Economic Incentives and Social Security Disability Entitlements in a Counting Process Model[*]

Lakshmi K. Raut
Social Security Administration
500 E. Street, SW, #833
Washington, DC 20254

mailto:Lakshmi.Raut@ssa.gov

May 29, 2013

Abstract

This paper estimates the incentive effects of replacement rate (measured as the ratio of disability benefits to an estimate of potential earnings), non-employment and disability policies on the hazard rate of disability entitlements. The incentives are formulated in a dynamic programming model within a counting process framework and estimated using the one percent Continuous Work History Sample (CWHS) of the Administrative data on around 3 million individuals. The paper examines the argument that the widening income gap between the rich and the poor in the eighties created strong incentives for workers to get onto the rolls by raising the replacement rates of the poor. The paper finds that while the replacement rate of the poor grew rapidly in the eighties, its effect on enrollment was not strong. The paper also shows that the effect of non-employment on the likelihood of one's getting onto the program depends on the health status of the worker: a healthier worker (as inferred from the proxy variable for health constructed in this paper using work history) has a much lower probability of getting onto the program than a worker with adverse health shocks. Among various disability policies, the addition of Medicare benefits contributed most to the growth of disability entitlements; the mental listings and broadening of coverage to musculoskeletal and mental disorders have significant positive effect and the D&A and Welfare reforms have significant negative effect in explaining the pattern of disability entitlements in the 80's and 90's.

JEL Classifications: I12, C41, C51.
Keywords: Pathways to disability, OASDI, multistate duration model, aging.

[*]I am grateful to Robert V. Gesumaria, Michael V Leonesio, Scott Muller, David Pattison, Kalman Rupp and Bernard Wixon for many comments or discussions at various stages of this paper.

1 Introduction

Since 1980, the number of people on the disability rolls has been sharply increasing. Policy debates on these issues surround three causes of this growth. (1) Because of growing income disparities, the replacement rate (defined as the ratio of disability benefits to recent earnings) has been growing for the poorer workers. Some argue that this gave them strong incentives to get onto the rolls. (2) Unemployment rates in the eighties were also growing, which created strong incentives for the unemployed workers to get onto the disability program, especially because the disability beneficiaries become eligible for Medicare benefits after being in the program for 24 months whereas unemployment insurance does not provide health insurance and benefits are for very short time. (3) Policies were introduced in the mid eighties that eased the program eligibility and added new categories of disabilities such as musculoskeletal and mental disorders to be eligible for benefits.

In this paper, I use the Social Security Administration's 1 percent random sample of continuous work history survey (CWHS) data to address the importance of these factors for disability entitlements. I compute the replacement rate for all individuals using the benefit formula under the current law I formulate a dynamic programming model of disability entitlements within a Cox regression framework to estimate the individual incentive effects of replacement rate, non-employment, and changes in the policy rules on the likelihood of one's getting onto the disability rolls. Unobserved health shocks are correlated with some of the regressors and thus the parameter estimates of those regressors will be biased. An innovation of this paper is that past work history is used to create a proxy for unobserved health shocks to circumvent this problem and to study if workers with good health has negligible probability of getting onto the program.

The previous studies addressed these issues mostly using aggregate data on disability application or award. I argue that the individual data on disability entitlement as more appropriate to study the causality effects of the economic variables. For, when an individual's economic incentives change, while his decision to apply for disability is contemporaneous, his date of award could be any time in the future, depending on if the time it took to get the award if he was initially denied and then decided to appeal. On the other hand, entitlement date for the award is retroactive and is close to the onset of the disability. The replacement rate is sometimes defined at the Social Security Administration as the ratio of a worker's initial monthly benefit to the worker's average indexed monthly earnings (AIME), henceforth will be referred to as R^{aime}. The average to calculate AIME is taken over up to 35 years of maximum indexed earnings, the actual number of years in the calculation depends on his age. The DI program tries to achieve intergenerational equity and within generation redistribution from the rich to the poor by using a piece-wise linear concave benefit formula whose bend points are adjusted upward over time to keep up with the growth in the average wages of the economy. Because of concavity of the benefit formula, if a worker becomes relatively poorer, as for instance due to growing earnings gap, the argument of the concave function decreases. The replacement rate R^{aime} will be higher, at least for individuals who are slightly above the bend points.

As a measure of incentive to quit working and getting onto the disability rolls, however, it is more relevant to define the replacement rate R by using in the denominator the average potential earnings of the worker. This is because while making the decision whether he wants to get onto the rolls, he compares his wellbeing from two alternatives: one, if he gets onto the rolls, he gets disability benefits, Medicare and leisure. Two, if he works, he gets his potential average future earnings that he expects in the labor market. Potential earnings are difficult to measure. Therefore, like many previous studies, I use the average of recent earnings in the denominator of the replacement rate R in my calculation. Because of the growing earnings disparities, the average earnings is growing faster than the average recent earnings of the poor, and thus the replacement rate R is growing even more than the growth in R^{aime}. Autor and Duggan, 2003; Autor and Duggan, 2006 have used the county-level cross section time series data to show that the replacement rate R has a significant positive effect on the rate of disability application. They attributed the surges of the disability rolls in the eighties to growth in bend points relative to applicant earnings and to policy changes in the eighties relaxing the determination process and expanding the impairment categories. They did not, however, examine any alternative policy remedying the pitfalls of the upward adjustments of the bend points in the benefit formula. Muller, 2008 used the 1 percent CWHS data to calculate the replacement rates for individuals and confirmed the rising replacement rates. He proposed an alternative benefit formula based on the median wage index, instead of the average wage index, and showed that it did not produce the sharper rise in the replacement rate. He also computed the effect of his proposed policy on benefit payments. He did not consider any behavioral econometric model to estimate the incentive effect of his suggested policy on the rate of disability entitlements.

The rising unemployment rate in the eighties might have led to a higher rate of disability entitlements. When a worker loses his job, he loses his employer-provided health insurance. For a worker with disability, it is very expensive to buy new health insurance in the private health insurance market. If he gets onto the DI program, and remains there for two years, however, he becomes entitled to Medicare, which provides public health insurance. Thus, during persistent high unemployment, an unemployed worker with borderline disabilities may find the prospect of finding a job with health insurance quite bleak and thus have a higher incentive to enroll onto the disability program.

Rupp and Stapleton, 1995 used the state-level pooled time series cross section data for the period 1980-1993 to estimate the effect unemployment rate on disability applications and awards. They found a strong contemporaneous positive effect of unemployment on disability applications, and a strong positive two-year-lagged effect on the awards. They also compiled previous studies on this issue and found their estimates are comparable to others. Lando et al., 1979 used the district-level quarterly cross section time series data for the period 1964-78. They defined the replacement rate using the aggregate benefits and aggregate current earnings at the district level. They found both the unemployment rate and the replacement rate have a strong positive effect on disability applications. Autor and Duggan, 2003 also found in their aggregate analysis a significant positive effect of unemployment rate on applications.

The effect of unemployment on disability enrollment is very tricky to analyze. Two most relevant policy issues in this context are: **First,** whether there is a lot of malingering of the type that when a worker gets unemployed and does not have serious or any disability may use it as a long-term unemployment insurance program because of its Medicare provision and generous replacement rate. If one could observe the disability health shocks that individuals encounter, one could then estimate if individuals without disability shocks have much lower or negligible probability of getting onto the program than those with disability shocks. Only individuals have information about his health shocks and some survey data have information on individual health status. But neither other aggregate datasets nor the SSA's administrative data on individuals have this information. In such econometric situations, one generally applies proxy variable to control for such an unobserved variable. I use the employment history of each worker over the past few years as a proxy variable to throw lights on his health status.

Second, whether the generosity of the disability program is causing workers with milder disabilities to quit work and enroll in the program, i.e., whether the existence of disability program has caused higher unemployment. Autor and Duggan, 2003 analyzed this using county level aggregate data and found some evidence for it. Individual level analysis using survey data is more appropriate but the administrative data cannot throw any light on this issue and thus not addressed in this paper.

Lahiri et al., 2008 used the SIPP data matched with the Social Security Administration data to get information on disability health status and the individual probability of getting approved for DI entitlement and aggregate unemployment rate. They used a four-equation Tobit model with correlated errors in the equations for DI application, medical determination, earnings for non-applicant and earnings for denied applicants. They estimated the effect of aggregate unemployment rate and the benefits relative to expected earnings given that DI entitlement could be denied with positive probability. Their study used self-reported health status to control for evolving health conditions and an estimate of the opportunity cost of applying for DI entitlements adjusting for an estimated likelihood of being denied and imputing the value of Medicare and health insurance in case one is entitled to DI. Their focus was not, however, on studying the individual incentive effects of employment shocks and replacement rates on disability entitlements.

Section 2 summarizes the pattern of the first-time disability entitlement that emerges from the dataset. Section 3 formulates the counting process model of disability entitlement and discusses the estimation procedure. Section 4 describes the dataset and the variables used in the econometric estimation of the counting process model. Section 5 reports the estimates. Section 6 concludes the paper.

