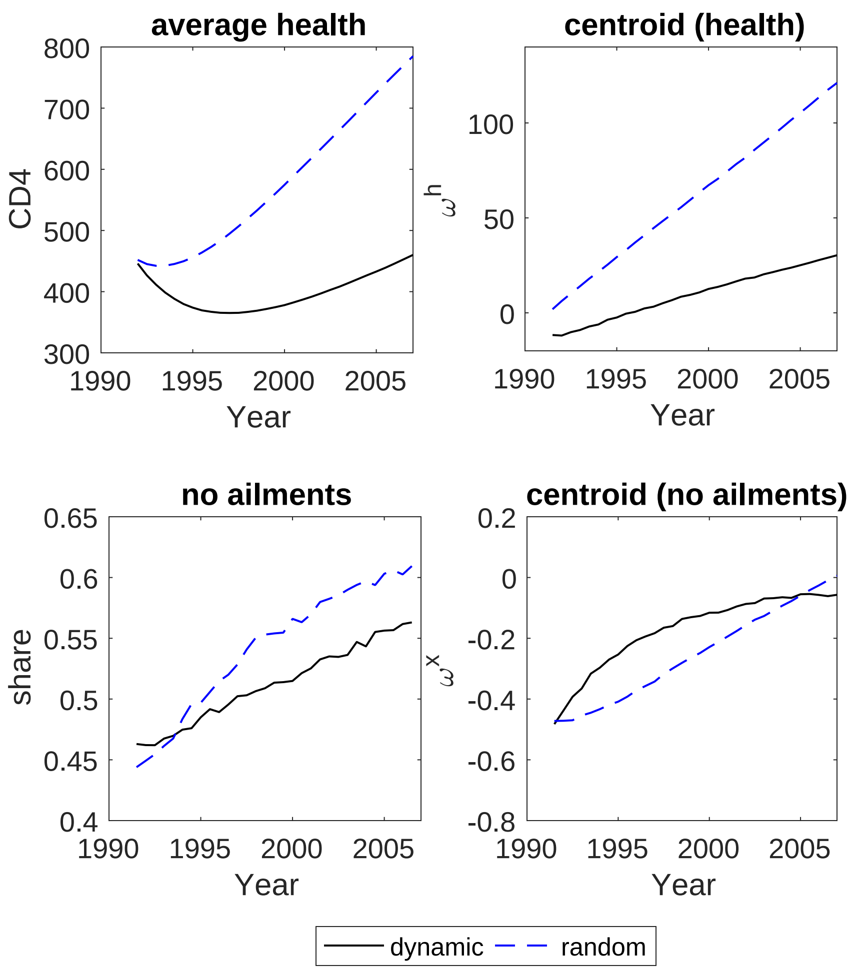


Figure 3.15: Alternative Choice Regimes: Health and No Ailments

Notes: Alternative choice regimes are: (i) optimal dynamic choice and (ii) random choice. Mean over 1000 simulated paths of the relevant statistic conditional on the state of the world at 1991.

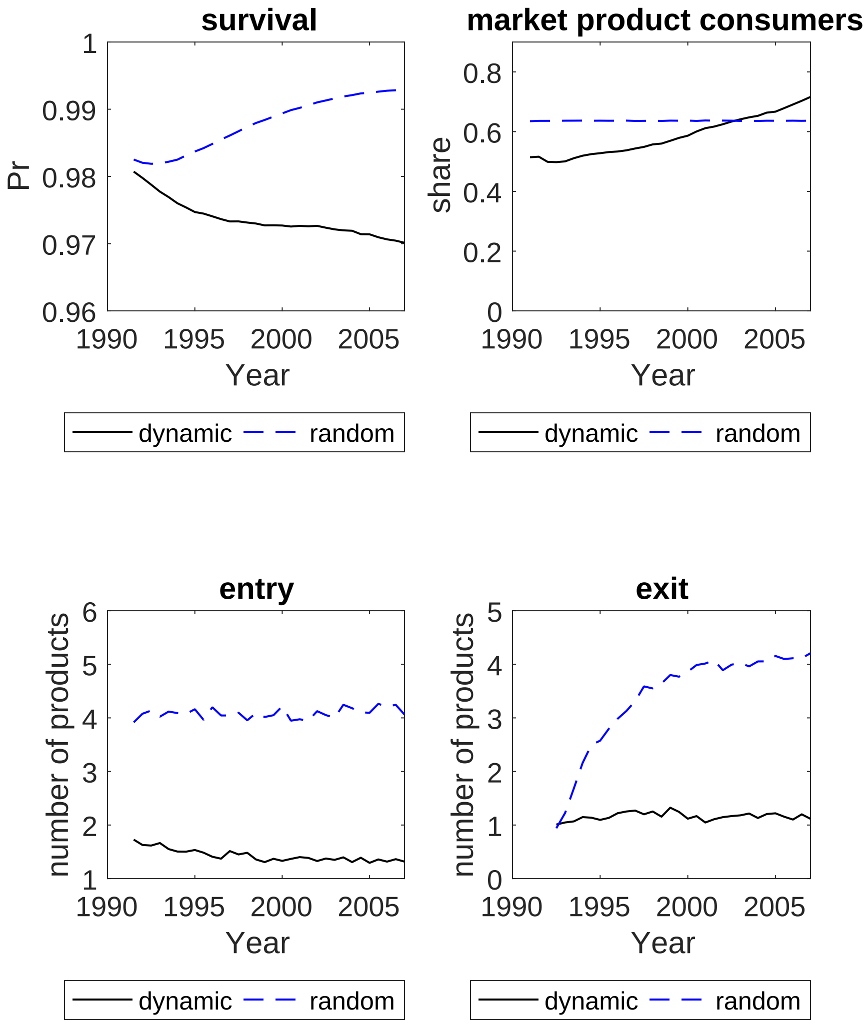


Figure 3.16: Alternative Choice Regimes: Survival, Consumption, Entry, and Exit

Notes: Alternative choice regimes are: (i) optimal dynamic choice and (ii) random choice. Mean over 1000 simulated paths of the relevant statistic conditional on the state of the world at 1991.

choice regimes that are inconsistent with individual dynamic optimization. We explore these possibilities for welfare improvement in the next section.

3.7 Demand Externalities and a Constrained Planner

Results in the previous section provide evidence suggestive of an externality whereby individually rational, optimal choices slow the path of technological progress in the direction of treatment efficacy. Ultimately, this can translate into welfare loses because individuals do not incorporate the impact of their choices in the process of innovation. By definition, an unrestricted social planner acting at every period can achieve at least the same level of welfare that atomistic individuals attain: she can let individuals choose whatever they find individually optimal. In general, an unrestricted planner will do better than that. However, given the size of the individual's state space, numerically solving the problem of an unrestricted planner quickly becomes intractable. Instead, we hope to explore the nature of the externality using a one-period planner that is constrained to act only on the basis of a subset of the information contained in the individual's state.[35] Because the planner is constrained, she does not necessarily do better. Hence, if she was to do better, this fact would constitute an initial measure of the size of the externality.

The first constrained one-period planner we consider is one that assigns individuals to choices on the basis of their health and their previous treatment. In particular, when assigning individuals to choices, the planner considers two levels of health (high and low) and two types of previous treatment choice (market treatment and no treatment/trial treatment). Compared to the amount of information available to the individual in his state $\mathcal{Z}_{it}$, the amount of information that the constrained planner bases her assignment rules on is minimal. Nevertheless, the number of assignment rules that we need to evaluate is already relatively large. For every individual in a $\langle$*health*, *previous treatment*$\rangle$ category, the planner assigns one of the six choices available or she lets the individual act freely.[36] This set up amounts to $7^2 * 6^2 = 1764$ assignment rules.[37] Notice that, because the constrained planner has less information than the individual, it may be optimal for her to let some groups to act freely even though those groups will not be internalizing the externality. This set up is

[35] We consider one-period planners because they constitute an unexpected shock to the individuals. Therefore, we can still use the same ccps that we estimated in Section 3.4 because individuals do not have time to adjust to the new regime before it is already gone.

[36] If the individual did not consume a market treatment last period there are only five options the planner can assign to him.

[37] The planner can only impose to stay on treatment to those groups that consumed a market treatment last period.

attractive because it nests the dynamic optimal problem of atomistic individuals when the planner lets every group to act freely. We solve the problem of this one-period planner in the first semester of 1991.

Table 3.10: Constrained Planner: Assignment Rules

		Groups			
	Group Share	0.50	0.26	0.05	0.19
	Average Welfare ($1000)	highH, no/trial	highH, mk	lowH, no/trial	lowH, mk
Top ten rules	82.16	6	7	6	7
	81.81	6	7	6	4
	81.58	6	7	7	4
	81.56	6	7	7	6
	81.54	6	4	6	7
	81.41	6	4	6	4
	81.41	6	7	6	6
	81.23	6	7	5	7
	81.22	6	4	6	6
	81.19	6	4	7	7
Atomistic	69.48	7	7	7	7
Bottom ten rules	43.36	1	5	1	2
	43.33	3	5	3	2
	43.23	1	5	3	2
	43.20	1	5	2	2
	43.20	3	5	1	2
	43.09	2	5	2	5
	43.00	2	5	3	2
	42.92	2	5	1	5
	42.90	2	5	1	2
	42.89	2	5	3	5

Notes: The table shows the best and worst assignment rules, and the average welfare they generate. The groups are determined by health status (high or low) and by previous treatment status (consumed a market treatment or not). The population shares of each of the groups are shown on top of their labels. Numbers 1 to 3 correspond to the three clusters available at that period. 4 corresponds to staying in previous market treatment. 5 stands for trial and 6 stands for no treatment. Finally, 7 stands for individually optimal choice; in other words, the planner renounces to her right to impose a choice and lets the individual in the group decide based on their richer information.

Results from this exercise, in Table 3.10, suggest that the constrained planner can do better than atomistic agents who make individually rational choices. In fact, the constrained planner can increase welfare by about 20%, which is a first measure of the size of the externality. The best rules in the planner portfolio have certain characteristics. The constrained planner sends the healthy individuals who did not consume a market treatment last period, to the no treatment option. In most cases she does the same with the sick individuals who also did not consume a market treatment last period. In general, she uses the richer information, in terms of product quality, of those who consumed a market treatment last period, to either set them free to act or to make them stay in their current treatment. One of the top ten rules includes trial participation. In this rule the planner sends sick individuals who did not consumed a treatment last period to trials. The worst rules also contain some information. They are characterized by sending healthy individuals who had consumed market treatments to trials. This imposes a high experimentation cost on those who dislike experimentation the most (the healthy) and disregards the information they had acquired about the quality of the treatment they consumed last period. In general, the constrained planner attempts to use the treatment quality information of those who have it, setting them free, and assigns healthy individuals with no treatment information (who are also the majority) to the no treatment alternative, which is consistent with the low quality of products in the early nineties.

In our second exercise we explore to what extent experimentation, in this case, experimentation in clinical trials, is one of the channels through which the externality acts. In other words, if individuals are not incorporating the future benefits of experimenting, and they rationally expect a certain level of aggregate experimentation given the distribution of consumer characteristics, then they may not participate in clinical trials and the level of experimentation resulting from this process may be suboptimal. We consider another constrained planner to get a sense of this problem. This planner's only tool is to randomly send individuals to clinical trials. In other words, all she can do is to set a parameter q, which is the probability that everyone faces of being sent to a clinical trial. If an individual is not sent to a trial by the planner, he gets to decide freely what to do, which may entail joining a clinical trial. Therefore, this constrained planner's problem nests the dynamic optimal problem of atomistic individuals when the planner sets $q = 0$. We solve the problem of this

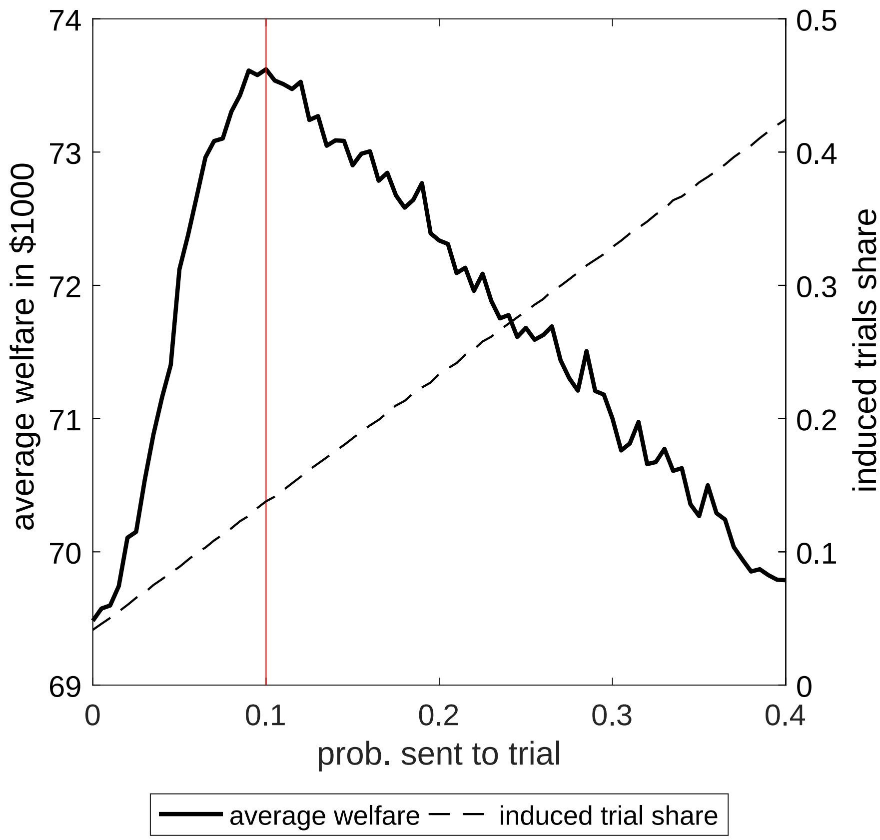


Figure 3.17: Constrained Planner: Random Assignment to Clinical Trials

Notes: Average welfare generated by a constrained planner that sends individuals to trials randomly. The constrained planner sets the value of the probability q that any individual is sent to a trial. If an individual is not sent to a trial, he may still choose to do so freely. This is reflected in the dotted line which represents the total share of trial participation yielding from a given planner policy. Beyond the value of the probability shown in the figure ($q = 0.4$) average welfare keeps declining. The lack of smoothness in the figure is a consequence of the low number of simulations relative to the size of the step from one value of q to the next.

one-period planner in the first semester of 1991. Figure 3.17 shows that the constrained planner can increase welfare by increasing experimentation randomly. However, if she imposes too much experimentation on the population, welfare will start declining because people do not like to experiment. According to Figure 3.17, the optimal level of q for the constrained planner is 0.10, which generates a share of trial participation of approximately 0.14.

This section underscores how, even with limited information or scope of action, a constrained planner can internalize the externality and substantially increase welfare per capita.

Moreover, we show that one of the mechanisms through which the externality acts is experimentation in clinical trials. Results in Section 3.5.1 showed that experimentation in clinical trials increases the expected value of the number of new products to be introduced to the market. Our one period-planner result shows that individuals may be experimenting sub optimally. In fact, one measure of the magnitude of the externality, acting through the channel of experimentation in clinical trials, is the increase in welfare per capita from a planner that optimally sets a probability with which to send individuals to trials randomly. The optimal rule chosen by this constrained planner generates an increase in welfare of about $4,000 per capita.[38]

3.8 Conclusion

It is often thought that new products that appear in the market shift consumer demand. New products with better characteristics obtain higher shares overtime as individuals discover them and decide to continue consuming them abandoning old products. However, the direction may also be reversed: how much innovation will happen tomorrow and on what dimension innovations are more likely to occur may depend on consumer demand.

We explore this idea, usually called "demand pull" in the literature, using data from the market for HIV treatments, which we observe basically since its beginning. We build a structural model with which we assess the role of demand pull and some of the mechanisms through which it operates. One of those mechanisms is experimentation. In the context of the market we study, experimentation takes a very specific form as individuals can join clinical trials. By joining a trial, individuals gain access to experimental products that may be breakthroughs of great quality, but that may also be less efficacious or unexpectedly toxic. Trial participation pushes the path of technology allowing new products

[38]Recall income is measured in six-months periods.

to appear. Moreover, the decisions of all consumers, favoring some products over others, bend the technological path in a certain direction as firms avoid innovating over unpopular products. Because individuals do not incorporate the consequences of their choices on the technological path, externalities appear and can be manifested, for instance, in less than optimal experimentation—individuals find experimenting to be costly.

Our results show that the data are consistent with demand pull and underscores the importance of accounting for the feedback from consumer demand into the innovation process. Consistent with previous literature, our results also show that consumers have multidimensional preferences when assessing their quality of life and how products can improve it. In particular, consumers of medical treatments care, not only about how efficacious treatments are, but also about how much side effects they cause. We show that consumer preferences along these multiple dimensions of quality can have effects on the direction and the speed of the technological path: consumer preferences towards less side effects slow down innovation on the dimension of efficacy which, in the long term, can have welfare consequences.

The reason why alternative technological paths may increase welfare is that individuals do not incorporate the consequences of their choices on the evolution of technology. We explore this issue by considering constrained planner problems that keep computation tractable, and provide initial measures of the size of the externality and to what extent it operates through experimentation. We find that a constrained planner can increase welfare by around 20 percent. We also find that a constrained planner that randomly sends people to trials would increase trial participation by about 10 percent points, which yields an increase in welfare of 5 percent.

As a consequence of the nature of our data, we focus on the demand side and on the consequences of consumer heterogeneity on the technological path. Other authors have paid more attention to the supply side of the problem (Carranza, 2010; Goettler and Gordon, 2011; Gowrisankaran and Rysman, 2012). A natural step forward, although by no means a simple one, is to model the competition of firms and their decisions to innovate by investing in R&D, while allowing for an acceptable level of consumer heterogeneity and its corresponding demand pull.

Bibliography

Aguirregabiria, V. and P. Mira, "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 2002, *70* (4), 1519–1543.

Allcott, Hunt, Rebecca Diamon, and Jean-Pierre Dubé, "The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States," Mimeo, New York University 2015.

Altonji, Joseph G. and Nicolas Williams, "Do Wages Rise with Seniority? A Reassessment," *ILR Review*, 2005, *58* (2), 370–397.

Altuğ, Sumru and Pamela Labadie, *Dynamic Choice and Asset Markets*, Academic Press, 1994.

__ **and Robert A. Miller**, "The Effect of Work Experience on Female Wages and Labour Supply," *Review of Economic Studies*, 1998, *65*, 45–85.

Anderson, T.W. and I. Olkin, "Maximum-likelihood Estimation of the Parameters of a Multivariate Normal Distribution," *Linear Algebra and its Applications*, 1985, *70*, 147 – 171.

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

Arcidiacono, Peter 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," Working Paper 2015.

__ **, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom**, "College Attrition and the Dynamics of Information Revelation," Working Paper, NBER 2016.

Bass, Frank M., "A New Product Growth for Model Consumer Durables," *Management Science*, 1969, *15* (5), 24–52.

Baumol, William J., "Entrepreneurship, Innovation and Growth: The David-Goliath Symbiosis," *Journal of Entrepreneurial Finance*, 2002, *7* (2).

Bayer, Patrick and Robert McMillan, "Tiebout Sorting and Neighborhood Stratification," *Journal of Public Economics*, 2012, *96* (11), 1129–1143.

Bhaskaran, K, O Hamouda, M Sannes, and et al, "Changes in the Risk of Death after HIV Seroconversion Compared with Mortality in the General Population," *Journal of the American Medical Association*, 2008, *300* (1), 51–59.

Bhattacharya, Jay and Mikko Packalen, "The Other Ex Ante Moral Hazard in Health," *Journal of Health Economics*, 2012, *31* (1), 135–146.

Bolton, Patrick and Christopher Harris, "Strategic Experimentation," *Econometrica*, 1999, *67* (2), 349–374.

Buera, Francisco J., "A dynamic model of entrepreneurship with borrowing constraints: theory and evidence," *Annals of finance*, 2009, *5* (3-4), 443–464.

Busenitz, Lowell W., "Entrepreneurial Risk and Strategic Decision Making: It?s a Matter of Perspective," *The Journal of Applied Behavioral Science*, 1999, *35* (3), 325–340.

Camerer, Colin and Dan Lovallo, "Overconfidence and Excess Entry: An Experimental Approach," *The American Economic Review*, 1999, *89* (1), 306–318.

Carranza, Juan Esteban, "Product Innovation and Adoption in Market Equilibrium: The Case of Digital Cameras," *International Journal of Industrial Organization*, 2010, *28*, 604–618.

Chan, T.Y. and B.H. Hamilton, "Learning, Private Information, and the Economic Evaluation of Randomized Experiments," *Journal of Political Economy*, 2006, *114* (6), 997–1040.

Chatterji, Aaron, Edward L. Glaeser, and William R. Kerr, "Clusters of Entrepreneurship and Innovation," in Josh Lerner and Scott Stern, eds., *Innovation Policy and the Economy, National Bureau of Economic Research*, Vol. 14, University of Chicago Press, 2013, chapter 5, pp. 129–166.

Crawford, G.S. and M. Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 2005, pp. 1137–1173.

DeGroot, Morris H., *Optimal Statistical Decisions*, McGraw Hill, 1970.

Dillon, Eleanor W. and Christopher T. Stanton, "Self-Employment Dynamics and the Returns to Entrepreneurship," Working Paper 2016.

Dranove, David, Craig Garthwaite, and Manuel Hermosilla, "Pharmaceutical Profits and the Social Value of Innovation," NBER Working Paper 2014.

Duda, R.O. and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

Erdem, Tülin and Michael P Keane, "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 1996, *15* (1), 1–20.

Evans, David S. and Boyan Jovanovic, "An estimated model of entrepreneurial choice under liquidity constraints," *The Journal of Political Economy*, 1989, pp. 808–827.

Fairlie, Robert W. and Javier Miranda, "Taking the Leap: The Determinants of Entrepreneurs Hiring Their First Employee," *Journal of Economics & Management Strategy*, Forthcoming.

Finkelstein, Amy, "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics*, 2004, *119* (2), 527–564.

Gayle, George-Levi and Robert A. Miller, "Has Moral Hazard Become a More Important Factor in Managerial Compensation?," *American Economic Review*, 2009, *99*, 1740–1769.

_ , Limor Golan, and Robert A. Miller, "Promotion Turnover and the Compensation in the Executive Labor Market," *Econometrica*, 2015, *83* (6), 2293–2369.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent, "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 2005, *23* (4), 681–724.

Goettler, Ronald L. and Brett Gordon, "Does AMD Spur Intel to innovate More?," *Journal of Political Economy*, 2011, *119* (6), 1141–1200.

Gottlieb, Joshua D., Richard R. Townsend, and Ting Xu, "Experimenting with Entrepreneurship: The Effect of Job Protected Leave," Working Paper, NBER 2016.

Gowrisankaran, Gautam and Marc Rysman, "Dynamics of Consumer Demand for New Durable Goods," *Journal of Political Economy*, 2012, *120* (6), 1173–1219.

Green, Edward J., "Lending and the Smoothing of Uninsurable Income," in E. Prescott and N. Wallace, eds., *Contractual Arrangements for Intertemporal Trade*, University of Minnesota Press, 1987, chapter 1, pp. 3–25.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Does Local Financial Development Matter," *The Quarterly Journal of Economics*, 2004, pp. 929–969.

Hamilton, Barton H., "Does entrepreneurship pay? An empirical analysis of the returns to self-employment," *The Journal of Political Economy*, 2000, *108* (3), 604–631.

_ , Nidhi Pande, and Nicholas W. Papageorge, "The Right Stuff? Personality and Entrepreneurship," Mimeo, Johns Hopkins University 2016.

