Essays in Labor and Health Economics

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Neha Bairoliya

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

José-Víctor Ríos-Rull, Adviser

July 2015

Acknowledgments

I want to take this opportunity to express my gratitude to all those who have helped me and inspired me during my graduate studies. First and foremost, I would like to thank my adviser Prof. José-Víctor Ríos-Rull for his invaluable guidance and help. I greatly value his patience and understanding as well as the freedom that he gave me in shaping my research work. It has been an enjoyable journey and at the same time, it has been a great learning experience.

I would like to thank my committee members Prof. Amil Petrin, Dr. Jonathan Heathcote and Prof. Pinar Karaca-Mandic for their feedback and suggestions. I would like extend special thanks to Prof. Pinar Karaca-Mandic for helping me procure the Restricted Health and Retirement Survey Data. I am grateful to Prof. Eric French (University College London) for his help and his many suggestions during the initial phase of this project and for sharing his data and computer code with me. I would also like to thank Prof. Naoki Aizwa, Prof. Mariacristina De Nardi (University College London), Dr. Melissa Tartari (University of Chicago) and participants of the University of Minnesota Labor Workshop for their helpful comments and suggestions.

I would like to thank my parents for their love and encouragement. Finally, I would like to thank my husband, Amartya, for his love and encouragement, for providing a warm home and for always being there through the hard times. This thesis would not have been possible without his untiring support and his immense help with computational issues.

Dedicated to my parents and my husband, Amartya.

Abstract

This dissertation consists of two essays. The first essay studies the effect of different kinds of pension plans on the labor market decisions of the older workers. Due to the aging population, Social Security's projected annual cost is expected to increase to about 6.2 percent of the Gross Domestic Product by 2035, thus posing significant challenges to the U.S. policy makers. This has fueled an interest in research geared towards understanding the determinants of retirement. Past research has shown that pensions have a significant effect on retirement decisions. But the pension landscape in the U.S. has changed dramatically in the last few decades. From being once dominated by the traditional annuity-based Defined Benefit (DB) plans, the trend has now moved towards account-based Defined Contribution (DC) plans. This change has been accompanied by a reversal in the participation trend of older workers resulting in an increasing labor force participation of the elderly in the United States over the last thirty years. This essay investigates the link between the two by building a life cycle model of retirement and pension plan types. By conducting counterfactual experiment which changes all DB plans to DC plans, I hope to understand the role played by the differences in the nature of pension wealth accumulation under different pension plans in explaining the differences in retirement behavior observed across different pension plan holders.

The second essay explores policy questions pertinent to the aging population in the healthcare field. Medicare Part D is a government program introduced in 2006 to offer outpatient drug benefits to Medicare beneficiaries. A lot of the brand-name drugs covered under Medicare Part D are also available in generic versions and it has been argued by policy makers that a higher level of utilization of these generic drugs would result in significant cost savings for the government. However, the cost savings of forcing consumers onto generics may lead to large welfare losses for consumers of non-generic alternatives if they highly value them. This issue is addressed in this essay where a structural model of drug demand that allows for heterogeneity in match quality between consumers and drugs and also allows for consumer learning about the stochastic match quality of the drug is estimated. The 2007-2008 administrative claims data for the 5% Medicare Sample is used for demand estimation. By conducting counterfactual experiment which eliminates branded drugs for which the generic is available from the choice-sets of consumers, I hope to understand the effect of generic substitution on consumer welfare and the resulting cost savings for the government.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It is a well known fact now that the United States is experiencing an aging population. According to the U.S Census Bureau estimates, the country will experience considerable growth in it's population between 2012 and 2050. The population aged 65 and over is expected to almost double (increase from 43.1 million in 2012 to 83.7 million in 2050) during this time period. The aging population will pose serious challenges to policy makers and federal entitlement programs such as Social Security and Medicare which become available at age 65. This dissertation explores policy scenarios for alleviating the fiscal burden faced by these two federal programs by simulating quantitative models of retirement and drug demand. It is structured as follows:

- Chapter 2 presents the essay "Effect of Pension Plan Type on Retirement Behavior". This essay studies the effect of different kinds of pension plans on the labor market decisions of the older workers. Due to the aging population, Social Security's projected annual cost is expected to increase to about 6.2 percent of the Gross Domestic Product by 2035, thus posing significant challenges to the U.S. policy makers. This has fueled an interest in research geared towards understanding the determinants of retirement. Past research has shown that pensions have a significant effect on retirement decisions. But the pension landscape in the U.S. has changed dramatically in the last few decades. From being once dominated by the traditional annuity-based Defined Benefit (DB) plans, the trend has now moved towards account-based Defined Contribution (DC) plans. This change has been accompanied by a reversal in the participation trend of older workers resulting in an increasing labor force participation of the elderly in the United States over the last thirty years. This essay investigates the link between the two by building a life cycle model of retirement and pension plan types. By conducting counterfactual experiment which changes all DB plans to DC plans, I hope to understand the role played by the differences in the nature of pension wealth accumulation under different pension plans in explaining the differences in retirement behavior observed across different pension plan holders.

- Chapter 3 presents the essay "Consumer Learning, Product Differentiation, and The Value of Generic Pharmaceutical Entry" [1]. The research outlined in this essay explores policy questions pertinent to the aging population in the healthcare field. Medicare Part D is a government program introduced in 2006 to offer outpatient drug benefits to Medicare beneficiaries. A lot of the brand-name drugs covered under Medicare Part D are also available in generic versions and it has been argued by policy makers that a higher level of utilization of these generic drugs would result in significant cost savings for the government. However, the cost savings of forcing consumers onto generics may lead to large welfare losses for consumers of non-generic alternatives if they highly value them. This issue is addressed in this essay where a structural model of drug demand that allows for heterogeneity in match quality between consumers and drugs and also allows for consumer learning about the stochastic match quality of the drug is estimated. By conducting counterfactual experiment which eliminates branded drugs for which the generic is available from the choice-sets of consumers, I am hoping to understand the effect of generic substitution on consumer welfare and the resulting cost savings for the government..

[1] The research outlined in this chapter was conducted in collaboration with Prof. Amil Petrin, Prof. Pinar Karaca-Mandic and Prof. Jeffery McCullough. For this research, we use a 5% sample of Medicare claims requested from the Center of Medicare and Medicaid services (CMS). We have the necessary Data Use Agreement (DUA) for publishing the results of this research in our upcoming paper of the same title. However, CMS requires a separate Data Use Agreement for publishing any results obtained using the data as part of a dissertation. Due to these constraints, this essay only contains a brief overview of the project. The author would like to direct the readers to the forthcoming paper "Consumer Learning, Product Differentiation, and The Value of Generic Pharmaceutical Entry" for results and more details.

Chapter 2

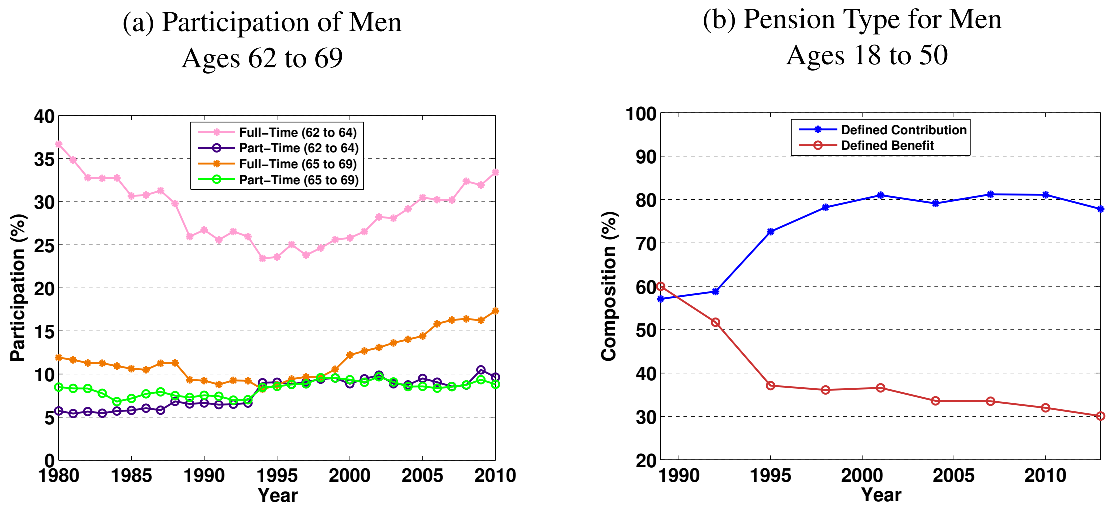
Effect of Pension Plan Type on Retirement Behavior

2.1 Introduction

An understanding of the determinants of retirement behavior is important both from the point of individual well-being as well as policy making. With nearly 76 million people in the baby boom population (those born between 1946 and 1964) beginning to retire in the U.S., Social Security's projected annual cost is expected to increase to about 6.2 percent of the Gross Domestic Product[1] by 2035, thus posing significant challenges to the U.S. policy makers [[De Nardi et al., 1999], [Galasso, 2008], [Bohn, 1999]]. This has fueled an interest in research geared towards understanding the determinants of retirement. Past research has shown that pension wealth is crucial in governing retirement decisions [[Stock and Wise, 1988], [Kotlikoff and Wise, 1987], [Kotlikoff and Wise, 1989], [Samwick, 1998], [Chan and Stevens, 2004]]. But the pension landscape in the U.S. has undergone a major overhaul in the last few decades, changing the very nature of these retirement plans. From being once dominated by the traditional annuity-based Defined Benefit (DB) plans, the trend has now moved towards account-based Defined Contribution (DC) plans. This change has been accompanied by a reversal in the participation trend of older workers resulting in an increasing labor force participation of the elderly in the United States over the last thirty years. The juxtaposition of these two trends suggest a potential link between the two. This has made it important to re-visit the role of pensions on retirement. More specifically, it sparks an interest in research trying to understand if the change in the pension plan

[1]Figure taken from 2013 Social Security Annual Report (Social Security and Medicare Boards of Trustees).

Figure 2.1: Changing Participation and Pension Plan Composition

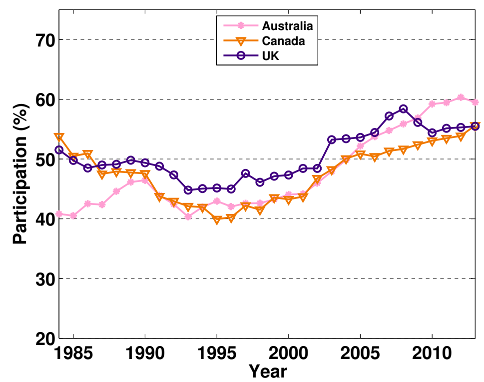


composition has been responsible for the recent increase in the labor market participation of the older workers.

The long trend of declining participation of older workers reversed in the late 1980's and the participation of the age group 62 to 69 increased by over 15%. The pension plan composition also changed during the same time. This resulted in a more than 20% increase in Defined Contribution (DC) pension plans and an even sharper decline in Defined Benefit (DB) pension plans (Figure 2.1a and 2.1b). More importantly these changes have not been confined to the United States alone. Broadbent et al. [2006] finds empirical evidence that indicates a shift from the traditional Defined Benefit to Defined Contribution plans for several other OECD countries like Canada, Australia and U.K. The striking observation is that these countries also underwent a reversal in the trend of declining labor force participation of the older workers around the same time as the change in pension landscape (Figure 2.2). It has been well documented that Defined Benefit pension holders exit the labor market two to three years earlier than people on Defined Contribution plans. I find three important margins of variation in the retirement behavior across different pension plan holders (DB, DC and those without any pension plans). These differences in retirement behavior are potentially tied to the nature of pension wealth accumulation under these plans.

This essay provides a quantitative analysis of the effect of different pension plans on retirement behavior. A life cycle model of retirement, savings and heterogeneity in pension wealth accrual is estimated using data from the Health and Retirement Survey. The key details of the two pension plans – DB and DC are built into a dynamic programming model

Figure 2.2: Participation of Men
Ages 62 to 65



of retirement which generates the differences in wealth accumulation patterns. This in turn generates the differences in retirement behavior in the model as observed in the data. Hence the model developed in this here produces rich variations in employment patterns across different pension groups without resorting to any kind of unobserved heterogeneity in preferences. Such a model has two important contributions. First, it can be used to understand the role of the recent pension plan phase-out on the increase in labor force participation of the elderly. Second, the variation in employment patterns across different pension plan types can be exploited to get a better understanding of retirement behavior by obtaining sharper estimates of some key preference parameters.