2 The Replacement Rate, Unemployment Rate and Major Policy Changes over Time

When a worker applies for disability, the worker is either awarded benefits initially, denied initially but awarded on appeal, or denied initially without reversal on appeal. If an initially denied applicant wins an award later, the date of entitlement is retroactive and hence is close to the time of the disability onset. Since the focus of this paper is the effects of changes in DI policy and in individual economic incentives on the number of people getting onto the DI rolls, the entitlement data instead of award data is more appropriate to analyze. Figure 2 shows the plots of these two series. In spite of the conceptual differences, both series show similar pattern. For instance, the figure shows that around an upward trend, both series rapidly grew in the late sixties and early seventies and then dropped in the late seventies and again started to grow in the eighties to 1992. In Figure 1, I also mark the periods with labels of major DI policy changes that took place.

What explains such surges and the time trend in the first time DI entitlements?

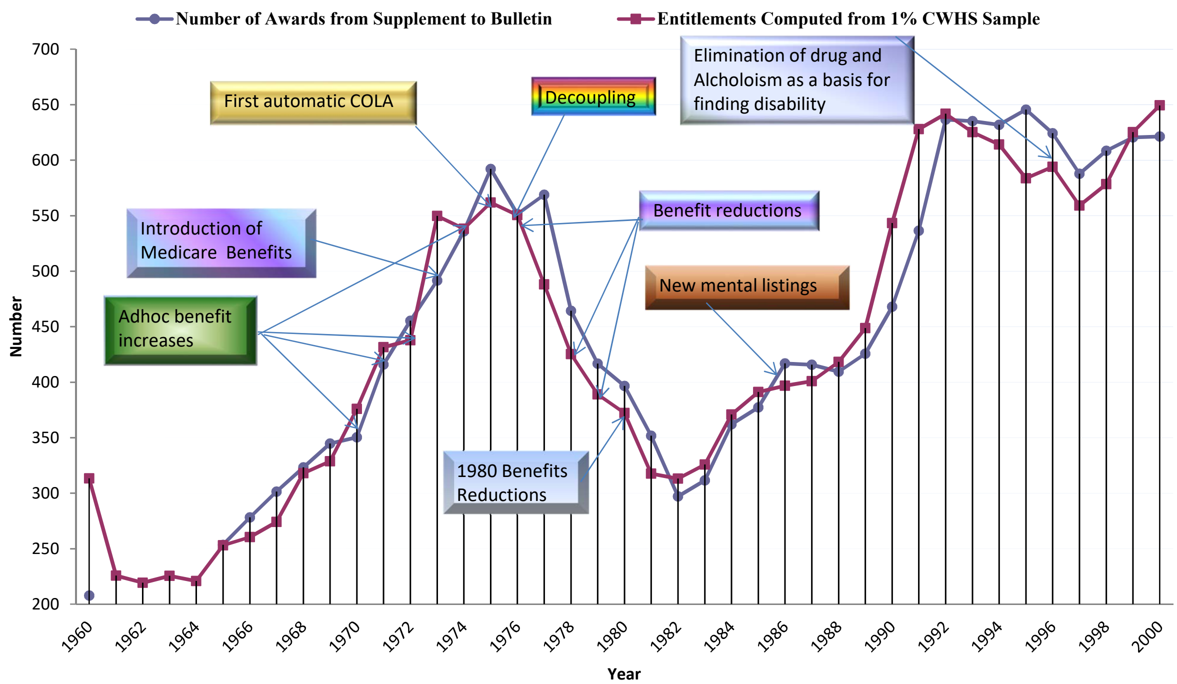


Figure 1: Comparison of published disability awards data with entitlements tabulated from the CWHS sample.
Source: Estimates of the number of Awards come from the Social Secruity Administration, 2005, Table6.C7, calculated using 100% sample of the Master Beneficiary Data.
The estimates of the number of entitlements are based on author's calculation using the 1 % CWHS Sample.

There could be many possible reasons. The hazard rate of the first time disability entitlement might be high for certain age groups, or for certain race and gender. The demographic shocks such as baby boom fertility shocks can lead to a surge of the high-risk population in the 1970s and the 1990s. Another important source of these surges could be that there were substantial changes to disability policies in the 1970's and 1980's. I examine these sources next.

At the aggregate level, the size of the first time disability enrollment N_t at time t can be computed as $N_t = h_t \cdot P_t$, where h_t is the hazard rate of a disability insured worker getting onto the disability rolls at time t, and P_t is the size of the DI insured population at time t. From this, it is clear that N_t can change due to a change in P_t which in turn mostly due to changes in the demographics of the labor force. Or it can change due to a change in the hazard rate h_t, which is mostly due to behavioral changes in a particular period t.

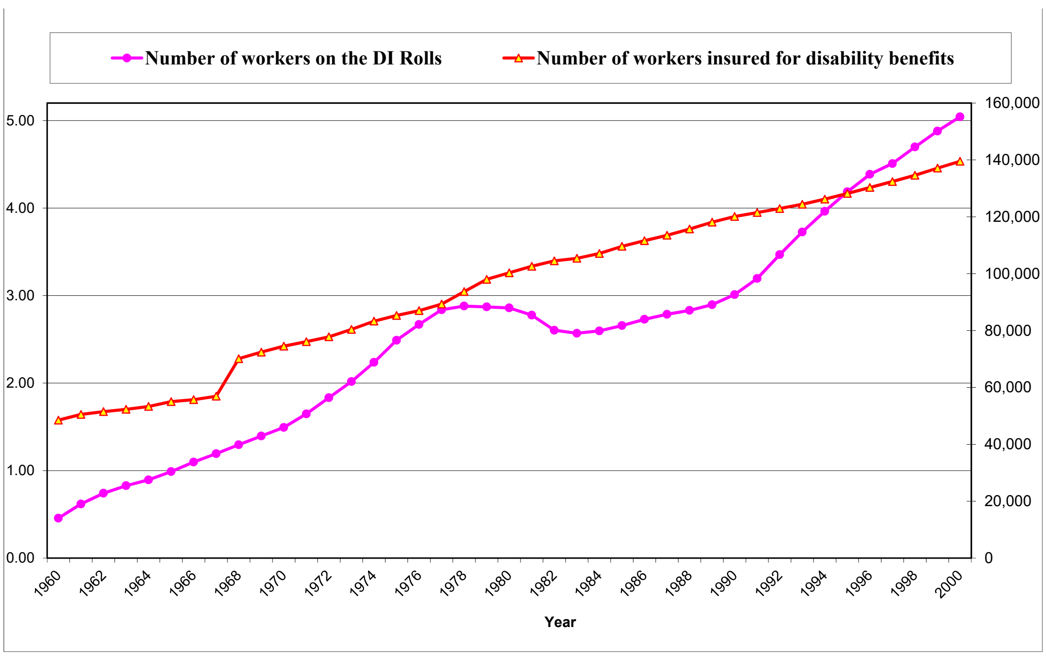


Figure 2: Number of workers (in thousands) disability insured (right y-axis), and on the disability rolls (left y-axis)
Source: The data on the number of workers insured for disability benefits came from Social Secruity Administration, 2007, Table C.1, The data on the number on the disability rolls came from Social Security Administration, 2009, Table 3.

Figure 2 depicts the number of disability insured workers, and the number of workers receiving DI benefits over the 1960-2001 period. It shows that the disability insured population P_t is growing at a constant rate over time - the growth rate is pretty much at the same rate as the number of enrollments in Figure 1. Moreover, it does not show fluctuations of the nature seen in Figure 2 The fluctuations in Figure 1 must be then due to the component h_t which is influenced by economic incentives arising from disability policy changes and economic circumstances of the individual, as well as changes in the age structure of the disability insured population.

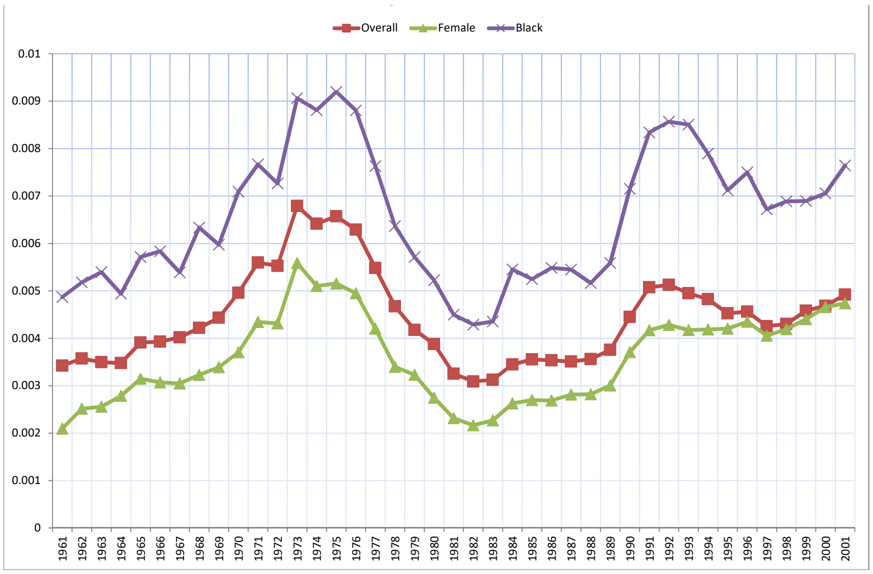


Figure 3: Incidence Rate of First Time Disability Entitlement among disability Insured workers.
Source: Author's calculation using the 1 percent CWHS sample.