Harris, Yvonne, Philip B Gorelick, Patricia Samuels, and Isaac Bempong, "Why African Americans May Not Be Participating in Clinical Trials.," *Journal of the National Medical Association*, 1996, *88* (10), 630.

Heckman, James J. and Bo E. Honoré, "The Empirical Content of the Roy Model," *Econometrica*, 1990, *58* (5), 1121–1149.

Hicks, John R, *The Theory of Wages*, London: Macmillan, 1932.

Hincapié, Andrés, "Where are the Young Entrepreneurs? A Study of Entrepreneurship over the Life Cycle," Mimeo, Washington University in St. Louis 2016.

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.

Hotz, V Joseph, Robert A Miller, Seth Sanders, and Jeffrey Smith, "A Simulation Estimator for Dynamic Discrete Choice Models," *The Review of Economic Studies*, 1994, *61*, 265–289.

Humphries, John Eric, "The Causes and Consequences of Self-Employment over the Life Cycle," Mimeo, University of Chicago 2016.

Hurst, Erik and Annamaria Lusardi, "Liquidity constraints, household wealth, and entrepreneurship," *The Journal of Political Economy*, 2004, *112* (2), 319–347.

_ **and Benjamin W. Pugsley**, "What Do Small Businesses Do?," *Brookings Papers on Economic Activity*, 2011, *2.*

_ **and** _ , "Wealth, Tastes, and Entrepreneurial Choice," Working Paper, NBER 2015.

Ickovics, Jeannette R and Andrew W Meisler, "Adherence in AIDS Clinical Trials: A Framework for Clinical Research and Clinical Care," *Journal of Clinical Epidemiology*, 1997, *50* (4), 385–391.

Iyigun, Murat F. and Ann L. Owen, "Risk, Entrepreneurship, and Human-Capital Accumulation," *The American Economic Review*, 1998, *88* (2), 454–457.

James, Jonathan, "Ability Matching and Occupational Choice," SSRN Scholarly Paper ID 1937879, Social Science Research Network, Rochester, NY October 2011.

Jamshidian, Mortanza and Robert I. Jennrich, "Standard Errors for EM Estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2000, *62* (2), 257–270.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *The Journal of Political Economy*, 1979, *87* (5), 972–990.

— **and Glenn M. MacDonald**, "Competitive Diffusion," *Journal of Political Economy*, 1994, *102* (1).

— **and Rafael Rob**, "Demand-Driven Innovation and Spatial Competition Over Time," *Review of Economic Studies*, 1987, *54* (1), 63–72.

— **and Yaw Nyarko**, "Learning by Doing and the Choice of Technology," *Econometrica*, 1996, *64* (6), 1299–1310.

Keane, Michael P. and Kenneth I. Wolpin, "The Career Decisions of Young Men," *The Journal of Political Economy*, 1997, *105* (3), 473–522.

Kerr, William R., Ramana Nanda, and Matthew Rhodes-Kropf, "Entrepreneurship as Experimentation," *Journal of Economic Perspectives*, July 2014, *28* (3).

Kreider, Rose M. and Renee Ellis, "Number, Timing, and Duration of Marriages and Divorces: 2009," Technical Report, U.S. Census Bureau. 2011.

Lafontaine, Francine and Kathryn Shaw, "Serial Entrepreneurship: Learning by Doing?," *Journal of Labor Economics*, 2016, *34* (S2), S217–S254.

Lancaster, K.J., "A New Approach to Consumer Theory," *The Journal of Political Economy*, 1966, *74* (2), 132–157.

Lazear, Edward P., "Entrepreneurship," *Journal of Labor Economics*, 2005, *23* (4), 649–680.

Levine, Ross and Yona Rubinstein, "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?," *Quarterly Journal of Economics*, Forthcoming.

Liang, James, Hui Wang, and Edward P. Lazear, "Demographics and Entrepreneurship," Working Paper, NBER 2014.

Magnac, Thierry and David Thesmar, "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 2002, *70* (2), 801–816.

Manso, Gustavo, "Experimentation and the Returns to Entrepreneurship," SSRN Scholarly Paper ID 2527034, Social Science Research Network, Rochester, NY November 2014.

March, James G. and Zur Shapira, "Managerial Perspectives on Risk and Risk Taking," *Management Science*, 1987, *33* (11), 418–438.

Margiotta, Mary M. and Robert A. Miller, "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review*, 2000, *41* (3), 418–438.

Miller, Robert A., "Job Matching and Occupational Choice," *The Journal of Political Economy*, December 1984, *92* (6), 1086–1120.

— , "Innovation and Reputation," *Journal of Political Economy*, 1988, *96* (4), 741–765.

Munk, Robert, "Does Self-Employment Pay? The Role of Timing and Voluntary Mobility," Mimeo, Ohio State University 2015.

Nanda, Ramana and Jesper B. Sørensen, "Workplace Peers and Entrepreneurship," *Management Science*, 2010, *56* (7), 1116–1126.

OECD, "Supporting Youth in Entrepreneurship," Seminar Summary, OECD, Paris 2013.

Oosterbeek, Hessel, Mirjam van Praag, and Auke Ijsselstein, "The Impact of Entrepreneurship Education on Entrepreneurship Skills and Motivation," *European Economic Review*, 2010, *54*, 442–454.

Papageorge, Nicholas W., "Why Medical Innovation is Valuable: Health, Human Capital, and the Labor Market," *Quantitative Economics*, 2016, *7* (3), 671–725.

Peterman, Nicole E. and Jessica Kennedy, "Enterprise Education: Influencing Students? Perceptions of Entrepreneurship," *Entrepreneurship Theory and Practice*, 2003, *28* (2), 129–144.

Petrin, A., "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 2004, *110* (4), 705–729.

Roll, Richard, "The Hubris Hypothesis of Corporate Takeovers," *The Journal of Business*, 1986, *59* (2), 1116–1126.

Romer, Paul M, "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 1986, pp. 1002–1037.

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *The Journal of Political Economy*, 1974, *82* (1), 34–55.

Roy, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers (New Series)*, 1951, *3* (2), 135–146.

Rust, John, "Structural Estimation of Markov Decision Processes," in R. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, Amsterdam: North Holland, 1994, chapter 51, pp. 3081–3143.

Ryan, Camille and Julie Siebens, "Educational Attainment in the United States: 2009," Technical Report, U.S. Census Bureau. 2012.

Sanders, Carl, "Skill Uncertainty, Skill Accumulation, and Occupational Choice," Mimeo, Washington University in St. Louis 2010.

Scherer, Frederic M, "Demand-Pull and Technological Invention: Schmookler Revisted," *The Journal of Industrial Economics*, 1982, pp. 225–237.

Schmookler, Jacob, *Invention and Economic Growth* 1966.

Schumpeter, Joseph A., *The Theory of Economic Development*, Cambridge, MA: Harvard University Press, 1911.

Shane, Scott, "Why Encouraging More People to Become Entrepreneurs is Bad Public Policy," *Small Business Economics*, 2009, *33*, 141–149.

Souitaris, Vangelis, Stefania Zerbinati, and Andreas Al-Laham, "Do Entrepreneurship Programmes Raise Entrepreneurial Intention of Science and Engineering Students? The Effect of Learning, Inspiration and Resources," *Journal of Business Venturing*, 2007, *22*, 566–591.

Stigler, G.J., "The Cost of Subsistence," *Journal of Farm Economics*, 1945, *27* (2), 303–314.

von Graevenitz, Georg, Dietmar Harhoff, and Richard Weber, "The Effects of Entrepreneurship Education," *Journal of Economic Behavior & Organization*, 2010, *76*, 90–112.

Waldfogel, Joel, "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets," *RAND Journal of Economics*, 2003, pp. 557–568.

Appendix A

Appendix to Chapter 2

A.1 Data Appendix

A.1.1 PSID Data

Chapter 2 uses data from the Panel Study of Income Dynamics (PSID). The PSID started in 1968 with a representative sample of about 18,000 individuals in 5,000 families in the United States. Information about these individuals and their descendants was collected yearly up to 1996, year after which the study became biennial. The study is restricted to white and black men between years 1968 and 1996. Survey information used include data on occupation, self-employment status, business ownership, incorporation status, labor income, business income, working hours, completed education, age, race, and marital status.

Setting the beginning of labor market careers. In order to account for the process of belief formation individuals must be observed from their entrance into the labor market. Potential experience is defined as

$$PotentialExperience = Age - CompletedEducation - 6$$

to set the beginning of individuals' labor market careers. First, the minimum potential experience for each individual is computed. Only those individuals whose minimum potential experience is at most 3 are kept. Then, the beginning of the individual's labor market career is set whenever

$$PotentialExperience = \begin{cases} 0 \quad if \quad \min PotentialExperience \leq 0 \\ k \quad if \quad \min PotentialExperience = k \in \{1, 2, 3\} \end{cases}$$

Self-employment. At any period, conditional on having declared to be working (or working for money) or only temporarily laid off, individuals answer the following question (or a slightly modified version of it):

"*On your main job, are you self-employed, are you employed by someone else, or what?*"

The answer options are "Someone else," "Both someone else and self," "Self-employed only," and "Don't Know." Entrepreneurs as defined as those individuals who have positive working hours and declare to be self-employed only. All other individuals with positive working hours are catalogued into one of the salaried occupations.

Occupation. The PSID provides the 3-digit occupation code from 1970 Census of Population which is build using the Alphabetical Index of Industries and Occupations issued June 1971 by the U.S. Department of Commerce and the Bureau of the Census was used for this variable. The PSID provides the following categorization of occupations

- Occupation 1: 1 - 195 Professional, Technical, and Kindred Workers
- Occupation 2: 201 - 245 Managers and Administrators, Except Farm
- Occupation 3: 260 - 285 Sales Workers

- Occupation 4: 301 - 395 Clerical and Kindred Workers
- Occupation 5: 401 - 600 Craftsmen and Kindred Workers
- Occupation 6: 601 - 695 Operatives, Except Transport
- Occupation 7: 701 - 715 Transport Equipment Operatives
- Occupation 8: 740 - 785 Laborers, Except Farm
- Occupation 9: 801 - 802 Farmers and Farm Managers
- Occupation 10: 821 - 824 Farm Laborers and Farm Foremen
- Occupation 11: 901 - 965 Service Workers, Except Private Household
- Occupation 12: 980 - 984 Private Household Workers

Observations corresponding to "members of the armed forces" (coded as 600) are dropped as well as observations of farm related occupations and observations of private household workers (occupations 9, 10 and 12). The remaining PSID categories are grouped into

- Blue Collar: Craftsmen and Kindred Workers; Operatives, Except Transport; Transport Equipment Operatives; Laborers, Except Farm; Service Workers, Except Private Household.
- White Collar: Professional, Technical, and Kindred Workers; Managers and Administrators, Except Farm; Sales Workers; Clerical and Kindred Workers

Individuals provide their occupation regardless of their self-employment status. However, provided that in the model entrepreneurship is an occupation on its own, the occupation data is disregarded whenever an individual is self-employed. Also, occupations were the individual reports working for more than 2.5% of the total amount of available hours in the year ($365.25 * 24$).

Up to 1980, the occupational data provided by the PSID is coded retroactively in order to correct for spurious transitions. PSID officials use original PSID reports and the three-digit 1970 Census occupation codes for a selected sample of PSID heads and spouses. Therefore, only part of the individuals' careers in the sample have been further corrected for spurious transitions. To the extent that the categories used in Chapter 2 are broad enough and that survey officials get more accurate cataloguing occupations over time, this problem should be minor in the sample. Figure A.1 shows the switching trends among collar occupations, entrepreneurship, and unemployment, computed with data before and after 1980; the decrease in switching corresponds to the fact that individuals in the sample are acquiring more experience. More importantly, there seems to be no evidence of jumps in the trend of switching before and after at this level of aggregation.

Labor Income. The PSID labor income variable is computed equally for employed and self-employed individuals. Up to 1993, it corresponds in general to the sum of wages (before taxes or other deductions) and "actual amounts of labor part of farm income and business income, bonuses, overtime, commissions, professional practice, labor part of income from roomers and boarders, and market gardening" (PSID Codebook). From this variable the following components are subtracted: the labor part of business income, of farm income, and of income from roomers and boarders when available. Starting from 1994, the labor part of farm income and that of business income are not included in the variable. Labor income is bracketed for 1968 and 1969. The midpoint value of the bracket is assigned; however, less than 1% of the individual-year observations correspond to those years. Also, the PSID labor income variable is censored at different upper values at different years. Less than 0.2% percent of the observations correspond to censored observations. The labor part of farm

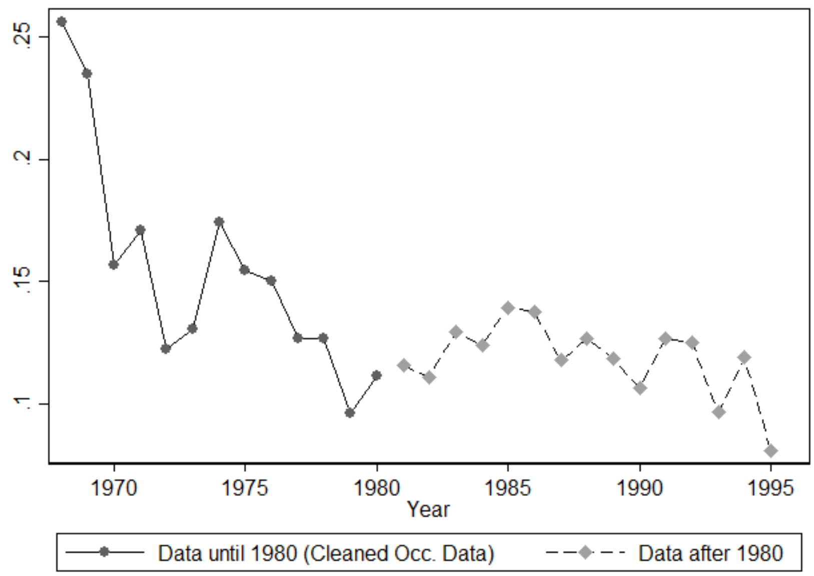


Figure A.1: Proportion of Individuals Switching between t and $t+1$ Over Years in the Labor Market (Collar)

income is bracketed until 1975. Again, the value of the midpoint of the bracket is assigned.

Business income. Business income is gathered for those individuals that satisfy the following two conditions:

- 1. They answer "yes" to the following question (or a modified version of it)

"*Did you or any other member of your family own a business at any time in year YYYY, or have a financial interest in any business enterprise?*"

Not all self-employed individuals answer "yes" to this question and not all individuals who answer yes to this question are self-employed. While about 82 percent of self-employed answer "yes" to this question less than 8 percent of salaried workers do. Regardless of this numbers, this may still be a drawback of how Chapter 2 treats the data.

- 2. They then proceed to say that the business mentioned was not uniquely a corporation. In other words, they proceed to say that the business was either (i) unincorporated or (ii) they have an interest in both types or (iii) they do not know.

If those two conditions are satisfied they then answer the question

"*How much was (your/his/her/their share of the total income from business in YYYY– that is, the amount (you/he/she/they) took out plus profit left in? [If zero: did you have a loss? How much was it?*"

Business income is computed as the sum of the labor and asset part of head's business income as reported in the PSID data. The labor part and asset part of business income are bracketed until 1975. Again, the value of the midpoint of the bracket is assigned. After computed, business income is added to the labor income measure only for unincorporated self-employed individuals.

Income. In summary, for salaried workers and incorporated self-employed individuals:

$$Income = LaborIncome$$

For self-employed unincorporated individuals:

$$Income = LaborIncome + BusinessIncome$$

Individuals who are not working any hours are assigned zero income. All values are in constant dollars of 2000.

Incorporated and unincorporated status. Following an affirmative answer to the business ownership question (above), individuals are asked about their incorporation status in all years in the PSID (denote this question IQ1). Also, in years 1975, 1976 and from 1985 onward, individuals are asked about their incorporated status after the self-employment question (denote this question IQ2). Even though question IQ2 seems closer to Chapter 2's definition of entrepreneurship, not all years are available for this question. An imputation algorithm is followed in order to determine the incorporated status of entrepreneurs.

In the imputation algorithm more relevance is given to stability and consistency of the measure across years. The imputation entails the following steps: **(i)** Initially, the incorporation status of entrepreneurs is determined from question IQ1; **(ii)** If incorporated status for entrepreneurs is missing or ambiguous (individual reported "Both," "Other," or "Do not know") in IQ1, the value from question IQ2 is assigned insofar as it corresponds to "Incorporated" or "Unincorporated;" **(iii)** If data is still missing or ambiguous, the $t-5,\dots,t-1,t+1,\dots,t+5$ (past and future) answers from IQ1 and IQ2 are used to assign the incorporated status at t; **(iv)** all remaining ambiguous observations are imputed as "unincorporated." Out of 2201 observations of entrepreneurs, this method imputes 551 observations: 406 from step **(ii)**, 120 from step **(iii)** and 25 from step **(iv)**.

Working Hours. The study uses the individual level PSID variable for working hours. It counts the actual hours worked by the individual during year YYYY. Missing data were not assigned.

Hourly income. Hourly income is computed simply as annual income divided by annual working hours.

Education. Consistent with the procedure for setting the beginning of individuals' labor market careers, the education variable corresponds to the value of completed education. Education data are discretized into: high school (12 years of education or less), some college

(13 to 15 years of education), college (16 years of education), more than college (more than 16 years of education). Education is censored at 17 years which may affect the potential experience criteria used above for setting the beginning of individuals' labor market careers.

Age and marital status. Reported age and marital status of individual.

Experience variables. Experience variables are computed using occupation data over the individual's career.

Wealth. The PSID includes a measure of wealth for selected years: 1984, 1989, 1994, and every two years starting in 1999. The wealth measure in the PSID is constructed as the sum of six types of assets (farm business, checking or savings accounts, real estate other than main home, stocks, vehicles, and other assets) net of debt value plus the value of home equity. Since the survey does not include data on wealth at every period, in the current analysis a measure of permanent wealth will be considered instead. In order to obtain the individual measure, the following fixed effects regression is run:

$$Wealth_{it} = \gamma_0 + \gamma_1 age_{it} + \gamma_2 age_{it}^2 + u_i + \epsilon_{it}$$

The individual measure for permanent wealth is then obtained as

$$\omega_i = \hat{\gamma}_0 + \hat{u}_i$$

In estimation only individuals with at least three wealth data points are used. Figure A.2 shows the age profile of wealth accumulation from pooling all available data.

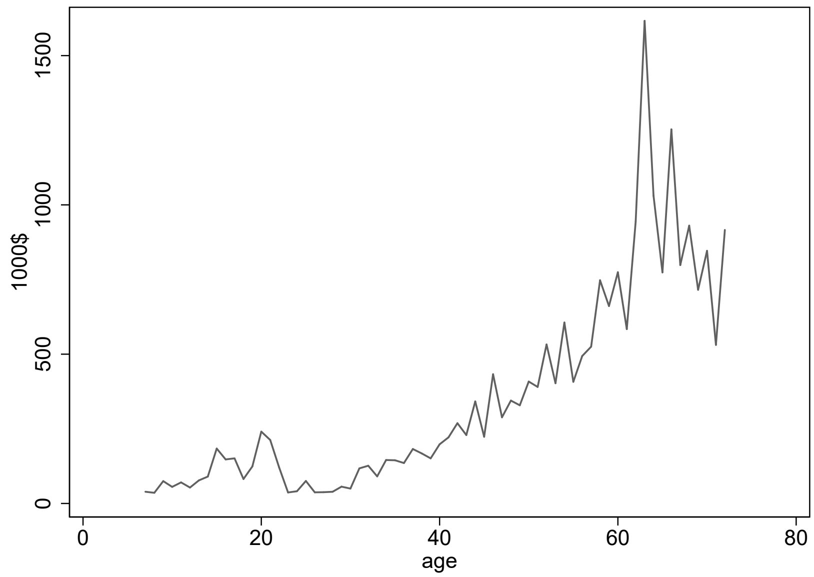


Figure A.2: Average Wealth

Notes: Average wealth in thousands of dollars of 2000.

Table A.1: Parameters of the Wealth Profile Equation

	coeff	se
γ_0	417.78	85.16
γ_1	-23.95	4.40
γ_2	0.47	0.06

Notes: wealth in thousands of dollars of 2000.

Full time vs part time workers. There is no differentiation in the treatment of the data between full-time workers and part time workers. In fact, only about 6% of individual-year observations for working individuals are part-time observations (less than 20 hours per week). Part-time individual-year observations are not dropped because they would create gaps in the careers of 36% of the individuals (See Table A.2 bellow).

Data Gaps. the histories of individuals with data gaps of more than 2 years are dropped.

Table A.2: Full time and Part Time observations (1968-1996)

	Full time	Part Time
% of All individual-year worker observations	0.95	0.05
% of unique individuals ever in part time		0.35

Notes: Table shows proportion of individual-year observations with less than 1044 working hours in a year (part time) and proportion of unique individuals that were ever part time workers by this criterion. Selection rule was defined as not dropping part time observations since they would create gaps by about 40 percent of all unique individuals' labor histories.

For those with data in t and $t+2$ but not in $t+1$, Time is then redefined by making $t+1=t+2$ and so forth. Similarly, for those with data in t and $t+3$ but not in $t+1$ and $t+2$, time is redefined by making $t+1=t+3$ and so forth.

Dropping Data Process. Initial number of individuals: 75,260. Individuals remaining after dropping individuals with no information on age, 75,153 for 3'457,038 individual-year observations. Individual-year observations remaining after keeping only household heads and their spouses: 446,242;[1] individual-year observations remaining after keeping black or white individuals: 424,497; individual-year observations remaining after dropping years after 1996: 326,455; individual-year observations remaining after dropping females: 146,083; individual-year observations remaining after dropping missing participation info: 132,248; individual-year observations after dropping missing marital status: 132,242. Individual-year observations satisfying potential experience criterion: 37,759; individual-year observations remaining after dropping data on missing occupations, farm related occupations, and private household workers: 30,006; individual-year observations remaining after dropping missing income: 29,676; individual-year observations remaining after dropping military occupations: 28,683; individual-year observations remaining after dropping jumps in data: 26,087; individual-year observations satisfying potential experience criterion after previous droppings: 25,152; individual-year observations of people who never worked: 47. After drop-

[1]Relevant data on income and occupation is only collected for household heads.

ping observations of individuals who lack data on relevant variables except wealth, data set contains 2,057 individuals and 25,105 individual-year observations. With this data set the first stage of the estimation procedure is undertaken. For the second stage, an extra dropping criterion is added to exclude those individuals with less than three data points of wealth. The final data set for estimation of the second stage contains 1,506 individuals and 21,334 individual-year observations.

A.1.2 Bond Price Data

Following Gayle and Miller (2009) the price of a bond is computed as the present value (in real terms) of a security (T-bill) which pays \$1 annually. Denote r_{it} the marginal annuitized yield from lengthening the bond one period by extending the maturity date from $t+i$ to $t+i+1$. Data comes from the Federal Reserve's Economic Research Data Base and is based on Treasury bills with maturities 1, 2, 3, 5, 7, 10, 20, and 30. Assume the marginal annuitized yield rate for any bond maturing over 30 years is the same as the 30-year rate. This yields b_t defined as

$$
\begin{aligned}
b_t &\equiv \sum_{s=1}^{\infty}\prod_{i=1}^{s}(1+r_{it})^{-1} \\
&= \sum_{s=1}^{30}\prod_{i=1}^{s}(1+r_{it})^{-1} + \prod_{i=1}^{30}(1+r_{it})^{-1}\sum_{s=31}^{\infty}(1+r_{30,t})^{s-30} \\
&= \sum_{s=1}^{30}\prod_{i=1}^{s}(1+r_{it})^{-1} + \frac{1}{r_{30,t}}\prod_{i=1}^{30}(1+r_{it})^{-1} \qquad \text{(A.1)}
\end{aligned}
$$

Then, for each date t, impute a yield curve using the data on newly issued bonds for various maturities. Then use a cubic spline for each date-maturity combination in the data to obtain imputations $\hat{r}_{it}$ for each date t and for all $i \in \{1, \ldots, 30\}$.