The existing structural models of retirement can be broadly classified into three groups with respect to modeling pension wealth. The first is a class of models having pension wealth accrual through a single kind of pension plan which combines features of both Defined Benefit and Defined Contribution [French, 2005], [French and Jones, 2011], [Blau and Gilleskie, 2008]. The models in these papers cannot generate the different retirement patterns across different pension plan types and hence, are not suited for understanding the effect of changing pension plan composition on retirement behavior. For instance, the pension benefits in [Blau and Gilleskie, 2008] depends on an individual's age, experience and employment status which closely resembles the Defined Benefit plan structure. This benefit formula is not a good approximation for wealth under a Defined Contribution plan for two reasons. First, DB benefits are distributed as an annuity whereas DC wealth is distributed as a onetime lump some transfer to a non-pension account. Second, this formula is unable capture the most important difference between the two plans – DB plans max out after a certain age incentivizing exit from the labor market whereas the returns to staying with the provider for an additional year stays the same for DC plan holders. This is key to explaining the difference in retirement behavior observed for people on these two different pension plans. For this reason, the pension benefits in this work for DB and DC plans are calculated using two different mechanisms. DB benefits are approximated on the same lines as the above mentioned papers but DC wealth is modeled as a separate account based wealth with fixed contributions from employer and employee every period.

The second group of papers abstract away from modeling pension wealth accrual [Casanova, 2010], [Van der Klaauw and Wolpin, 2008], [Rust and Phelan, 1997]. The models in these papers miss a key source of variation in retirement behavior. Finally [Blau, 2011] models pension wealth accrual through both Defined Benefit and Defined Contribution plans for understanding the effect of pension wealth in crowding out private savings. This paper abstract away from some important ingredients pinning down retirement behavior.

The model developed in here mainly builds on [French, 2005] and [Blau, 2011]. It extends the former by introducing three new features –1) pension wealth accumulation through DC plans 2) allowing for people to have no pension wealth and 3) allowing for job switches from pension providing to non-pension providing jobs along with an exogenous job destruction resulting in a choice between new job and non-participation. The first two extensions allow the dynamic dynamic programming model to generate the difference in participation and hours (as observed in data) across these pension groups. The third extension has varying degrees of importance for the three pension groups in the model. It allows the DB agents in the model to reduce hours (in the event of bad health) by switching to non-pension employer. Hence, this mechanism helps them to mitigate the loss in pension wealth in the event of reducing hours. In the absence of this mechanism the model generates sharper drop in participation for this group. For instance, just before age 62 (the maturity of DB pension in the model), a bad health shock results in labor market exit for DB pension holders. This is due to the fact that reduction in hours result in pension loss for this group. This mechanism is relatively less important for DC pension group and does not affect the behavior of the people with no pensions (no one in this group chooses to switch employers endogenously). Agents with DC pension plans in the model choose to switch employers mainly between ages 60 to 65 to get access to their DC account (before age 65, DC account in the model is accessible only upon switching employers). This is observed in the event of a bad health shock reducing hours worked and hence earnings (both through hours worked and effect of health on wages) in that period.

This work builds on the latter work by introducing – 1) intensive margin and 2) health shocks affecting the total amount leisure available in a given period. Introducing the choice of hours worked in a retirement model generates a new margin of adjustment for the agents. Since hours worked before retirement are observed to be different for the three pension groups, this can generate a new source of variation in the model which can be exploited to get sharper estimates of participation cost parameter and labor supply elasticities. Health is an important determinant of retirement [Dwyer and Mitchell, 1999], [Sickles and Taubman, 1984]. The health shocks in the model are crucial for generating reduction in hours worked and non-participation.

The primary goal of the model developed here is to understand the effect of the recent shift from DB to DC pension plans in the increase in participation of the older workers. The model uses two parameters for fitting the participation rates across the pension groups (others fixed from French [2005]). The model can generate the systematic difference in participation rates across the two pension groups at older ages . The counterfactual experiment involves changing all DB plans to DC plans in the model (keeping all other initial conditions like pension wealth the same). This results in a 17% increase in participation for age group 57 to 69. So according to this model roughly 43% of the recent increase in participation can be attributed to the change in the pension plan composition.

The dynamic programming model developed in this essay is estimated using Method of Simulated Moments. One novel feature about the model is it's ability to use the variation in the retirement behavior across pension pension plans to get sharper estimates for the participation cost. The estimate of participation cost is 35% lower than other estimates obtained using a model with pension wealth accumulation through a single DB like plan.

2.2 Empirical Facts

Using the bi-annual panel data from the Health and Retirement Survey (1992-2010), I find three important sources of variation in the retirement behavior of older workers across the three pension groups - DB, DC and those with no pension plans. I first describe the structure of the two pension plans in the U.S. to demonstrate how this could be responsible for generating the differences in retirement behavior across different plan holders. I then document the main sources of variation in retirement behavior across the three pension groups as observed in the data. Finally, I lay out potential reasons for the change in pension landscape in the U.S..

2.2.1 Defined Benefit vs. Defined Contribution Plans

The two types of pension plans – Defined Benefit and Defined Contribution – have different wealth accrual patterns over the life cycle of an individual. The benefits in a defined benefit plan are based on tenure and earnings in the final years of service. As a result, pension wealth in a DB plan accrues non-linearly with age, the benefit increase being the greatest from working the year just before the eligibility for *early retirement* and declines sharply after the eligibility for *normal retirement*[2] benefits. This nature of wealth accrual in a DB plan has two important implications for labor market outcomes of the elderly - first, the present discounted value of pension benefits accruing from a DB plan decreases by staying with the employer longer than a certain age, providing a strong incentive to exit the labor market right after reaching the full potential of the pension plan (normal retirement age). Secondly it is expensive to cut work hours close to years before retirement as it affects the entire stream of benefits to be received after retirement. Pension wealth in a DC plan, on the other hand, is the market value of the current assets accumulated in a portable account, resulting in an age independent profile of pension wealth accrual. An additional year of work increases pension wealth by the same amount at any point in the life cycle and a reduction in hours reduces pension wealth only in the year in which reduced hours are observed.

In order to understand the age-specific work disincentive provided by DB plans, let's consider a simple example. Let the DB plan pay 1% of final year salary (before quitting) times the tenure at the firm. This is capped at tenure and wages at the normal retirement age (say 65 in this case) , i.e. if the worker works beyond age 65 at the same firm, then not only does he loose the benefits of that additional year but also his higher wage (on account of the inflation rate/wage growth and extra 1 year of tenure) does not count towards his benefits. For the DC plan, let's assume a preset contribution rate (fixed fraction of each year's salary) by the employer to a portable bank account which earns a real interest rate of 3%.

Table 2.1 gives a comparison of the nature and timing of wealth accrual under the two plans. The pension wealth under a DB plan accrues fast with age, reaching it's full-potential at age 65 (the normal retirement age of the plan). An additional year of service after age 65 results in roughly 1% loss in the pension wealth and continuing to work further leads to a significant 21% reduction in benefits. This happens because working after age 65 does not result in any further increase in the pension benefits but 1 year worth of benefits are lost as in-service distributions are not allowed. Switching to part-time work in the final year

[2] 62 and 65 are the most common early and normal retirement ages respectively

Table 2.1: Present Value of Accrued Benefits and Marginal Change in Benefits for Hypothetical Worker with a Defined Benefit or a Defined Contribution Plan

	DB					DC				
	Present Value of Accrued Benefits (Constant $)			Marginal Change in Present Value of Accrued Benefits from an Additional Year's Work (Constant $)		Present Value of Accrued Benefits (Constant $)			Marginal Change in Present Value of Accrued Benefits from an Additional Year's Work (Constant $)	
Current Age	Full-time	Part-time	% lost	Full-time	% of salary	Full-time	Part-time	% lost	Full-time	% of salary
30	$144	$72	50	$20	0.26	$8,430	$7,659	9.1	$1,274	8.5
35	$466	$233	50	$42	0.57	$14,124	$13,352	5.4	$1,053	7.02
45	$2,418	$1,209	50	$164	2.2	$22,719	$21,948	3.3	$719	4.8
55	$9,408	$4,704	50	$570	7.6	$28,019	$27,820	2.7	$491	3.28
64	$28,841	$14,420	50	$1,647	21.96	$32,267	$31,496	2.3	$348	2.33
65	**$32,538**	**$16,269**	50	**$ 1,848**	**24.65**	**$32,603**	**$31,832**	2.3	**$335**	**2.24**
66	$32,398	$16,199	50	$-70	-0.94	$32,926	$32,155	2.3	$323	2.16
67	$29,196	$14,598	50	$-1,600	-21.35	$33,238	$32,466	2.3	$311	2.08

The present discounted value of pension benefits at age 25 are computed if the worker were to quit (working full-time or part-time in the last year) at the age in the first column.
Assume worker is paid $7.5 hourly wage with no real wage growth.
Full-time is 2000 annual hours of work and part-time is 1000 hours of work
The normal retirement age of DB plan is 65
Contribution rate for DC plan is set to 14% to match the maximum wealth accruable under DB and DC plan
Inflation rate is 7% and real interest rate is 3%
Source: Columns 2,4 and 5 Adapted from Ellwood [1985]

of service with the pension providing firm results in a flat 50% reduction of DB pension benefits. This is very different from the wealth accrual under DC plans. After an initial increase of 8%, the marginal change in the present value of pension in DC plans remains constant at approximately 2% every year. Unlike DB plans, the wealth accrual under DC plans does not max out or become negative after a certain age.

2.2.2 Retirement Behavior

I now present empirical evidence of differences in retirement behavior across different pension plan holders.

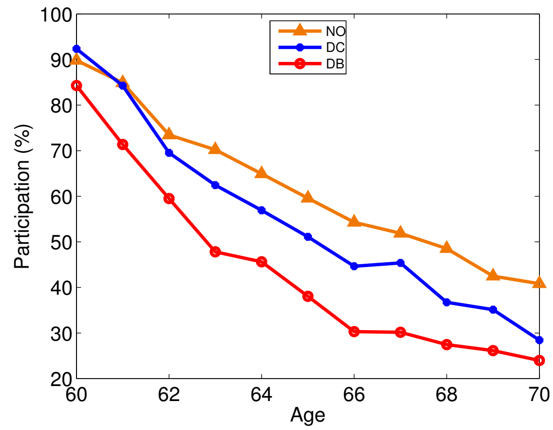
Extensive Margin

I observe that people on Defined Benefit plans exit the labor market on an average 3.1[3] years earlier than people on Defined Contribution plans. Figure (2.3) shows that participation rates decline over the life cycle for all three pension groups but the participation for DC and no pension groups remain systematically higher than DB pension holders.

Intensive Margin

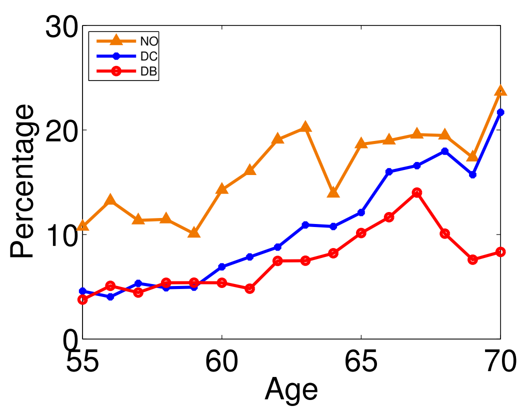
[3]Retirement is observed to be an absorbing state but this statistic takes into account the first observed exit from the labor market

Figure 2.3: Participation of Men
By Pension Plan Type

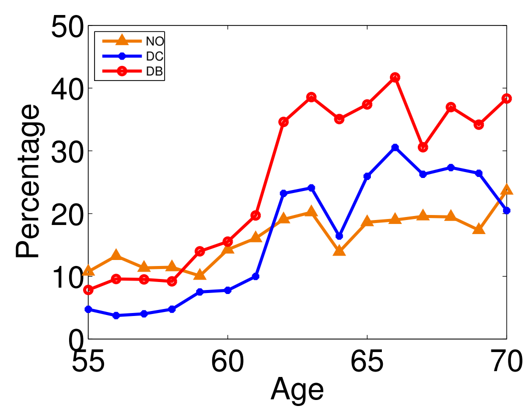


Along the intensive margin, I observe that people on Defined Benefit plans work on an average 2100 annual hours before switching to zero hours. People on Defined Contribution plans work on an average 1758 annual hours before switching to zero hours and finally people with no pension plans only work 1300 hours before retiring. Figure (2.4) indicates that DB pension holder are more likely to move into retirement directly from full-time jobs and DC pension holders switch into retirement more gradually by taking some part-time job.