The aggregate hazard rate h_t in period t is the aggregation of hazard rates by age, sex and race. Figure 2 shows that both sex and race components have the same pattern of fluctuations as the aggregate hazard rate. This is also true for age, for which the plots are not reported. Furthermore, females have lower hazard rates and blacks have higher hazard rates for disability entitlements. This still does not rule out that the pattern arises from demographic changes. To see what might be at work, I further disaggregate it as the sum over different age groups at time t as follows:

$$h_t = \sum_a h_t(a) \cdot \pi_t(a)$$

Where $h_t(a)$ is the hazard rate of disability entitlement at age a of an insured worker in period t, $\eth_t(a)$ is the proportion of the disability insured population of age a in period t. The formula is the same if we take a to denote more generally any fixed characteristics of a worker such as his gender and race even for each age group.

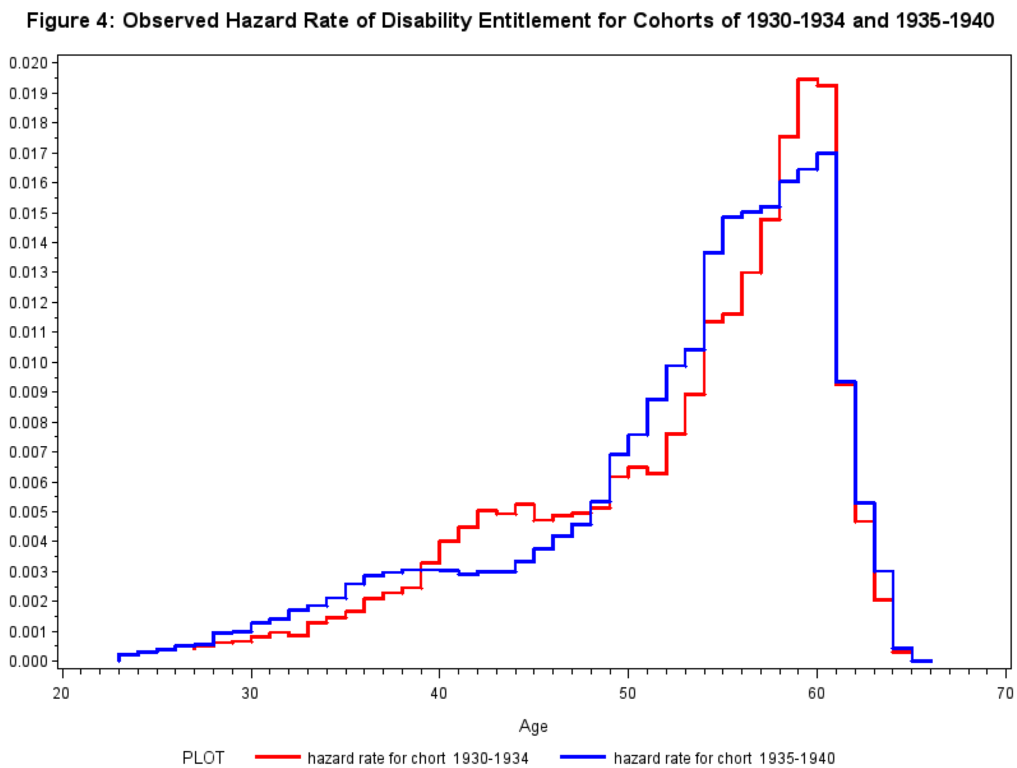


Figure 4: Observed hazard rate of disability entitlements for cohorts of 1930-1934 and 1935-1940.
Source: Produced using the 1 percent CWHS sample.

Figure 2 shows the observed h_t(a) over the full life-cycle of disability entitlement for two cohorts born between 1930-1934 and between 1935-1940. The cohorts born later than 1940 would be younger than age 65 and thus would be censored, and the hazard rates over the whole life-cycle could not be calculated for those. Figure 4 shows that the age specific disability entitlement rate rises slowly until around age 50 and then rapidly grows until age 59 for the first cohort and age 62 for the second cohort. There is no biomedical reason why suddenly the hazard rate increases rapidly around age 50. Nor are there good grounds to expect that the economic incentives suddenly become strong around at age 50. A possible reason is the "vocational grid" as it is known at the Social Security Administration. This pertains to the ease of approval of an application by the Disability Determination Service if the worker is lower educated and too old to find a significantly gainful activity. Alternatively, it could be that when they were in their fifties, the disability policies were liberalized and became more generous (see the description of policy evolution that follows), inducing more people to get onto the program. Whatever might be the reason, the figure shows that workers in their fifties have significantly high probability of getting onto the disability rolls. This figure cannot however explain the swings in the aggregate hazard rate of disability entitlements, especially because in any period t, these two cohorts just constitute a few of the age-groups, the remaining age groups consist of younger cohorts.

To explore further the source of those swings, note that even when $h_t(a)$ is constant over time, a demographic shock may cause a high risk age group population share $\eth_t(a)$ to go up in period t and in the next period the same will be the case for $\eth_{t+1}(a+1)$ and this can generate an upward swing. From Figure 2, however, we see that except for in the 1967 the growth in the insured population has been stable. The aggregate fluctuations in age-specific rates, on the other hand, have the same pattern over time (Figure 2). The likely reason for age specific rates to move in parallel are either common economic changes like recessions that affect all ages simultaneously or changes in disability determination policy that apply over many ages. The next section will examine the first of these, the economic factors.

2.1 The Replacement Rate

As pointed out in the introduction, to capture the incentives effect of the disability program, an ideal measure of replacement rate would be to take the ratio of benefits over potential stream of future earnings that he expects in the labor market. There is, however, no good estimate of the potential stream of future earnings, especially using the administrative data.

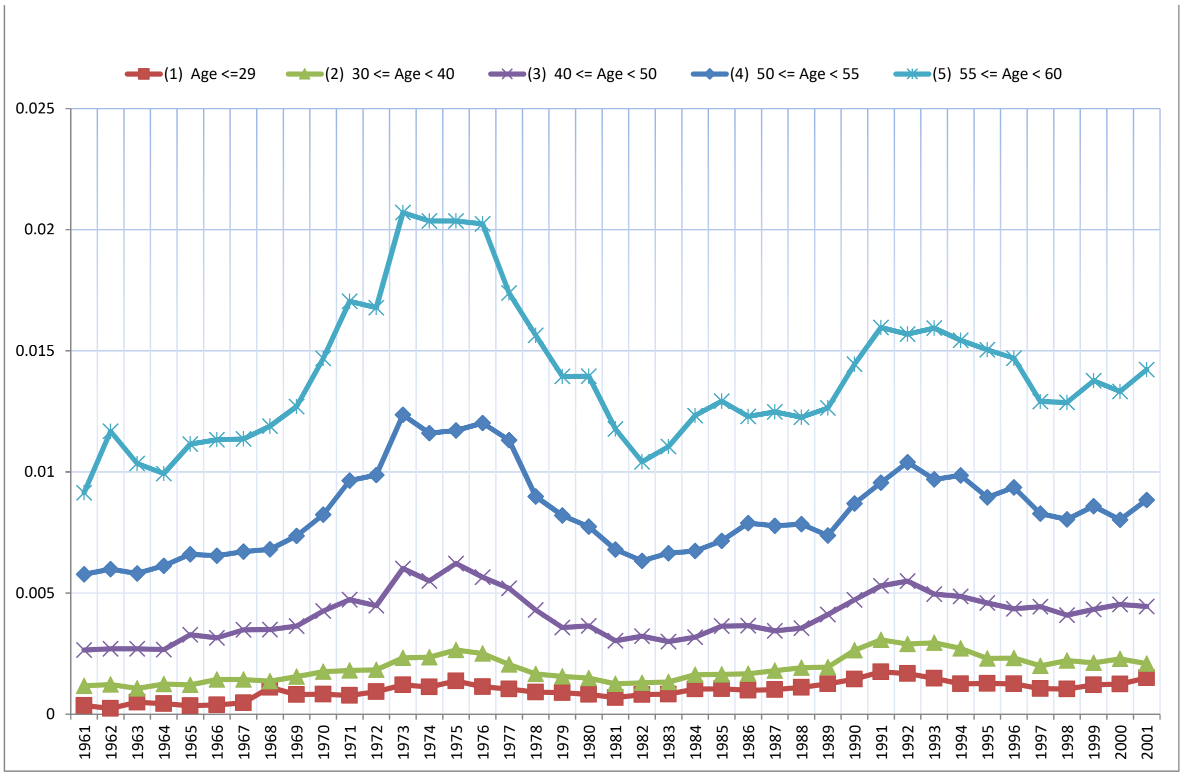


Figure 5: Incidence Rate of First Time Disability Entitlement among DI Insured workers by Age Group.
Source: Author's calculation using the 1 percent CWHS sample.