Step 1:

Use the annual compounding interest rate $\widetilde{r}_{st}$ (from the interpolated yield curve) to obtain b_t as

$$b_t = \sum_{s=1}^{30} \left(\frac{1}{1+\widetilde{r}_{st}} \right)^s + \frac{1}{r_{30,t}} \left(\frac{1}{1+\widetilde{r}_{30,t}} \right)^{30} \tag{A.2}$$

Step 2:

Given that r_{it} and $\widetilde{r}_{it}$ are nominal interest rates, b_t is adjusted by the deflator based on base year 2000. To reflect inflation let

$$\widetilde{b}_t = \frac{b_t}{deflator2000} \tag{A.3}$$

The series of $\widetilde{b}_t$ is the one used in estimation. Given the sample, the earliest bond price needed is for year 1968 and the last bond price needed is for year 2033. The last yield curve available is for year 2015. Hence in-sample bond prices can be obtained up to 2015. Given the bond prices in sample $\widetilde{b}_t$ for $t = 1954, \ldots, 2015$, a regression is run for $\widetilde{b}_{t+1}$ on $\widetilde{b}_t$ and in order to obtain out-of-sample prices $\hat{b}_t$ for $t = 2016, \ldots, 2033$.

A.2 Model Appendix

A.2.1 Claim about belief variance.

Claim: For any $t > t_{i0}$ the prior variance $\mathbb{V}_{it}$ is a deterministic function of the parameters of the population distribution variance matrix $\mathbf{\Delta}$ and the experience vector x_{it}.

Proof: wlog let $t_{i0} = 1$ and drop the i index. First, let $\tilde{\mathbf{\Delta}} = \mathbf{\Delta}^{-1}$ with characteristic component $\tilde{\delta}_{k,k'}$. For $t = 2$, after choosing occupation $j \in \{1, \ldots, 4\}$, from the updating rule in equation (2.7) we know that

$$\mathbb{V}_2^{-1} = \mathbf{\Delta}^{-1} + \Sigma_1 \tag{A.4}$$

where Σ_t is defined in (2.5). The off-diagonal components of $\mathbb{V}_2^{-1}$ are simply $\tilde{\delta}_{k,k'}$. The diagonal components can be written as $\tilde{\delta}_{k,k} + d_{k1}/\sigma^2_{\eta_k} = \tilde{\delta}_{k,k} + x_{k2}/\sigma^2_{\eta_k}$ characteristic components. Now suppose it is also true that the off-diagonal components of $\mathbb{V}_t^{-1}$ are $\tilde{\delta}_{k,k'}$ and the diagonal components are $\tilde{\delta}_{k,k} + x_{kt}/\sigma^2_{\eta_k}$. Then, using the updating rule in equation (2.7) again we obtain that

$$\mathbb{V}_{t+1}^{-1} = \mathbb{V}_t^{-1} + \Sigma_t \tag{A.5}$$

From the previous equation it is clear that the diagonal and off-diagonal components of $\mathbb{V}_{t+1}^{-1}$ can also be written as deterministic functions of the components of $\mathbf{\Delta}^{-1}$ and the vector of experience, x_{it+1}. An induction argument finishes the proof. *Q.E.D.*

A.2.2 Proof of proposition 2.1

Proof: The proof works by backwards induction. Consider the set up of his problem in the last period of his labor market career, T, in present value terms. Suppose that he has chosen alternative k at period T. His consumption and savings choice maximizes

$$\begin{aligned} - \quad & \alpha_{Tk}(h_T)\beta^T \exp\{-\rho c_T - \varepsilon_{Tk}\} - E_T\left[\lambda_{\tau(T+1)} b_{\tau(T+1)} v_{kT+1} \exp\left(\frac{-(\rho\xi_{T+1} + a_{\tau(T+1)})}{b_{\tau(T+1)}}\right)\Bigg| \mathbb{E}_T, h_T\right] \\ & s.t.\ \ E_T[\lambda_{\tau(T+1)}\xi_{T+1} | d_{Tk}, h_T, \mathbb{E}_T] + \lambda_{\tau(T)} c_T = \lambda_{\tau(T)}\xi_T \end{aligned} \tag{A.6}$$

His budget constraints shows the relation between the value of his wealth today, his consumption choice, and the expected value of his wealth tomorrow. If he works in occupation k he obtains income $\bar{L}_k y_{kt+1}$ at the beginning of his retirement age which is simply added to his wealth in equation (2.9). Following a similar procedure as in Margiotta and Miller

(2000, p. 680) the conditional value function of choosing alternative k is obtained as

$$V_{kT}(h_T, \mathbb{E}_T, \xi_T, a_{\tau(T)}, b_{\tau(T)}, \varepsilon_{kT}) =$$
$$-\lambda_{\tau(T)} b_{\tau(T)} \alpha_{kT}(h_T)^{1/b_{\tau(T)}} e^{-\varepsilon_{kT}/b_{\tau(T)}} E_T[v_{kT+1}|\mathbb{E}_T]^{1-1/b_{\tau(T)}} \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) \tag{A.7}$$

Integrating over ε_T and averaging over the 5 choices using the conditional choice probabilities yields

$$\begin{aligned} & V_T(h_T, \mathbb{E}_T, \xi_T, a_{\tau(T)}, b_{\tau(T)}) = \\ & \quad -\sum_{k=0}^{4} p_{kT}(h_T, \mathbb{E}_T) \lambda_{\tau(T)} b_{\tau(T)} \alpha_{kT}(h_T)^{1/b_{\tau(T)}} E_\varepsilon[e^{-\varepsilon^*_{kT}/b_{\tau(T)}}] E_T[v_{kT+1}|\mathbb{E}_T]^{1-1/b_{\tau(T)}} \\ & \quad \times \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) \\ = & \quad -\lambda_{\tau(T)} b_{\tau(T)} \exp\left(\frac{-(\rho\xi_T + a_{\tau(T)})}{b_{\tau(T)}}\right) A_T(h_T, \mathbb{E}_T) \end{aligned} \tag{A.8}$$

where

$$E_\varepsilon[e^{-\varepsilon^*_{kT}/b_{\tau(T)}}] \equiv E_\varepsilon[e^{-\varepsilon_{kT}/b_{\tau(T)}}|d_{kT} = 1]$$

and $A_T(h_T, \mathbb{E}_T)$ is defined as in equation (2.11) with $A_{T+1}(h_{T+1}, \mathbb{E}_{T+1}) \equiv 1$.

To finish the proof suppose that equations (2.10) and (2.11) hold for $t+1$. Then, at age t an individual who has chosen alternative k selects consumption and savings to maximize

$$
\begin{aligned}
- \quad & \alpha_{kt}(h_t)\beta^t \exp\{-\rho c_t - \varepsilon_{kt}\} \\
- \quad & E_t\left[\lambda_{\tau(t+1)} b_{\tau(t+1)} A_{t+1}(h_{t+1}, \mathbb{E}_{t+1}) v_{kt+1} \exp\left(\frac{-(\rho\xi_{t+1} + a_{\tau(t+1)})}{b_{\tau(t+1)}}\right)\middle| \mathbb{E}_t, h_t, d_{kt} = 1\right] \\
& \quad s.t.\ \ E_t[\lambda_{\tau(t+1)}\xi_{t+1}|d_{kt}, h_t, \mathbb{E}_t] + \lambda_{\tau(t)} c_t = \lambda_{\tau(t)}\xi_t
\end{aligned}
$$

Which yields an equation similar to equation (A.7):

$$
\begin{aligned}
& V_{kt}(h_t, \mathbb{E}_t, \xi_t, a_{\tau(t)}, b_{\tau(t)}, \varepsilon_{kt}) = \\
& -\lambda_{\tau(t)} b_{\tau(t)} \alpha_{kt}(h_t)^{1/b_{\tau(t)}} e^{-\varepsilon_{kt}/b_{\tau(t)}} E_t[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) v_{kt+1}|\mathbb{E}_t, h_t]^{1-1/b_{\tau(t)}} \\
& \times \exp\left(\frac{-(\rho\xi_t + a_{\tau(t)})}{b_{\tau(t)}}\right) \qquad \text{(A.9)}
\end{aligned}
$$

The proof is finished by integrating over ε_t and averaging over the 5 choices using the conditional choices probabilities. *Q.E.D.*

A.2.3 Proof of Proposition 2.2

Proof: Assuming that the taste shocks are distributed Extreme Value Type-I renders the expression in equation (2.12) as a standard logit. Hence, the odds ratio can be written as

$$
\frac{p_{0t}(h_t, \mathbb{E}_t)}{p_{kt}(h_t, \mathbb{E}_t)} = \alpha_{kt}(h_t) E_t\left[\frac{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1})}{A_{t+1}(h_t, \mathbb{E}_t)} v_{kt+1}\middle| \mathbb{E}_t, h_t\right]^{b_{\tau(t)}-1} \qquad \text{(A.10)}
$$

Equation (A.10) describes the likelihood ratio of any choice relative to the choice of not working. The reason why the arguments of the index in the denominator are subscripted with t is that neither the individual's human capital vector nor his beliefs change if he decides not to work. Use equation (A.10) to write

$$E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t),\mathbb{E}_{kt+1})v_{kt+1}\middle|\,\mathbb{E}_t,h_t\right]^{1-1/b_{\tau(t)}} = \alpha_{kt}(h_t)^{-1/b_{\tau(t)}}A_{t+1}(h_t,\mathbb{E}_t)^{1-1/b_{\tau(t)}}\left(\frac{p_{kt}(h_t,\mathbb{E}_t)}{p_{0t}(h_t,\mathbb{E}_t)}\right)^{-1/b_{\tau(t)}} \tag{A.11}$$

From page 3 in the online appendix of Gayle et al. (2015):

$$E_\varepsilon[e^{-\varepsilon^*_{kt}/b_{\tau(t)}}] = p_{kt}(h_t,\mathbb{E}_t)^{1/b_{\tau(t)}}\Gamma\left(\frac{b_{\tau(t)}+1}{b_{\tau(t)}}\right) \tag{A.12}$$

where $\Gamma(\cdot)$ denotes the complete gamma function. Substitute equations (A.11) and (A.12) in equation (2.11) to obtain

$$A_t(h_t,\mathbb{E}_t) = p_{0t}(h_t,\mathbb{E}_t)^{1/b_{\tau(t)}}\Gamma\left(\frac{b_{\tau(t)}+1}{b_{\tau(t)}}\right)A_{t+1}(h_t,\mathbb{E}_t)^{1-1/b_{\tau(t)}} \tag{A.13}$$

Using equation (A.13) we can write the ratio of human capital and beliefs indexes as

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t),\mathbb{E}_{kt+1})}{A_{t+1}(h_t,\mathbb{E}_t)} = \frac{p_{0t+1}(h^{(1)}_{kt},\mathbb{E}^{(1)}_{kt})^{1/b_{\tau(t)+1}}A_{t+2}(h^{(1)}_{kt},\mathbb{E}^{(1)}_{kt})^{1-1/b_{\tau(t)+1}}}{p_{0t+1}(h^{(1)}_{0t},\mathbb{E}^{(1)}_{0t})^{1/b_{\tau(t)+1}}A_{t+2}(h^{(1)}_{0t},\mathbb{E}^{(1)}_{0t})^{1-1/b_{\tau(t)+1}}} \tag{A.14}$$

where $h^{(1)}_{kt}$ and $\mathbb{E}^{(1)}_{kt}$ indicate the value of the state variables at future age $t+1$, conditional on the decision path described by making $d_{kt}=1$. In general, define $h^{(s)}_{kt}$ and $\mathbb{E}^{(s)}_{kt}$ as the value of the state variables at future age $t+s$, conditional on the decision path described by making $d=1$ for all $d\in\{d_{kt},d_{0t+1},\ldots,d_{0T}\}$ and define

$$\phi_t(s) = \frac{1}{b_{\tau(t)+s}}\prod_{r=1}^{s-1}(1-1/b_{\tau(t)+r}) \tag{A.15}$$

Iterative substitution of equation (A.13) in (A.14) up to retirement age yields

$$\frac{A_{t+1}(\bar{H}_{kt+1}(h_t),\mathbb{E}_{kt+1})}{A_{t+1}(h_t,\mathbb{E}_t)} = \prod_{s=1}^{T-t}\left(\frac{p_{0t+s}(h^{(s)}_{kt},\mathbb{E}^{(s)}_{kt})}{p_{0t+s}(h^{(s)}_{0t},\mathbb{E}^{(s)}_{0t})}\right)^{\phi_t(s)} \tag{A.16}$$

Plugging equation (A.16) into equation (A.10) and applying logarithms finishes the proof.

Q.E.D.

A.3 Estimation Appendix

A.3.1 First Stage Detailed

The first stage uses an Expectation-Maximization algorithm. The EM algorithm is an iterative method that yields maximum likelihood estimates when a portion of the relevant data is unobserved. In the model, the unobserved part of the data is $\mathcal{M}_i$. In order to implement the EM algorithm assume $\mathcal{M}_i$ is observed for all i. Hence, the log likelihood of the data for individual i is

$$\begin{aligned} \ln \mathcal{L}_i &= \sum_{t=t_{i0}}^{T_i} \sum_{k=0}^{4} d_{kit} \ln \Pr\left[d_{kit} = 1 | h_{it}, \mathbb{E}_{it}; \Lambda, \Theta\right] \\ &\quad + \sum_{t=t_{i0}}^{T_i} \sum_{j=0}^{4} d_{jit} \ln \Pr\left[y_{jit+1} | h_{it}, \mu_j; \Theta\right] \\ &\equiv \ln \mathcal{L}_i^P + \ln \mathcal{L}_i^y \end{aligned} \tag{A.17}$$

Since $\ln \mathcal{L}_i$ is additively separable, $\ln \mathcal{L}_i^y$ is used to consistently estimate Θ and $\mathbf{\Delta}_s$ using the EM algorithm. Implementation of the EM algorithm iterates over two steps to obtain maximum likelihood estimates. The expectation step at the mth iteration requires computation of the expectation of $\ln \mathcal{L}_i^y$ conditional on the observed data and the parameters at the mth iteration. The maximization step finds the new iterated value of the vector of parameters by maximizing the expression obtained in the expectation step.

Expectation Step. Using Bayes' rule (see Ch. 9 in DeGroot (1970) and James (2011)), the conditional distribution of $\mathcal{M}_i$ for an individual with education level s at the m^{th} iteration, based on the observed data, is $N(\mathbb{E}_i^m, \mathbb{V}_i^m)$ where

$$\mathbb{E}_i^m = \left((\boldsymbol{\Delta}_s^m)^{-1} + \Psi_i\right)^{-1} \mathbf{W}_i \tag{A.18}$$

$$\mathbb{V}_i^m = \left((\boldsymbol{\Delta}_s^m)^{-1} + \Psi_i\right)^{-1} \tag{A.19}$$

where the kth component of the $\mathbf{W}_i$ vector is

$$\mathbf{W}_{i\{k\}} = \frac{\sum_{t=1}^{T} d_{kit}\left(y_{kit} - h_{it}'\theta_k\right)}{\sigma_k^{2,m}}$$

and the diagonal components of the square matrix Ψ_i are

$$\Psi_{i\{k,k\}} = \frac{\sum_{t=1}^{T} d_{kit}}{\sigma_k^{2,m}}$$

The off-diagonal terms of Ψ_i are all zeros. Given μ_{ki} and the distribution of η_{kit}

$$\begin{aligned} \log \Pr\left[y_{it}|h_{it}, \mu_k; \Theta\right] &= \log\left(\frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{\frac{-\left(y_{kit} - h_{it}'\theta_k - \mu_{ki}\right)^2}{2\sigma_k^2}\right\}\right) \\ &= -\frac{1}{2}\log\left(2\pi\sigma_k^2\right) - \frac{1}{2\sigma_k^2}\left(y_{kit} - h_{it}'\theta_k - \mu_{ki}\right)^2 \end{aligned}$$

Therefore, the expectation step of the EM algorithm yields

$$\begin{aligned} E_m\left[\log \mathcal{L}_i^y\right] &= -\sum_{t=1}^{T}\sum_{k=1}^{4} d_{kit} \cdot E_m\left[\frac{1}{2}\log\left(2\pi\sigma_k^2\right) + \frac{1}{2\sigma_k^2}\left(y_{kit} - h_{it}'\theta_k - \mu_{ki}\right)^2\right] \\ &= -\sum_{t=1}^{T}\sum_{k=1}^{4} d_{kit}\left[\frac{1}{2}\log\left(2\pi\sigma_k^2\right) + \frac{1}{2\sigma_k^2}\left(\mathbb{V}_{i\{k,k\}}^m + \left(y_{kit} - h_{it}'\theta_k - \mathbb{E}_{i\{k\}}^m\right)^2\right)\right] \end{aligned} \tag{A.20}$$

where $E_m[\cdot]$ stands for the expectation over $\mathcal{M}_i$ using the distribution characterized by the parameters of the mth iteration conditional on the observed data.

Maximization Step. Following the expectation step, the maximization step entails maximizing (A.20) in order to obtain Θ^{m+1}. In fact, each θ_k^{m+1} is given by

$$\theta_k^{m+1} = \arg\min_{\theta_k} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{kit} \left(y_{it} - h'_{it}\theta_k - \mathbb{E}^m_{i\{k\}} \right)^2 \tag{A.21}$$

which yields

$$\theta_k^{m+1} = (H'W_kH)^{-1}H'W_kY_k$$

where H is the $[NT \times \# (\theta_k)]$ matrix that stacks together all values of h'_{it}, Y_j is the $[NT \times 1]$ matrix that stacks together all values of $y_{it} - \mathbb{E}^m_{i\{k\}}$, and W_k is the $[NT \times NT]$ diagonal matrix with d_{kit} in its diagonal. Using the FOCs from (A.20) and the estimated values of θ_j^{m+1}, $\sigma_k^{2,m+1}$ has the closed form solution

$$\sigma_k^{2,m+1} = \frac{\sum\limits_{i=1}^{N} \sum\limits_{t=1}^{T} d_{kit} \left(\mathbb{V}^m_{i\{k,k\}} + \left(y_{it} - h'_{it}\theta_k^{m+1} - \mathbb{E}^m_{i\{k\}} \right)^2 \right)}{\sum\limits_{i=1}^{N} \sum\limits_{t=1}^{T} d_{kit}} \tag{A.22}$$

A summary of the EM algorithm is

- Step 1: Given mth iteration values $\left\{\theta_k^m, \sigma_k^{2,m}\right\}_{k\in\{1,\ldots,4\}}$ and $\{\mathbf{\Delta}_s^m\}_{s\in\{1,\ldots,4\}}$, solve for $\mathbb{E}_i^m$ and $\mathbb{V}_i^m$ using (A.18) and (A.19).

- Step 2: Update population parameter $\mathbf{\Delta}_s^{m+1}$ for education level s as

$$\mathbf{\Delta}_s^{m+1} = \frac{1}{N_s} \sum_{i=1}^{N} \sum_{s=1}^{4} \delta_{is} \left(\mathbb{V}_i^m + \mathbb{E}_i^m \mathbb{E}_i^{m\prime} \right) \tag{A.23}$$

 where δ_{is} is an indicator of individual i having education level s and $N_s = \sum_i \delta_{is}$. Equation (A.23) follows from maximization of the expected value of the log likelihood of $\mathcal{M}_i$, $E_m[\log f(\mathcal{M}_i)]$.[2]

- Step 3: For each occupation $k > 0$, new iteration values θ_k^{m+1} are obtained using equation (A.21) and new iteration values $\sigma_k^{2,m+1}$ are obtained using equation (A.22).

The algorithm is initialized with arbitrary values and the steps are repeated until convergence under the criterion

$$\left\| \sum_{i=1}^{N} \log \tilde{\mathcal{L}}_i^{y,m+1} - \sum_{i=1}^{N} \log \tilde{\mathcal{L}}_i^{y,m} \right\| < \epsilon$$

where

$$\tilde{\mathcal{L}}_i^{y,m} = \int_{\tilde{\mathcal{M}}} \left\{ \prod_{t=t_{i0}}^{T_i} \prod_{j=1}^{4} \Pr\left[y_{jit+1} | h_{it}, \tilde{\mu}_j; \Theta^m\right]^{d_{jit}} \right\} dF\left(\tilde{\mathcal{M}}; \boldsymbol{\Delta}_s^m\right) \tag{A.24}$$

is computed using Monte Carlo integration. ϵ is set to be 1×10^{-4}

A.3.2 Second Stage Detailed

The second stage of the estimation procedure is initialized with flexible parametric versions of the future conditional choice probabilities estimated from the data, where the beliefs, estimated in the first stage, are also treated as data. In the model, individuals have perfect foresight about their marital status. However, their entire marital status vector up to period T is not always observed. Hence, their marital status histories are completed using a single marital status path constructed using the median age of first marriage at 1970 from the U.S. Census Bureau, Current Population Survey and the median marriage duration presented in Kreider and Ellis (2011).[3] Effectively it amounts to individuals getting married at age 23 and remaining married until age 50.

[2] See Anderson and Olkin (1985).

[3] For the median age at first marriage visit: `http://www.census.gov/hhes/families/data/marital.html`.

Maximization Step

At any iteration of the second stage, for a given set of estimated ccps, utility parameters are obtained from maximization of the log likelihood

$$\frac{1}{NT}\sum_{i}\sum_{t}\sum_{k=0}^{4} d_{kit}\ln p_{kit}(h_t, \mathbb{E}_t) \tag{A.25}$$

The expectation in the expression for $V_k(h_t, \mathbb{E}_t)$ in equation (2.17) can be written as

$$\begin{aligned}
&E_t\left[v_{kit+1}\prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(s)})}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{0it}^{(s)})}\right)^{\phi(s)}\middle|\, \mathbb{E}_{it}, h_{it}\right]\\
&= \int_{\zeta_k}\left\{\exp\left(\frac{-\rho\bar{L}_k y_{kit+1}(h_{it})}{b_{\tau(t+1)}}\right)\prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})}\right)^{\phi(s)}\right\} dF(\zeta_k|\mathbb{E}_{it}, h_{it})\\
&= \int_{\zeta_k}\left\{\exp\left(\frac{-\rho\bar{L}_k(f_k(h_{it}, \omega_i; \theta_k)+\zeta_k)}{b_{\tau(t+1)}}\right)\prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})}\right)^{\phi(s)}\right\} dF(\zeta_k|\mathbb{E}_{it}, h_{it})
\end{aligned} \tag{A.26}$$

I compute the value of (A.26) using Monte Carlo integration.