Figure 2.4: Labor Market Transitions for Men
by Pension Plan Type



Full-Time to Part-Time



Full-Time to Retirement

Job Switches Before Retirement

Finally I observe that conditional on reducing hours (switching from full-time to part-time

Table 2.2: Nature of Full Time to Part-Time Switch
by Pension type (55-65 years)

Pen. Type	Same (%)	Diff. (%)	Mean Age of Switch	
			Same	Diff.
DB	36.43	63.57	60.25	59.68
DC	58.29	41.71	60.40	59.80
No	69.15	30.85	60.48	59.78

work), DB pension holders are more likely to switch jobs and people on DC pension plans are more likely to reduce hours with the same employer (pension provider) as shown in Table (2.2). The third row indicates that switching jobs is not preferable at older ages since people without any pension plans are most likely to stay with the same employers.

2.2.3 Recent Changes

A series of structural, regulatory and demographic changes have taken place in the United states in the last three decades that are responsible for the shift from DB to DC pension plans.

First several government regulations were enacted starting early 1980's which made DC plans more attractive over DB plans for both for the employer and employee (refer to Rajnes [2002], Iams et al. [2009], Hustead [1998], Gebhardtsbauer and Turner [2004], Gustman and Steinmeier [1992]). For instance, the 1978 Revenue Act implemented a provision which allowed employees to make voluntary contributions to employer-sponsored retirement plans with pretax dollars. Prior to this act, employer contribution to DC accounts were tax-exempt but employee's contributions were not tax deductible. Later, Tax Equity and Fiscal Responsibility Act of 1982 and the Tax Reform Act of 1986, reduced incentives for employers to maintain their DB plans.

The increase in life expectancy and the aging population are other reasons for the decline in the Defined Benefit plans. Since DB plans are annuitized with a fixed stream of payments starting upon retirement, an increase in life expectancy directly increases the cost of these plans relative to DC plans. According to the 2003 report of the Pension Benefit Guarantee Corporation, years in retirement for an average U.S. male worker increased from 11.5 years in 1950 to 18.1 years in 2003. Hence the seven additional years of funding increased the cost of DB plans by a significant amount.

These regulatory and demographic changes were accompanied by structural changes in the economy which resulted in occupational shift away from manufacturing towards service and technology sectors. These newly growing sectors of the economy adopted DC plans due its lesser regulatory burden. Further, DB plans have long vesting periods and best suited for jobs with long tenure with the pension provider. The structural changes in the economy increased job mobility and led to a decline in job tenure. These made Defined Contribution plans more popular due to their shorter vesting periods and portability across jobs.

2.3 Theoretical Model

This section presents a stochastic dynamic programming model of retirement, social security and private pensions. In order to capture the true nature of retirement incentives for older workers, retirement benefits from private pension programs and social security are modeled in great detail to match that of the current U.S system.

Labor supply, savings and social security benefit application decisions of a male household head is modeled close to the years before retirement. Individuals make these decisions in every time period t and adjust their behavior in response to uncertainty pertaining to wages, health, survival, employment (job loss) and rate of return on assets.

At every time period t, $t = 55, 56, ..., 95$, given an initial stock of assets, pension wealth, social security wealth and wages, households choose optimal consumption, labor supply and benefit application (if possible at that age) to maximize the present discounted value of life-time utility. The dynamic programming model has various components. The following sections describe each key ingredient in detail.

2.3.1 Preferences

Agents in period t derive utility from consumption c_t and leisure l_t where the amount of leisure consumed is dictated by the choice of hours worked h_t and the decision to work with a new or an old employer d_t. The within period utility is non-separable[4] between

[4]This work follows [French and Jones, 2011], [French, 2005], [Casanova, 2010] and others in addressing the "Retirement-Consumption puzzle". A decline in consumption at retirement is caused by both- 1) unexpected health shocks to leisure causing unplanned retirement and 2) non-separability of preferences between consumption and leisure.

consumption and leisure and is given by:

$$U(c_t, h_t, d_t) = \frac{1}{1-\rho}(c_t^{\nu} l_t^{1-\nu})^{(1-\rho)}$$

Where ρ is the coefficient of relative risk aversion and ν is the weight on consumption. The total amount of leisure in period t is given by:

$$l_t = \bar{l} - h_t - \chi I\{d_t = 1\} - \phi_P I\{h_t > 0\} - \phi_H I\{m_t = \text{bad}\} \tag{2.1}$$

Where $\bar{l}$ is the total endowment of leisure each period, h_t is hours worked, χ is the psychic cost of switching employer, ϕ_H is the amount of leisure lost due to a bad health shock and ϕ_P is participation cost incurred if hours worked h_t are positive. Upon dying an individual values bequests of any leftover assets a_t according to the utility function developed by De Nardi [2004]

$$b(a_t) = \frac{\theta_{beq}}{1-\rho}(a_t + \kappa_{beq})^{(1-\rho)\nu}$$

The coefficient θ_{beq} measures the strength of bequest motive and κ_{beq} measures the curvature of bequest function. Increase in θ_{beq} increases the marginal utility of a unit of bequest and increase in κ_{beq} indicate that the bequest is valued more like a luxury good.

2.3.2 Health and Mortality

Every period individuals are subject to an exogenous health shock which can take two values $m_t \in \{good, bad\}$. Bad health affects individuals in multiple ways – it lowers the survival probability for the next period, lowers the wages and affects the amount of leisure consumed. The transition probability for health depends on current health status and age in the next period. A typical element in the two health transition matrix is given by:

$$\pi^m_{good,bad,t+1} = prob(m_{t+1} = good | m_t = bad, t+1)$$

Individuals are also subject to mortality shocks in each period. The survival probability for the next period depends on age next period and current health status:

$$\pi^s_{t+1} = prob(s_{t+1} = 1 | m_t)$$

2.3.3 Wages

The logarithm of wages in every time period is a function of health and age specific profile $\omega(m_t, age_t)$ and an autoregressive component η_t.

$$\begin{aligned} \log w_t &= \omega(m_t, t) + \eta_t \\ \eta_t &= \rho_w \eta_{t-1} + \epsilon_t^w \\ \epsilon_t^w &\sim N(0, \sigma^2_{\epsilon^w}) \end{aligned} \tag{2.2}$$

Changing employers does not affect wages in this version of the model.

2.3.4 Budget-Constraint

An individual's income consist of various components. He receives income through hours worked in the labor market $w_t h_t$, spousal income y_{st}, interest on assets $\bar{r}a_t$, pension benefits pb_t^{DB} from Defined Benefit plan, social security benefits ss_t (if applied for it) and government transfers tr_t if eligible.

Let $y(., \tau)$ be the level of post-tax income,then the asset accumulation equation is given by:

$$a_{t+1} = \begin{cases} a_t + y(w_t h_t, y_{st}, \bar{r}a_t, pb_t^{DB}, \tau) + b_t \times ssb_t + tr_t - c_t & \text{if } pen = DB \\ a_t + y((1 - cr_w) w_t h_t, y_{st}, \bar{r}a_t, \tau) + b_t \times ssb_t + tr_t - c_t & \text{if } pen = DC \\ a_t + y(w_t h_t, y_{st}, \bar{r}a_t, \tau) + b_t \times ssb_t + tr_t - c_t & \text{if } pen = NO \end{cases} \tag{BC}$$

Where cr_w is the contribution made by the worker to the DC account [5]. While the rate of return on assets is a risk-free one, there is a stochastic rate of return r_t on the balances in a DC account given by a mean reverting stochastic process [Blau, 2011].

$$1 + r_t = (1 + \bar{r}) \exp\{\psi_t\}$$

$\bar{r}$ is the mean rate of return and $\psi_t \sim N(0, \sigma^2_\psi)$ is an iid (over time and across individuals) normal shock. The stochastic rate of return on DC balances captures the key difference in uncertainty between the two types of pension plans. The rate of return heterogeneity captures the heterogeneity in portfolio allocation choice which is not modeled here. There is a borrowing constraint on non-pension assets given by:

$$a_{t+1} \geq 0 \; \forall t \tag{LC}$$

[5] Employee contribution is only subject to federal payroll taxes.

and a consumption floor which captures the income and asset tested government programs like Supplemental Security Income (SSI) and Medicaid that guarantees a minimum level of consumption [Hubbard et al., 1995].

$$c_t \geq \bar{c} \tag{CF}$$

Government transfers tr_t bridge the gap between this minimum level of consumption and individual's liquid resources.

$$tr_t = \min\{0, \bar{c} - (a_t + y_t + ss_t)\} \tag{2.3}$$

2.3.5 Job-Loss

Every period a working individual faces an exogenous probability π^λ of being laid off from the current job. Naturally retired individuals do not face any such uncertainty pertaining to job loss. These job shocks are realized at the end of the period. So at the beginning of any period, an individual who was working in the previous period can be laid off or not. Individuals who were not working in the previous period remain retired at the beginning of the next period. So depending on the exogenous job destruction shock, an individual has following employment choices to make every period:

$$d_t = \begin{cases} \{0,1\} & \text{working in } t-1 \ \& \ \lambda_t = 0 \\ \{1\} & \text{working in } t-1 \ \& \ \lambda_t = 1 \\ \{1\} & \text{not working in } t-1 \end{cases}$$

The job loss probability is a function of health status and age. It is given by:

$$\pi^\lambda_{t+1} = \begin{cases} prob(\lambda = 1 | m_t, t+1) & \text{if} \quad h_t > 0 \\ 0 & \text{if} \quad h_t = 0 \end{cases}$$

2.3.6 Pensions

Like social security, private pensions provide important retirement incentives. Hence I model the private pension program in the U.S. in detail. Even though there is a wide variation in the employer provided pension design, they can be broadly classified into two types – Defined Benefit (DB) and Defined Contribution plans (DC) [6].

[6] There are hybrid plans like such as cash balance plans, and money purchase pension plans, which are defined contribution plans with a predefined contribution formula.

Defined Benefit (DB)

DB plans pay a sequence of benefits computed using a predefined formula commencing upon normal retirement age of the plan until death. Even though there is a wide heterogeneity across firms in the formula determining the benefits under a DB plan, the formula typically depends on age at exit, years of service and the average of the five highest earnings (last 5 years of service) at the firm.

The model captures some of these important features of the DB pension plan. The DB pension plans are illiquid until age 62 [7]. In service distributions are not allowed under DB plans i.e. a worker has to quit the job with the pension provider (retire or work with a new employer) to start drawing benefits from the pension plan [8]. The pension benefits are based on tenure, age at exit and Average Indexed Monthly Earnings (AIME) like social security. The restricted Health and Retirement Survey data has pension plan information from the employers of HRS respondents who gave permission. The data comes with a pension calculation software. I use pension software to compute the DB benefits for the respondents in my sample. I then run a regression with pension benefits as the explanatory variable and tenure, and age dummies to back out the dependence of pension benefits on tenure for an average pension plan. I use the estimates from Gustman and Steinmeier [1999] for the AIME related coefficients. Pension benefits are computed as:

$$\begin{aligned} pb_t^{DB} = \alpha_0 ten_t + \alpha_1 ssb_t + (\alpha_{2,0} + \alpha_{2,1} t + \alpha_{2,2} t^2) \cdot \max\{0, ssb_t - \kappa_1\} \\ + (\alpha_{3,0} + \alpha_{3,1} t + \alpha_{3,2} t^2) \cdot \max\{0, ssb_t - \kappa_2\} \end{aligned}$$

Defined Contribution (DC)

Pension wealth under a DC plan is characterized by an account balance with employer and worker contribution rates. As long as the individual works for the DC pension provider, both the employer and the employee contribute a fixed fraction of the employee's pre-tax labor earnings to the account. The stochastic rate of return on assets in this account captures

[7] Gustman et al. [2010] reports that the average normal retirement age for the HRS cohort using the 1992 employer data is 61.8