I use the average of the past five years' indexed positive earnings[1], which I denote by $\bar{w}_t$. In the numerator of the replacement rate I use the benefits that a worker is entitled to under the current law.

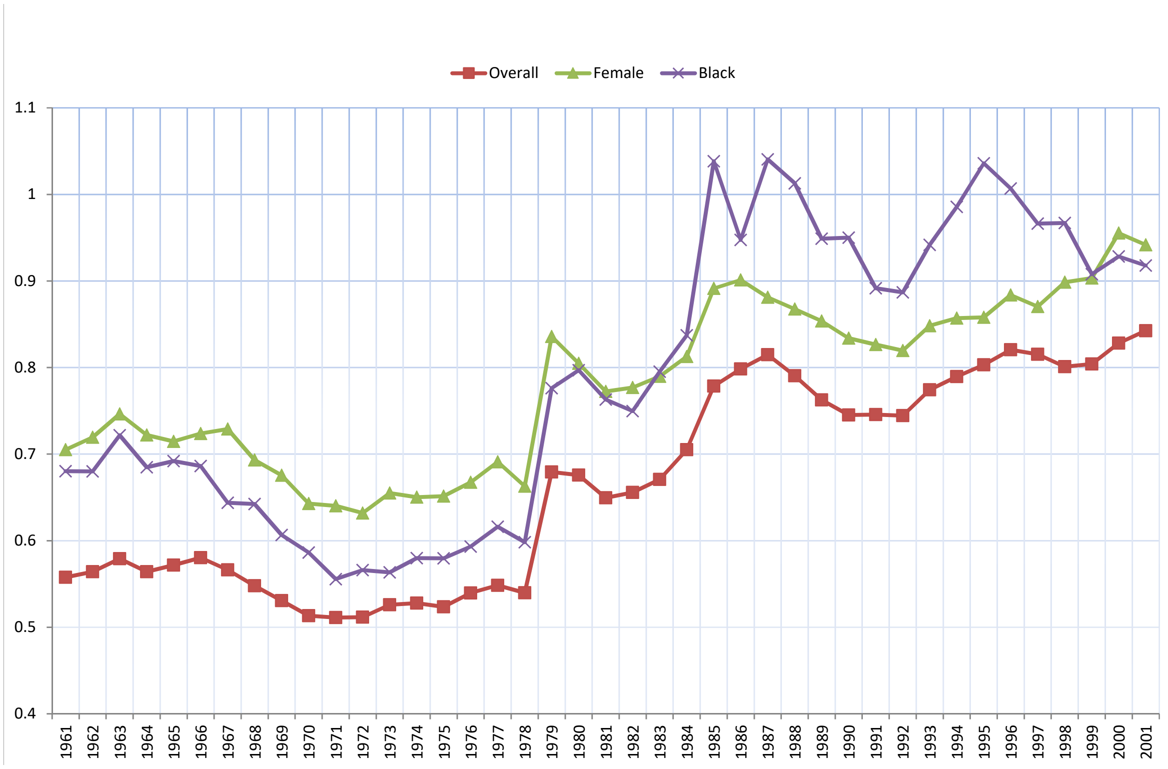


Figure 6: Average Replacement Rate of the Insured Population if became disabled for the first time in the year.
Source: Author's calculation using the 1 percent CWHS sample.

Figure 2.1 shows the replacement rate R_t for the black, female and overall population. All three rates are increasing since 1979; moreover, blacks have higher replacement rates than non-blacks, and females have higher replacement rates than males.

Why has the replacement rate been increasing since 1980? Could that have generated significantly strong incentives for workers to get onto the program and thus explaining the growth in the disability entitlement rate and the number of disability entitlements since 1980?

[1] I also used two other measures of recent earnings, giving two alternative measures of replacement rate: (1) the last year's earnings, and (2) the average of the past five years' indexed earnings including zero earnings. The time trend and the parameter estimates of the econometric model presented later are very similar with respect to each of these three measures of replacement rates.

To address the second question, one needs a behavioral econometric model of disability entitlement, which I consider in the next section. I examine the first issue in details here.

Since 1979, the SSA calculates disability benefits using what is known as bend points and the AIME (Average Indexed Monthly Earnings). The average to compute AIME is taken up to highest 35 years of indexed maximum taxable earnings, the number of years in the computation of average depends on his age. Let x_t be the AIME in year t, with 1979 as the base period and thus year 0 in our description. The disability benefits y_t given his AIME x_t is calculated by the Social Security Administration as $y_t = g_t(x_t)$, defined with two bend points b_t^1 and b_t^2,$b_t^1 < b_t^2$ in each year as follows:

$$g_t\left(x_t\right) = \begin{cases} 0.90x_t & if \quad 0 < x_t \leq b_t^1 \\ 0.90b_t^1 + 0.32(x_t - b_t^1) & if \quad b_t^1 < x_t \leq b_t^2 \\ 0.90b_t^1 + 0.32(b_t^2 - b_t^1) + 0.15\left(x_t - b_t^2\right) & \quad otherwise, \end{cases}$$

where the base period values $b_0^1 = 180$ and $b_0^2 = 1085$; and b_t^1 and b_t^2 are increasing over time at the same rate as the growth rate of the average wage index lagged two years.

Internally at the Social Security Administration, the replacement rate is defined as the benefits to AIME ratio, which I denote by R_t^{aime} in period t. The replacement rate R_t that I use in this paper to capture incentive effect is then given by

$$R_t = R_t^{aime} \cdot \xi_t, \text{ where } R_t^{aime} = g_t(x_t)/x_t, \text{ and } \xi_t = x_t/\bar{w}_t.$$

Note that there are two sources of growth in the replacement rate R_{t-1} from the replacement rate R_t^{aime} and from the factor ξ_t. With the income gap increasing between the rich and the poor, the poor workers in the eighties became relatively poorer and hence their replacement rate is growing faster than the richer people whose replacement rate is dropping but at a slower rate.

Figure 2.1 shows the replacement rates in the period after 1979 for three earnings groups: group 1 with average earnings over the past five years less than the median value, group 2 with average earnings over the past five years between the median and the 95$^{\text{th}}$ percentile,, and group 3 is with average past five years earnings above the 95$^{\text{th}}$ percentile. It appears that for group 1, the replacement rate is indeed growing faster than the other two groups. For group 2, it is more or less constant, but for group 3, it is slightly declining.

Autor and Duggan, 2003 have argued that the rise in the replacement rate in the eighties due to growing wage gap between the rich and the poor has created strong incentives to get onto the program. I will address this incentive effect after estimating the econometric model.

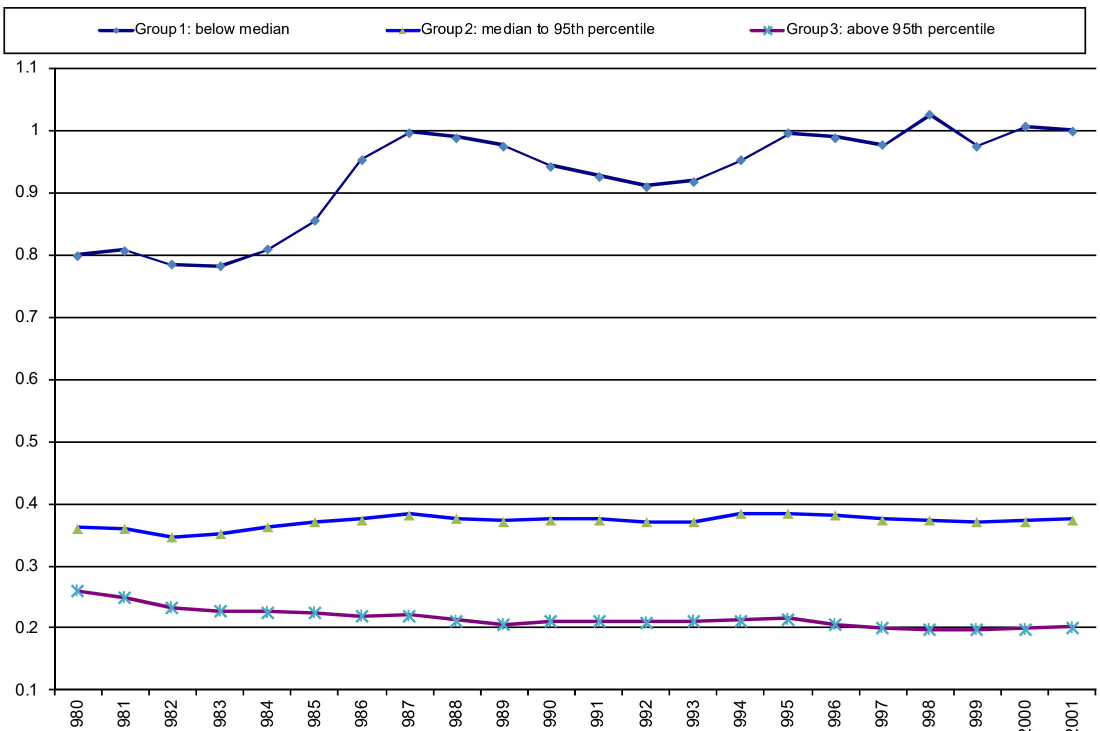


Figure 7: Average replacement Rate of three income groups.
Source: Author's calculation using the 1 percent CWHS sample.