Given a value for ρ the model becomes a simple logit in the α parameters and the Monte Carlo integral is[4]

$$B_{kit}(\rho) = \frac{1}{S}\sum_{s}\left\{\exp\left(\frac{-\rho\bar{L}_k(f_k(h_{it}; \theta_k)+\zeta_k^s)}{b_{\tau(t+1)}}\right)A_{kit}(\zeta_k^s)\right\} \tag{A.27}$$

where

$$A_{kit}(\zeta_k^s) = \prod_{s=1}^{T-t}\left(\frac{p_{0it+s}(h_{kit}^{(s)}, \mathbb{E}_{kit}^{(1)}(\zeta_k^s))}{p_{0it+s}(h_{0it}^{(s)}, \mathbb{E}_{it})}\right)^{\phi(s)}$$

is a value which varies across signals, ζ_k^s, drawn for integration. The draws come from the

[4]Recall that the scale of parameters Θ, $\boldsymbol{\Delta}_s$, and ρ depends on the units in which income and consumption are measured. I express hourly income in \$10 units and consumption in \$1,000 units. Therefore, in estimation, instead of $\bar{L}_k$ I write $\bar{L}_k/100$.

distribution of the signal conditional on beliefs $\zeta_k^s = \mu_{ki} + \eta_{kit} \sim N\left(\mathbb{E}_{\{k\}it}, \mathbb{V}_{\{k,k\}it} + \sigma^2_{\eta_k}\right)$. Using the definition of $\alpha_{kt}(h_t)$ in equation (2.3), equation (2.17) can be rewritten as

$$V_k(h_t, \mathbb{E}_t) = -h'_{it}\alpha_k - C_{kit}(\rho) \tag{A.28}$$

where

$$C_{kit}(\rho) = (b_{\tau(t)} - 1)\ln B_{kit}(\rho)$$

Equation (A.28) is then substituted into (2.16). In estimation, the log likelihood is maximized conditional on a value of ρ. Search over values for ρ is then undertaken an the value that maximizes the log likelihood is selected. This procedure is faster that searching over all the parameter space at once because it avoids computing the Monte Carlo integral in (A.26) more than once for each value of ρ.

CCP Step

For a given value of utility parameters the model is solved backwards and new model-generated ccps are obtained. This new ccps are fed into the maximization step and new utility parameters are obtained. Given that each iteration is computationally intensive, the iterative process is stopped after 5 iterations because the minimum log likelihood is achieved in iteration 4. The Euclidean distance between the parameter vectors in iteration 4 and 5 is 9.4. Future version of the model may entail more iterations to ensure that the solution is not a local minimum. The relatively small distance between the parameter vectors and the fact that the first search was initialized from 10 different initial points suggests that the solution may not be local. Notice that the estimated parameters at each iteration are consistent since the ccps that initialize the process are themselves consistent.

A.4 Results Appendix

A.4.1 Solving the Model

As mentioned in the Estimation section, solution of the model is needed in order to provide new estimates of the conditional choice probabilities. The model is solved using the same representation obtained in Proposition 2.2 and summarized in equation (2.17). Notice that this representation is obtained as a function of the probability of not working in the future conditional on specific choice paths. Consistent, with this representation, for a given vector of estimated parameters, the value function is solved with the following recursive algorithm starting at $t = T$:

- *Step 1.* Obtain the value of the mapping $V_k(h_{it}, \mathbb{E}_{it})$ for a grid spanning the relevant state space using equation (2.17) and the future choice paths described in Proposition 2.2.[5]

- *Step 2.* Obtain relevant ccps for period t using equation (2.16).

- *Step 3.* Obtain parametric versions of the ccps for period t, Ω_t. Noting that only the not working ccps are needed, the parametric version is obtained using a non linear regression that minimizes the distance between the model ccps, $p_{0it}(h_{it}, \mathbb{E}_{it})$, from Step 2 and a parameterization given by $\exp(X_{it}'\Omega_t)/(1 + \exp(X_{it}'\Omega_t))$. X_{it} includes multiple interactions of components of the state.

- *Step 4.* If $t = t_0$, stop. Otherwise, set $t = t - 1$ and go back to Step 1 using the collection of parametric ccps obtained so far for the representation.

As a final product from the previous algorithm, a collection of future ccps, $\{\Omega_t\}_{t=t_0}^{T}$, characterizing the value function at any period t is obtained.

[5]Notice that at period T there is no future value of human capital and beliefs. Only the value of income to be received at $T + 1$. Hence, no future ccps are needed as the occupational choice becomes static.

A.4.2 Model Fit

In order to assess goodness of fit, an initial state is generated and the model is simulated forward using the collection of future ccps implied by the model, $\{\Omega_t\}_{t=t_0}^{T}$, that characterize the value function. For comparison against the data initial states are obtained drawing from the data under certain restrictions. First, a collection of initial states is drawn from the initial states observed in the data (race, education, entry age, year of entry, permanent wealth). To avoid the high volatility of bond prices before 1980, only years after 1980 are considered for the comparison. Second, only one marital status path is allowed: the one constructed in the Estimation Appendix. Third, ability for each individual drawn from the data is set to be the mean beliefs conditional on all the information available for him. Using Bayes' rule, this is equal to his beliefs in the last period the individual is observed. In order to increase precision, only individuals that are observed for at least ten years are used in the comparison against the data.

An initial measure of model fit is presented in Figure A.3. It shows that the estimated model replicates the choices well given the observed state. Table A.3 compares the transition matrices. In the model, occupations are less absorbing than in the data. However, consistent with the data, entrepreneurial occupations are on average less sticky than salaried occupations. Notably, the not working alternative is much less absorbing in the model, which suggests that there are barriers to exit unemployment that are not captured in the model. In terms of switching behavior, the model successfully captures the fact that most switching from salaried occupations happens within the salaried group. It also captures the fact that, whereas unincorporated individuals tend to switch in similar percentages into either salaried occupation, incorporated entrepreneurs tend to overwhelmingly switch into white collar work. Table A.4 compares descriptive statistics of occupational spells. Although, consistent with transition results, the model under-predict spell durations, it performs well in terms of the distribution of occupational spells across-occupations. At the beginning of their careers, the model over predicts the number of individuals starting as unemployed or blue collar workers, and under predicts the number starting as white collar workers. .

Table A.3: Transition Patterns: Observed and Simulated

Data

	blue collar	white collar	unincorporated	incorporated	not working
blue collar	0.87	0.08	0.02	0.00	0.02
white collar	0.07	0.89	0.02	0.01	0.01
unincorporated	0.11	0.10	0.73	0.04	0.01
incorporated	0.03	0.15	0.06	0.75	0.01
not working	0.37	0.15	0.03	0.00	0.44

Model

	blue collar	white collar	unincorporated	incorporated	not working
blue collar	0.74	0.18	0.04	0.01	0.03
white collar	0.22	0.71	0.05	0.02	0.01
unincorporated	0.23	0.21	0.53	0.02	0.01
incorporated	0.08	0.17	0.03	0.72	0.00
not working	0.62	0.25	0.05	0.01	0.07

Notes: Matrix entry i, j represents the proportion of people in occupation in row i who move into occupation in column j between t and $t+1$.

Table A.4: Spells: Observed and Simulated

Data

	all	blue collar	white collar	unincorporated	incorporated	not working
Total	4294	1707	1652	453	194	288
Percent		39.75	38.47	10.55	4.52	6.71
Duration	4.97	5.21	6.03	3.10	3.10	1.63
First		52.06	42.56	2.19	0.27	2.92
Tried		68.73	69.92	20.05	9.03	14.54

Model

	all	blue collar	white collar	unincorporated	incorporated	not working
Total	282999	114681	104919	34842	11183	17374
Percent		40.52	37.07	12.31	3.95	6.14
Duration	3.03	3.54	3.13	2.02	3.02	1.07
First		56.84	29.02	3.46	0.74	9.94
Tried		77.06	76.74	32.04	16.44	39.99

Notes: **Duration** is the average duration of spells in years. **First** is the percentage of first spells that belong to a particular occupation. **Tried** is the percentage of individuals who tried the occupation during their observed careers.

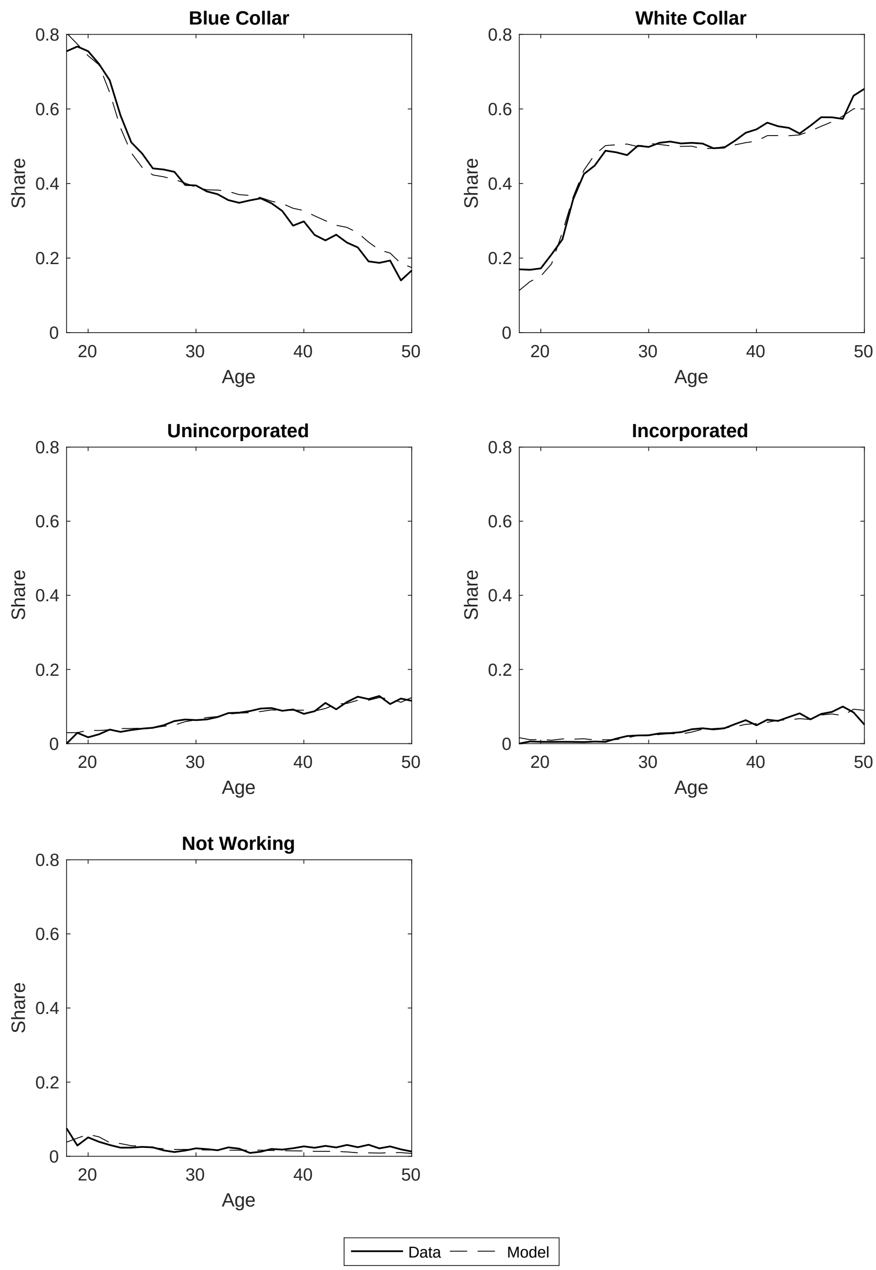


Figure A.3: Model Fit by Age

Notes: Actual and simulated choices by age.

Table A.5 shows the mean and variance of hourly income across all individuals who participate in each of the four occupations. With the exception of incorporated entrepreneurship, the model captures well the first two moments of the income distribution. For incorporated entrepreneurs, the model over predicts mean and variance. Notwithstanding this over prediction, the model respects the relative order in terms of which occupations generate higher income variance and which ones offer higher average income. Additionally, the model is able to capture the trend in the relation between the probability of switching into entrepreneurship from salaried occupations and the current income signal (see Figure 2.2). Figure A.4 shows that better signals in salaried occupations are negatively associated with the probability of switching into unincorporated entrepreneurship. The model captures the relative flatness of the relation between the signal and the probability of switching into incorporated entrepreneurship from blue collar work. Moreover, the increase in the probability of switching into incorporated entrepreneurship from white collar, for those receiving the best signals, is also captured by the model. Consistent with the excess switching shown in the simulated transition matrix, the model is unable to capture the level of the relations found in the data. Highlighting the role of correlated learning, Figure A.5 shows that a model that does not allow for correlated learning is unable to capture neither the level not the trend of this relation.

Table A.5: Income: Observed and Simulated

Data

	blue collar	white collar	unincorporated	incorporated
mean income	14.14	21.17	21.00	37.48
variance income	7.94	14.25	22.77	51.17

Model

	blue collar	white collar	unincorporated	incorporated
mean income	15.21	22.35	23.88	51.04
variance income	5.83	13.04	27.47	89.57

Notes: Quantities in dollars of 2000.

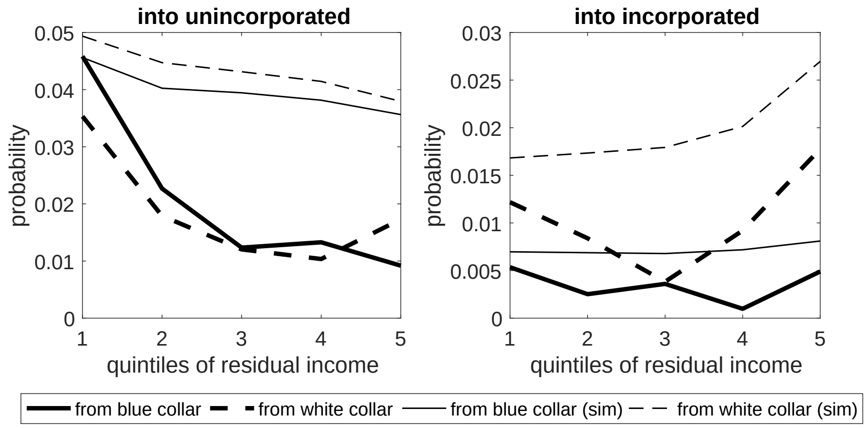


Figure A.4: Switching into Entrepreneurial Occupations (Correlated Learning)

Notes: Probability of switching into entrepreneurial occupations in $t+1$ by decile of residual income in t. Residual income is computed from occupation-specific regressions of hourly income on occupation-specific experience, general experience squared, race, education and marital status.

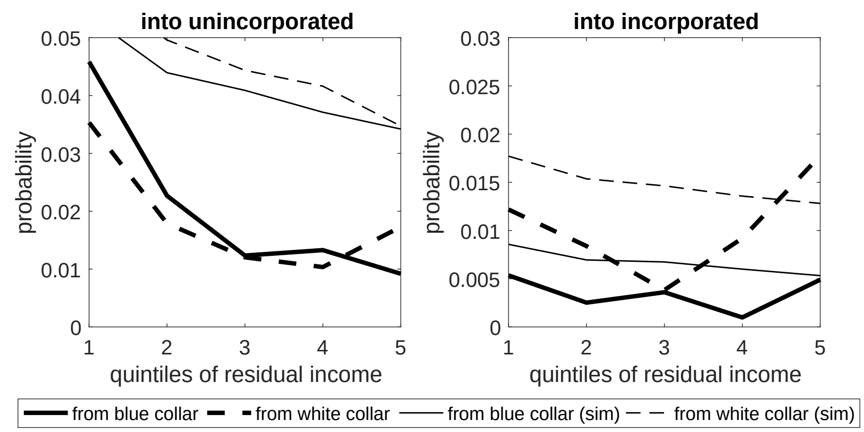


Figure A.5: Switching into Entrepreneurial Occupations (Uncorrelated Learning)

Notes: Probability of switching into entrepreneurial occupations in $t+1$ by decile of residual income in t under the counterfactual that learning about ability is uncorrelated. Residual income is computed from occupation-specific regressions of hourly income on occupation-specific experience, general experience squared, race, education and marital status.

A.4.3 Certainty Equivalent

Static

In order to get a sense of the magnitude of the estimated risk aversion parameter consider a static individual with beliefs $\mathbb{B}_t = \{\mathbb{E}_t, \mathbb{V}_t\}$. His expected annual income from working in occupation k at age t is

$$\bar{y}_{kt+1} = f_k(h_{it}; \theta_k) + \mathbb{E}_{t\{k\}}$$

and he considers the variance of his hourly income to be

$$\sigma^2_{kt} = \mathbb{V}_{t\{k,k\}} + \sigma^2_{\eta_k}$$

Therefore, his certainty equivalent at occupation k, y^c_k, solves

$$-\exp\left\{-\rho \bar{L}_k y^c_k\right\} = -\exp\left\{-\rho \bar{L}_k \bar{y}_{kt+1} + \frac{\rho^2 \bar{L}^2_k}{2}\sigma^2_{kt}\right\}$$

which yields

$$y^c_k = \bar{y}_{kt+1} - \frac{\rho \bar{L}_k}{2}\sigma^2_{kt} \tag{A.29}$$

In estimation $\bar{L}_k$ is substituted with $\bar{L}_k/100$.

Dynamic

In order to obtain the dynamic version of the certainty equivalent use equation (A.9) to find the quantity y^c_k such that

$$E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right) |\mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}} = E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t) \exp\left(\frac{-\rho \bar{L}_k y^c_k}{b_{\tau(t+1)}}\right) |\mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}}$$

As opposed to the static case, the future value of human capital and beliefs also determine the certainty equivalent:

$$y_k^c = -\left(\frac{b_{\tau(t+1)}}{\rho \bar{L}_k}\right) \ln \left(\frac{E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) v_{kt+1} | \mathbb{E}_t, h_t\right]}{A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_t)}\right) \tag{A.30}$$

A.4.4 Monetary Value of Entry Costs

The monetary value of entry costs is obtained using equation (A.9). From equation (2.3) one can separate the non-pecuniary costs in two factors, one corresponding to the entry costs, $\alpha_{kt}^e(h_t)$ and the other corresponding to all other non-pecuniary costs. Hence, let $\alpha_{kt}(h_t) = \alpha_{kt}^o(h_t) \times \alpha_{kt}^e(h_t)$. Next, use equation (2.3) to figure out the quantity that should be taken out of annual income in the budget constraint in order to equalize the conditional value functions. In other words, find the quantity ψ that must be given to the individual to leave him indifferent between (a) receiving ψ and facing entry costs and (b) not receiving ψ but facing no entry costs. It solves:

$$\begin{aligned}
&\alpha_{kt}^e(h_t)^{1/b_{\tau(t)}} E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho \bar{L}_k y_{kt+1}(h_t)}{b_{\tau(t+1)}}\right) | \mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}} \\
&\quad = E_t\left[A_{t+1}(\bar{H}_{kt+1}(h_t), \mathbb{E}_{kt+1}) \exp\left(\frac{-\rho(\bar{L}_k y_{kt+1}(h_t) - \psi)}{b_{\tau(t+1)}}\right) | \mathbb{E}_t, h_t\right]^{1-1/b_{\tau(t)}}
\end{aligned}$$

which yields

$$\psi = \frac{\ln \alpha_{kt}^e(h_t)}{\rho} \frac{b_{\tau(t+1)}}{b_{\tau(t)} - 1} \tag{A.31}$$

Since the quantity $\bar{L}_k y_{kt+1}$ was written in thousands of dollars in estimation, the value of ψ is in thousands of dollars.

A.4.5 Alternative Regimes

In order to increase precision and facilitate comparison across options, in this section ability is not approximated from the data. Instead, individuals' ability vectors are drawn from the estimated distributions in Table 2.7. Rather than being replicated from the data, individuals are simulated using the empirical joint distribution of initial states. Simulations are undertaken using a fictional economy in which there is no aggregate variation in bond prices. In this stationary environment the bond price is set to remain constant at the 1990 level.[6] Marital status paths follow the same restriction specified in the Estimation Appendix. The initial state and bond price sequence used in the decomposition is also used for the policy counterfactuals.

These counterfactual regimes are described below:

- C1: *No Learning-by-doing.* In this counterfactual, individuals receive a fix hourly return regardless of how much experience in the occupation they have accumulated. The fixed hourly return provided to individuals is constructed as an approximation of the average returns from experience in the occupation. It is computed as the average of the returns to experience, computed during the first 20 years in the labor market, of an individual that works exclusively in the occupation. Let $R_k(x)$ be the returns to experience in occupation k for somebody that has worked x years in occupation k and zero years in any other occupation (see Figure 2.4). Then, the fixed hourly return to individuals in occupation k under this counterfactual is

$$\bar{y}_k = \sum_{x=0}^{20} R_k(x)$$

This exercise yields the following values:

[6] An alternative way of dealing with the aggregate variation is to undertake a partial equilibrium analysis that fixes the sequence of bond prices observed in the data across counterfactual regimes.

Table A.6: Average Income for Counterfactual: No Learning-by-doing

	blue collar	white collar	unincorporated	incorporated
$\bar{y}_k$	0.378	0.879	0.725	1.234

Notes: Computed using the profiles in Table 2.6.

Individuals under this counterfactual still have differential returns based on their education, race, marital status, and ability.

- C2: *Isolated full information about ability.* In this counterfactual, individuals have full information about their ability. In order to isolate the effect of sorting on ability, the value of the idiosyncratic income variance of their income is set to equal its original value (see Table 2.8) plus the value of the ability variance (see Table 2.7). In terms of equation (2.4), this amounts to changing the value of the idiosyncratic income variance in occupation k from just σ_{η_k} to $\sigma_{\eta_k} + V_{\{k,k\}}$.

- C3: *No Cross-occupation Returns.* In this counterfactual, the returns in occupation k from experience accumulated in occupation $k' \neq k$ (see Figure 2.5) are set to be zero.

- C4: *Uncorrelated Learning.* In this counterfactual, individuals use an alternative variance-covariance matrix in order to update their beliefs. This variance-covariance is formed setting at zero the off-diagonal terms of the variance-covariance matrix of the distribution of ability.

- C5: *No Uncertainty.* In this counterfactual, individuals have full information about their ability and they face no extra uncertainty coming from the idiosyncratic variance. In other words, this is the same as counterfactual C2 plus setting the idiosyncratic variance to zero.

- C6: *Uniform Entry Costs.* In this counterfactual, individuals of all ages pay the same entry cost, provided they have the same permanent wealth. This cost equals the one faced by a 35 year old individual with their education level (see Table 2.9).