[8] Under the Internal Revenue Code and ERISA, a defined benefit plan could only permit a distribution of benefits at termination of employment, retirement, termination of the plan or total and permanent disability of the participant. Later in 2006 some of these restrictions were relaxed. The Pension Protection Act of 2006 (PPA) and it's finalization in 2007 (the "Final PPA Regulations") provided rules permitting distributions from DB plans upon normal retirement age and after age 62. But since most of the individuals in my estimation sample are older than 70 years in 2006, these laws do not affect them

the risk that the worker bears as opposed to benefits in a DB plans which has no risk. Even though in practice, individuals can choose to claim benefits and roll over the funds into a tax-sheltered Individual Retirement Account (IRA) (or transfer the money to the new employer's plan or take a lump-sum payment at a penalty), I do not model any claiming decisions. I also assume for simplicity that in-service distributions are not allowed. That is as long as the worker works for the pension provider, the wealth is illiquid[9] and there are fixed contributions to the account every period. Once the worker leaves the pension provider, then the funds are transferred as a lump-some to the non-pension account of the worker and there is no uncertainty pertaining to the rate of return.

$$q_{t+1}^{dc} = \begin{cases} [q_t^{dc} + (cr_w + cr_e) w_t h_t](1 + r_{t+1}) & \text{if with DC provider} \\ q_t^{dc}(1 + r_{t+1}) & \text{if } \; age_t \leq 59 \;\&\; \text{with new employer} \end{cases} \tag{2.4}$$

2.3.7 Social Security

The Social Security system in the U.S. provides retirement incentives at the time when these retirement benefits become available. The benefits are computed in several steps. First the earnings of the 35 highest earnings years are averaged into an index. It increases by working an extra year if earnings in that year is higher than the lowest earnings embedded in the index. Indexed lifetime earnings are also capped at some threshold. Then this index is converted to obtain the primary insurance amount (PIA) which determines the social security benefits. Let e_t be the social security wealth in the model (measure of indexed lifetime earnings). Then the social security wealth evolution is approximated in the model in the following simple way:

$$e_{t+1} = \max\{(e_t + \max\{0, (w_t h_t - e_t)/35\}), e^{\max}\}$$

The social security benefits are a piecewise linear function of social security wealth. It is computed in the following way:

$$ssb_t = \kappa_t \Big[0.90 \times \min\{e_t, b_0\} + 0.32 \times \min\{\max\{e_t - b_0, 0\}, b_1 - b_0\} \\ + 0.15 \times \max\{e_t - b_1, 0\} \Big]$$

[9] In practice there is a penalty for claiming the DC account balance before age $59\frac{1}{2}$. If my model, the DC wealth is illiquid until age 60

The social security benefits provide three major work dis-incentives after age 62. First the Average Indexed Monthly Earnings is only recomputed upwards if current earnings are greater than previous year of work. For instance staying longer in the labor market by working part-time does not increase the benefits. Secondly benefits can be first claimed at the early retirement age (ERA) which is 62. For every year before the normal retirement age (NRA) which is age 65 that the benefits are claimed, they are reduced by 6.7%. This is actuarially fair. But for every year after the NRA that the benefit application is delayed, benefits are increased by 3% which is actuarially unfair. Hence there is a strong incentive to draw benefits by age 65. Finally the social security earnings test taxes the labor income for the social security beneficiaries at a very high rate if the earnings are above a certain threshold till the age of 70. All these features are captured in the model.

2.3.8 State-Space

The state space x_t consists of a permanent pension type ($pen \in \{DB, DC, NO\}$), age of the individual at time t, continuous variables related to wealth and earnings – assets (a_t), wages (w_t), social security wealth (e_t), DC balances (q_t^{dc}) and DB pension wealth (q_t^{DB}). There is a discrete health related variable indicating bad health ($m_t \in \{good, bad\}$), social security application status ($b_{t-1} \in \{0, 1\}$) and two discrete variable indicating the labor market status. Previous period participation status ($p_{t-1} \in \{0, 1, 2\}$) indicating non-participation, participation with pension employer and participation with non-pension employer respectively. And finally an indicator for being laid off or not ($\lambda_t \in \{0, 1\}$).

2.3.9 Timing

The timing in the model is as follows: individuals wake up and observe their fixed pension plan type pen, age and current state $x_t = (a_t, e_t, w_t, q_t^{DC}, ten_t, m_t, b_{t-1}, \lambda_t, p_{t-1})$ at the beginning of every period t. If currently laid off, he chooses whether to retire, hours to work for the new employer conditional on participation decision, consumption and whether to receive social security if eligible. If not laid off currently, then he also makes a decision to switch employers. Shocks to wages, health, rate of return and job loss are realized

2.3.10 Recursive Formulation

Individuals solve a finite-horizon Markovian decision problem where they choose a sequence of consumption $\{c(\mathbf{x_t})\}_{t=1}^{T}$, hours $\{h(\mathbf{x_t})\}_{t=1}^{T}$, social security benefit application $\{b(\mathbf{x_t})\}_{t=1}^{T}$ and employment $\{d(\mathbf{x_t})\}_{t=1}^{T}$ rules to maximize the expected discounted lifetime utility subject to the exogenous processes for health transition, survival, job loss and wage determination, a set of budget (DB), borrowing (LC) and time constraint (2.1), government transfer rule (2.3), private pension wealth accrual and policies for taxes and Social Security.

The value function is a solution to a bellman equation given below. For exposition purposes, the bellman equation for each discrete employment scenario has been written separately.

<u>**Not Laid off**</u>

If an individual was working in the previous period and not laid off at the end of the period, then in the current period he chooses both hours to work in the labor market and whether to work for period $t-1$ employer or not. So he has three employment choices - work with the same employer, work with a different employer or retire. A job with a new employer is available in every period. If hours worked in the current period are positive, then he faces the uncertainty of losing his current job in the future period. If he chooses to work zero hours in the current period (retire), then there is no uncertainty about job loss in the future period.

Work with same employer

$$V(x_t) = \max_{\{c_t, h_t \in \mathbb{R}^+, b_t\}} \Bigg\{ U(c_t, h_t, d_t = 0) + \beta \pi^s_{t+1} \int V(x_{t+1}|x_t) dF(x_{t+1}|x_t)$$
$$+ \beta(1 - \pi^s_{t+1}) b(a_{t+1}) \Bigg\} \quad \text{s.t} \quad \text{(BC), (LC) and (CF)}$$

Work with different employer

$$V(x_t) = \max_{\{c_t, h_t \in \mathbb{R}^+, b_t\}} \Bigg\{ U(c_t, h_t, d_t = 1) + \beta \pi^s_{t+1} \int V(x_{t+1}|x_t) dF(x_{t+1}|x_t)$$
$$+ \beta(1 - \pi^s_{t+1}) b(a_{t+1}) \Bigg\} \quad \text{s.t} \quad \text{(BC), (LC) and (CF)}$$

Not work

$$V(x_t) = \max_{\{c_t, b_t\}} \Bigg\{ U(c_t, h_t = 0, d_t = 0) + \beta \pi^s_{t+1} \int V(x_{t+1}|x_t) dF(x_{t+1}|x_t)$$
$$+ \beta(1 - \pi^s_{t+1}) b(a_{t+1}) \Bigg\} \quad \text{s.t} \quad \text{(BC), (LC) and (CF)}$$

Laid off

If an individual was working in the previous period and laid off at the end of the period, then in the current period he only chooses the amount of hours to work in the labor market (with a new employer). A job with a new employer is available in every period. If hours worked in the current period are positive, then he faces the uncertainty of losing his current job in the future period. If he chooses to work zero hours in the current period (retire), then there is no uncertainty about job loss in the future period.

Work with different employer

$$V(x_t) = \max_{\{c_t, h_t \in \mathbb{R}^+, b_t\}} \Bigg\{ U(c_t, h_t, d_t = 1) + \beta \pi^s_{t+1} \int V(x_{t+1}|x_t) dF(x_{t+1}|x_t)$$
$$+ \beta(1 - \pi^s_{t+1}) b(a_{t+1}) \Bigg\} \quad \text{s.t} \quad \text{(BC), (LC) and (CF)}$$

Not work

$$V(x_t) = \max_{\{c_t, b_t\}} \Bigg\{ U(c_t, h_t = 0, d_t = 0) + \beta \pi^s_{t+1} \int V(x_{t+1}|x_t) dF(x_{t+1}|x_t)$$
$$+ \beta(1 - \pi^s_{t+1}) b(a_{t+1}) \Bigg\} \quad \text{s.t} \quad \text{(BC), (LC) and (CF)}$$

Individuals who were retired in the previous period solve the same problem as working individuals who were laid off at the end of the period.

2.4 Data

The model is estimated using data from the Health and Retirement Survey (HRS) [10] for the years 1992-2010. The HRS is a longitudinal sample of non-institutionalized individuals, over the age of 50. The first cohort (born in 1931-1941) was interviewed in 1992 and subsequently every two years. Along with the age-eligible respondents, the survey also interviewed the spouses or partners of the respondents. The HRS has a rich source of information on demographics, different measures of health status (subjective and objective), financial wealth, private pensions, social security, government transfers, income from work, labor market activity and retirement. For those respondents who gave permission to access their administrative records, the HRS data can be matched to the Social Security Administration (SSA) earnings data and pension plan information from the employers. These provide very accurate measure of social security and pension wealth accrued from both DB and DC plans. I use the restricted SSA administrative data and pension plan data to construct a measure of social security and pension wealth held by the individuals in the model at the beginning of the life-cycle.

I use data on male household heads with either a DB, DC or no pension plan. Since the theoretical model in this essay does not allow for combination pension plans, individuals on hybrid plans are dropped from the sample. I also drop observations on account of missing values for hours, wages and assets. Since self-employed workers face different financial incentives for reducing hours (not captured in my model), I drop self employed individuals. This reduces my sample to 5,130 individuals.

In order to make sure that the individuals in my model face similar social security and pension rules, I fit my model to the initial HRS cohort aged 51-61 in 1992. But I use all 10 waves of data to estimate the stochastic processes faced by the individuals.

2.5 Estimation

This section describes the estimation strategy. I use a two-step estimation procedure on the lines of Gourinchas and Parker [2002]. In the first step, I estimate or calibrate parameters which can be cleanly identified without using the structural model. This parameter vector is given by $\Phi = (\pi^m, \pi^s, \pi^\lambda, \omega(.), \rho^w, \sigma_\epsilon, \sigma_\psi, cr_w, cr_e, \bar{r}, \tau, ssb_t, pb_t^{DB}, \bar{c})$. I estimate the health transitions, wage process, survival and job loss probability from the HRS data in the first step. In the second step, I estimate the vector of preference parameters $\Theta =$

[10] The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan

$(\bar{l}, \beta, \rho, \nu, \theta_{beq}, \phi_H, \phi_P, \chi)$ using Method of Simulated Moments (MSM).

2.5.1 Initial Conditions

Initial conditions (state vector at the start of life cycle) are generated by taking random draws from the empirical joint distribution of household assets, wages, pension wealth (held under both Defined Benefit and Defined Contribution plans), health status, tenure and pension plan type for those who are working with their pension provider or do not have any pension plan. No one is laid off or retired at the beginning of the life cycle.

2.5.2 Method of Simulated Moments

Given the vector of exogenous data generating processes Φ and some vector of preference parameters Θ, I solve for the decision rules $c(x_t, \Phi, \Theta)$, $d(x_t, \Phi, \Theta)$, $b(x_t, \Phi, \Theta)$ and $d(x_t, \Phi, \Theta)$. I then use the estimated Φ and initial conditions x_0 to simulate the life cycle profiles of hypothetical individuals. Finally a MSM criterion function is used to find $\hat{\Theta}$ that minimizes the distance between aggregated simulated profiles and data profiles.