2.2 Unemployment

A worker is non-employed if he/she has zero earnings during the period. The difference between non-employment and unemployment is that an unemployed worker is still looking for work. The administrative data do not have that information about a worker. So the focus in this paper is on non-employment. Figure 2.2, left y-axis, depicts the average non-employment rate over time for the black, female and the overall populations, and the right y-axis plots the macro unemployment rate, published by the Bureau of Labor Statistics. It is clear in the figure that all four series co-move over time in a similar way except for females whose labor force participation has increased over time. While in the eighties we see the same type of movements of the non-employment as the disability entitlement rates in Figure 2, this co-movements are not always true in earlier period.

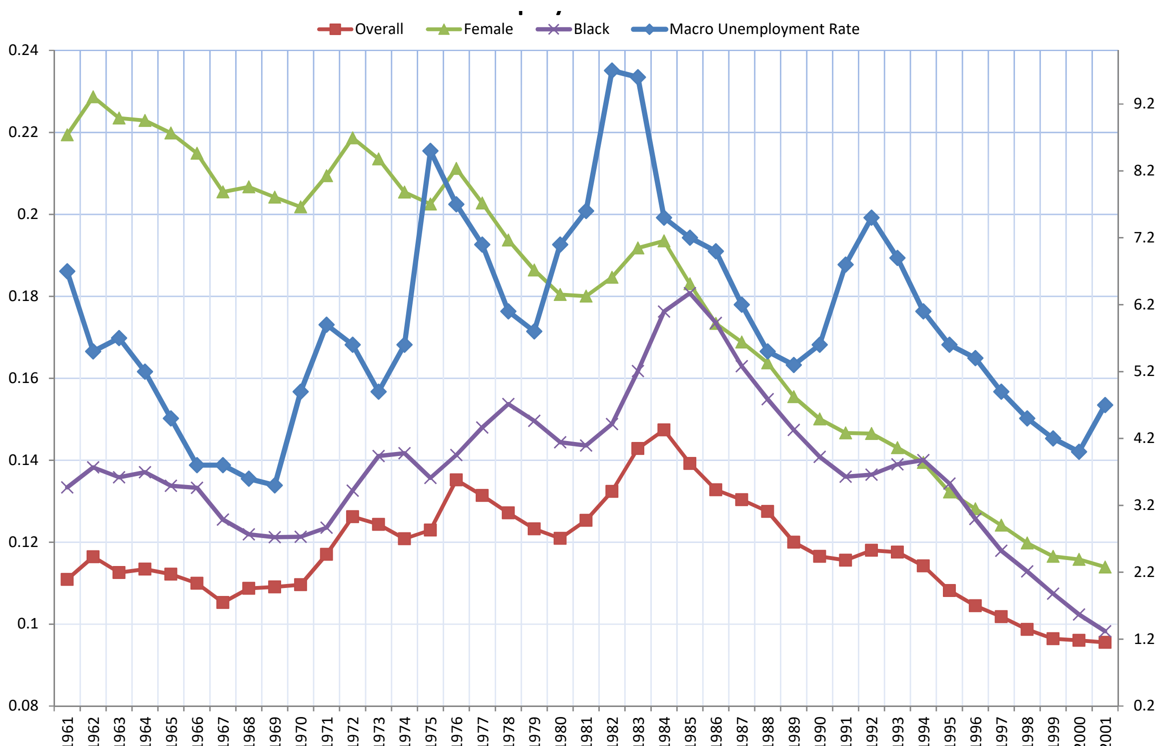


Figure 8: Average replacement Rate of three income groups.
Source: The aggregate unemployment rate comes from Bureau of Labor Statistics, Table A-1, the other series are author's calculation using the 1 percent CWHS sample.

2.3 Major DI Policy Changes

Even though the amendments of the 1954 Social Security Policy Act initiated the Disability Insurance (DI) program, it was not until 1960 the Congress extended the benefits to workers of age less than 50 and to widows and widowers. For this reason, I do the analysis restricting to the period after 1960.

The major policy changes are summarized from the SSA Publication No. 13-11831 below and are juxtaposed in Figure 1 for easier visualization.

In the early 1970s, the congress made many changes to the levels of disability benefits such as ad hoc, across-the-board benefit increases of 15 percent in 1970, 10 percent in 1971, 20 percent in 1972, and 11 percent in 1974. In June 1975, an automatic cost - of - living adjustments (COLAs) was introduced. Starting in 1973, Medicare benefits was added to the disability program. The increasing benefits (higher cash benefits and Medicare) probably contributed to the growth of the disability program in the early to mid-1970s (see Figure 1).

In response to the high growth in the disability rolls during the early to mid-1970s, the Congress legislated policies to reduce benefits. For instance, in 1977 the Congress "decoupled" the cost of living adjustment by making the earnings in the benefit formula to be adjusted with the average wage index and adjust the benefits amount to adjust with the consumer price index (CPI). In the 1980 Amendments, the Congress applied additional benefit reductions by capping the family benefit amount, reducing the number of dropout years in the benefit calculation and introducing more stringent reviews of the applications. In 1981, the Congress dropped the Social Security minimum benefit, imposing a cap on the replacement rate from all public disability program benefits. These policies persisted in the post-1980 period and are labeled in Figure 2 as 1980 Benefit Reductions.

The Congress introduced the 1984 Amendments mandating that “SSA develop new disability standards for individuals with mental disorders, evaluate pain as part of the decision process, place emphasis on evidence from treating physicians in the decision process, and consider the impact of multiple non-severe impairments in determining disability”. New mental listings put in place in 1986 seem to have led to continuing increases in the number of awards for mental disorders other than mental retardation, initially from the 1984 backlog and then continuing for several more years.

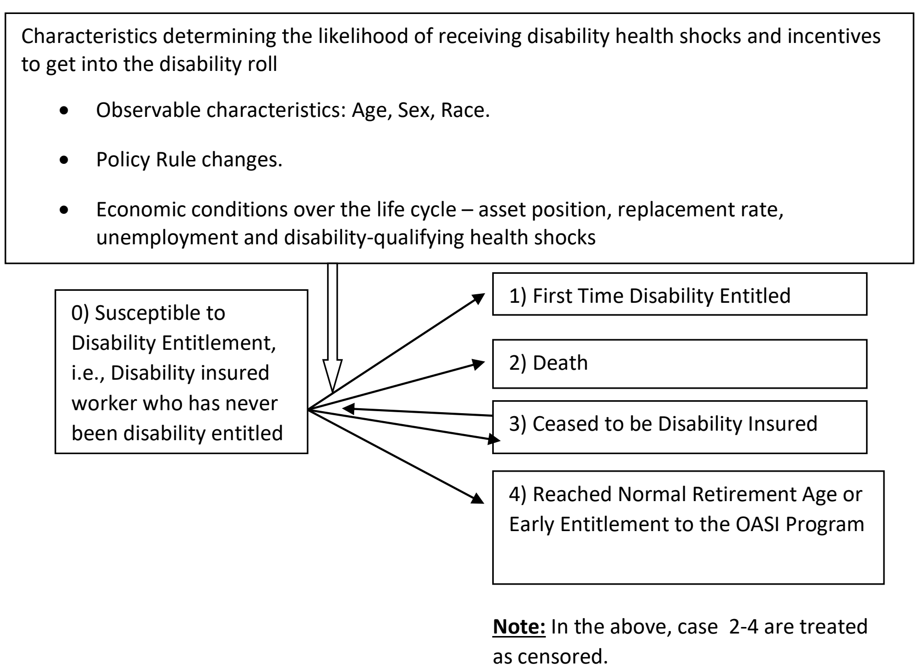
In 1996, the welfare reform legislation mandated a stricter definition of disability for children, and for exclusion of non-citizens from disability benefits. Furthermore, the same year the Congress eliminated drug addiction and alcoholism (DA&A) as the basis for a finding of disability.

3 A Counting Process Model of Disability Entitlement

I formulate a competing relative risk Cox hazard model of a worker's first disability entitlement and use the Cox partial likelihood method to statistically estimate the model.