Table A.7: Comparison of Counterfactual Regimes

Unincorporated

	Baseline	C1	C2	C3	C4	C5	C6
Ever tried	0.31	0.38	0.34	0.28	0.32	0.35	0.54
Ever tried in first 5 years	0.08	0.08	0.11	0.06	0.08	0.12	0.31
PVI if ever tried	518	510	663	493	481	666	618
Spell duration	2.17	2.16	2.81	2.37	2.07	2.89	3.10
Participation rate at age 40	0.10	0.13	0.14	0.10	0.10	0.15	0.26
At first entry							
Ability (10$ per hour)	0.05	0.05	0.59	0.09	-0.09	0.57	0.03
Belief (10$ per hour)	0.04	0.04	-	0.07	0.00	-	0.01
Age	34.07	34.88	32.80	34.91	33.99	32.53	28.04
exp_{bc}	6.76	8.37	6.01	9.76	6.47	5.90	0.85
exp_{wc}	4.77	3.86	4.25	2.55	5.09	4.04	3.88
Overall							
Ability (10$ per hour)	0.37	0.39	1.16	0.45	0.21	1.12	0.23
College or more	0.55	0.58	0.65	0.63	0.49	0.65	0.44

Incorporated

	Baseline	C1	C2	C3	C4	C5	C6
Ever tried	0.15	0.33	0.20	0.14	0.13	0.26	0.26
Ever tried in first 5 years	0.02	0.03	0.05	0.01	0.02	0.07	0.12
PVI if ever tried	757	624	1115	622	576	1115	840
Spell duration	2.88	2.98	5.11	2.91	2.62	5.57	3.39
Participation rate at age 40	0.04	0.09	0.10	0.04	0.03	0.14	0.13
At first entry							
Ability (10$ per hour)	0.55	0.43	1.73	0.50	-0.25	1.52	0.20
Belief (10$ per hour)	0.64	0.49	-	0.54	0.00	-	0.26
Age	38.62	40.52	36.06	40.59	38.74	35.43	29.19
exp_{bc}	6.95	11.43	5.98	13.50	7.03	6.07	0.45
exp_{wc}	8.14	5.68	6.52	3.97	8.09	5.85	5.18
Overall							
Ability (10$ per hour)	1.18	0.92	2.70	1.12	0.24	2.37	0.42
College or more	0.70	0.62	0.66	0.64	0.61	0.63	0.37

Notes: Average of several summary statistics across alternative regimes. *Rows*: **PVI** stands for the present value of income in thousands of dollars. This average is computed only over those who tried the occupation. **At first entry** indicates that quantities are computed at first entry. **Ability** contains the ability of those entering the occupation. **Belief** contains the mean of the belief regarding ability. Both ability and the mean of the belief are in 10$ per hour. $\mathbf{exp_{bc}}$ and $\mathbf{exp_{wc}}$ stand for blue and white collar experience. **Overall** indicates that quantities are computed across all observations of individuals participating in the occupation. *Columns*: **Baseline** is the model specification used in Chapter 2. Columns C4 to C7 correspond to the solution and simulation of the model under alternative regimes. **C1** shuts down accumulation of human capital through experience. All individuals going into occupation k receive the equivalent of the average return from experience of somebody who always works in occupation k. The average is computed over the first 20 years of his labor market career. **C2** is a full information model where the overall level of initial uncertainty is maintained in order to isolate the effect of sorting on ability from risk aversion. In this counterfactual, the idiosyncratic variance is set to be $\sigma_{\eta_k} + V_{\{k,k\}}$. **C3** sets the cross-occupation returns to experience to be zero. **C4** shuts down correlated learning. **C5** is the full information model without uncertainty. In this counterfactual, the idiosyncratic variance σ_{η_k} is set to be zero. **C6** keeps the entry costs constant relative to age. Entry costs are always those of a 35 years old person. However, entry costs still vary with permanent wealth. The same simulated individuals, including their ability vector, is keep constant across counterfactual regimes.

Appendix B

Appendix to Chapter 3

B.1 Data Appendix

Beginning in 1984, the Multi-Center AIDS Cohort Study (MACS) started gathering information regarding natural and treated histories of HIV infection in homosexual and bisexual men. The study is conducted in Baltimore, Chicago, Pittsburgh and Los Angeles. At each semi-annual visit, data are collected on: demographics, psychosocial characteristics, sexual behavior, and specially important for our purposes, antiretroviral (AV henceforth) drugs consumption and trial participation. In addition, blood tests are administered to measure health status and serostatus (whether the individual is HIV+). Data collection started with 4,954 men enrolled. Two more enrollments have taken place: one in 1987-1991 (668 additional men) and another in 2001-2003 (1,350 additional men). We only use data from the first two enrollments.

B.1.1 Main Variables

Health (h_{it-1}): At every visit individuals go through a physical examination in which several health measurements are taken. As our measure of underlying health status, we use the

CD4 count obtained from a blood sample. “CD4 is a glycoprotein found on the surface of immune cells [...]. If CD4 cells become depleted, for example in untreated HIV infection, or following immune suppression prior to a transplant, the body is left vulnerable to a wide range of infections that it would otherwise have been able to fight. [...] Normal blood values are usually expressed as the number of cells per microliter (or cubic millimeter, mm^3) of blood, with normal values for CD4 cells being 500-1200 cells/mm” (Wikipedia). We denote as h_{it-1} the CD4 count at of the individual at the start of period t.

Labor supply (l_{it-1}): Whether the individual was working full time (35 hours or more) in visit t.

Income (m_{jit}): Starting at visit 14, individuals answer the following question: “Which of the following categories describes your annual individual gross income before taxes”? For visit 14, categories are: less than 10000, 10000-19999, 20000-29999, 30000-39999, 40000-49999, 50000-59999, 60000-69999, 70000 or more, Doesn’t wish to answer. For visits 15 to 35, categories are: less than 10000, 10000-19999, 20000-29999, 30000-39999, 40000-49999, 50000 or more, Doesn’t wish to answer. For visits 36 to 41, categories are: less than 10000, 10000-19999, 20000-29999, 30000-39999, 40000-49999, 50000-99999, 60000 or more, Doesn’t wish to answer.

We censor all periods at 50000 or more to obtain a uniform question over time. Then we assign the middle point to individuals in the bracket. For the highest bracket we assign the upper limit (50000). In our model gross income is divided by two since the survey asks about annual income. Gross income as well as out-of-pocket payments (below) are in constant dollars of 2000.

Out-of-pocket payments (o_{jit}): Starting from visit 14, individuals are given the following direction “Please estimate the TOTAL out-of-pocket expenses that you or other personal sources (your lover, family or friends) paid for prescription medications since your last visit.”[1]

[1] Wording changes slightly in visits 14 and 15.

As opposed to the gross income question, this question is open so values are not categorized.

Ailments (x_{jit}): Starting from at visit 4, individuals are asked about physical symptoms. Even though other ailments are recorded, we focus on unusual bruises lasting at least two weeks, unintentional weight loss of at least 10 pounds, fatigue, diarrhea, fever, night sweats, and tender/enlarged glands. The last 5 ailments must be felt for at least 3 days.

Even though individuals are asked explicitly about side effects starting from visit 13, we choose not to use such data because it is less consistent and, more importantly, because we do not think individuals are able to differentiate correctly between side effects and symptoms. Therefore, in our model x_{it} takes the value of 1 if individual reports having any of the problems mentioned above.

Race (b_i): Individuals in the sample are either white, black or hispanic.

Age (a_{it}): Age of the individual at the beginning of period t.

B.1.2 Products and Product Components

At every visit after visit 6, individuals are asked whether they took any medication to fight AIDS. Starting from visit 13, as the number of medications becoming available for HIV exploded, separate surveys were administered for antiretroviral drugs (ARVs) and non antiretroviral drugs (NARVs). We focus on ARVs since these are the drugs used to treat HIV infection. Further, since our analysis includes estimating the health and ailments of people using different drugs, we focus on observations where individuals have reported a treatment along with h_{it}, h_{it-1}, and x_{it}.

Individuals are asked to name specifically which drugs they took as well as whether or not they took the drug as part of a research study (the exact wording of the question regarding research studies changes slightly over time). Some of the reported drugs are themselves coded as trials; we regard these instances as individuals participating in trials (see Table B.1). If at individual i at period t is consuming one of his drugs as part of a trial we regard individual i as consuming a trial product at period t.

Table B.1: Trial Components

Name	Observations
AZT/ddI Blinded Trial	91
AZT/ddC Blinded Trial	69
ddI/ddC Blinded Trial	6
AZT/ddI/ddC Blinded Trial	31
AZT/d4T trial	4
AZT/3-TC Blinded Trial	23
AZT/ddI/protease inhibitor Blinded Trial	1
AZT/protease inhibitor Blinded Trial	2
d4T/protease inhibitor Blinded Trial	1
AZT/3-TC/protease inhibitor Blinded Trial	1
Combivir/Trizivir Blinded Trial	5
Trizivir + Sustiva/Combivir + Sustiva Blinded Trial	3
Generic AIDS Vaccine Trial	1

Next, we define market products as treatments with no components consumed in trial. Given that the sum of effects of individual drugs is not equal to the effect of a treatment formed by the sum of the drugs, the relevant market products consumed in our data correspond to combinations of components. For instance a product is AZT and another is AZT plus 3TC plus ddI. Table B.2 describes the individual components of market products. Some components, listed separately in Table B.3, are in fact fixed-dose combinations of other components. In our sample, if individual i is consuming the fixed-dose combination $(A + B)$ and individual i' is consuming components A and B, we assign consumers i and i' to the same treatment. One of the coded components in the data corresponds to "other ARVs". We add all uncoded components (96 instances) to "other ARVs" which results in 158 instances of "other ARVs". Finally, we treat α and β Interferons (177 instances and 33 instances, respectively) as one single component.

Our definition of market products, as combinations of drug components, generates 1835 different market treatments. We reduce the number of market products using the following algorithm:

1. We select our core market products as those treatments that overall have more than 40 instances.[2] We acknowledge that our definition of core treatments is biased against treatments appearing near the end of the time period studied. We address this issue by excluding the last 4 periods of data. Our core treatments are listed in Table B.4 which shows that there are 70 core products overall and they have at most five components. Out of 20142 subject-visit observations of consumers taking market products, 13767 are covered by core treatments and 6375 correspond to non-core treatments.

2. Second, we assign non-core treatments to core treatments in the following fashion. Each step is used sequentially to assign remaining non core treatments that were not assigned in previous steps.

   (a) Assignment of Non-core: Non core treatment A is assigned to core treatment B if B is the core treatment with the highest number of components that is contained by A. This procedure yields both non-unique assignments or null assignments. Of the remaining 6375 subject-visit observations of non core treatments, 2963 are assigned uniquely in this step. This means that we are left with 3412 subject-visit observations with non core treatments, 1647 that are assigned to multiple core treatments and 1765 that are not assigned to any core treatment.

   (b) Assignment of Multiple Assignments:

      i. First, we use the past history of the individual. If at period t individual i is consuming non core treatment W that was assigned to both core treatments A and B in previous steps, and he was observed consuming core treatment A in period $t-1$, then his treatment at t is assigned uniquely as A. We repeat

[2] We can change this to a different number and main results remain robust.

this procedure until no further gains are obtained. Out of the remaining 1647 subject-visit observations of non core treatments with multiple assignments, 428 are assigned uniquely in this step.

ii. Second, we use the future history of the individual. If at period t individual i is consuming non core treatment W that was assigned to both core treatments A and B in previous steps, and he was observed consuming core treatment B in period $t+1$, then his treatment at t is assigned uniquely as B. We repeat this procedure until no further gains are obtained. Out of the remaining 1219 subject-visit observations of non core treatments with multiple assignments, 274 are assigned uniquely in this step.

iii. Third, we assign the reaming 945 subject-visit observations of non core treatments with multiple assignments using the core treatment with the highest share at t: if at period t individual i is consuming non core treatment W that was assigned to both core treatments A and B in previous steps, and treatment A's market share at t is greater than B's, his treatment at t is assigned uniquely as A. This final step assigns uniquely the remaining 945 observations.

(c) Next, we regard all 1765 not assigned treatment observations as "fringe" treatments since they do not contain any core treatment. We aggregate them in the following fashion. We aggregate all fringe treatments that appear at period t and assign to this "cohort" fringe treatment, all users consuming this product over time. Similarly as we do with core treatments, we only consider fringe cohort treatments that have at least 40 users. This reduces the number of observations by 345 (which represents 1.6% of the number of observations of treatment consumers). This aggregations leads to 17 fringe cohort treatments that we will treat in the same way we treat core treatments: as innovations from the trials distribution. This amounts to a total of 87 treatments over all. From this point on fringe treatments are included in the denomination of core treatments.

3. We have specified that a treatment gets withdrawn from the market if it has zero share for $X = 2$ consecutive periods. However, in the data, a treatment may have zero share for $Y > X$ periods and then reappear again. 78 out of 87 core treatments have unique spells; we regard the remaining treatments with multiple spells as measurement error and follow the next procedure to ensure that treatments have one single spell from entry to exit. Consider a core treatment with multiple spells B.

   (a) We identify all spells that treatment B has in the data.

   (b) Among treatment B's spells, we select the spell that contains the period in which treatment B's share was the highest. We drop all observations of market consumers of treatment B that are not in this spell.

   (c) We follow the same steps for all 9 core treatments with multiple spells. Out of 19797 (20142 $-$ 345) subject-visit observations of consumers taking market products, this smoothing procedure drops 42 observations leaving 19755 subject-visit observations of consumers taking market products.

As evidence of the relevance of the spells selected by this procedure we compute the difference between the maximum share in the selected spell and the maximum share in each of the other spells, as a percentage of the maximum share in the other spell. The mean value of this measure is 2401, which suggests that the maximum share in the selected spell is on average about 24 times larger that the maximum share in other spells. We also try the following criteria: (i) selecting the spell with the highest average share and (ii) selecting the spell with the highest sum of shares. All criteria result in closely similar entry and exit dates so we stick to the maximum-share criteria.

Table B.2: Chemical Formulae of Product Components

Component	Chemical Formula	Observations
Isoprinosine	$C_{52}H_{78}N_{10}O_{17}$	87
Ribavirin	$C_8H_{12}N_4O_5$	62
Interferons (α and β)		210
Foscarnet	CH_3O_5P	92
AZT	$C_{10}H_{13}N_5O_4$	7436
ddC	$C_9H_{13}N_3O_3$	1123
AL-721 egg lecithin		147
Dextran-Sulfate	$H(C_6H_{10}O_5)_xOH$	65
Acyclovir	$C_8H_{11}N_5O_3$	2550
ddI	$C_{10}H_{12}N_4O_3$	3069
d4T	$C_{10}H_{12}N_2O_4$	3807
Nevirapine	$C_{15}H_{14}N_4O$	2210
Delavirdine	$C_{22}H_{28}N_6O_3S$	176
3TC	$C_8H_{11}N_3O_3S$	5250
Saquinavir	$C_{38}H_{50}N_6O_5$	1279
Ritonavir	$C_{37}H_{48}N_6O_5S_2$	3230
Indinavir	$C_{36}H_{47}N_5O_4$	2255
Nelfinavir	$C_{32}H_{45}N_3O_4S$	1278
Kaletra	$C_{37}H_{48}N_4O_5$	1883
Abacavir	$C_{14}H_{18}N_6O$	1549
Agenerase	$C_{25}H_{35}N_3O_6S$	372
Efavirenz	$C_{14}H_9CIF_3NO_2$	3362
Adefovir	$C_8H_{12}N_5O_4P$	44
Enfuvirtide (T-20)	$C_{204}H_{301}N_{51}O64$	160
Tenofovir	$C_9H_{14}N_5O_4P$	2488
Emtricitabine	$C_8H_{10}FN_3O_3S$	263
Atazanavir	$C_{38}H_{52}N_6O_7$	1583
Lexiva	$C_{25}H_{36}N_3O_9PS$	418
Etravirine	$C_{20}H_{15}BrN_6O$	155
Darunavir	$C_{27}H_{37}N_3O_7S$	315
Raltegravir	$C_{20}H_{21}FN_6O_5$	384
Ampligen	Double-stranded RNA compound	25
Peptide T	$C_{35}H_{55}N_9O_{16}$	30
DTC	$C_5H_{10}NS_2Na$	10
CD4		2
Other protease		31
Vistide (cidofovir)	$C_8H_{14}N_3O_6P$	2
Tipranavir (PNU-140690)	$C_5H_{10}NS_2Na$	30
Other Avs		158

Notes: Source: Wikipedia (November, 2014)

Table B.3: Combination Components

Name	Combination	Instances
Combivir	AZT + 3TC	2673
Trizivir	AZT + 3TC + Abacavir	778
Truvada	Emtricitabine + Tenofovir	1933
Epzicom	Abacavir + 3TC	724
Atripla	Efavirenz + Emtricitabine + Tenofovir	968

Table B.4: Products in the Market, Entry and Exit

Treatment Id	Treatment	Haart	Entry (visit)	Exit (visit)
3	AZT	0	6	-
13	Interferons (α and/or β), AZT	0	7	23
9	AL-721 egg lecithin	0	7	15
34	AZT, Acyclovir	0	11	32
33	Acyclovir	0	11	32
47	AZT, Acyclovir, ddI	0	12	26
51	Acyclovir, ddI	0	12	32
14	AZT, ddC	0	12	35
39	AZT, ddI	0	12	41
46	ddI	0	12	-
69	AZT, ddC, Acyclovir, ddI	0	14	26
65	AZT, ddC, Acyclovir	0	14	31
67	AZT, ddC, ddI	0	14	23
63	ddC, Acyclovir	0	14	27
64	ddC	0	14	30
85	d4T	0	18	-
117	AZT, Acyclovir, 3TC	0	21	32
124	AZT, 3TC	0	22	-
146	Acyclovir, d4T, 3TC	0	23	32
161	AZT, 3TC, Saquinavir	1	24	42
157	d4T, 3TC	0	24	-
185	AZT, 3TC, Saquinavir, Ritonavir	1	25	-
164	AZT, Acyclovir, 3TC, Indinavir	1	25	32
171	Acyclovir, d4T, 3TC, Indinavir	1	25	32
169	AZT, 3TC, Ritonavir, Indinavir	1	25	45
214	d4T, 3TC, Ritonavir, Indinavir	1	25	45
254	d4T, 3TC, Saquinavir, Ritonavir	1	25	41
202	ddI , d4T, Indinavir	1	25	41

175	d4T, 3TC, Indinavir	1	25	48
165	AZT, 3TC, Indinavir	1	25	-
242	d4T, Nevirapine, 3TC	1	26	-
236	AZT, Nevirapine, 3TC	1	26	-
268	AZT, 3TC, Nelfinavir	1	26	-
377	ddI , d4T, Nelfinavir	1	26	43
292	d4T, 3TC, Nelfinavir	1	27	-
349	ddI , d4T, Nevirapine	1	27	-
311	ddI , 3TC, Nelfinavir	1	27	-
615	ddI , d4T, Efavirenz	1	29	48
644	3TC, Abacavir, Efavirenz	1	29	-
573	AZT, Nevirapine, 3TC, Abacavir	1	30	-
720	AZT, 3TC, Abacavir, Efavirenz	1	30	-
548	AZT, 3TC, Efavirenz	1	30	-
701	AZT, 3TC, Abacavir	0	30	-
532	d4T, 3TC, Efavirenz	1	30	44
581	Nevirapine, 3TC, Abacavir	1	31	-
782	d4T, 3TC, Kaletra	1	34	44
940	3TC, Kaletra, Abacavir	1	35	-
869	AZT, 3TC, Kaletra	1	35	-
987	AZT, 3TC, Kaletra, Abacavir	1	36	-
963	3TC, Abacavir, Efavirenz, Tenofovir	1	36	-
921	AZT, 3TC, Abacavir, Tenofovir	1	36	-
909	AZT, 3TC, Kaletra, Tenofovir	1	36	-
923	Nevirapine, 3TC, Tenofovir	1	36	46
949	3TC, Kaletra, Tenofovir	1	36	-
919	Kaletra, Efavirenz, Tenofovir	0	36	-
926	3TC, Efavirenz, Tenofovir	1	36	-
1010	AZT, 3TC, Kaletra, Abacavir, Tenofovir	1	37	-
1020	ddI , Kaletra, Tenofovir	1	37	-
976	ddI , Efavirenz, Tenofovir	1	37	-

1011	Abacavir, Efavirenz, Tenofovir	1	37	-
994	Kaletra, Abacavir, Tenofovir	1	37	-
1230	3TC, Ritonavir, Abacavir, Atazanavir	1	39	-
1071	Efavirenz, Tenofovir, Emtricitabine	1	39	-
1227	Ritonavir, Efavirenz, Tenofovir, Emtricitabine, Atazanavir	1	40	-
1245	3TC, Ritonavir, Abacavir, Tenofovir, Atazanavir	1	40	-
1303	ddI , Ritonavir, Tenofovir, Atazanavir	1	40	-
1222	Ritonavir, Tenofovir, Emtricitabine, Atazanavir	1	40	-
1128	Nevirapine, Tenofovir, Emtricitabine	1	40	-
1253	Kaletra, Tenofovir, Emtricitabine	1	41	-
1342	Ritonavir, Tenofovir, Emtricitabine, Lexiva	1	42	-
10006		0	6	16
10026		0	26	46
10027		0	27	45
10028		0	28	45
10030		1	30	43
10031		0	31	-
10035		0	35	49
10037		1	37	-
10038		0	38	-
10040		0	40	-
10041		1	41	-
10042		1	42	-
10043		1	43	-
10046		1	46	-
10048		1	48	-
10049		1	49	-

B.2 Estimation Appendix

B.2.1 k-means Clustering Algorithm

We implement the following version of the k-means algorithm. At every period t:

1. We select the products that have not yet being applied the *exit switching* rule. In other words, we select products that are still available for people to swith into at period t. Denote this set of products available for clustering at t, $\mathcal{A}_t$.

2. We re-scale the characteristics of all products available for clustering at t. In order to do this we compute
$$\tilde{\theta}^r = \frac{\theta^r}{\max_{\delta \in \mathcal{A}_t} |\delta^r|}, \text{ for } r = h, x$$
Therefore, by construction $\tilde{\theta}^r \in [-1, 1]$.

3. We choose the first k centroids using k initial $\tilde{\theta}$'s in $\mathcal{A}_t$ randomly selected.

4. We allocate all remaining points in $\mathcal{A}_t$ sequentially. At each step the point selected is the one that is closest to one of the existing clusters. This point is then allocated to the correspondent cluster and the centroid of the cluster is updated. This process is repeated until all points are allocated to a cluster.

5. We undertake a reallocation step in which, taken the centroids as given, all points are allocated to their closest centroid.

6. We calculate the value of (3.8) for the current allocation.

7. We repeat the process 200 times using different random initial $\tilde{\theta}$'s in $\mathcal{A}_t$. The allocation with the lowest value of (3.8) is chosen. When simulating clusters in estimation we only repeat the process 50 times to speed up the process.

B.2.2 Product Characteristics

We estimate product characteristics using data on individual treatment usage and subsequent reports of health and ailments. Our estimation equations mimic equations (3.19) and (3.20), which individuals use to form expectations over their health and ailments conditional on their choice. The key difference between equations (3.19) and (3.20) and our estimation equations is that here our aim is to obtain characteristics of each individual treatment.