Moments Matched

I match the following moments to estimate the elements of Θ

1. Participation by pension plan type and age resulting in $3 \times T$ moment conditions.
2. Log of hours worked conditional on working by pension plan type and age resulting in $3 \times T$ moment conditions.
3. Employer switches by pension plan type and age resulting in $3 \times T$ moment conditions.
4. Log of hours worked conditional on working by health status and age resulting in $2 \times T$ moment conditions.
5. Participation by health status and age resulting in $2 \times T$ moment conditions.
6. Mean assets by age resulting in T moment condition

This gives a total of $14 \times T$ moment conditions. Formally the MSM estimate $\hat{\Theta}_{MSM}$ is one that solves:

$$\hat{\Theta}_{MSM} = \text{argmin } \tilde{g}(\Theta, \Phi) W_T \tilde{g}(\Theta, \Phi)$$

Where

$$\underbrace{\tilde{g}(\Theta,\Phi)}_{14T\times 1} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\{\log h_{it}^{j} - \log \tilde{h}_{t}^{j}(x_{it},\Theta,\Phi)\} \\ \frac{1}{N}\sum_{i=1}^{N}\{p_{it}^{j} - \tilde{p}_{t}^{j}(x_{it},\Theta,\Phi)\} \\ \frac{1}{N}\sum_{i=1}^{N}\{d_{it}^{j} - \tilde{d}_{t}^{j}(x_{it},\Theta,\Phi)\} \\ \frac{1}{N}\sum_{i=1}^{N}\{\log h_{it}^{m} - \log \tilde{h}_{t}^{m}(x_{it},\Theta,\Phi)\} \\ \frac{1}{N}\sum_{i=1}^{N}\{p_{it}^{m} - \tilde{p}_{t}^{m}(x_{it},\Theta,\Phi)\} \\ \frac{1}{N}\sum_{i=1}^{N}\{a_{it} - \tilde{a}_{t}(x_{it-1},\Theta,\Phi)\} \end{bmatrix}$$

$$t = \{1,...,T\} \; j \in \{DB, DC, NO\} \; m \in \{good, bad\}$$

W_T could be an optimal weighting matrix given by the inverse of a consistent estimate of the covariance matrix of data moments. However efficient choice of weighting matrix could introduce finite sample bias [Altonji and Segal, 1996]. Hence I use the following non-optimal weighting matrix for the structural estimation:

$$\underbrace{W_T}_{14T\times 14T} = \left[diag\Big(var\big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_{it}\big)\Big)\right]^{-1}$$

Where m_{it} is a vector of data moments

2.5.3 Selection

The main issue with this analysis is the assumption of the exogeneity of pension plans. In other words if agents with some unobserved preferences for early retirement are systematically selecting into jobs with DB pension plans, then the resulting retirement behavior will not be driven by the pension plan but these preferences themselves. In that case the model developed here will be overstating the impact of the different pension plans in generating differences in retirement behavior. One way to solve this problem would be to account for self selection in the model by building in unobserved preference heterogeneity on the lines of French and Jones [2011]. But one of the strengths of this model lies in it's ability to generate the differences in retirement behavior solely by a more accurate description of the budget constraint. Hence without a strong empirical evidence of selection into pension plans, it seems undesirable to resort to a more complicated description of preferences.

The Health and Retirement Survey Data contains subjective answers to some preference based questions asked to the retirement age population. I use the answers to three of these questions in a multinomial logit analysis to predict the probability of being in a particular pension plan. Table 2.3 reports the predicted probability of being in a particular pension plan conditional on answering yes to the respective preference based questions. The results indicate mixed evidence for selection. For instance the predicted probability of being in a DB plan is the lowest conditional on answering yes to the question "Do you look forward to retirement?". If people with high preference for leisure or early retirement were self selecting into DB plans then this probability should have been the highest. On the other hand, column 2 of the same table indicate that a person is most likely to be without any pension plan and least likely to be in a DB plan conditional on answering yes to the question "Would you like to retire gradually". This seems to be somewhat consistent with selection.

Table 2.3: Predicted Probability of having a DB, DC or NO Pension Plan from Multinomial Logit Regression

Pension Group	Q1 (Yes)	Q2 (Yes)	Q3 (Yes)
DB	0.2521	0.1682	0.1959
DC	0.3794	0.3387	0.3712
NO	0.3683	0.4930	0.4328

[Q1] Do you look forward to retirement?
[Q2] Would you like to retire gradually?
[Q3] Would you like to work even if you don't need the money?

2.6 Results

This section reports the results related to the structural estimation of preference parameters using MSM, model-fit and simulations from the estimated model.

2.6.1 Preference Parameter Estimates

The structural model is used to estimate two key parameters currently, the participation cost and the cost of switching employer. The value of other parameters are currently fixed (taken from French [2005])[11] in the estimation. Both ϕ_P and χ are measured in terms of hours. The estimates in Table (2.4) indicate that 106 hours are lost in switching an

[11]Details of all other parameters are presented in Appendix A.2

employer, i.e, choosing to work with a new employer. Participation in the labor market is expensive at older ages and cost 845 hours of time. This estimate of participation cost is 30% smaller than the model estimated with a single DB pension plan for all. In this model the average participation and hours for each age are matched to the average in the data without exploiting the variation in retirement behavior across pension plan types.

Table 2.4: Preference Parameter Estimates

Parameter and Definition	(1)	(2)
ϕ_P fixed cost of work	845	1178
	(12)	(15)
χ cost of switching employer	106	-
	(4)	

(1) Model with pension plan heterogeneity
(2) Model with a single DB plan for all (without employer switching)

Identification of Participation Cost

Participation cost is identified from the hours worked and participation profiles or in other words hours worked before switching to non-participation (zero hours of work). Usually a high participation cost ($\approx$ 1300 annual hours) is required to generate the same drop in hours in the model as observed in the data. This means that there is some unobservable component of preferences (not captured by the model ingredients) which are responsible for this big discontinuity in the life cycle hours profile. The data exhibits interesting variation in both hours worked and participation across three groups of retirement age population – DB, DC and No pension holders.

The model in this essay is constructed (capturing details of the different pension plans) such that these differences are originating solely from the differences in the budget constraint of these groups (and not due to differences in preferences). I match hours, participation and job switches for the three pension groups to jointly estimate two parameters - participation cost and cost of job switch in terms of annual hours. The variation in these moments for these different types carry identifying information. First the model cannot generate any incentive for the no pension group to endogenously switch employers but the DB pension groups use switching to mitigate the loss in pension wealth in the event of reducing hours. Hence, allowing for job switches from pension providing firms result in more variation in hours across the DB pension holders. This mechanism results in a small fraction of DB pension types working less than 2000 hours with a different employer in the event of bad health shock. Second, only a small fraction (DB pension holders) of the data population sees the sharp drop in hours. The behavior is more gradual for the other two groups. So when we match the behavior of the three groups in the model jointly to their counterparts in the data (1), the averaging behavior results in a participation cost which is smaller than the one estimated on DB population solely (2). And finally there is variation in the magnitude of the drop in hours generated by the budget constraint across the three groups. The budget constraint cannot generate most of the observed drop in hours for the no pension group. The only mechanism that can generate abrupt retirement for this pension group is participation cost (other than health shocks, switching cost and social security which are also common for all). On the other hand, the budget constraint for the DB pension group can generate part of the drop in hours worked before retirement. For all these reasons, earlier estimates of participation cost, obtained by matching the average behavior in the data to that of a DB pension holder in the model, potentially suffers from an upward bias.

The participation cost has implications for the labor elasticities along the intensive and extensive margin. A high participation cost generates large reservation hours. This results in most of the response of labor supply to wage changes along the participation margin and very little action along the intensive margin. The model in this essay has three groups of agents facing a common relatively lower participation cost (in time) of work but different monetary costs (in terms of loss in pension wealth) of reducing hours. Since the model generates more variation in hours for these pension groups, the resulting elasticities are different from those computed using a standard model with no pension wealth heterogeneity. Participation elasticities are 29% lower and hours elasticities are 58% higher as seen in Table (2.5).

Table 2.5: Elasticities

Wage Elasticity at age 60	(1)	(2)
Extensive Margin	0.8743	1.248
Intensive Margin	0.4562	0.1904

Wages at age 60 are reduced by 15% to compute the participation and hours elasticity

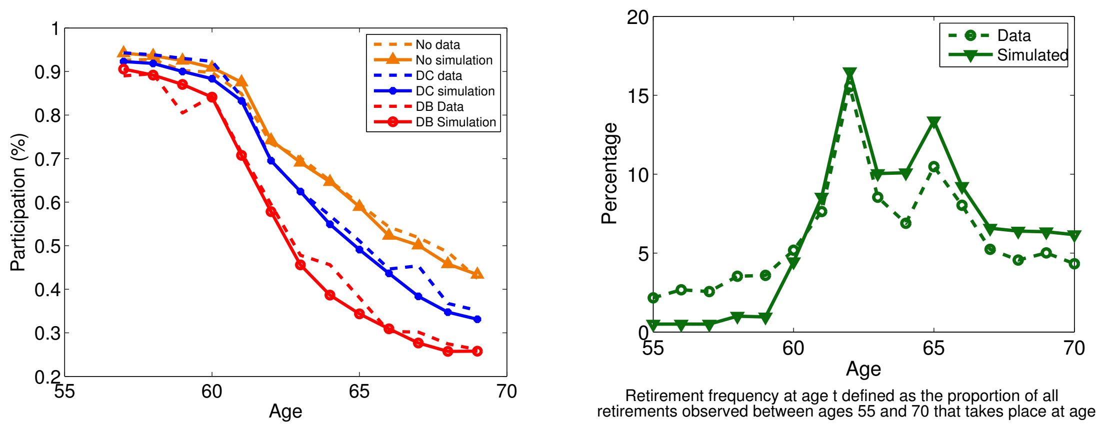
2.6.2 Model-fit

The model is able to fit participation rates by DB and DC pension plans well for the fixed parameter values and reasonable values of participation cost and switching cost. Figure 2.5a shows the participation rates in the data and model for Defined Benefit and Defined Contribution pension holders over ages 55 to 69. Defined Contribution plan holders in the model have higher participation rates on account of two reasons. First they hold less in pension wealth as compared to those on Defined Benefit pension plans. Second the marginal returns to staying a year longer with an employer for them remains fixed at 9% of their labor earnings at all ages. The marginal returns to staying a year longer for DB pension holders drop sharply at age 62. This results in most of the exits observed at age 62 from the labor market. Social security also provides strong work dis-incentives at older ages. The model is able to generate the age 62 and 65 peaks in retirement as observed in the data (Figure 2.5b).

Figure 2.5: Model vs. Data

(a) Participation of Men Ages 62 to 69

(b) Pension Type for Men Ages 18 to 50



2.6.3 Simulations

The estimated model is used to simulate the effect of a complete phase out from DB to DC pension plans on participation rates. To this effect, all DB pension plans at the beginning of the life cycle are changed to DC plans without changing the pension wealth. The estimated model predicts that increasing the DC pension composition by 32% result in a 17% increase in the participation rates for the age group 57 to 69 and 13% increase for the 60 to 65 years old. The increase is mainly due the the removal of age 62 work disincentive provided by the DB plans to exit the labor market.

2.7 Conclusion

In this essay, a stochastic dynamic programming model of retirement, savings, social security and pension wealth is estimated. The model allows for a very rich and precise formu-

Table 2.6: Pension Plan Composition

Pension Plan	Benchmark	Policy	Change
DB	52%	0%	$\downarrow$ 52%
DC	32%	84%	$\uparrow$ 52%
NO	16%	16%	-

Table 2.7: Labor force participation

Age Group	Benchmark	Policy	Change
57-69	35%	52%	$\uparrow$ 17%
60-65	31%	44%	$\uparrow$ 13%

lation of the budget constraint with respect of the retirement wealth. More specifically, it allows for pension wealth accrual through both Defined Benefit and Defined Contribution pension plans. DB plans in the model provide age specific incentive to retire as seen in practice whereas DC plans provide a fixed age-independent profile of wealth accrual with uncertainty pertaining to the rate of return.

Estimation results show that it is expensive to change employers. The participation cost estimated here is 35% lower as compared to some other estimates in the literature. The value of this model lies in it's ability to predict labor supply changes in response to changes in pension plan composition. An increase in DC plan composition by 32% in the estimated model, result in an increase in the participation rates of the 57 to 69 years old by 17% and 60 to 65 years old by 13%. This indicates that 43% of the recent increase in the labor force participation of the older workers can be mapped to the change in pension plan composition.

Ignoring pension plan variation a model can account for none of the variation in retirement behavior observed in the data while the model developed in this essay with homogenous preferences and pension plan heterogeneity can account for 67% of the variation across pension plans in participation and 71% variation in hours conditional on participation.

Chapter 3

Consumer Learning, Product Differentiation, and The Value of Generic Pharmaceutical Entry

The research outlined in this essay was conducted in collaboration with Prof. Amil Petrin, Prof. Pinar Karaca-Mandic and Prof. Jeffery McCullough. For this research, we use a 5% sample of Medicare claims requested from the Center of Medicare and Medicaid services (CMS). We have the necessary Data Use Agreement (DUA) for publishing the results of this research in our upcoming paper of the same title. However, CMS requires a separate Data Use Agreement for publishing any results obtained using the data as part of a dissertation. Due to these constraints, this essay only contains a brief overview of the project. The author would like to direct the readers to the forthcoming paper "Consumer Learning, Product Differentiation, and The Value of Generic Pharmaceutical Entry" for results and more details.