Entitlement to DI program requires a worker to satisfy a few conditions. First, the worker's disability should meet medical qualification for the severity of the disability. Second, the worker also satisfies the condition that the worker cannot have engaged into significant gainful activities (SGA) after the disability onset. Third, the worker must have acquired disability-insured status by working a minimum required amount of time which depends on his age at disability onset. A worker ceases to be susceptible to disability entitlement due to one of the following many causes: (1) Entitlement to the DI program, (2) Death; (3) Ceased to be DI Insured, (4) Reached Normal Retirement Age or Early Entitlement to the OASI Program See the schematic representation in Chart 1. Our interest is in the cause (1).

Chart 1 : A Schematic Representation of the Disability Entitlement Model



Characteristics determining the likelihood of receiving disability health shocks and incentives to get into the disability roll

• Observable characteristics: Age, Sex, Race.

• Policy Rule changes.

• Economic conditions over the life cycle - asset position, replacement rate, unemployment and disability-qualifying health shocks η.

Unobserved heterogeneity - disability related health status, and personal attributes controlling health related activities such as smoking, exercising, healthy eating habits and all other characteristics not included above which affect disability health status.

I sketch this decision problem in a discrete time dynamic programming framework. Let S_t denote the set of possible status that a worker at the beginning of time t can be in - he is disability insured and not entitled to disability before, dead, ceased to be disability insured, already in the disability program or reached the early retirement age. At the beginning of period t, the worker knows the current as well as all the past values of the replacement rate R_t, his unemployment shocks u_t, the outlook of the overall economic conditions represented by the aggregate unemployment rate U_t, his disability-qualifying health shocks H_t, and his general asset position A_t. His taste for leisure, his nature of healthy living style, and his general morbidity to get disability type of health shocks are denoted by a vector of parameters ç, which I assume to be constant over time. Given the health shocks he received and given the policies and practices for disability entitlements, he assesses his probability of getting entitled to disability program to be $ð_t$. The choice d_t that he can make depends on the values of S_t, H_t,and u_t and denoted by the set $D(S_t, H_t, u_t)$, the set of those choices. For instance, if he is disability insured and got the disability shocks that he will qualify for entitlement, he can apply and get onto the program if he is already not working; else if he is working he can first quit his job and then apply for disability benefits. If he is not approved, which is a possibility with probability (1- $ð_t$), he loses earnings during the period if he quit his job or did not try to find a job to qualify for disability benefits. His indirect utility during period t thus depends on A_t, d_t and $ð_t$. It is denoted by $u(A_t, d_t\ ð_t)$. Denote the state variable of the dynamic programming by $z_t = (S_t, A_t, R_t, \{u_t\}, \{U_t\}, \{H_t\}, \pi_t)$. The dynamic programming model of his choice problem is defined as follows.

$$V(z_t) = \max_{d_t \in D(S_t, H_t, u_t)} u\left(A_t, d_t, \pi_t\right) + \beta \int V(z_{t+1}) p\left(z_{t+1} | z_t, d_t\right).$$

Under general conditions, there exists a solution to the above problem, and takes the form

$$d_t = \Psi\left(S_t, A_t, R_t, \{u_t\}, \{U_t\}, \{H_t\}, \pi_t, \eta\right).$$

It is known that the above dynamic programming problem cannot be solved analytically for the function Ψ() but generally used to guide the specification of econometric models. Even though d_t is a deterministic function to the individual at time t, it also depends on the realization of the random variable z_t and ç which are known to the individual but most components of these are unknown to us. I treat d_t as a random variable for each t, i.e., as a stochastic process, and treat time as continuous.

Let $N_i(t)$ be a counting-process random variable for each worker i, which takes value 1 at time t if he gets onto the disability rolls for the first time at t, otherwise it takes value 0. At the beginning of his working life (t=0), $N_i(0) = 0$ for all i.

Denote the vector valued process $\vec{X}_i(t) \equiv X_i(s), 0 \leq s \leq t$, where each $X_i(s)$ is a vector of the socio-economic and health characteristics of individual i at age s. The Doob-Myer decomposition theorem asserts that under some general conditions $N_i(t)$ can be decomposed uniquely into two components as follows:

$$N_i(t) = E_i(t) + M_i(t),$$

where $E_i(t)$ is a function of the ***intensity process*** $\lambda_i(t; \vec{X}_i(t), \eta_i)$ (also known as the ***hazard rate of failure*** in the duration analysis literature**)** of the counting process $N_i(t)$, and $M_i(t)$ is a mean zero Martingale process. Drawing a similarity with the regression analysis, the above can be stated also as

Data = Model + Error

The biomedical process of disability health shocks coupled with an economic decision as to apply and get onto the DI rolls together determines the $E_i(t)$ process. It is not possible in general to derive a closed-form solution of the dynamic programming a model. I follow the reduced form approach of parameterizing the intensity process directly. In this paper, I further assume the absence of the unobserved heterogeneity variable η in the specification of hazard functions, and restrict to the following relative risk specification, also known as the **Cox regression model** for the hazard function of exit due to the event of interest, disability enrollment:

$$\lambda_i(t; \vec{X}_i(t), \eta_i) = \lambda_0(t) \cdot \exp^{X_i(t)\beta},$$

where $\exp^{X_i(t)\beta}$ measures how the history of individual characteristics $\vec{X}_i(t)$ of individual i affects the natural hazard rate either through (unobserved) individual choices or through some other natural phenomena. The term $\lambda_0(t)$ captures any duration dependence (i.e., the dependence of the probability of exit at any time t on the time spent being susceptible to disability entitlement). This term is known in the literature as baseline hazard function of getting out of the susceptible state due to disability entitlement. The baseline hazard function could also be interpreted as the natural hazard rate that we mentioned above, however, it is normalized up to a multiplicative constant, calibrating it to the hazard rate of the population with all characteristics X's taking value 0. In this framework, the natural hazard rate of first time disability entitlement can be identified only up to a scale like this. I use the Statistical Analysis System to estimate the model.

4 The Dataset and Variables

I use the one percent CWHS (Continuous Work History Sample) dataset of 2003 for this analysis. It contains information on 3,146,138 individuals covering birth cohorts of 1951 to 2003. The CWHS was created by merging variables from other administrative datasets containing longitudinal information on individual's earnings, selected claims and benefits information on OASDI related information. I use the active CWHS data file, which provides information on individual workers with some history of non-zero covered earnings. Prior to 1978, CWHS had information only on covered workers. So to be consistent throughout the period of our analysis, I restricted to covered workers.

Table 1 provides the definition of the variables that are used in this paper.

5 Empirical Findings

I estimate the incentive effect of replacement rate R and non-employment at the individual level, controlling for macroeconomic conditions and policy variables. In the first model, I use last year's non-employment dummy variable U1, and replacement rate variable, rrate2_1 (this is R in our earlier notation). The rest of the variables are common across all specifications, see the column with heading Model 1a in Table 2 for their listing.

Parameter estimates for both rrate2_1 and U1 are positive and statistically significant. An estimate of $\hat{a} = 0.4624$ for U1 would mean that if a worker does not work, his probability of getting onto the rolls is 46 percent higher than if he works. This may not give much useful policy inference. As I mentioned in the introduction, what is important is to get an estimate of how non-employment affects the probability of getting onto the program if one has disability-qualifying health shocks as compared to one who does not have such health shocks. For instance, if the disability determination process does a good job of screening the real disability case from fake ones, one would expect that a worker who has not encountered any disability-qualifying health shocks has a negligible probability of getting onto the rolls as compared to one who has such a disability shock.

Table 1: Definition of variables used in the econometric estimation of the hazard model.

Variables	Definition
DOB_Y	Year of Birth
AgUemp_1	Aggregate unemployement last year
AgUemp_2	Aggregate unemployement two year lag
DplusU	One year forward change in aggregate unemployment last year
Black	Dummy Variable taking the value 1 if one's race is black, else it is zero.
Female	Dummy Variable taking the value 1 if one's gender is female, else it is zero
rrate2_1	one year lag of the replacement rate defined as benefits over average of last five years indexed non-zero earnings.
U1, U2, U3, and their products	Dummy variable taking the value 1 if the individual was not working last year, else taking value 0. U2 is one year lag of U1 and U3 is two year lag of U1 and * denotes the multiplication
DY1970, DY1971, DY1972, DY1974	Policy dummy variables for the indicated years during which there were one time adhoc benefit increases.
DY1976, DY1979	Policy dummy variables for the indicated years during which there were one time adhoc benefit reductions.
BenRed_80	Policy time varying dummy variable taking value 0 before 1980 and taking value 1 during and after 1980, indicating the permanent benefit reductions that were introduce in 1980.
Medicare_73	Policy time varying dummy variable taking value 0 before 1973 and taking value 1 during and after 1973, indicating the permanent inclusion of medicare benefits to disability benefits.
Decoupled_77	Time varying dummy variable representing the decoupling of cos-of-living adjustments in the benefits formula which takes value 0 prior to 1977 and value 1 there after.
Mental_86	Time varying dummy variable representing the new mental listings that were put in effect in 1986, taking value 0 prior to 1986 and value 1 otherwise. 1977 and value 1 there after.
DA_Welfare_96	Time varying dummay variable taking value 0 prior to 1996 and value 1 there after. It represents the policy of disallowing eligibilities of disabilities arising from drug and alcohol abuse and welfare reform regarding the eligibility of disability benefits.