Let δ_{rit} be an indicator that treatment r was used by individual i at time t. The characteristics of treatment r are denoted

$$\theta_r = \left\{\theta_r^x, \theta_r^h\right\} \in \mathbb{R}^2 \tag{B.1}$$

The components of θ_r are estimated as the coefficients of δ_{rit} in the health and no-ailments regressions

$$h_{it} = \sum_{m=0}^{5} \alpha_m^h h_{it-1}^m + \sum_r \theta_r^h \delta_{rit} + \epsilon_{it} \tag{B.2}$$

$$\Pr\left[x_{it} = 1 | \cdot\right] = \frac{\exp\left(\sum_{m=0}^{5} \alpha_m^x h_{it-1}^m + \sum_r \theta_r^x \delta_{rit}\right)}{1 + \exp\left(\cdot\right)} \tag{B.3}$$

B.2.3 GMM Estimation Algorithm

Using the fact that we observe the underlying stochastic process that generates the stochastic process of cluster characteristics we can write the moment condition in equation (3.25) can be written as

$$\mathbb{E}\left\{ w\left(z_{it}\right)\otimes\left[\begin{array}{c} \vdots \\ \ln\left(\frac{p_{oit}(z_{it})}{p_{jit}(z_{it})}\right)+\mathbb{E}_{\mathcal{P}}\left[y_{jit}\right]-\mathbb{E}_{\mathcal{P}}\left[y_{oit}\right] \\ +\sum_{s=1}^{T^*}\beta^s P_j^{o(s-1)}\left(z_{it}\right) \\ \mathbb{E}_{\mathcal{P}}\left[D_{it+s}\left(z_{it+s}\right)\sum\limits_{k\in C_{t+s}} p_{kit}\left(z_{it+s}\right)\left[y_{kit+s}\left(z_{it+s}\right)+\psi_{kit+s}\left(z_{it+s}\right)\right]\middle|\cdot,j,\mathcal{P}_t\right] \\ -\sum_{s=1}^{T^*}\beta^s P_o^{o(s-1)}\left(z_{it}\right) \\ \mathbb{E}_{\mathcal{P}}\left[D_{it+s}\left(z_{it+s}\right)\sum\limits_{k\in C_{t+s}} p_{kit}\left(z_{it+s}\right)\left[y_{kit+s}\left(z_{it+s}\right)+\psi_{kit+s}\left(z_{it+s}\right)\right]\middle|\cdot,o,\mathcal{P}_t\right] \\ \vdots \end{array}\right]\right\}=0 \tag{B.4}$$

Equation (B.4) is crucial for our simulation estimation method explained below. The key fact is that we observe the characteristics of the underlying process of product evolution and we are then able to use it to generate the stochastic evolution of clusters. We undertake simulation in order to obtain the value of

$$\begin{array}{c} \sum_{s=1}^{T^*}\beta^s P_j^{o(s-1)}\left(z_{it}\right)\times \\ \mathbb{E}_z\left[D_{it+s}\left(z_{it+s}\right)\sum\limits_{k\in C_{t+s}} p_{kit+s}\left(z_{it+s}\right)\left[y_{kit+s}\left(z_{it+s}\right)+\psi_{kit+s}\left(z_{it+s}\right)\right]\middle| z_{it},d_{jit}=1,S_{it}^{(s-1)}=1,d_i^o,\mathcal{P}_t\right] \end{array} \tag{B.5}$$

for each individual i and choice j at every period t. Let NS denote the number of simulated technology paths for each individual at every period and let the superscript ns indicate that a quantity has being simulated. For individual i and decision j at period t we write the simulated counterpart of equation (B.5) as

$$\begin{aligned} & \frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s P_j^{o(s-1),ns}\left(z_{it}\right)D_{it+s}\left(z_{it+s}^{ns,j}\right)\sum_{k\in C_{t+s}^{ns,j}} d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,j}\right)+\psi_{kit+s}\left(z_{it+s}^{ns,j}\right)\right] \\ = & \frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s\left(\prod_{\tau=1}^{s}D_{it+\tau}\left(z_{it+\tau}^{ns,j}\right)\right)\sum_{k\in C_{t+s}^{ns,j}} d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,j}\right)+\psi_{kit+s}\left(z_{it+s}^{ns,j}\right)\right] \end{aligned} \tag{B.6}$$

For a given vector of parameters of the utility function, the above simulation must be undertaken NS times for each individual i available at period t, and for all t, and for $J-1$ choices as well as for choice o, which means it must be repeated at least $NS \times T \times N \times J$. Further, notice that within each individual simulation we must simulate N optimal paths, one for every person, in order to obtain the aggregate behavior. In other words, even though we simulate only $NS \times T \times N \times J$ technology paths, we simulate $NS \times T \times N \times J \times N$ individual paths. Given our numbers we will be simulating at most $NS \times 33 \times 1669 \times 6 = NS \times 330,462$ technology paths of length T^* and $NS \times 33 \times 1669 \times 6 \times 1669 = NS \times 551,541,078$ individual paths of length T^*. Relying on Hotz et al. (1994) we could set $NS = 1$ and still obtain consistency. We set $NS = 10$ after trying different values of NS for robustness.

The sample moment conditions will then be

$$\frac{1}{\sum_i \sum_t \delta_{it}} \sum_{i=1}^{N} \sum_{t=1}^{T} \delta_{it} w\left(z_{it}\right) \otimes \begin{bmatrix} \vdots \\ \ln\left(\frac{p_{oit}(z_{it})}{p_{jit}(z_{it})}\right) + y_{jit} - y_{oit} \\ +\frac{1}{NS} \sum_{ns} \sum_{s=1}^{T^*} \beta^s \left(\prod_{\tau=1}^{s} D_{it+\tau}\left(z_{it+\tau}^{ns,j}\right)\right) \times \\ \sum_{k \in C_{t+s}^{ns,j}} d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,j}\right) + \psi_{kit+s}\left(z_{it+s}^{ns,j}\right)\right] \\ -\frac{1}{NS} \sum_{ns} \sum_{s=1}^{T^*} \beta^s \left(\prod_{\tau=1}^{s} D_{it+\tau}\left(z_{it+\tau}^{ns,o}\right)\right) \times \\ \sum_{k \in C_{t+s}^{ns,j}} d_{kit+s}^{ns,o}\left(z_{it+s}^{ns,o}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,o}\right) + \psi_{kit+s}\left(z_{it+s}^{ns,o}\right)\right] \\ \vdots \end{bmatrix} = 0 \tag{B.7}$$

where δ_{it} is an indicator of availability of individual i at period t. Estimation follows the simulation strategy described below. Simulation will be undertaken in order to obtain

$$\sum_{s=1}^{T^*} \beta^s P_j^{o(s-1)}\left(z_{it}\right) \mathbb{E}_z \left[D_{it+s}\left(z_{it+s}\right) \sum_{k \in C_{t+s}} p_{ikt+s}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \psi_{ikt+s}\left(z_{it+s}\right)\right] \middle| z_{it}, d_{jit} = 1, S_{it}^{(s-1)} = 1, d_i^o \right] \tag{B.8}$$

and

$$\sum_{s=1}^{T^*} \beta^s P_o^{o(s-1)}\left(z_{it}\right) \mathbb{E}_z\left[D_{it+s}\left(z_{it+s}\right) \sum_{k \in C_{t+s}} p_{ikt+s}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right)+\psi_{ikt+s}\left(z_{it+s}\right)\right] \middle| z_{it}, d_{iot}=1, S_{it}^{(s-1)}=1, d_i^o\right] \tag{B.9}$$

for each individual i at every period t. Let the superscript ns indicate that a quantity has been simulated. Also let subscript j denote the decision made at time t to be compared against the base choice o.

For individual i at period t who chose j, the simulation algorithm to obtain (B.8) entails the following steps for each simulated path ns (again, we set the number of simulated paths for every data point (i,t) at $NS=1$):

1. **Number of new products**. If $s=1$, define $Q_{t+s-1}^{ns} \equiv Q_t$. Using Q_{t+s-1}^{ns} we draw number of new products, New_{t+s}^{ns}, using a negative binomial process. First we draw

   $$\mu_{t+s}^* \sim Gamma\left(1/\alpha, \alpha\mu_{t+s}\right)$$

   where

   $$\mu_{t+s} = \beta_0^N + \beta_1^N Q_{t+s-1}$$

   Then we draw

   $$New_{t+s}^{ns}\Big|_{\mu^*} \sim Poisson\left(\mu_{t+s}^*\right)$$

   $\left(\alpha, \beta_0^N, \beta_1^N\right)$ are parameters estimated in a first stage.

2. **Characteristics of new products**. If $New_{t+s}^{ns} > 0$, for each simulated new product we obtain simulated product characteristics. Consistent with our model, new products at $t+s$ are characterized by simulated realizations of the bivariate random vector

   $$\omega_{t+s-1} + \nu_{t+s-1} \tag{B.10}$$

where ω_{t+s-1} is the centroid at $t+s-1$, $\nu_{t+s-1} \sim F_\nu$ and F_ν is our innovations distribution which is estimated non parametrically.

As a by-product of steps 1 and 2 we obtain Q^{ns}_{t+s} using equation (3.5).

3. **Exit**.

   - *Overall exit rule.* If the ratio of people consuming product k (either by staying or switching) relative to the number of people consuming a market product falls bellow $\tilde{\sigma}_2$ during the last 2 consecutive periods (i.e. $t+s-1, t+s-2$), the product is withdrawn from the market and cannot be consumed at any $\tau \geq t+s$. $\tilde{\sigma}_2$ is chosen as the minimum conditional share observed in the data.

   - *Switching exit rule.* If the product satisfies the overall exit rule or if the ratio of people switching and being assigned product k relative to the number of people switching falls bellow $\tilde{\sigma}_1$ during the last 3 consecutive periods (i.e. $t+s-1, t+s-2, t+s-3$), the product is no longer available for switchers and therefore cannot be used to form clusters at any $\tau \geq t+s$. $\tilde{\sigma}_1$ is chosen as the minimum conditional share observed in the data. These products may still be used by "staying" individuals who consumed the product last period.

   Old products minus exits plus simulated new products yields the simulated set of products in period $t+s$, $\mathcal{P}^{ns}_{t+s}$.

4. **Clusters.** From the simulated set of products $\mathcal{P}^{ns}_{t+s}$, we select those products that can be used for clustering and along with the grouping algorithm we obtain simulated clusters $\mathcal{G}^{ns}_{t+s}$. We then compute characteristics for the simulated clusters, W^{ns}_{t+s}.

5. **Centroid.** If $s=1$ define $\mathcal{P}^{ns}_{t+s-1} \equiv \mathcal{P}_t$. Using the characteristics of products in $\mathcal{P}^{ns}_{t+s-1}$, unconditional choice probabilities ($\mathbb{E}_i \left[p_{jit+s-1}\left(z_{it}\right)\right]$), within-cluster product weights at

$t+s-1$, and $t+s-1$ shares of products conditional on staying, we compute the simulated centroid ω_{t+s}^{ns} using equation (3.1).

Steps 1 through 5 provide the aggregate part of the simulated state, $z_{t+s}^{\mathcal{P},ns}$. Denote the future choice set induced by the simulated evolution of products as $\mathcal{C}_{t+s}^{ns}$.

6. **Future state for** i. **(i)** If $s=1$, define $h_{jit+s-1}^{ns}$as the observed $h_{jit+s-1}$. If $s>1$, draw $\epsilon_{it+s-1}^{h,ns}$ from the non parametric distribution of ϵ^{h}; then, using d_{it+s-1}^{ns}, and when necessary, the realization of the within cluster treatment assigned at $t+s-1$, we compute simulated health at the beginning of period $t+s$, $h_{jit+s-1}^{ns}$, using equation (B.2). If d_{it+s-1}^{ns} involves the trial alternative, trial-product characteristics for computing equation (B.2) are drawn from the trial distribution at $t+s-1$, $F_{\theta|\omega_{t+s-1}}$; which is equivalent to using equation (B.10) and the innovations distribution, F_{ν}. **(ii)** We draw a simulated out-of-pocket payment shock $\epsilon_{it+s}^{o,ns}\sim N\left(0,\sigma_{o}^{2}\right)$. **(iii)** We draw a simulated labor state l_{it+s}^{ns} using equation (3.17). **(iv)** We compute deterministic state variables for i.

7. **Future state for all** $i'\neq i$. **(i)** If $s=1$, define $h_{i't+s-1}^{ns}$as the observed $h_{i't+s-1}$. If $1<s<T^{*}$, draw $\epsilon_{i't+s-1}^{h,ns}$ from the non parametric distribution of ϵ^{h}. Then, using $d_{i't+s-1}^{ns}$, and when necessary, the realization of the within cluster treatment assigned at $t+s-1$, we compute simulated health at the beginning of period $t+s$, $h_{i't+s-1}^{ns}$, using equation (B.2). If $d_{i't+s-1}^{ns}$ involves the trial alternative, trial-product characteristics for computing equation (B.2) are drawn from the trial distribution at $t+s-1$, $F_{\theta|\omega_{t+s-1}}$. We have deliberately written $h_{i't+s-1}^{ns}$ instead of $h_{i'jt+s-1}^{ns}$ as it is explained below. **(ii)** We draw a simulated labor state $l_{i't+s}^{ns}$ using equation (3.17). **(iii)** We compute deterministic state variables for i'.

Steps 6 and 7 provide the relevant pieces of the individual-specific part of the simulated state, $z^{o,ns}_{jit+s}$ for i and $z^{o,ns}_{i't+s}$ for all $i' \neq i$.

8. **Probability of Survival up to $t+s-1$.** If $s=1$, by definition, $P^{o(s-1)}_{j}\left(z_{it}\right)=1$ for all i available at t. If $s>1$, using $z^{o,ns}_{jit+s-1}$, and $P^{o(s-2),ns}_{j}\left(z_{it}\right)$ we obtain $P^{o(s-1),ns}_{j}\left(z_{it}\right)$ using

$$\begin{aligned} P^{o(s-1),ns}_{j}\left(z_{it}\right) &= \prod_{\tau=1}^{s-1} D_{it+\tau}\left(z^{ns}_{it+\tau}\right) \\ &= D_{it+s-1}\left(z^{ns}_{jit+s-1}\right) P^{o(s-2),ns}_{j}\left(z_{it}\right) \end{aligned} \tag{B.11}$$

9. **CCPs and simulated choice for i.** Using $z^{\mathcal{P},ns}_{t+s}$, $z^{o,ns}_{jit+s}$, and equations (B.14), (B.16), and (B.15), we compute simulated $t+s$ ccps, $p^{ns}_{ikt+s}\left(z^{ns}_{jit+s}\right)$, for every alternative $k \in C^{ns}_{t+s}$. Then, using the simulated ccps we draw a decision $d^{ns}_{it+s}\left(z^{ns}_{jit+s}\right)$ for i.

10. **CCPs and simulated choice for all $i' \neq i$.** Using $z^{\mathcal{P},ns}_{t+s}$, $z^{o,ns}_{i't+s}$ for all $i' \neq i$, and equations (B.14), (B.16), and (B.15), we compute simulated $t+s$ ccps, $p^{ns}_{i'kt+s}\left(z^{ns}_{i't+s}\right)$, for every alternative $k \in C^{ns}_{t+s}$. Then, using the simulated ccps we draw a decision $d^{ns}_{i't+s}\left(z^{ns}_{i't+s}\right)$ for all $i' \neq i$.

11. **Static payoff for i. (i)** We compute $\bar{m}^{s}_{it+s} = X^{m,ns}_{it+s}\theta^{m} + \nu^{m}_{i}$ using equation (3.15). Even though individuals know their idiosyncratic shocks in the income equation, ϵ^{m}_{it}, we do not need to simulate these as they are iid and have mean zero an enter linearly in the flow utility, which will result in them averaging out to zero in the moment condition. **(ii)** Using the simulated choice $d^{ns}_{it+s}\left(z^{ns}_{jit+s}\right)$ we compute expected simulated out-of-

pocket payments using

$$o_{it+s}\left(d_{it+s}^{ns}\right)=\begin{cases} o_{it+s}^{*,ns} & \text{if } o_{it+s}^{*,ns}>0 \\ 0 & \text{if } o_{it+s}^{*,ns}\leq 0 \end{cases}$$

where

$$o_{it+s}^{*,ns}\left(d_{it+s}^{ns}\right)=X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o}+\epsilon_{it+s}^{o,ns}$$

and $X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)$ are given in equation (3.16). Hence

$$\mathbb{E}\left[o_{it+s}\left(d_{it+s}^{ns}\right)\middle|\, d_{it+s}^{ns}\right]=\Phi\left(X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o}/\sigma^{o}\right)X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o}+\sigma^{o}\phi\left(X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o}/\sigma^{o}\right)$$

(iii) We compute the expected probability of no-ailments as

$$\mathbb{E}\left[x_{it+s}\middle|\, d_{it+s}^{ns}\right]$$

using equation (3.19) and the relevant distribution: cluster, trial, or degenerate. Notice that here we exploit again the fact that we observe the underlying stochastic process. Whenever the choice is a cluster, we use the within cluster weights. **(iv)** Using above components and i's simulated decision we compute flow payoffs $y_{it+s}^{ns}\left(z_{it+s}^{ns},d_{it+s}^{ns}\right)$ using equation (3.13). **(v)** We compute the probability of survival from $t+s-1$ into $t+s$, $D_{it+s}\left(z_{it+s}^{ns}\right)$, using equation (3.21) and the term $\psi_{it+s}\left(z_{it+s}^{ns},d_{it+s}^{ns}\right)$ using equation (3.27).

12. **Repeat all steps above until $s=T^{*}$.**

In order to obtain all other simulated counterparts of (B.8) for individual i at period t we do not repeat all the steps above. Instead, we use the same simulated aggregate evolution of the market and repeat only those steps involving individual i's path conditional on choice

$j' \neq j$ at t; this is the reason why we deliberately write $h^{ns}_{i't+s-1}$ instead of $h^{ns}_{i'jt+s-1}$ for all $i \neq i'$, as their simulated individual paths do not depend on i's decision at period t. We abstain from generating a path of product innovation following counterfactual choice k by individual i as the impact of his decision at period t on the overall aggregate evolution of the market is negligible.

When simulating the path following counterfactual choice j' we need counterfactual health when $s = 1$, $h^{ns}_{ij't+s-1}$; for this we need to compute the realized residuals of the health equation at t

$$\hat{\epsilon}^h_{it} = h_{it} - \sum_{m=0}^{5} \alpha^h_m h^m_{it-1} - \sum_r \theta^h_r \delta_{it-1r}$$

Then, using the realized residual $\hat{\epsilon}^h_{it}$ and equation (B.2) we obtain $h^{ns}_{ij't}$. When individual i is in a trial in period t we do not observe the characteristics of the trial ex post; hence, we draw a health shock as well as trial characteristics and compute future simulated health, $h^{ns}_{ij'}$.

Current period payoffs. On the one hand, in order to obtain y_{jit} we need $\mathbb{E}_j\left[x_{jit+s}\right]$. Here, when j corresponds to a cluster alternative, we exploit again the fact that we observe the underlying stochastic process and use the within cluster weights. On the other hand, in order to obtain counterfactual y_{ikt} we need the realized error term of the out-of-pocket payment equation at t given by

$$\hat{\epsilon}^o_{it} = o^*_{jit} - X^o_{jit}\theta^o$$

However, we only observe o^*_{jit} if $o^*_{jit} > 0$. Hence, if $o^*_{jit} \leq 0$, we need to draw a simulated error $\epsilon^{o,ns}_{it}$ from a truncated normal conditional on

$$\epsilon^{o,ns}_{it} \leq -X^o_{jit}\theta^o$$

The sample simulated counterpart of (B.8) is

$$\frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s P_j^{o(s-1),ns}\left(z_{it}\right) D_{it+s}\left(z_{it+s}^{ns,j}\right)\sum_{k\in C_{t+s}^{ns,j}} d_{ikt+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{ikt+s}\left(z_{it+s}^{ns,j}\right)+\psi_{ikt+s}\left(z_{it+s}^{ns,j}\right)\right] \quad \text{(B.12)}$$

One potential issue with our simulation algorithm is that in reality individuals die and others become potential consumers. This two phenomena are likely to affect the aggregate joint distribution of individual characteristics and therefore the ccps and the evolution of the market. In order to control for death when computing i's continuation value we could simulate death conditional on optimal behavior for all $i' \neq i$, i.e. some people will leave the sample in the simulated paths. However, we would also need to create people to be introduced into the market. We decide to simulate neither people into the absorbing state nor the stream of people into the sample. Instead, we condition on the aggregate distribution of characteristics at any period t in order to simulate ahead and on optimal future behavior.

Also, a related issue is that our sample is refreshed at least once as new subjects are surveyed. Figures not shown here present no special effect of this refreshing in terms of the aggregate ccps suggesting that the aggregate distribution of characteristics of the new surveyed people matches that of the surveyed individuals at the time.

B.2.4 Estimator

We use a GMM estimator to obtain our structural parameters. Define B as the $K-$dimensional vector of parameters.Following Hotz et al. (1994) we want to obtain the parameter vector that solves

$$\left((NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_{it}w\left(z_{it}\right)\otimes\bar{v}_{it}(z_{it},B)\right)' W_n\left((NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_{it}w\left(z_{it}\right)\otimes\bar{v}_{it}(z_{it},B)\right) \quad \text{(B.13)}$$

where

$$\bar{v}_{it}(z_{it},B)=\begin{bmatrix} \vdots \\ \ln\left(\frac{p_{oit}(z_{it})}{p_{jit}(z_{it})}\right)+y_{jit}-y_{oit} \\ +\frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s\left(\prod_{\tau=1}^{s}D_{it+\tau}\left(z_{it+\tau}^{ns,j}\right)\right)\times \\ \sum_{k\in C_{t+s}^{ns,j}}d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,j}\right)+\psi_{kit+s}\left(z_{it+s}^{ns,j}\right)\right] \\ -\frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s\left(\prod_{\tau=1}^{s}D_{it+\tau}\left(z_{it+\tau}^{ns,o}\right)\right)\times \\ \sum_{k\in C_{t+s}^{ns,j}}d_{kit+s}^{ns,o}\left(z_{it+s}^{ns,o}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,o}\right)+\psi_{kit+s}\left(z_{it+s}^{ns,o}\right)\right] \\ \vdots \end{bmatrix}$$

and W_n is a square weighting matrix. Using the linear structure of our utility function we collect and factor terms in order to write the jth component of the vector $\bar{v}_{it}(z_{it},B)$ as the linear form

$$\tilde{y}_{jit}-\tilde{x}_{jit}'B$$

Define Y as the $[(J-1)NT\times 1]-$dimensional vector that stacks all $\tilde{y}_{jit}$, X the matrix of dimensions $[(J-1)NT\times K]$ that stacks all $\tilde{x}_{jit}$. Define Z as the $[NT\times R]-$dimensional matrix the columns of which contain the R instruments orthogonal to the difference in alternative representations—which renders W_n as a $(J-1)R-$dimensional square matrix.

Finally, let $\mathbf{I}_{[J-1]}$ be a $(J-1)-$dimensional identity matrix

$$Y = \begin{bmatrix} \tilde{y}_{1,1,1} \\ \tilde{y}_{1,1,2} \\ \vdots \\ \tilde{y}_{1,N,T-1} \\ \tilde{y}_{1,N,T} \\ \vdots \\ \tilde{y}_{J-1,1,1} \\ \tilde{y}_{J-1,1,2} \\ \vdots \\ \tilde{y}_{J-1,N,T-1} \\ \tilde{y}_{J-1,N,T} \end{bmatrix}, \quad X = \begin{bmatrix} \tilde{x}_{1,1,1,1} & \dots & \tilde{x}_{1,1,1,K} \\ \tilde{x}_{1,1,2,1} & \dots & \tilde{x}_{1,1,2,K} \\ \vdots & & \vdots \\ \tilde{x}_{1,N,T-1,1} & \dots & \tilde{x}_{1,N,T-1,K} \\ \tilde{x}_{1,N,T,1} & \dots & \tilde{x}_{1,N,T,K} \\ \vdots & & \vdots \\ \tilde{x}_{J-1,1,1,1} & \dots & \tilde{x}_{J-1,1,1,K} \\ \tilde{x}_{J-1,1,2,1} & \dots & \tilde{x}_{J-1,1,2,K} \\ \vdots & & \vdots \\ \tilde{x}_{J-1,N,T-1,1} & \dots & \tilde{x}_{J-1,N,T-1,K} \\ \tilde{x}_{J-1,N,T,1} & \dots & \tilde{x}_{J-1,N,T,K} \end{bmatrix}, \quad Z = \begin{bmatrix} w(z_{11})_1 & \dots & w(z_{11})_R \\ w(z_{12})_1 & \dots & w(z_{12})_R \\ \vdots & & \vdots \\ w(z_{NT})_1 & \dots & w(z_{NT})_R \end{bmatrix}$$

And define

$$\tilde{Z} = \mathbf{I}_{[J-1]} \otimes Z$$

Then we can write the objective function in (B.13) as

$$\left((NT)^{-1}\tilde{Z}'(Y - XB)\right)' W_n \left((NT)^{-1}\tilde{Z}'(Y - XB)\right)$$

From where we can obtain a close form solution for $\hat{B}$ as the optimal GMM estimator. It entails a first stage estimator given by

$$\hat{B}^{1S} = \left(X'\tilde{Z}\tilde{Z}'X\right)^{-1}\left(X'\tilde{Z}\tilde{Z}'Y\right)$$

and a second stage estimator given by

$$\hat{B}^{2S} = \left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'X\right)^{-1}\left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'Y\right)$$

where

$$\hat{S} = \frac{1}{N^*}\tilde{Z}' D \tilde{Z}$$

and D is the $N(J-1) \times N(J-1)$ diagonal matrix with elements $\hat{u}_{jit}^2 = \left(y_{jit} - x_{jit}'\hat{B}^{1S}\right)^2$ in its diagonal. The variance-covariance matrix of the second stage estimator is

$$\hat{V}^{2S} = N^* \left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'X\right)^{-1}$$

and

$$N^* = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J-1} 1\left\{\text{Decision } j \text{ available for } i \text{ at } t\right\}$$

which accounts for the fact that some individuals cannot stay in their lagged treatments at some periods (for instance, if lagged decision was no treatment or trial treatment).