3.1 Introduction

Prescription pharmaceuticals constitute a large and growing portion of medical expenditures. Pharmaceutical costs - and incentive innovation incentives - are generated by patent protection. Patent expiration results in generic entry for most products and a market reduction in prices. Generic products have chemically equivalent active ingredients (i.e., the exact same molecule) and typically yield the same clinical benefits. Generic entry generally results in lower health expenditures, increased access, and higher short-run consumer surplus. The benefits of generic entry may be reduced by both firm strategies and consumer information barriers.

Pharmaceutical firms may develop follow-on products, often referred to as line extensions. These products often constitute real clinical innovation including formulations that can increase patient compliance or lower side effects. Line extensions may, however, reduce or delay consumers' adoption of generic competitors. Given the difficulty consumers face in learning about pharmaceutical value, line extensions may effectively differentiate products in excess of any efficacy or side effect differences. This is particularly important as the line extensions often face later generic entry dates.[1] These products are employed to effectively extend patent life.

Medicare Part D (created under the Medicare Modernization Act of 2003) provides outpatient prescription drug coverage to seniors and to people under the age of 65 with certain disabilities. Ever since it went to effect in 2006, it has been the center of much policy debate due to the substantial cost of the program. For instance, Part D cost the government $62 billion in 2010, constituting 12% of total federal spending for Medicare for that year [2]. The spending is projected to rise to 14% of total Medicare spending in 2015 and to 17% by 2023. This would result in a 6% per capita growth rate of federal spending on the Part D program [3].

A large fraction of the federal government's spending on the Medicare Part D program is incurred on the payment for brand name drugs. According to the report by the Congressional Budget Office[4], generic drugs which formed 73% of the total number of prescription filled under Part D in 2010, constituted only 13% ($21 billion) of the total prescription drug costs. Whereas brand name drugs, which only formed 27% of the prescription filled, resulted in 87% ($141 billion) of the total cost. A lot of these brand-name drugs are also available in generic versions and it has been argued by policy makers that a higher level of utilization of these generic drugs would result in significant cost savings for the government [Haas et al. [2005] Shrank et al. [2010], Frank [2007] Fischer and Avorn [2004], Rizzo and Zeckhauser [2009]]. However, the cost savings of forcing consumers onto generics may lead to large welfare losses for consumers of non-generic alternatives if they highly value

[1] Later generic dates often occur when line extensions are delayed beyond patent expiration when introduced late in a patent period followed by a lengthy FDA approval process. In relatively small product markets line extensions may never face generic competition.

[2] Department of Health and Human Services, Centers for Medicare and Medicaid Services, Office of the Actuary, 2011 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds (May 2011)

[3] Kaiser Family Foundation, The Medicare Prescription Drug Benefit Fact Sheet, September 2014

[4] July 2014 CBO report "Competition and the Cost of Medicares Prescription Drug Program"

them. A significant literature in Medicine and Public Health report that generic substitution can result in non-adherence, reduction in medication effectiveness or even adverse health outcomes in certain treatments [Crawford et al. [2006], Van Wijk et al. [2006], Kjoenniksen et al. [2006], Kanis et al. [2012], Duerden and Hughes [2010]].

We address this issue by analyzing the welfare implications of generic substitution. For our welfare analysis, we estimate a structural model of drug demand that allows for heterogeneity in match quality between consumers and drugs and also allows for consumer learning about the stochastic match quality of the drug. In doing so we build on and extend a rich literature on discrete choice demand model [McFadden et al. [1973], Rosen and Small [1979b], Trajtenberg [1985], Berry [1994], Petrin [2001]] and Bayesian learning models [insert citations]. Our framework is similar to that of [Crawford and Shum, 2005] and [Shin et al., 2012] .

For our analysis we focus on the osteoporosis (OP) market in the U.S. over the period 2007 and 2008. The osteoporosis market includes the presence of several branded drugs and one generic drug which is an exact bio-equivalent for one of the brand drugs Fosamax. This makes it an interesting case for analyzing the welfare implications of generic substitution. The OP market is also a significant market in terms of disease prevalence and pharmaceutical revenues [5].

3.2 Model

3.2.1 Overview

We estimate a structural model of pharmaceutical demand where patients are uncertain about drug match quality, are risk neutral and myopic. Patients receive signals of the match quality through drug use and update their expectations about match quality in a Bayesian manner as they receive future signals.

The model specification and the assumptions underlying the Bayesian learning model of drug demand are laid out in the rest of the section.

[5] Blume and Curtis [2011] report that roughly 30% of Medicare beneficiaries were treated for either fractures or OP without fractures with an estimated health care cost of $16 billion in the U.S. in 2002

3.2.2 Utility Function

Consumer i's utility from purchasing brand j of the osteoporosis drug at time t is given by:

$$U_{ijt} = \beta_{ij,t-1} + \alpha X_{ij,t} + \epsilon_{ij,t} \tag{3.1}$$

where $\beta_{ij,t}$ denotes patient i's beliefs about the stochastic match quality of drug j at time t. $X_{ij,t}$ is the out-of pocket price paid by patient i for drug j at time t and α is the price coefficient. Finally $\epsilon_{ij,t}$ is a patient-drug and time- specific shock. Notice that the utility is stochastic because the belief about the match quality is stochastic but $\epsilon_{ij,t}$ while unobserved by the econometrician, is not stochastic from the point of view of the patient. We assume that $\epsilon_{ij,t}$ is *iid* Type I Extreme Value distribution. The patient is assumed to maximize her expected utility given by:

$$U^E_{ijt} = E[U_{ijt}] = E[\beta_{ij,t-1}] + \alpha X_{ijt} + \epsilon_{ij,t} \tag{3.2}$$

3.2.3 Learning Process

Following the literature on Bayesian learning process (Crawford and Shum [2005], Erdem and Keane [1996], Ackerberg [2003], Shin et al. [2012] and so on), we assume that a single instance of drug use does not reveal it's match quality. Patients learn about the match quality by updating their beliefs over successive use of alternative drugs. More specifically, they receive a noisy signal of match quality after each experience of a particular drug and combine the information contained in the signal with their prior beliefs to obtain posterior beliefs about the drug in accordance with the Bayes's rule.

Quality Signal

Let $q_{ij,t}$ be the quality signal about drug j's match that patient i received after consuming drug j at time t. We assume that quality signals follow a normal distribution:

$$q_{ij,t} \sim N(\beta_{ij}, \sigma^2_{q_{ij}}) \tag{3.3}$$

Where β_{ij} is patient i's true mean match quality for drug j and $\sigma^2_{q_{ij}}$ is the variance of the quality signal for drug j faced by patient i. This implies that learning happens over multiple drug experience occasions and the quality signal is a noisy measure of the true mean match quality .

Prior Beliefs

Every patient is newly diagnosed with osteoporosis at time $t = 0$ and has no experience with osteoporosis treatment drugs. Hence the patient has no information about her match quality with different drugs. But the patient has prior beliefs about the match quality of the drug. Patient $i's$ belief about the match quality of drug j has the following normal distribution:

$$\beta_{ij,0} \sim N(\mu_{\beta_{ij,0}}, \sigma^2_{\beta_{ij,0}}) \tag{3.4}$$

Where $\mu_{\beta_{ij,0}}$ and $\sigma^2_{\beta_{ij,0}}$ are patient i's initial beliefs of the mean and variance of drug j's match quality.

Posterior Beliefs

When the patient experiences drug j at time $t = 1$ she receives a quality signal which she uses to update her beliefs about the match quality in a Bayesian manner. Since the prior belief at time $t = 0$ and all subsequent quality signals are assumed to be normally distributed, it follows that the posterior belief at every time period is also normally distributed. Hence the evolution of posterior beliefs can be fully characterized by the laws of motion for posterior mean and variance.

The posterior mean and variance are updated in the following recursive fashion using Bay's rule:

$$\mu_{\beta_{ij,t}} = \frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij,t-1}}} \mu_{\beta_{ij,t-1}} + d_{ij,t} \frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij}}} q_{ij,t} \tag{3.5}$$

$$\sigma^2_{\beta_{ij,t}} = \frac{1}{\frac{1}{\sigma^2_{\beta_{ij,t-1}}} + d_{ij,t} \frac{1}{\sigma^2_{\beta_{ij}}}} \tag{3.6}$$

where $d_{ij,t}$ is an indicator variable such that:

$$d_{ij,t} = \begin{cases} 1 & \text{if drug } j \text{ taken in period } t, \\ 0 & \text{otherwise} \end{cases}$$

Iterating forward Equations 3.5 and3.6 we get the following expressions for the mean and variance of the match quality at time t

$$\mu_{\beta_{ij,t}} = \frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij,0}}}\mu_{\beta_{ij,0}} + \sum_{\tau=1}^{t} d_{ij,\tau} q_{ij,\tau} \frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij}}} \tag{3.7}$$

$$\sigma^2_{\beta_{ij,t}} = \frac{1}{\frac{1}{\sigma^2_{\beta_{ij,0}}} + \frac{\sum_{\tau=1}^{t} d_{ij,\tau}}{\sigma^2_{\beta_{ij}}}} \tag{3.8}$$

Following Shin et al. [2012] we do the following change of variables to write down an alternate expression for the Bayesian learning process:

$$\nu_{ij,t} = \mu_{\beta_{ij,t}} - \beta_{ij} \tag{3.9}$$

$$\eta_{ij,t} = q_{ij,t} - \beta_{ij} \tag{3.10}$$

Where $\nu_{ij,t}$ is referred as the "perception bias " and represents deviation of patients i's belief about the mean quality from the true mean quality of the match. $\eta_{ij,t}$ is the signal noise. This transformation results in the following expression for the evolution of the posterior mean of the match quality of drug j with patient i:

$$\mu_{\beta_{ij,0}} = \beta_{ij} + \nu_{ij,0} \tag{$t=0$}$$

$$\mu_{\beta_{ij,1}} = \beta_{ij} + \nu_{ij,1} = \beta_{ij} + \frac{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}}\nu_{ij,0} + d_{ij,1}\eta_{ij,1}}{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}} + d_{ij,1}} \tag{$t=1$}$$

$$\vdots$$

$$\mu_{\beta_{ij,k}} = \beta_{ij} + \nu_{ij,k} = \beta_{ij} + \frac{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}}\nu_{ij,0} + \sum_{\tau=1}^{k} d_{ij,\tau}\eta_{ij,\tau}}{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}} + \sum_{\tau=1}^{k} d_{ij,\tau}} \tag{$t=k$}$$

$$\vdots$$

$$\mu_{\beta_{ij,t*}} = \beta_{ij} + \nu_{ij,t*} = \beta_{ij} \tag{$t=t*$}$$

Where $t*$ is the time when a patient's belief about match quality converges to the the true match quality and learning stops. Using this expression for the evolution of posterior beliefs about the mean match quality of the drug, the expected utility in equation 3.2 can be rewritten as:

$$
\begin{aligned}
U^E_{ijt} =&\beta_{ij} + \nu_{ijt-1} + \alpha X_{ij,t} + \epsilon_{ij,t} \\
=&\beta_{ij} + \frac{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}}\nu_{ij,0} + \sum_{\tau=1}^{t-1} d_{ij,\tau}\eta_{ij,\tau}}{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}} + \sum_{\tau=1}^{t-1} d_{ij,\tau}} + X_{ij,t}\alpha + \epsilon_{ij,t}
\end{aligned}
$$

Notice that the distributional assumptions on $\epsilon_{ij,t}$ result in a random coefficient multinomial logit model where product-specific intercepts are composed of true mean match quality (β_{ij}) and perception bias $(\nu_{ij,t-1})$.

3.3 Identification

The parameters of the Bayesian learning process $\{\beta_{ij}, \sigma^2_{\beta_{ij}}, \nu_{ij,0}, \sigma^2_{\beta_{ij,0}}\}$ $\forall i, j$ are not identifiable in the current form. Following [Shin et al., 2012] we make the following identifying assumptions. First we normalize the quality signal variance $\sigma^2_{\beta_{ij}}$ to 1 as both the quality signal variance$\sigma^2_{\beta_{ij}}$ and the variance of the initial belief about the match quality $\sigma^2_{\beta_{ij,0}}$ cannot be identified separately but only their ratio $\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}}$ can be identified. Individual-level mean and variance of initial quality perception $\{\nu_{ij,0}, \sigma^2_{\beta_{ij,0}}\}$ is generally identified by using survey information or advertising data collected prior to the choices being observed. Since this information is not available to us, we restrict the priors to be homogeneous across consumers. More specifically we restrict $\nu_{ij,0} = \bar{\nu_j}, (\bar{\nu}_J = 0)$ and $\sigma^2_{\beta_{ij,0}} = \bar{\sigma}^2_{\beta_0}$.