Table 2: Estimates of the Model using data for the period 1961-2001

Parameters	Model 1a		Model 1b	
		t-stat		t-stat
DOB_Y	0.02675	21.57	0.02648	21.35
AgUemp_1	0.00639	1.42	0.0047	1.05
AgUemp_2	-0.02458	-5.40	-0.02322	-5.10
DplusU	0.01227	2.27	0.01206	2.23
Black	0.61339	90.74	0.61829	91.46
Female	-0.30208	-57.43	-0.29467	-56.02
rrate2_1	0.0001921	1.88	0.000255	2.89
U1	0.4624	62.23	0.79049	76.97
U1*U2			-0.16503	-10.00
U1*U3			-0.372	-10.70
U1*U2*U3			-0.32766	-8.31
DY1970	0.10282	5.05	0.10489	5.15
DY1971	0.1984	9.92	0.20226	10.12
DY1972	0.20802	10.30	0.21035	10.41
DY1974	-0.01742	-0.95	-0.01729	-0.94
DY1976	-0.08312	-3.55	-0.08193	-3.50
DY1979	-0.22516	-10.90	-0.22525	-10.91
BenRed_80	-0.5119	-33.09	-0.51102	-33.03
Medicare_73	0.36402	19.20	0.36826	19.43
Decoupled_77	-0.24704	-13.52	-0.24574	-13.45
Mental_86	0.04858	3.48	0.05276	3.78
DA_Welfare_96	-0.26917	-21.30	-0.26863	-21.25

Disability-related health shocks are observed to the individuals but not to us. So it cannot be directly controlled for. The administrative dataset has no information about the worker's health status unless they are enrolled in the disability program. I use, however, the work history as an instrument (alas it's a noisy instrument!!) to infer one's health status. I use the past two year's non-employment variables U2 and U3 to construct the interaction terms U1*U2, U1*U3, U1*U2*U3 to include as regressors together with U1. The rationale for this is, fluctuations in the past non-employment history gives ideas about one's disability related health status. For instance, consider an individual who has (U1,U2,U3) = (1,1,1) but he has not got onto the program before. It is highly likely that he had not encountered a disability-qualifying health shock. For, if he had such a shock, he would have got onto the rolls earlier when he was not working. Thus, this non-employment pattern is most likely due to other factors than disability-qualifying health shocks. However, if the worker has (U1,U2,U3) = (1,0,0), he has a higher chance as compared to the previous case of having disability-qualifying health shocks which is why he is not working, and we would expect these workers to have higher probability of getting onto the rolls as compared to the others. There are other two intermediary possibilities in our setup: (U1,U2,U3) = (1,1,0) and (1,0,1). I included these interaction terms and kept the other variables intact in the final model. Table 3 shows the parameter estimates of this final model under the column heading Model 1b.

To check how good the model is, I used the estimated final model to compute the predicted proportions of disability insured workers who will be getting onto the rolls for each year, and compared them with the observed proportion of disability entitlements in the dataset. The predictions in Table 5 are the mean over the sample of the predicted hazard rate of each worker given his characteristics. Note that these predicted proportions are very close to the observed proportions.

I discuss the policy implications of this final model in the next section. In a following section, I carry out sensitivity analysis of these estimates to various model specifications and to sampled period.

5.1 Sensitivity Analysis

I carried out sensitivity analysis of estimated effects with many other specifications to see how robust the parameters estimates are. Table 4 shows these estimates. Model 2a in Table 4 is the same as the final model with the exception that I included an interaction term rrate2_high which is defined as the product rrate2_1*high, where high is a dummy

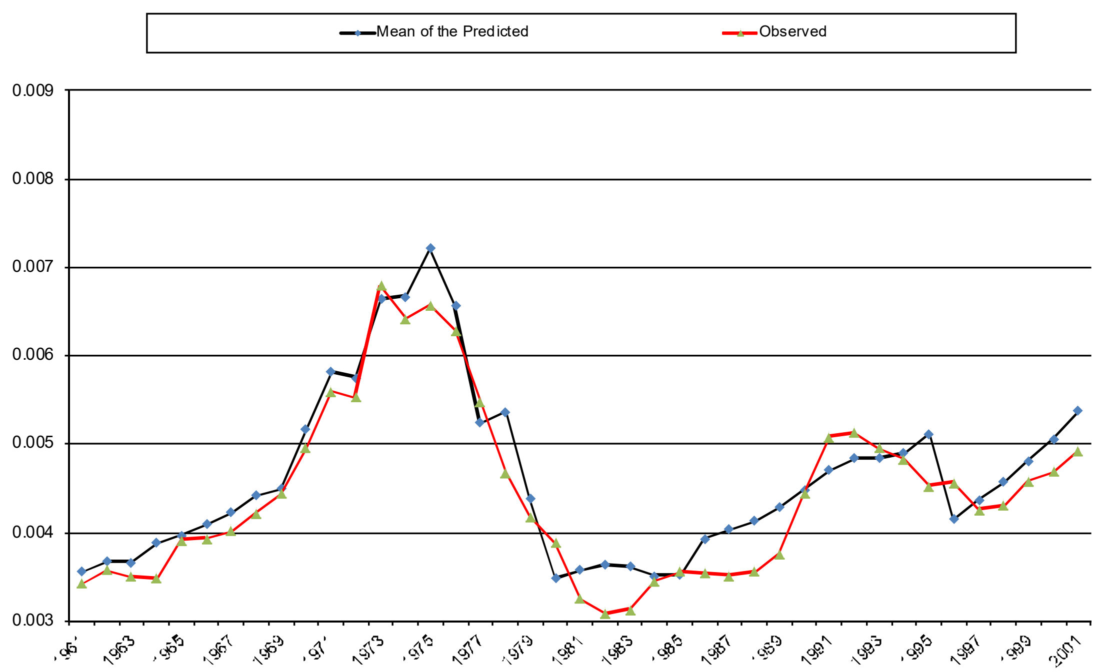


Figure 9: Plot of the actual and the model predicted incidence of DI entitlements over time. Source: Author's calculation using the 1 percent CWHS sample.

Table 3: Sensitivity Analysis of the Estimated Model using data for the period 1961-2001

parameters	Model 2a	t-stat	Model 2b (used sabsampling)	t-stat (sub)	t-stat (asymp)	Model 2c	t-stat
DOB_Y	0.02534	20.44	0.01731	10.31	11.27	0.02313	17.26
AgUemp_1	0.00190	0.42	-0.00158	-0.35	-0.34	-0.01549	-3.34
AgUemp_2	-0.02373	-5.22	-0.02801	-6.87	-5.95	-0.02142	-4.18
DplusU	0.00931	1.72	0.00856	1.61	1.52	0.01456	1.76
Black	0.52680	77.47	0.54715	21.38	38.05	0.65343	78.82
Female	-0.47507	-87.17	-0.28795	-33.73	-52.95	-0.24384	-37.92
rrate2_1	0.00016	1.21	0.00231	8.17	6.23	0.00021	1.89
rrate2_rich	-2.02320	-115.28					
U1	0.64368	62.31	0.78413	65.30	74.11	0.80909	64.47
U1*U2	-0.18383	-11.14	-0.16342	-10.74	-9.61	-0.14515	-7.25
U1*U3	-0.46523	-13.38	-0.41493	-11.04	-10.98	-0.35765	-8.64
U1*U2*U3	-0.28460	-7.22	-0.29335	-6.66	-6.92	-0.28349	-6.02
DY1970	0.08448	4.15	-0.00130	-0.01	0.00		
DY1971	0.17472	8.74	0.22719	9.40	10.00		
DY1972	0.18459	9.14	0.25208	10.48	10.80		
DY1974	-0.01919	-1.04	-0.02742	-1.54	-1.37		
DY1976	-0.08996	-3.84	-0.05299	-2.25	-2.13		
DY1979	-0.24471	-11.86	-0.23454	-10.25	-10.70		
BenRed_80	-0.47860	-30.94	-0.51243	-24.57	-29.59		
Medicare_73	0.34338	18.13	0.42677	18.37	18.94		
Decoupled_77	-0.26858	-14.71	-0.20896	-10.67	-10.46		
Mental_86	0.03174	2.28	-0.05274	-2.77	-3.38	0.01267	0.77
DA_Welfare_96	-0.28113	-22.24	-0.22559	-14.57	-16.51	-0.27424	-17.82

Notes: Model 2b assumes heterogeneity in the baseline hazard for different generations grouped into three broader cohort and since SAS went out of memory to estimate the Cox regression model, I have used the subsampling method to estimate the parameters and two estimates of standard errors producing two estimates of t-stat.
Model 2c is based on the sample restricted years 1980 and up.

variable taking value 1, if one's earnings is above median, otherwise taking value 0. This is to see if the above-median earners have significantly lower effects of an increase in the replacement rate. The estimate shows that this is true. Notice that all other estimates are very close to the estimates of the final model.