B.2.5 CCP Estimation and Fit

The probability that an individual chooses one of the options depends on the elements of his state. As such, the CCPs needed to simulate choices in our estimation method are functions of individual-specific variables as well as market-level variables.

Individuals decide between one of $\mathcal{G}$ clusters, yesterday's product (if any), a trial product, and no product. Let W_{jit} be the characteristics describing alternative j for individual i at period t: mean health, mean ailments, and the variance matrix. Let $W_{jit}W_{jit}$ denote a vector of interactions between the elements of W_{jit}. Let $\tilde{x}_{it}$ and $\tilde{z}_{it}$ be subsets of the individual-specific components of the state.[3] Let $\omega_t W_{jit}$ denote a vector of interactions between the centroid and the elements of W_{jit}. Similarly, let $W_{jit}\tilde{z}_{it}$ be a vector of interactions between the components of W_{jit} and individual-specific state components and let $\omega_t W_{jit}\tilde{z}_{it}$ be defined in a similar fashion. Finally, let $\tilde{\mathcal{F}}_t$ denote a set of non parametric moments describing the

[3] $\tilde{z}_{it}$ includes h_{it-1}, a_{it-1}, b_i, l_{it} while $\tilde{x}_{it}$ includes a constant, a_{it-1}, b_i.

joint distribution of aggregate characteristics, $\mathcal{F}_t$.[4]

For each of the options, the CCPs are expressed as follows:

Cluster ccps $(j = 1, \ldots, \mathcal{G})$

$$p_{jit} = \frac{\exp\left(\gamma_0 \tilde{x}_{it} + \beta_0 W_{jit} + \beta_1 W_{jit} W_{jit} + \beta_2 \omega_t W_{jit} + \beta_3 W_{jit} \tilde{z}_{it} + \beta_4 \omega_t W_{jit} \tilde{z}_{it} + \beta_5 W_{jit} \tilde{\mathcal{F}}_t\right)}{1 + \sum_{k=1}^{\mathcal{G}+2} \exp(\cdot)} \tag{B.14}$$

γ_0 is constant across clusters and over time. For a given cluster j and period t, W_{jit} is in fact constant across individuals so $W_{jit} = W_{jt}$.

Trial ccps $(j = \mathcal{G} + 1)$

$$p_{jit} = \frac{\exp\left(\gamma_j \tilde{x}_{it} + \beta_0 W_{jit} + \beta_1 W_{jit} W_{jit} + \beta_3 W_{jit} \tilde{z}_{it} + \beta_5 W_{jit} \tilde{\mathcal{F}}_t\right)}{1 + \sum_{k=1}^{\mathcal{G}+2} \exp(\cdot)} \tag{B.15}$$

For the trial alternative, W_{jit} is constant across individuals so $W_{\mathcal{G}+1it} = W_{\mathcal{G}+1t}$. In fact, two of the components of W_{jt} are $\omega_{t-1} + \mu_\nu$, where μ_ν is the mean of the innovations distribution. Therefore, to avoid collinearity we do not include terms $\omega_t W_{jt}$ and $\omega_t W_{jt} \tilde{z}_{it}$ in the trials ccps.

Staying ccps $(j = \mathcal{G} + 2)$

$$p_{jit} = \frac{\exp\left(\gamma_j \tilde{x}_{it} + \beta_0 W_{jit} + \beta_1 W_{jit} W_{jit} + \beta_2 \omega_t W_{jit} + \beta_3 W_{jit} \tilde{z}_{it} + \beta_4 \omega_t W_{jit} \tilde{z}_{it} + \beta_5 W_{jit} \tilde{\mathcal{F}}_t\right)}{1 + \sum_{k=1}^{\mathcal{G}+2} \exp(\cdot)} \tag{B.16}$$

When individuals choose to stick to their previous product $W_{\mathcal{G}+1it}$ becomes heterogeneous—individuals may have consumed different products last period.

No product ccps $(j = 0)$

$$p_{jit} = 1 - \sum_{k=1}^{\mathcal{G}+2} p_{kit} \tag{B.17}$$

[4] We specify these moments as shares of people with different sets of characteristics.

Even though the characteristics of the choice set are non stationary, by interacting our time-varying regressors $\tilde{z}_{it}$ with the characteristics of the choice for individual i, W_{jit}, we are able to control for the state of the world inside the ccps. As a consequence of this we do not have to run period-specific logits and we can have ccps for any simulated world as long as our observed worlds cover the space of possible worlds reasonably well. We also include parameters that are invariant to the state of the technology, γ, which capture stationary taste differences between staying in current choice, trying a new market product, going to a trial, or not consuming anything. Also, since all clusters correspond to the action of "trying a market product" we impose $\gamma_j = \gamma_{j'} = \gamma_0$ for any $j, j' = 1, \ldots, \mathcal{G}$.

Figures B.1, B.2, and B.3 display the mean predicted conditional choice probability using equations (B.14), (B.16), (B.15) and (B.17) over time against the correspondent share of the population who chose the alternative. Our ccps map the choices in the data relatively well. In fact, we further explore the fit of our ccp estimates comparing the relatives shares that clusters received in reality against our the predictions from our estimated ccps. We do this by ranking the three clusters at every period by the share they received and comparing this ranking with the ranking obtained from our estimated ccps. A cross tabulation of these rankings—not shown here—suggests that the predicted ranks match the real ranks in more than 79 percent of the periods. In fact, the lowest-ranking cluster matches the predicted lowest-ranking cluster 88 percent of the times.

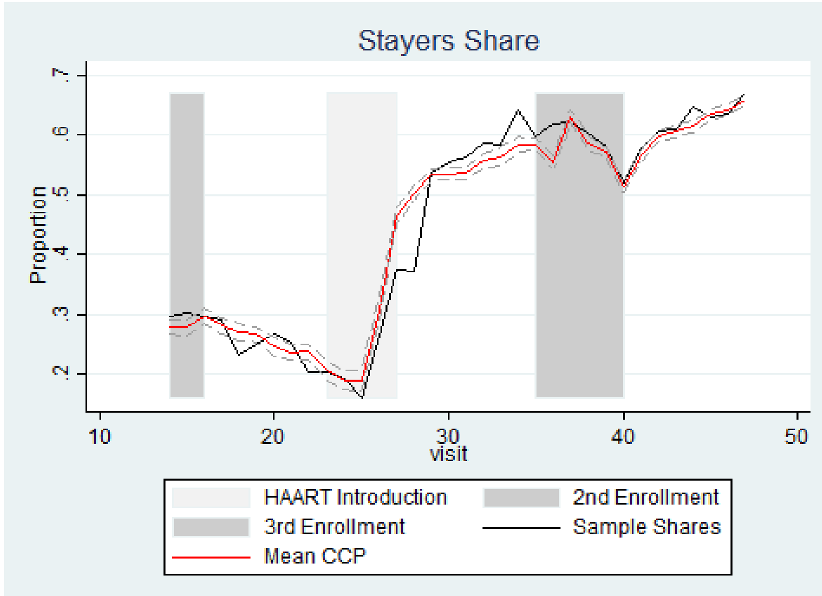


Figure B.1: CCPs Goodness of fit: Stayers

Notes: Figure shows the average estimated conditional choice probability against the share of people choosing the alternative. Dashed lines represent 95 percent confidence intervals around the predicted CCPs. Figure highlights three periods that are of special relevance in the period we study. It depicts periods during which enrollment into the sample was undertaken and more importantly, it displays the period in which products belonging to the HAART family were introduced.

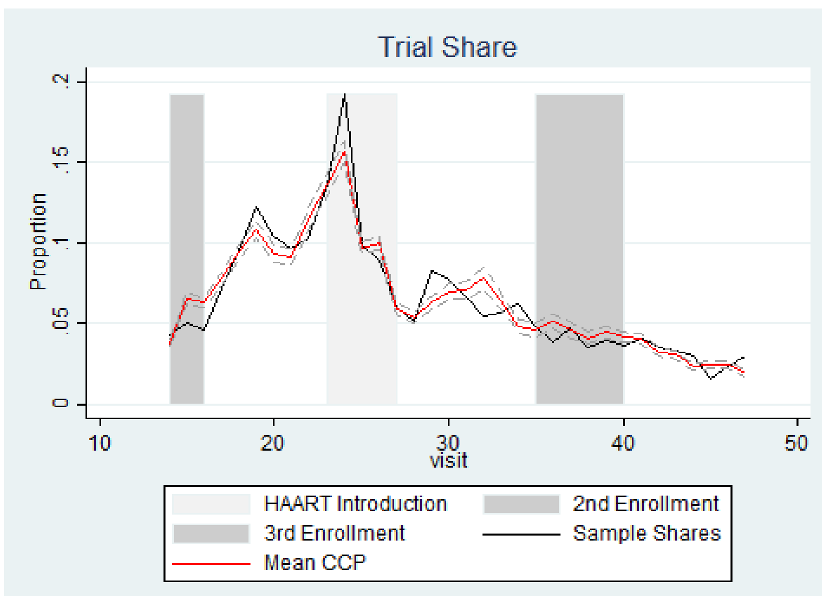


Figure B.2: CCPs Goodness of fit: Trial

Notes: Figure shows the average estimated conditional choice probability against the share of people choosing the alternative. Dashed lines represent 95 percent confidence intervals around the predicted CCPs. Figure highlights three periods that are of special relevance in the period we study. It depicts periods during which enrollment into the sample was undertaken and more importantly, it displays the period in which products belonging to the HAART family were introduced.

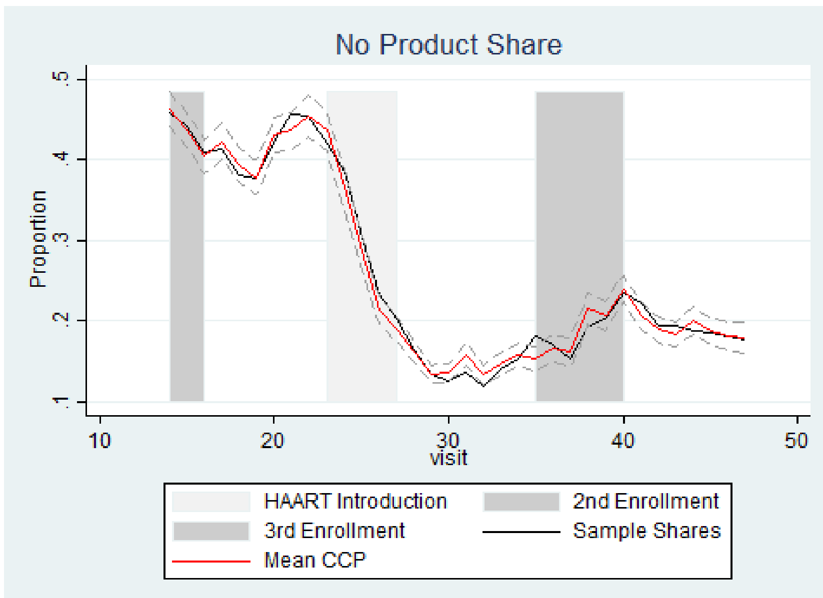


Figure B.3: CCPs Goodness of fit: No Product

Notes: Figure shows the average estimated conditional choice probability against the share of people choosing the alternative. Dashed lines represent 95 percent confidence intervals around the predicted CCPs. Figure highlights three periods that are of special relevance in the period we study. It depicts periods during which enrollment into the sample was undertaken and more importantly, it displays the period in which products belonging to the HAART family were introduced.

B.2.6 Proof of Proposition 3.1

$$
\begin{aligned}
v_{jit}(z_{it}) &= y_{jit} + \beta \mathbb{E}\left[V\left(z_{it+1}, \varepsilon_{it+1}\right) | z_{it}, j\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \varepsilon_{ikt+1}\right]\right] | z_{it}, j\right] \\
&\quad + \beta^2 \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} \mathbb{E}_{\varepsilon}\left[d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \varepsilon_{ikt+1}\right]\right| d^{o}_{ikt+1}\left(z_{it+s}\right) = 1\right]\right] | z_{it}, j\right] \\
&\quad + \beta^2 \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \mathbb{E}_{\varepsilon}\left[\varepsilon_{ikt+1}\right| d^{o}_{ikt+1}\left(z_{it+s}\right) = 1\right]\right]\right] | z_{it}, j\right] \\
&\quad + \beta^2 \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \psi_{ikt+s}\left(z_{it+s}\right)\right]\right] | z_{it}, j\right] \\
&\quad + \beta^2 \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+1}\right)\left[y_{ikt+1}\left(z_{it+1}\right) + \psi_{ikt+1}\left(z_{it+1}\right)\right]\right] | z_{it}, j\right] \\
&\quad + \beta^2 P^{o(2-1)}_{j}\left(z_{it}\right) \mathbb{E}\left[V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, S_{it+2-1} = 1, d^{o}_{i}\right] \\
&= y_{jit} \\
&\quad + \sum_{s=1}^{T^*} \beta^s P^{o(s-1)}_{j}\left(z_{it}\right) \mathbb{E}_{z}\left[D_{it+s}\left(z_{it+s}\right) \sum_{k \in C_{t+s}} p_{ikt+s}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \psi_{ikt+s}\left(z_{it+s}\right)\right] \middle| z_{it}, j, S_{it+s-1} = 1, d^{o}_{i}\right] \\
&\quad + \beta^{T^*+1} P^{o(T^*)}_{j}\left(z_{it}\right) \mathbb{E}_{z}\left[D_{it+T^*+1}\left(z_{it+T^*+1}\right) V\left(z_{it+T^*+1}, \varepsilon_{it+T^*+1}\right)\right| z_{it}, j, S_{it+T^*} = 1, d^{o}_{i}\right]
\end{aligned}
$$

That

$$\psi_{kit}\left(z_{it}\right) = \gamma - \ln\left(p_{kit}\left(z_{it}\right)\right) \tag{B.18}$$

follows from the joint distribution of the taste shifter ε_{it}, which is Extreme Value Type-I. *Q.E.D.*

B.3 Additional Tables and Figures

Table B.5: Health Characteristics of Treatments

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	cd4	cd4	cd4	cd4	cd4	cd4
$CD4_{t-1}$	0.834***	1.064***	1.009***	1.021***	1.152***	1.136***
	(0.006)	(0.013)	(0.014)	(0.026)	(0.032)	(0.051)
$CD4^2_{t-1}/10^3$		-0.174***	-0.097***	-0.121**	-0.519***	-0.456**
		(0.012)	(0.018)	(0.059)	(0.098)	(0.212)
$CD4^3_{t-1}/10^7$			-0.274***	-0.112	4.375***	3.363
			(0.060)	(0.430)	(1.123)	(3.518)
$CD4^4_{t-1}/10^{10}$				-0.031	-2.016***	-1.288
				(0.080)	(0.502)	(2.598)
$CD4^5_{t-1}/10^{14}$					2.803***	0.482
					(0.718)	(8.325)
$CD4^6_{t-1}/10^{18}$						2.623
						(9.398)
DT3	-21.583***	-11.174***	-11.950***	-11.909***	-12.004***	-11.983***
	(2.840)	(2.707)	(2.700)	(2.699)	(2.697)	(2.697)
DT9	-20.319*	-19.561*	-19.089	-19.165	-19.655*	-19.623*
	(11.743)	(11.879)	(11.865)	(11.865)	(11.861)	(11.860)
DT13	-72.170***	-53.988***	-55.512***	-55.437***	-55.796***	-55.726***
	(11.798)	(12.504)	(12.348)	(12.364)	(12.455)	(12.455)
DT14	-15.506**	-4.315	-5.164	-5.115	-5.155	-5.140
	(6.241)	(6.105)	(6.100)	(6.098)	(6.094)	(6.094)
DT33	-21.629**	0.985	-2.034	-1.689	-0.017	-0.112
	(9.038)	(9.161)	(9.112)	(9.123)	(9.108)	(9.113)
DT34	-21.810***	-12.450***	-13.310***	-13.219***	-12.752***	-12.779***
	(4.900)	(4.755)	(4.761)	(4.764)	(4.764)	(4.765)
DT39	-31.246***	-15.492***	-17.057***	-16.924***	-16.615***	-16.607***
	(5.805)	(5.715)	(5.691)	(5.691)	(5.687)	(5.688)
DT46	7.741*	15.348***	14.581***	14.678***	15.263***	15.229***
	(4.679)	(4.566)	(4.577)	(4.581)	(4.574)	(4.573)
DT47	-34.371***	-15.583**	-17.510**	-17.327**	-16.474**	-16.521**
	(7.318)	(7.033)	(7.027)	(7.028)	(7.040)	(7.037)
DT51	-22.630***	-3.664	-6.022	-5.740	-4.159	-4.252

	(7.962)	(7.693)	(7.690)	(7.693)	(7.669)	(7.670)
DT63	-16.743	2.384	-0.183	0.162	2.415	2.275
	(14.637)	(13.594)	(13.649)	(13.659)	(13.746)	(13.735)
DT64	-37.988***	-17.449*	-19.583**	-19.387**	-18.630**	-18.656**
	(8.900)	(9.076)	(8.991)	(8.996)	(9.035)	(9.032)
DT65	-27.913***	-12.409*	-13.823**	-13.704*	-13.186*	-13.220*
	(7.203)	(7.007)	(7.006)	(7.006)	(6.993)	(6.994)
DT67	-50.755***	-31.179**	-33.087**	-32.938**	-32.700**	-32.673**
	(15.300)	(14.998)	(14.986)	(14.990)	(15.052)	(15.044)
DT69	-26.741*	-11.827	-13.331	-13.215	-13.351	-13.275
	(14.379)	(13.908)	(13.950)	(13.945)	(13.973)	(13.973)
DT85	34.619***	40.457***	39.721***	39.790***	39.776***	39.792***
	(6.424)	(6.319)	(6.311)	(6.308)	(6.299)	(6.299)
DT117	33.323***	42.736***	41.910***	41.991***	42.267***	42.277***
	(12.098)	(11.819)	(11.837)	(11.834)	(11.819)	(11.818)
DT124	33.711***	33.910***	33.804***	33.864***	34.398***	34.364***
	(6.284)	(6.229)	(6.234)	(6.235)	(6.227)	(6.228)
DT146	27.752*	34.323**	33.694**	33.761**	33.792**	33.831**
	(14.796)	(14.611)	(14.644)	(14.639)	(14.625)	(14.625)
DT157	34.353***	37.455***	37.258***	37.282***	37.173***	37.208***
	(6.945)	(6.861)	(6.862)	(6.862)	(6.856)	(6.857)
DT161	33.496**	38.559***	38.364***	38.340***	38.283***	38.259***
	(13.411)	(13.215)	(13.201)	(13.205)	(13.193)	(13.198)
DT164	55.089**	64.825**	64.314**	64.302**	63.734**	63.798**
	(27.560)	(27.365)	(27.365)	(27.370)	(27.414)	(27.419)
DT165	60.168***	64.722***	65.220***	65.337***	65.041***	65.045***
	(7.077)	(6.246)	(6.215)	(6.225)	(6.220)	(6.222)
DT169	33.129**	34.545**	34.182**	34.289**	35.032**	35.012**
	(16.388)	(16.316)	(16.316)	(16.317)	(16.339)	(16.334)
DT171	73.104***	78.825***	78.453***	78.478***	78.559***	78.548***
	(17.682)	(18.040)	(18.017)	(18.010)	(17.950)	(17.943)
DT175	44.728***	52.470***	52.730***	52.619***	53.128***	53.153***
	(8.770)	(8.233)	(8.202)	(8.198)	(8.176)	(8.172)
DT185	50.899***	58.842***	57.833***	57.922***	57.776***	57.825***
	(12.661)	(12.659)	(12.642)	(12.638)	(12.608)	(12.608)
DT202	32.648**	32.522**	33.226**	33.107**	32.286**	32.338**
	(14.544)	(14.584)	(14.576)	(14.576)	(14.573)	(14.574)
DT214	33.330***	33.541***	34.154***	34.057***	33.510***	33.535***
	(12.245)	(12.162)	(12.166)	(12.164)	(12.163)	(12.161)

DT236	48.886***	46.186***	46.239***	46.267***	46.275***	46.281***
	(7.172)	(7.130)	(7.121)	(7.123)	(7.123)	(7.124)
DT242	47.980***	46.484***	46.110***	46.240***	46.846***	46.863***
	(9.266)	(9.144)	(9.154)	(9.160)	(9.161)	(9.162)
DT254	42.775***	42.587***	42.568***	42.601***	42.631***	42.656***
	(13.502)	(13.436)	(13.444)	(13.443)	(13.446)	(13.445)
DT268	47.316***	52.474***	51.249***	51.353***	50.776***	50.855***
	(10.457)	(10.436)	(10.415)	(10.407)	(10.417)	(10.415)
DT292	42.030***	48.796***	48.057***	48.109***	48.018***	48.069***
	(10.638)	(10.242)	(10.225)	(10.230)	(10.212)	(10.215)
DT311	39.770*	50.740**	49.168**	49.215**	47.816**	47.922**
	(23.206)	(22.720)	(22.698)	(22.709)	(22.774)	(22.780)
DT349	50.575***	42.467**	43.413**	43.408**	44.240***	44.156***
	(16.958)	(17.083)	(17.108)	(17.098)	(17.046)	(17.046)
DT377	56.809***	57.778***	57.727***	57.716***	57.227***	57.259***
	(19.614)	(19.566)	(19.530)	(19.538)	(19.552)	(19.555)
DT532	49.321***	47.631***	47.537***	47.614***	47.978***	47.990***
	(10.952)	(10.886)	(10.899)	(10.899)	(10.893)	(10.892)
DT548	45.842***	43.345***	43.281***	43.348***	43.526***	43.551***
	(5.368)	(5.331)	(5.329)	(5.330)	(5.327)	(5.329)
DT573	40.314***	39.595***	39.432***	39.485***	39.379***	39.426***
	(10.981)	(10.960)	(10.901)	(10.909)	(10.919)	(10.922)
DT581	19.612	18.387	18.413	18.417	17.866	17.935
	(14.499)	(14.316)	(14.341)	(14.341)	(14.376)	(14.378)
DT615	44.239***	40.856***	41.087***	41.122***	41.280***	41.301***
	(11.769)	(11.592)	(11.603)	(11.604)	(11.622)	(11.622)
DT644	54.543***	53.883***	53.615***	53.650***	53.341***	53.368***
	(8.639)	(8.512)	(8.504)	(8.505)	(8.516)	(8.515)
DT701	52.853***	55.916***	54.878***	54.997***	54.824***	54.870***
	(11.144)	(10.997)	(10.979)	(10.975)	(10.999)	(11.003)
DT720	60.713***	79.231***	78.995***	78.688***	78.914***	78.726***
	(13.046)	(14.550)	(14.500)	(14.464)	(14.412)	(14.405)
DT782	28.143*	35.924**	35.177**	35.267**	35.611**	35.633**
	(15.627)	(14.945)	(14.959)	(14.967)	(15.077)	(15.068)
DT869	50.005***	50.037***	49.904***	49.946***	49.838***	49.885***
	(13.110)	(12.947)	(12.940)	(12.946)	(12.997)	(12.999)
DT909	27.964**	33.525***	33.367***	33.310***	32.227***	32.320***
	(10.898)	(10.827)	(10.764)	(10.781)	(10.842)	(10.846)
DT919	47.628***	48.846***	48.522***	48.536***	47.617***	47.722***