The true mean quality of the drug β_{ij} is identified by the steady state drug choice of the patient. As the patient receives quality signals over successive drug experiences, learning occurs and the beliefs about match quality evolves to the true match quality of the drug. In our data set, we observe drug choices of patients for 24 months after being diagnosed with osteoporosis. This gives us a long enough time series to correctly identify the true mean quality. Further, not all drug-specific β_{ij}'s are identified, hence we normalize one of them $(\beta_{iJ} = 0)$.

3.4 Estimation

We use the Bayesian method of inference to estimate the parameters of the model. We use a Markov Chain Monte Carlo (MCMC) scheme to draw from the stationary joint posterior density of the parameters. The details of the MCMC sampler are provided in Appendix B.1

The set of parameters to be estimated is given by $\Theta = \{\Phi_i\}_i^N \times \Psi$ where $\Phi_i = \{\beta_{i1}, \ldots, \beta_{iJ-1}, \{\eta_{ij,t}\}_{\tau=1}^{T_i-1}\}$ represent a set of individual level parameters and $\Psi = \{\bar{\nu}_1, \ldots, \bar{\nu}_{J-1}, \bar{\sigma}^2_{\beta_0}, \alpha\}$ are the aggregate level parameters. The liklihood function is given by:

$$L_i(\mathbf{d_i}|\mathbf{X_i}; \Phi_i, \Psi) = \prod_{t=1}^{T_i} \prod_{J=1}^{J} \left(\frac{\exp(\bar{U}^E_{ij,t})}{\sum_{j'} \exp(\bar{U}^E_{ij',t})} \right)^{d_{ij,t}}$$

3.5 Baseline Results

Please refer to the upcoming paper "Consumer Learning, Product Differentiation, and The Value of Generic Pharmaceutical Entry" by Neha Bairoliya, Amil Petrin, Pinar Karaca-Mandic and Jeffery McCullough for estimation results.

3.6 Counterfactual Policy Simulations

To understand the welfare implications of generic substitution we model demand under alternative cost sharing policies. Once the generic becomes available in 2008, individuals pay for the full price of the branded versions of Alendronate, Fosamax and Fosamax Plus D. The full market prices for one month supplies of Fosamax and Fosamax Plus D are approximately \$130 and \$165 respectively. We then evaluate expenditures and welfare changes by means of compensating variation. Following Train [1998] and Rosen and Small [1979a], our measure of compensating variation for household i is given by:

$$CV_i = \frac{\log[\sum_{j=1}^J \exp(V^1_{ij})] - \log[\sum_{j=1}^J \exp(V^0_{ij})]}{\alpha}$$

Where V^1_{ij} and V^0_{ij} are expected utilities as defined in equation (3.2) post and pre price change respectively. CV_i is then averaged across all agents to obtain an average dollar amount for each individual in the population. In this setting, CV could be interpreted as either a welfare loss or an information barrier. We compare the cost savings and the CV under alternative assumptions.

For the first policy scenario, we individuals pay for the full price of Fosamax after the generic Alendronate becomes available. For the second policy scenario, individuals pay for the full price of both Fosamax and it's line extension, Fosamax Plus D, once Alendronate becomes available. Please refer to the upcoming paper "Consumer Learning, Product Differentiation, and The Value of Generic Pharmaceutical Entry" by Neha Bairoliya, Amil

Petrin, Pinar Karaca-Mandic and Jeffery McCullough for results.

3.7 Conclusion

We estimate a structural model of drug demand allowing for heterogeneity in match quality between consumers and drugs and also allowing for consumer learning about the stochastic match quality of the drug. Estimates from the Bayesian learning model suggest that there are significant welfare losses to consumers from generic substitution. Furthermore, naive models seriously underestimate price elasticities for branded pharmaceuticals. We analyze a policy scenario where agents pay the full price of the brand drug once it's generic becomes available (Fosamax in our case). The compensating variations we compute may be interpreted as a welfare loss if consumers are fully informed or they may represent an information barrier. In either case, we find that the cost savings from increased generic substitution excede the welfare losses associated with this policy change.

Bibliography

D. A. Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination*. *International Economic Review*, 44(3):1007–1040, 2003.

J. G. Altonji and L. M. Segal. Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366, 1996.

S. T. Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

D. Blau. Pensions, household saving, and welfare: A dynamic analysis. 2011.

D. M. Blau and D. B. Gilleskie. The role of retiree health insurance in the employment behavior of older men*. *International Economic Review*, 49(2):475–514, 2008.

S. W. Blume and J. Curtis. Medical costs of osteoporosis in the elderly medicare population. *Osteoporosis International*, 22(6):1835–1844, 2011.

H. Bohn. Will social security and medicare remain viable as the us population is aging? In *Carnegie-Rochester Conference Series on Public Policy*, volume 50, pages 1–53. Elsevier, 1999.

J. Broadbent, M. Palumbo, and E. Woodman. The shift from defined benefit to defined contribution pension plans–implications for asset allocation and risk management. *Reserve Bank of Australia, Board of Governors of the Federal Reserve System and Bank of Canada*, 2006.

M. Casanova. Happy together: A structural model of couples joint retirement choices. *Unpublished Manuscript, Department of Economics, UCLA, Los Angeles*, 2010.

S. Chan and A. H. Stevens. Do changes in pension incentives affect retirement? a longitudinal study of subjective retirement expectations. *Journal of Public Economics*, 88(7): 1307–1333, 2004.

G. S. Crawford and M. Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, 2005.

P. Crawford, M. Feely, A. Guberman, and G. Kramer. Are there potential problems with generic substitution of antiepileptic drugs?: A review of issues. *Seizure*, 15(3):165–176, 2006.

M. De Nardi. Wealth inequality and intergenerational links. *The Review of Economic Studies*, 71(3):743–768, 2004.

M. De Nardi, S. Imrohoroğlu, and T. J. Sargent. Projected us demographics and social security. *Review of Economic Dynamics*, 2(3):575–615, 1999.

M. G. Duerden and D. A. Hughes. Generic and therapeutic substitutions in the uk: are they a good thing? *British journal of clinical pharmacology*, 70(3):335–341, 2010.

D. S. Dwyer and O. S. Mitchell. Health problems as determinants of retirement: Are self-rated measures endogenous? *Journal of health economics*, 18(2):173–193, 1999.

D. T. Ellwood. Pensions and the labor market: a starting point (the mouse can roar). In *Pensions, labor, and individual choice*, pages 19–54. University of Chicago Press, 1985.

T. Erdem and M. P. Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, 15(1):1–20, 1996.

M. A. Fischer and J. Avorn. Economic implications of evidence-based prescribing for hypertension: can better care cost less? *Jama*, 291(15):1850–1856, 2004.

R. G. Frank. The ongoing regulation of generic drugs. *New England Journal of Medicine*, 357(20):1993–1996, 2007.

E. French. The effects of health, wealth, and wages on labor supply and retirement behavior. *The Review of Economic Studies*, 72(2):395–427, 2005.

E. French and J. B. Jones. The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732, 2011.

V. Galasso. *The political future of social security in aging societies*. Mit press, 2008.

R. Gebhardtsbauer and J. Turner. The protection of pension-covered workers and beneficiaries: An analysis of the uk legislation establishing the pension protection fund. *prepared for Age Concern UK, March*, 2004.

P.-O. Gourinchas and J. A. Parker. Consumption over the life cycle. *Econometrica*, 70(1): 47–89, 2002.

A. L. Gustman and T. L. Steinmeier. The stampede toward defined contribution pension plans: fact or fiction? *Industrial Relations: A journal of Economy and Society*, 31(2): 361–369, 1992.

A. L. Gustman and T. L. Steinmeier. Effects of pensions on savings: Analysis with data from the health and retirement study. In *Carnegie-Rochester Conference Series on Public Policy*, volume 50, pages 271–324. Elsevier, 1999.

A. L. Gustman, T. L. Steinmeier, and N. Tabatabai. *Pensions in the Health and Retirement Study*. Harvard University Press, 2010.

J. S. Haas, K. A. Phillips, E. P. Gerstenberger, and A. C. Seger. Potential savings from substituting generic drugs for brand-name drugs: medical expenditure panel survey, 1997–2000. *Annals of internal medicine*, 142(11):891–897, 2005.

R. G. Hubbard, J. Skinner, and S. P. Zeldes. Precautionary saving and social insurance. Technical report, National Bureau of Economic Research, 1995.

E. C. Hustead. Trends in retirement income plan administrative expenses. *Living with defined contribution pensions*, pages 166–177, 1998.

H. Iams, B. A. Butrica, K. E. Smith, and E. J. Toder. The disappearing defined benefit pension and its potential impact on the retirement incomes of baby boomers. *Social Security Bulletin*, 69(3):1–28, 2009.

K. L. Judd. *Numerical methods in economics*. MIT press, 1998.

J. Kanis, J.-Y. Reginster, J.-M. Kaufman, J.-D. Ringe, J. Adachi, M. Hiligsmann, R. Rizzoli, and C. Cooper. A reappraisal of generic bisphosphonates in osteoporosis. *Osteoporosis International*, 23(1):213–221, 2012.

I. Kjoenniksen, M. Lindbaek, and A. G. Granas. Patients attitudes towards and experiences of generic drug substitution in norway. *Pharmacy World and Science*, 28(5):284–289, 2006.

L. J. Kotlikoff and D. A. Wise. The incentive effects of private pension plans. In *Issues in pension economics*, pages 283–340. University of Chicago Press, 1987.

L. J. Kotlikoff and D. A. Wise. Employee retirement and a firm's pension plan, 1989.

D. McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.

A. Petrin. Quantifying the benefits of new products: The case of the minivan. Technical report, National Bureau of Economic Research, 2001.

D. Rajnes. An evolving pension system: Trends in defined benefit and defined contribution plans. *EBRI Issue Brief*, (249), 2002.

J. A. Rizzo and R. Zeckhauser. Generic script share and the price of brand-name drugs: the role of consumer choice. *International journal of health care finance and economics*, 9 (3):291–316, 2009.

H. S. Rosen and K. A. Small. Applied welfare economics with discrete choice models, 1979a.

H. S. Rosen and K. A. Small. Applied welfare economics with discrete choice models, 1979b.

J. Rust and C. Phelan. How social security and medicare affect retirement behavior in a world of incomplete markets. *Econometrica: Journal of the Econometric Society*, pages 781–831, 1997.

A. A. Samwick. New evidence on pensions, social security, and the timing of retirement. *Journal of public economics*, 70(2):207–236, 1998.

S. Shin, S. Misra, and D. Horsky. Disentangling preferences and learning in brand choice models. *Marketing Science*, 31(1):115–137, 2012.

W. H. Shrank, N. K. Choudhry, J. Agnew-Blais, A. D. Federman, J. N. Liberman, J. Liu, A. S. Kesselheim, M. A. Brookhart, and M. A. Fischer. State generic substitution laws can lower drug outlays under medicaid. *Health affairs*, 29(7):1383–1390, 2010.

R. C. Sickles and P. Taubman. An analysis of the health and retirement status of the elderly, 1984.

J. H. Stock and D. A. Wise. Pensions, the option value of work, and retirement, 1988.

K. E. Train. Recreation demand models with taste differences over people. *Land economics*, pages 230–239, 1998.

K. E. Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

M. Trajtenberg. The welfare analysis of product innovations with an application to ct scanners. Technical report, National Bureau of Economic Research, 1985.

W. Van der Klaauw and K. I. Wolpin. Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145(1):21–42, 2008.

B. L. Van Wijk, O. H. Klungel, E. R. Heerdink, and A. de Boer. Generic substitution of antihypertensive drugs: does it affect adherence? *Annals of Pharmacotherapy*, 40(1): 15–20, 2006.

Appendix A

Supplemental Material to Effect of Pension Plan Type on Retirement Behavior

A.1 Mathematical Appendix

This section outlines the method for solving the dynamic programming problem of the individuals described earlier. It describes in detail the computation of the decision rules and methods for integrating the value function with respect to uncertainty.