I have assumed the common baseline hazard function for all the specifications. But different cohorts may have different baseline hazard function due to the fact that they are from different periods with different medical technology. In the specification of final model, I allowed the baseline hazard to vary for five different cohorts: those born (1) in or before1930, (2) after 1930 but in or before 1940, (3) after 1940 but in or before 1950, (4) after 1950 but in or before 1960, (5) after 1960.[2] The parameter estimates are in table 4 under the label Model 2b.Except for the effect of policy variable Mental_86, all other parameter estimates are similar to the estimates from the final model. The estimate for Mental_86 has now a significantly negative effect.

Autor and Duggan, 2003 and Muller, 2008 did their study for the period 1980 and later. I also estimated the model for the sub-period 1980-2001. The parameter estimates are in Table 3 under the label Model 2c. Most of the common parameters in the final model and this model are very close in sign and significance, except the estimate of the policy dummy Mental_86, which has become insignificant.

It appears that most of the important parameter estimates in the final model are robust to sample period and model specifications, except the parameter estimate of the policy variable representing mental listing policy of 1986.

5.2 Parameter Estimates from the final model

5.2.1 Replacement Rate

First I address how much difference the lower replacement rates would have made to the number of first time disability entitlements if the bend points were fixed as in the alternative policy that I proposed. The parameter estimate of $\hat{a} = 0.00025$ for rrate2_1 means that all other characteristics remaining constant, if a disability insured worker has 10 percentage point lower replacement rate, it would lower his hazard rate, i.e., the probability of getting

[2] Because this more detailed specification ran into computational problems, a subsampling technique was used that aggregated the overall parameter estimates and standard errors from estimates on a set of subsamples. The theory and statistical properties of this technique, and a test of its validity, will be reported in another paper. Two estimates of standard errors or t-statistics are provided by the technique. One (labeled "asymp") aggregates the estimated asymptotic variances from subsamples. The other (labeled "sub") is the variance of the parameter estimates from the subsamples.

onto the rolls will be reduced only by 0.0025 percent. This is a very small incentive indeed. Even the largest increase in the average replacement rate of the poor in the eighties which is about 20 percentage point had very little effect on the number getting onto the disability rolls, contrary to what Autor and Duggan, 2003 argued.

5.2.2 Unemployment

The coefficient estimates in the final model for the four different employment history considered above are as follows:

The estimates are as expected. For instance, if one is not working for the past three years, his probability of getting onto the program given that he has not been on the program so far is 7 percent lower than others. As I argued earlier, a (1,1,1) pattern (no earnings in the previous three years) may signify non-employment for reasons other than a disability-qualifying health shock. On the other hand, if one was working for the past two years and then he is not working, i.e., one with (U1,U2,U3) = (1,0,0), he has a 79 percent higher probability of getting onto the rolls. As I argued earlier this pattern is much more likely to signify that he has disability-qualifying health shocks. Similar interpretations go with the other not-working scenarios.

Workers may look at some characteristics of the economy to forecast the future outlook of the economy while deciding whether to get onto the disability rolls. We do not know have data on those variables. I proxy them with the one-year forward difference of unemployment rate as his expectations of the outlook of the economy in the future. The previous studies based on state level aggregate data found a positive effect of the unemployment rate on the rate of disability applications and awards (see for instance, Autor and Duggan, 2003 and Rupp and Stapleton, 1995). In our final model, the parameter estimates for last year's aggregate unemployment rate on the hazard rate turns out to be insignificant. But the effect of the aggregate unemployment rate a year earlier is negatively significant. The effect of the forward difference DPlusU is significantly positive. That is if a worker finds the future outlook of the economy not very promising, he is more likely get onto the disability program.

5.2.3 Policies

I now examine the effects of the policy dummy variables. The policy dummies DY1970-DY1974, capturing the effects of those years when there were ad hoc increase in the benefits that I described earlier. All except DY1974 have significant positive effects. For instance, the estimates for these policy dummies suggest that the effect of ad hoc increase in benefits might have increased the percentage of workers getting onto the roll by 20 percent in 1971 and 21 percent in 1972. Similarly, the ad hoc benefits reduction in 1976, 1979 and then more permanent reduction in benefits in and after 1980, all led to lower the number of new entitlements and the lower magnitude can be as much as 51 percent due to permanent benefit reduction of the policies in 1979. Also, notice that inclusion of Medicare benefits in 1973 has a significant strong effect on the number of people getting onto the rolls, increasing the number to as high as 36percent.

Two policies that are argued by many to have affected the number of workers getting onto disability rolls are the mental listings of 1986, and the Drug and Alcohol and the Welfare reform of 1996. The mental listing has a significant positive effect of increasing the number of enrollees by as much as 5 percent. The estimated effect is, however, not robust across specifications and samples. The effect of D&A and Welfare reform is estimated to have reduced the number of new enrollees by as much as 27 percent in the years 1996 and later.

Regarding the effects of gender and race, the final model shows that the black has much higher probability than the non-black and the female has much lower probability than the male in getting onto the program.

6 Conclusion

Recent policy debates on the causes of high growth in disability entitlements in the 1980's attribute them to the rise in the replacement rate (defined as disability benefits as a ratio of potential earnings), unemployment rate and generosity of disability program eligibilities, screening procedure, and to extension of the program to musculoskeletal and mental disorders. This paper analyzed these issues using the Social Security Administration's 1 percent random sample of continuous work history survey (CWHS) data. The paper used the work history to construct proxy variables for unobserved health shocks.

It has been argued that the widening earnings gap between the rich and the poor led to high increase in the replacement rates of the poor workers which strongly increased their incentives to get onto the program instead of attaching themselves to the labor market. The paper did not find evidence for this.

Within the limits of the data set, the paper analyzed the incentive effects of non-employment on the rate of disability entitlement. The effect of non-employment depends on the health status. A healthier worker (as inferred from the proxy variable) has a low probability of getting onto the program. It is an important policy issue to examine if the generosity of the disability program led workers quit their jobs and get onto the disability program. This cannot be analyzed using the administrative data. Future work along this line is useful.

Most of the policies introduced since the inception of the program have their expected effects. The most significant of all the policies is the addition of Medicare benefits, which according to our estimate has increased disability entitlements by 37 percent since its inception. Two most widely cited policies that many argued to have strongly affected the number of workers getting onto the disability rolls is the mental listings of 1986 in the 80's, and the Drug and Alcohol and the Welfare reform of 1996 in the 90's. The estimates in this paper show that the mental listing has a significant positive effect of increasing the number of enrollees by as much as 5 percent. This effect is, however, not robust across specifications and samples. More evidence from other datasets is needed. The effect of D&A and Welfare reform of 1996 is estimated to have reduced the number of new enrollees by as much as 27 percent, which is robust across specifications and sample.

References

Autor, D. H. and M. G. Duggan (2003). The Rise in the Disability Rolls and the Decline in Unemployment, *The Quarterly Journal of Economics*, **118**, no. 1, 157–206. DOI: 10.1162/00335530360535171 (cit. on pp. 3, 4, 13, 26, 27).

Autor, D. H. and M. G. Duggan (Sept. 2006). The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding, *Journal of Economic Perspectives*, **20**, no. 3, 71–96. DOI: 10.1257/jep.20.3.71 (cit. on p. 3).

Lahiri, K., J. Song, and B. Wixon (2008). A model of Social Security Disability Insurance using matched SIPP/Administrative data, *Journal of Econometrics*, **145**, no. 1, 4–20. DOI: https://doi.org/10.1016/j.jeconom.2008.05.006 (cit. on p. 5).

Lando, M. E., M. B. Coate, and R. Kraus (1979). Disability benefit applications and the economy, *Soc. Sec. Bull.*, **42**, 3 (cit. on p. 4).

Muller, L. S. (2008). The Effects of Wage Indexing on Social Security Disability Benefits, *Soc. Sec. Bull.*, **68**, 1 (cit. on pp. 3, 26).

Rupp, K. and D. Stapleton (1995). Determinants of the growth in the Social Security Administration's disability programs-an overview, *Soc. Sec. Bull.*, **58**, 43 (cit. on pp. 4, 27).

Social Secruity Administration (2005). *Annual Statistical Supplement, 2005*. Tech. rep. Washington, DC (cit. on p. 6).

Social Secruity Administration (2007). *Annual Statistical Supplement, 2007*. Tech. rep. Washington, DC (cit. on p. 7).

Social Security Administration (2009). Annual Statistical Report on the Social Security Disability Insurance Program, 2009, (cit. on p. 7).