	(13.308)	(13.619)	(13.486)	(13.499)	(13.453)	(13.466)
DT921	15.458	18.929	19.116	19.058	19.273	19.286
	(16.086)	(14.468)	(14.426)	(14.448)	(14.327)	(14.341)
DT923	30.619*	26.759	26.896	26.960	27.246	27.283
	(17.077)	(16.964)	(16.951)	(16.954)	(16.971)	(16.973)
DT926	47.776***	47.965***	47.952***	47.971***	47.790***	47.835***
	(10.277)	(9.953)	(9.995)	(9.993)	(10.024)	(10.021)
DT940	54.575***	50.873***	50.705***	50.821***	51.570***	51.537***
	(15.098)	(15.196)	(15.198)	(15.195)	(15.164)	(15.163)
DT949	53.752***	50.921***	51.203***	51.236***	51.672***	51.661***
	(11.818)	(11.668)	(11.706)	(11.701)	(11.705)	(11.701)
DT963	37.095***	30.953**	31.641**	31.628**	31.845**	31.829**
	(12.997)	(13.013)	(13.021)	(13.019)	(13.014)	(13.015)
DT976	-7.655	2.510	2.487	2.349	2.381	2.436
	(20.271)	(17.213)	(16.922)	(16.980)	(16.471)	(16.507)
DT987	4.115	10.199	9.583	9.654	9.855	9.870
	(14.474)	(14.526)	(14.496)	(14.498)	(14.503)	(14.500)
DT994	10.625	15.088	15.225	15.171	14.891	14.893
	(17.109)	(17.416)	(17.404)	(17.396)	(17.282)	(17.291)
DT1010	12.409	20.292**	19.891**	19.906**	19.980**	19.966**
	(9.769)	(9.701)	(9.687)	(9.686)	(9.676)	(9.676)
DT1011	43.817*	39.067*	38.498*	38.691*	39.457*	39.484*
	(22.599)	(22.024)	(22.150)	(22.142)	(22.163)	(22.157)
DT1020	20.973	17.603	18.253	18.220	18.396	18.375
	(13.204)	(13.121)	(13.141)	(13.135)	(13.111)	(13.111)
DT1071	58.721***	53.880***	54.012***	54.094***	54.798***	54.796***
	(4.578)	(4.475)	(4.469)	(4.478)	(4.453)	(4.453)
DT1128	40.782***	37.265***	37.397***	37.415***	37.227***	37.246***
	(9.919)	(9.662)	(9.682)	(9.683)	(9.722)	(9.720)
DT1222	53.239***	52.943***	52.719***	52.793***	53.028***	53.050***
	(5.409)	(5.301)	(5.306)	(5.307)	(5.309)	(5.310)
DT1227	83.132***	85.565***	84.437***	84.556***	83.823***	83.917***
	(20.468)	(21.138)	(21.037)	(21.031)	(20.842)	(20.859)
DT1230	25.312**	27.037**	27.504**	27.416**	26.850**	26.886**
	(12.040)	(12.112)	(12.093)	(12.093)	(12.079)	(12.081)
DT1245	37.822***	38.379***	38.357***	38.374***	38.313***	38.336***
	(13.418)	(13.334)	(13.326)	(13.328)	(13.347)	(13.347)
DT1253	46.066***	46.654***	46.213***	46.319***	46.723***	46.735***
	(7.946)	(7.735)	(7.745)	(7.747)	(7.767)	(7.765)

DT1303	51.942***	46.893***	47.631***	47.602***	47.800***	47.786***
	(16.434)	(16.225)	(16.276)	(16.269)	(16.284)	(16.279)
DT1342	32.840**	29.637**	30.481**	30.395**	30.226**	30.204**
	(14.069)	(14.124)	(14.137)	(14.132)	(14.115)	(14.116)
DT10006	-29.515	-21.709	-22.258	-22.209	-21.950	-21.961
	(19.930)	(20.278)	(20.256)	(20.251)	(20.191)	(20.194)
DT10026	64.425***	66.829***	65.194***	65.444***	65.353***	65.496***
	(24.136)	(22.649)	(22.765)	(22.778)	(23.021)	(23.023)
DT10027	0.210	7.741	7.165	7.179	6.457	6.562
	(14.742)	(14.802)	(14.701)	(14.715)	(14.739)	(14.748)
DT10028	26.450	32.481	31.616	31.629	30.293	30.414
	(19.796)	(20.521)	(20.367)	(20.371)	(20.276)	(20.282)
DT10030	20.031	19.261	19.703	19.654	19.278	19.315
	(17.621)	(17.572)	(17.592)	(17.586)	(17.546)	(17.548)
DT10031	34.205**	29.830**	30.781**	30.712**	31.044**	30.969**
	(15.053)	(14.878)	(14.911)	(14.901)	(14.838)	(14.843)
DT10035	44.294***	43.738***	43.648***	43.682***	43.495***	43.550***
	(15.320)	(15.368)	(15.337)	(15.340)	(15.357)	(15.356)
DT10037	28.525***	28.144***	28.115***	28.134***	27.893**	27.931**
	(11.035)	(10.856)	(10.855)	(10.857)	(10.893)	(10.892)
DT10038	48.755***	44.792***	44.735***	44.855***	45.683***	45.679***
	(13.769)	(13.693)	(13.692)	(13.695)	(13.673)	(13.674)
DT10040	26.314***	28.505***	28.742***	28.692***	28.440***	28.431***
	(8.531)	(8.329)	(8.333)	(8.335)	(8.334)	(8.337)
DT10041	43.028***	41.990***	42.154***	42.158***	42.050***	42.074***
	(12.657)	(12.467)	(12.470)	(12.470)	(12.474)	(12.473)
DT10042	30.254**	32.127**	31.743**	31.779**	31.824**	31.811**
	(12.900)	(13.008)	(13.007)	(12.996)	(12.924)	(12.922)
DT10043	28.589**	26.499**	26.919**	26.894**	26.678**	26.715**
	(12.874)	(12.711)	(12.704)	(12.707)	(12.739)	(12.739)
DT10046	36.196***	32.318**	32.455**	32.517**	32.865**	32.884**
	(13.618)	(13.720)	(13.651)	(13.662)	(13.672)	(13.678)
DT10048	33.017**	33.618**	33.783**	33.756**	33.352**	33.394**
	(15.278)	(14.856)	(14.884)	(14.888)	(14.943)	(14.943)
DT10049	49.474***	47.695***	48.034***	48.005***	47.736***	47.760***
	(11.002)	(10.687)	(10.698)	(10.703)	(10.757)	(10.753)
DT10050	38.954**	38.166**	38.394**	38.415**	38.933**	38.887**
	(17.897)	(18.043)	(18.036)	(18.037)	(18.096)	(18.090)
Constant	56.819***	-2.797	6.623**	5.233	-5.874*	-4.898

	(3.161)	(3.411)	(3.018)	(3.432)	(3.313)	(3.681)
Observations	33,258	33,258	33,258	33,258	33,258	33,258
R-squared	0.728	0.736	0.736	0.736	0.736	0.736

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table B.6: No Ailments Characteristics of Treatments

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	NoSfx	NoSfx	NoSfx	NoSfx	NoSfx	NoSfx
$CD4_{t-1}$	0.001***	0.003***	0.005***	0.006***	0.008***	0.009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
$CD4_{t-1}^2/10^3$		-0.001***	-0.005***	-0.007***	-0.013***	-0.017***
		(0.000)	(0.001)	(0.001)	(0.002)	(0.003)
$CD4_{t-1}^3/10^7$			0.013***	0.033***	0.109***	0.166***
			(0.003)	(0.004)	(0.019)	(0.040)
$CD4_{t-1}^4/10^{10}$				-0.005***	-0.040***	-0.084***
				(0.001)	(0.010)	(0.028)
$CD4_{t-1}^5/10^{14}$					0.054***	0.203**
					(0.016)	(0.092)
$CD4_{t-1}^6/10^{18}$						-0.186*
						(0.106)
DT3	-0.576***	-0.515***	-0.500***	-0.498***	-0.500***	-0.501***
	(0.040)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
DT9	-0.402*	-0.404*	-0.421**	-0.427**	-0.433**	-0.434**
	(0.208)	(0.212)	(0.213)	(0.213)	(0.212)	(0.212)
DT13	-0.738***	-0.627**	-0.594**	-0.593**	-0.600**	-0.603**
	(0.260)	(0.259)	(0.260)	(0.260)	(0.260)	(0.260)
DT14	-0.523***	-0.457***	-0.439***	-0.437***	-0.439***	-0.439***
	(0.099)	(0.100)	(0.100)	(0.101)	(0.101)	(0.101)
DT33	-1.039***	-0.900***	-0.819***	-0.799***	-0.783***	-0.781***
	(0.127)	(0.128)	(0.130)	(0.130)	(0.131)	(0.132)
DT34	-0.622***	-0.569***	-0.549***	-0.544***	-0.539***	-0.538***
	(0.079)	(0.079)	(0.080)	(0.080)	(0.080)	(0.081)
DT39	-0.713***	-0.618***	-0.580***	-0.573***	-0.571***	-0.571***
	(0.105)	(0.106)	(0.107)	(0.108)	(0.108)	(0.108)
DT46	-0.457***	-0.411***	-0.390***	-0.383***	-0.375***	-0.374***
	(0.069)	(0.070)	(0.070)	(0.071)	(0.071)	(0.071)
DT47	-1.016***	-0.907***	-0.866***	-0.858***	-0.851***	-0.850***
	(0.157)	(0.159)	(0.162)	(0.162)	(0.163)	(0.163)
DT51	-0.577***	-0.456***	-0.388**	-0.369**	-0.348**	-0.344**
	(0.161)	(0.164)	(0.166)	(0.167)	(0.168)	(0.168)
DT63	-0.567**	-0.445	-0.367	-0.341	-0.310	-0.304

	(0.269)	(0.274)	(0.280)	(0.282)	(0.284)	(0.285)
DT64	-0.566***	-0.436**	-0.379*	-0.367*	-0.358	-0.357
	(0.210)	(0.213)	(0.218)	(0.220)	(0.222)	(0.222)
DT65	-0.650***	-0.557***	-0.525***	-0.519***	-0.514***	-0.512***
	(0.133)	(0.133)	(0.135)	(0.135)	(0.136)	(0.137)
DT67	-1.582***	-1.473***	-1.439***	-1.436***	-1.440***	-1.442***
	(0.306)	(0.303)	(0.305)	(0.306)	(0.307)	(0.307)
DT69	-0.916***	-0.827***	-0.790***	-0.785***	-0.789***	-0.793***
	(0.261)	(0.261)	(0.262)	(0.263)	(0.263)	(0.264)
DT85	-0.772***	-0.737***	-0.718***	-0.715***	-0.717***	-0.718***
	(0.091)	(0.091)	(0.091)	(0.092)	(0.092)	(0.092)
DT117	-0.609***	-0.555***	-0.535**	-0.530**	-0.527**	-0.528**
	(0.205)	(0.208)	(0.210)	(0.211)	(0.212)	(0.212)
DT124	0.036	0.040	0.050	0.055	0.064	0.065
	(0.093)	(0.094)	(0.094)	(0.094)	(0.094)	(0.094)
DT146	-0.568**	-0.530**	-0.513**	-0.509**	-0.509**	-0.511**
	(0.223)	(0.226)	(0.227)	(0.227)	(0.227)	(0.227)
DT157	-0.132	-0.111	-0.103	-0.102	-0.104	-0.105
	(0.112)	(0.113)	(0.114)	(0.114)	(0.114)	(0.114)
DT161	-0.302*	-0.270	-0.270	-0.272	-0.271	-0.270
	(0.180)	(0.182)	(0.182)	(0.182)	(0.182)	(0.182)
DT164	-0.532***	-0.475**	-0.468**	-0.471**	-0.479**	-0.482**
	(0.184)	(0.185)	(0.186)	(0.187)	(0.187)	(0.187)
DT165	-0.106	-0.084	-0.078	-0.073	-0.075	-0.076
	(0.080)	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)
DT169	-0.604**	-0.600**	-0.585**	-0.577**	-0.567**	-0.566**
	(0.244)	(0.246)	(0.248)	(0.249)	(0.250)	(0.250)
DT171	-0.342*	-0.308	-0.299	-0.297	-0.295	-0.295
	(0.190)	(0.195)	(0.197)	(0.198)	(0.199)	(0.199)
DT175	-0.424***	-0.377***	-0.399***	-0.402***	-0.395***	-0.397***
	(0.099)	(0.100)	(0.100)	(0.100)	(0.100)	(0.100)
DT185	-0.671***	-0.621***	-0.592***	-0.588***	-0.591***	-0.594***
	(0.174)	(0.175)	(0.176)	(0.176)	(0.177)	(0.177)
DT202	0.003	-0.002	-0.027	-0.036	-0.048	-0.050
	(0.226)	(0.227)	(0.228)	(0.228)	(0.229)	(0.229)
DT214	-0.720***	-0.730***	-0.754***	-0.761***	-0.767***	-0.768***
	(0.267)	(0.268)	(0.269)	(0.269)	(0.269)	(0.268)
DT236	0.125	0.107	0.109	0.110	0.109	0.109
	(0.112)	(0.112)	(0.112)	(0.111)	(0.111)	(0.111)

DT242	-0.415***	-0.425***	-0.403***	-0.393***	-0.386***	-0.387***
	(0.121)	(0.120)	(0.121)	(0.120)	(0.121)	(0.121)
DT254	-0.446***	-0.450***	-0.447***	-0.444***	-0.444***	-0.445***
	(0.160)	(0.159)	(0.159)	(0.159)	(0.160)	(0.160)
DT268	-0.508***	-0.466***	-0.426***	-0.423***	-0.432***	-0.436***
	(0.122)	(0.122)	(0.122)	(0.122)	(0.122)	(0.122)
DT292	-0.941***	-0.900***	-0.882***	-0.879***	-0.881***	-0.883***
	(0.128)	(0.130)	(0.130)	(0.130)	(0.130)	(0.130)
DT311	-0.913***	-0.830***	-0.785***	-0.788***	-0.810***	-0.813***
	(0.214)	(0.213)	(0.214)	(0.213)	(0.213)	(0.213)
DT349	0.809**	0.756**	0.736**	0.740**	0.753**	0.757**
	(0.332)	(0.327)	(0.325)	(0.324)	(0.324)	(0.325)
DT377	-1.043***	-1.040***	-1.039***	-1.041***	-1.049***	-1.050***
	(0.219)	(0.220)	(0.219)	(0.220)	(0.220)	(0.220)
DT532	-0.353**	-0.366**	-0.356**	-0.350**	-0.346**	-0.346**
	(0.146)	(0.148)	(0.149)	(0.149)	(0.149)	(0.149)
DT548	0.340***	0.327***	0.337***	0.341***	0.342***	0.341***
	(0.084)	(0.085)	(0.085)	(0.085)	(0.085)	(0.085)
DT573	0.031	0.027	0.038	0.040	0.038	0.036
	(0.233)	(0.231)	(0.232)	(0.233)	(0.235)	(0.235)
DT581	-0.454**	-0.463**	-0.460**	-0.462**	-0.470***	-0.473***
	(0.181)	(0.180)	(0.180)	(0.180)	(0.180)	(0.179)
DT615	-0.605***	-0.631***	-0.632***	-0.628***	-0.626***	-0.627***
	(0.163)	(0.164)	(0.164)	(0.164)	(0.163)	(0.163)
DT644	0.106	0.102	0.113	0.113	0.108	0.107
	(0.128)	(0.126)	(0.125)	(0.125)	(0.125)	(0.125)
DT701	-0.508***	-0.481***	-0.444***	-0.438***	-0.442***	-0.444***
	(0.127)	(0.128)	(0.128)	(0.128)	(0.128)	(0.128)
DT720	0.263	0.387**	0.349*	0.337*	0.348*	0.355*
	(0.190)	(0.181)	(0.186)	(0.188)	(0.186)	(0.185)
DT782	-0.393	-0.345	-0.321	-0.315	-0.310	-0.311
	(0.262)	(0.265)	(0.269)	(0.270)	(0.271)	(0.271)
DT869	-0.661***	-0.664***	-0.656***	-0.653***	-0.655***	-0.657***
	(0.183)	(0.183)	(0.182)	(0.181)	(0.180)	(0.180)
DT909	-0.560***	-0.528***	-0.530***	-0.537***	-0.552***	-0.556***
	(0.197)	(0.196)	(0.194)	(0.193)	(0.193)	(0.193)
DT919	-0.974***	-0.966***	-0.951***	-0.952***	-0.966***	-0.971***
	(0.274)	(0.268)	(0.261)	(0.260)	(0.261)	(0.262)
DT921	-0.659***	-0.641***	-0.656***	-0.657***	-0.652***	-0.653***

	(0.204)	(0.207)	(0.206)	(0.206)	(0.205)	(0.205)
DT923	-0.244	-0.270	-0.266	-0.261	-0.258	-0.260
	(0.212)	(0.213)	(0.214)	(0.214)	(0.214)	(0.214)
DT926	-0.015	-0.013	-0.009	-0.008	-0.011	-0.013
	(0.153)	(0.153)	(0.153)	(0.153)	(0.154)	(0.154)
DT940	-0.940***	-0.967***	-0.952***	-0.943***	-0.934***	-0.933***
	(0.207)	(0.206)	(0.207)	(0.208)	(0.208)	(0.208)
DT949	-0.081	-0.100	-0.102	-0.098	-0.092	-0.092
	(0.171)	(0.173)	(0.174)	(0.174)	(0.174)	(0.174)
DT963	-0.251	-0.296*	-0.311*	-0.311*	-0.308*	-0.307*
	(0.178)	(0.179)	(0.179)	(0.179)	(0.178)	(0.178)
DT976	-0.463*	-0.396	-0.418	-0.422*	-0.420*	-0.424*
	(0.253)	(0.259)	(0.255)	(0.254)	(0.253)	(0.253)
DT987	0.214	0.262	0.287	0.294	0.298	0.297
	(0.260)	(0.266)	(0.271)	(0.272)	(0.274)	(0.275)
DT994	-0.814***	-0.798***	-0.813***	-0.817***	-0.820***	-0.820***
	(0.244)	(0.246)	(0.248)	(0.248)	(0.249)	(0.249)
DT1010	-0.778***	-0.739***	-0.738***	-0.739***	-0.738***	-0.738***
	(0.209)	(0.212)	(0.214)	(0.215)	(0.215)	(0.215)
DT1011	-0.793***	-0.820***	-0.785***	-0.770***	-0.762***	-0.763***
	(0.279)	(0.279)	(0.276)	(0.275)	(0.275)	(0.275)
DT1020	-0.234	-0.261	-0.278	-0.279	-0.276	-0.275
	(0.240)	(0.237)	(0.238)	(0.238)	(0.239)	(0.239)
DT1071	0.129*	0.098	0.102	0.110	0.118*	0.118*
	(0.066)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)
DT1128	-0.178	-0.203	-0.202	-0.202	-0.205	-0.205
	(0.134)	(0.134)	(0.133)	(0.133)	(0.132)	(0.132)
DT1222	0.114	0.116	0.130	0.135	0.138	0.137
	(0.088)	(0.089)	(0.090)	(0.090)	(0.090)	(0.090)
DT1227	0.254	0.275	0.315	0.318	0.306	0.303
	(0.310)	(0.311)	(0.310)	(0.309)	(0.309)	(0.310)
DT1230	-0.040	-0.032	-0.048	-0.054	-0.061	-0.063
	(0.209)	(0.211)	(0.211)	(0.211)	(0.210)	(0.210)
DT1245	-0.406**	-0.405**	-0.403*	-0.402*	-0.403*	-0.404*
	(0.204)	(0.205)	(0.206)	(0.207)	(0.208)	(0.208)
DT1253	-0.223*	-0.217	-0.196	-0.189	-0.183	-0.184
	(0.133)	(0.134)	(0.134)	(0.134)	(0.135)	(0.135)
DT1303	0.101	0.065	0.047	0.046	0.049	0.049
	(0.251)	(0.255)	(0.255)	(0.254)	(0.252)	(0.251)

DT1342	-0.310	-0.339*	-0.366*	-0.371*	-0.372*	-0.371*
	(0.194)	(0.197)	(0.199)	(0.199)	(0.199)	(0.199)
DT10006	-1.059***	-1.025***	-1.020***	-1.019***	-1.017***	-1.017***
	(0.260)	(0.266)	(0.269)	(0.269)	(0.269)	(0.269)
DT10026	-0.153	-0.126	-0.062	-0.048	-0.054	-0.061
	(0.332)	(0.330)	(0.324)	(0.323)	(0.325)	(0.326)
DT10027	0.010	0.063	0.080	0.079	0.068	0.063
	(0.291)	(0.288)	(0.285)	(0.285)	(0.285)	(0.285)
DT10028	-0.735***	-0.690***	-0.665**	-0.669**	-0.689***	-0.694***
	(0.268)	(0.267)	(0.264)	(0.263)	(0.261)	(0.262)
DT10030	-1.082***	-1.099***	-1.114***	-1.117***	-1.121***	-1.123***
	(0.276)	(0.277)	(0.278)	(0.279)	(0.279)	(0.279)
DT10031	-0.631**	-0.671**	-0.701**	-0.704**	-0.697**	-0.693**
	(0.291)	(0.300)	(0.303)	(0.303)	(0.302)	(0.301)
DT10035	-0.411*	-0.416*	-0.409*	-0.407*	-0.410*	-0.413*
	(0.242)	(0.242)	(0.243)	(0.243)	(0.243)	(0.243)
DT10037	-0.462***	-0.467***	-0.463***	-0.463***	-0.467***	-0.468***
	(0.170)	(0.172)	(0.173)	(0.173)	(0.173)	(0.172)
DT10038	-1.266***	-1.301***	-1.287***	-1.277***	-1.265***	-1.266***
	(0.231)	(0.233)	(0.232)	(0.231)	(0.230)	(0.230)
DT10040	-0.452***	-0.445***	-0.458***	-0.462***	-0.465***	-0.465***
	(0.164)	(0.165)	(0.166)	(0.166)	(0.167)	(0.167)
DT10041	-0.597***	-0.609***	-0.611***	-0.611***	-0.612***	-0.613***
	(0.228)	(0.229)	(0.229)	(0.229)	(0.229)	(0.229)
DT10042	-0.694***	-0.681***	-0.668***	-0.666***	-0.665***	-0.664***
	(0.190)	(0.193)	(0.193)	(0.192)	(0.191)	(0.191)
DT10043	-0.178	-0.194	-0.205	-0.206	-0.210	-0.212
	(0.225)	(0.226)	(0.226)	(0.226)	(0.226)	(0.226)
DT10046	0.084	0.060	0.064	0.068	0.072	0.071
	(0.327)	(0.328)	(0.327)	(0.326)	(0.325)	(0.324)
DT10048	0.042	0.046	0.042	0.039	0.032	0.030
	(0.233)	(0.234)	(0.234)	(0.234)	(0.234)	(0.234)
DT10049	-0.191	-0.206	-0.215	-0.217	-0.221	-0.222
	(0.167)	(0.168)	(0.168)	(0.167)	(0.167)	(0.167)
DT10050	-0.375	-0.386	-0.390*	-0.387	-0.379	-0.377
	(0.231)	(0.235)	(0.236)	(0.236)	(0.236)	(0.236)
Constant	0.091***	-0.320***	-0.638***	-0.754***	-0.929***	-0.981***
	(0.031)	(0.045)	(0.065)	(0.057)	(0.067)	(0.073)

Observations	33,258	33,258	33,258	33,258	33,258	33,258

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1