A.1.1 Numerical Method

The value function is a solution to the following finite-horizon Markovian decision problem:

$$V(x_t) = \max_{\{c_t,h_t,b_t,d_t\}} \Bigg\{ U(c_t,h_t,d_t) + \beta(1-\pi^s_{t+1}) \Bigg[\pi^\lambda_{t+1} \int V_{t+1}(x_{t+1}|x_t) dF(x_{t+1}|x_t) \\ + (1-\pi^\lambda_{t+1}) \int V_{t+1}(x_{t+1}|x_t) dF(x_{t+1}|x_t) \Bigg] + \beta \pi^s_{t+1} b(a_{t+1}) \Bigg\} \tag{A.1}$$

The solution to the above decision problem is given by a sequence of consumption $\{c(\mathbf{x_t})\}_{t=1}^T$, hours $\{h(\mathbf{x_t})\}_{t=1}^T$, employment $\{d(\mathbf{x_t})\}_{t=1}^T$and social security benefit application $\{b(\mathbf{x_t})\}_{t=1}^T$

rules which solve the bellman equation (A.1). Even though there is no closed form solution for optimal consumption, hours, employment choice and benefit application, these rules fully characterize the decisions of the individuals.

The decision rules are computed numerically by value function iteration, starting at time T and working backwards to the first period. Time T decision rules are found by maximizing equation (A.1) subject to the budget constraint (BC), liquidity constraint (LC) and consumption floor (CF) for each value of the state x_t and $V_{T+1} = b(a_{T+1})$.

The state variables are discretized into a finite number of points on a grid and the value function is evaluated at those points. More specifically I choose 30 asset states, 10 wage states, 10 AIME states, 10 defined contribution wealth states, 2 application states, 2 health states, 10 tenure states, 3 pension plan types, 2 states for being laid off or not and 3 states for last period participation status. This requires solving the value function at $30 \times 10 \times 10 \times 10 \times 2 \times 2 \times 2 \times 3 = 720000$ different points between ages 62 and 70 (when the individual is eligible to apply) and 360000 points before 62 and after age 70 for both DB and DC guys.

A.1.2 Quadrature Rules

This section provides the quadrature rules used to evaluate the uncertainty with respect to wages and rate of return on DC account.

Wages

$$
\begin{aligned}
w_{t+1} &= \exp(\omega(m_{t+1}, age_{t+1}))\exp(\rho_w \eta_t)\exp(\epsilon^w_{t+1}) \\
&= \exp(\omega(m_{t+1}, age_{t+1}))\exp(\rho_w \log(w_t) - \rho_w \omega(m_t, age_t))\exp(\epsilon^w_{t+1}) \\
&= \exp[\rho_w \log(w_t) + (1-\rho_w)\omega(m_t, age_t)]\exp[\rho_w \omega(m_{t+1}, age_{t+1}) \\
&- \rho_w \omega(m_t, age_t)]\exp(\epsilon^w_{t+1}) \\
&= \bar{w}_{t+1}\exp(\epsilon^w_{t+1}) \\
EV_{t+1}(w_{t+1}) &= \int V_{t+1}(w_{t+1}) f(w_{t+1}|w_t) dw \\
&= \int V_{t+1}(\bar{w}_{t+1} \cdot \exp(\epsilon^w_{t+1})) f(\exp(\epsilon^w_{t+1})|\exp(\epsilon^w_t)) d\epsilon_{t+1} \\
&= \int V_{t+1}(\bar{w}_{t+1} \cdot z) f(z) dz \quad (\text{Where } \; z \sim \log \text{Normal}(0, \sigma^2_\epsilon)) \\
&= \frac{1}{z\sigma_\epsilon \sqrt{2\pi}} \int V_{t+1}(\bar{w}_{t+1} \cdot z) \mathrm{e}^{\frac{-(\log(z))^2}{2\sigma^2_\epsilon}} dz
\end{aligned}
$$

Let $\frac{\log(z)}{\sqrt{2}\sigma_\epsilon} = x \implies z = \exp[x\sqrt{2}\sigma_\epsilon]$ This change of variable gives $dz = (z\sqrt{2}\sigma_\epsilon)dx$. Hence we get:

$$
EV_{t+1}(w_{t+1}) = \frac{1}{\sqrt{\pi}} \int V_{t+1}(\bar{w}_{t+1} \cdot \exp(x\sqrt{2}\sigma_\epsilon)) \mathrm{e}^{-x^2} dx \tag{A.2}
$$

The integral in Equation (A.2) can be approximated by Gauss Hermite Quadrature Rules:

$$
\frac{1}{\sqrt{\pi}} \int V_{t+1}(\bar{w}_{t+1} \cdot \exp(x\sqrt{2}\sigma_\epsilon)) \mathrm{e}^{-x^2} dx \approx \frac{1}{\pi} \sum_{1}^{N} V_{t+1}(\bar{w}_{t+1} \exp(x_i \sqrt{2}\sigma_\epsilon)) \omega_i
$$

Where $\{x_i, \omega_i\}_{i=1}^N$ are Quadrature nodes and weights respectively. More specifically, I use quadrature of order 5 to evaluate the above integrals.

i	1	2	3	4	5
x_i	-2.0202	-0.9586	0.0000	0.9586	2.0202
ω_i	0.0200	0.3936	0.9453	0.3936	0.0200

Defined Contribution Wealth

$$\begin{aligned}
DC^*_{t+1} &= DC_{t+1}(1 + r_{t+1}) \\
&= DC_{t+1}(1 + \bar{r}) \exp(\theta_{t+1}) \\
EV_{t+1}(DC^*_{t+1}) &= \int V_{t+1}(DC^*_{t+1}) f(DC^*_{t+1}) dDC^* \\
&= \int V_{t+1}(DC_{t+1}(1 + \bar{r}) \exp(\theta_{t+1})) f(\exp(\theta_{t+1})) d\theta \\
&= \int V_{t+1}(DC_{t+1}(1 + \bar{r}) \cdot z) f(z) dz \quad \text{(Where } z \sim \log \text{Normal}(0, \sigma^2_\theta)) \\
&= \frac{1}{z\sigma_\theta\sqrt{2\pi}} \int V_{t+1}(DC_{t+1}(1 + \bar{r}) \cdot z) \mathrm{e}^{\frac{-(\log(z))^2}{2\sigma^2_\epsilon}} dz
\end{aligned}$$

Let $\frac{\log(z)}{\sqrt{2}\sigma_\theta} = x \implies z = \exp[x\sqrt{2}\sigma_\theta]$ This change of variable gives $dz = (z\sqrt{2}\sigma_\theta)dx$. Hence we get:

$$EV_{t+1}(DC^*_{t+1}) = \frac{1}{\sqrt{\pi}} \int V_{t+1}(DC_{t+1}(1 + \bar{r}) \cdot \exp(x\sqrt{2}\sigma_\theta)) \mathrm{e}^{-x^2} dx \tag{A.3}$$

The integral in Equation (A.3) can be approximated by Gauss Hermite Quadrature Rules:

$$\frac{1}{\sqrt{\pi}} \int V_{t+1}(DC_{t+1}(1 + \bar{r}) \cdot \exp(x\sqrt{2}\sigma_\theta)) \mathrm{e}^{-x^2} dx \approx \frac{1}{\sqrt{\pi}} \sum_1^N V_{t+1}(DC_{t+1}(1 + \bar{r}) \exp(x_i\sqrt{2}\sigma_\theta)) \omega_i$$

Where $\{x_i, \omega_i\}_{i=1}^N$ are Quadrature nodes and weights respectively. More specifically, I use quadrature of order 5 to evaluate the above integrals.

i	1	2	3	4	5
x_i	-2.0202	-0.9586	0.0000	0.9586	2.0202
ω_i	0.0200	0.3936	0.9453	0.3936	0.0200

2-D Quadrature

In the case where there is uncertainty pertaining to both DC wealth and wages, I use 2-dimensional quadrature to evaluate the expectation. More specifically, I use product rule based on 1-dimensional Gauss Hermite Quadrature rules for both wages and DC wealth

[Judd, 1998].

$$EV_{t+1}(w_{t+1}, DC^*_{t+1}) \approx \frac{1}{\pi} \sum_{i=1}^{5} \sum_{j=1}^{5} \omega_i \omega_j V_{t+1}(\bar{w}_{t+1} \exp(x_i \sqrt{2} \sigma_\epsilon), DC_{t+1}(1+\bar{r}) \exp(x_j \sqrt{2} \sigma_\theta))$$

A.2 Parameters

The preference parameters in Table A.1 are going to be estimated in the next version of this using the structural model described in Section (3.2). Currently these are taken from French [2005]. Table A.2 reports the value of some other parameters fixed in the analysis:

Table A.1: Preference Parameters

Parameter and Definition	Value
β discount factor	0.992
$\bar{l}$ leisure endowment	4466
ρ coefficient of relative risk aversion	3.34
ν consumption weight	0.578
θ_{beq} bequest weight	1.69
ϕ_H hours of leisure lost due to bad health	318

Table A.2: Other Parameters

Parameter	value
$\bar{r}$	0.04
σ_ψ	0.014
cr_w	0.06
cr_e	0.03
ρ^W	0.013
σ_{ϵ^w}	0.977
$\bar{c}$	1000

Appendix B

Supplemental Material to The Costs and Benefits of Increased Generic Substitution

B.1 MCMC Sampler

The set of parameters to be estimated is given by $\{\beta_{i1}, ..., \beta_{iJ-1}, \{\eta_{ij,t}\}_{\tau=1}^{T_i-1}, \bar{\nu}_1, ..., \bar{\nu}_{J-1}, \bar{\sigma}^2_{\beta_0}, \alpha\}$ We use Bayesian inference to estimate this parameter vector which requires an unconditional prior distribution of the parameter vector along with the liklihood function.

Prior Distribution

We specify the following prior distribution for our parameters:

$$
\begin{aligned}
[\beta_i|\bar{\beta}, \Omega_{\bar{\beta}}] &\sim MVN(\bar{\beta}, \Omega_{\bar{\beta}}) \\
[\bar{\beta}|\bar{\beta}_0, S_0] &\sim MVN(\bar{\beta}_0, S_0) \\
[\Omega_{\bar{\beta}}|K, I] &\sim IW(K, I) \\
[\alpha|\alpha_0^m, \alpha_0^v] &= N(\alpha_0^m, \alpha_0^v) \\
[\bar{\nu}|\nu_0^m, \nu_0^v] &\sim MVN(\nu_0^m, \nu_0^v) \\
[\frac{1}{\sigma^2_{\beta_0}}|\sigma_0^m, \sigma_0^v] &\sim N(\sigma_0^m, \sigma_0^v) \\
[\eta_{ijt}|\mu_\eta, \sigma_\eta] &\sim N(\mu_\eta, \sigma_\eta)
\end{aligned}
$$

Hence joint prior distribution of our parameter vector $\Theta = \{\Phi_i\}_i^N \times \Psi$ is given by:

$$k(\Theta) = \prod_{i=1}^{N}\left([\beta_i|\bar{\beta}, \Omega_{\bar{\beta}}]\prod_{\tau=1}^{T_i}[\eta_{ij\tau}|\mu_\eta, \sigma_\eta]\right) \times [\bar{\beta}|\bar{\beta}_0, S_0] \times [\Omega_{\bar{\beta}}|K, I] \times [\alpha|\alpha_0^m, \alpha_0^v]$$
$$\times[\bar{\nu}|\nu_0^m, \nu_0^v] \times [\frac{1}{\sigma^2_{\beta_0}}|\sigma_0^m, \sigma_0^v]$$

Joint Posterior Distribution

Given the prior distribution of the parameter vector and the liklihood function, the joint posterior distribution of the parameter vector conditional on the data is given by:

$$K(\Theta|\{\mathbf{d_i}, \mathbf{X_i}\}_{i=}^N) \propto \prod_{i=1}^{N} L_i(\mathbf{d_i}|\mathbf{X_i}; \Phi_i, \Psi)k(\Theta)$$

Bayesian inference requires drawing from this joint posterior distribution of the parameter vector. While it is possible to draw directly from this distribution using the Metropolis Hastings algorithm, it computationally very slow. Hence draws from this distribution are obtained through Gibbs sampling. This is an iterative process which requires drawing each parameter conditional on other parameters. Following [Train, 2009], we specify the following Gibbs sampling procedure for drawing from the joint posterior distribution.

1. Update $\beta_i = \{\beta_{i1}, ..., \beta_{iJ-1}\}$ using a Metropolis-Hastings sampler.
2. Update $\bar{\beta}$ and $\Omega_{\bar{\beta}}$ using a Gibbs sampler
3. Update α by a Metropolis-Hastings sampler
4. Update $\{\bar{\nu}_1, ..., \bar{\nu}_{J-1}, \bar{\sigma}^2_{\beta_0}\}$ using a Metropolis-Hastings sampler
5. Update $\{\eta_{ij,\tau}\}_{\tau=1}^{t-1}$ by a Metropolis-Hastings sampler