Essays on Macroeconomics and Health

Essays on Macroeconomics and Health

Tobias Laun

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics 2012

KEYWORDS: Unemployment insurance; Disability insurance; Optimal contracts; Life cycle; Retirement; Pension reform; Health investment; Labor supply; Savings decision; Screening stringency

Essays on Macroeconomics and Health

ISBN 978-91-7258-866-0.

PRINTED IN SWEDEN BY:
Ineko, Göteborg 2012

DISTRIBUTED BY:
The Research Secretariat
Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden
www.hhs.se

Preface

This report is a result of a research project carried out at the Department of Economics at the Stockholm School of Economics (SSE).

This volume is submitted as a doctor's thesis at SSE. The author has been entirely free to conduct and present his research in his own ways as an expression of his own ideas.

SSE is grateful for the financial support which has made it possible to fulfill this project.

Göran Lindqvist
Director of Research
Stockholm School of Economics

Acknowledgement

First and foremost, I would like to thank my supervisor, Lars Ljungqvist, for his guidance, support and encouragement throughout these years. He was the perfect supervisor for me and each meeting with him provided me with new energy and inspiration to pursue my research. David Domeij always had an open door and provided me with very valuable feedback during all stages of my projects. Johanna Wallenius is not only an excellent coauthor and a constant source of advice, she is also as great a friend as anyone could wish for. I am looking forward to continuing to work with her. Juanna Schrøter Joensen was always available for a chat when I tried avoiding to work. I enjoyed discussing the latest soccer results with Tore Ellingsen and Magnus Johannesson. Yoichi Sugita provided me with very useful support during the job hunt. Finally, I am deeply indebted to Friedrich Breyer for introducing me to the world of academia and convincing me to pursue a PhD.

Many thanks go to the administrative staff. Without the help of Carin Blanksvärd, Ritva Kiviharju and Lilian Öberg the department could not function. I am also grateful for the financial support from the Jan Wallander and Tom Hedelius Foundation.

I was lucky to meet many fellow PhD students who turned my graduate studies into a fun experience. André Romahn started out as my neighbor and became a great friend, office mate, lunch date as well as a fellow Bundesliga fan. He is by far the funniest person I know, we shared countless laughs, and watched plenty of TV-Shows and soccer games together. Margherita Bottero deserves a special thanks for making sure that we all got out every once in a while. She was the social glue of our group and she has been dearly missed during the last year. Many other PhD students deserve to be mentioned for making this journey more enjoyable: Henrik Lundvall, Ettore Panetti, Björn Wallace, Amanda Jakobsson, Elena Mattana, Claudia Wolff, Taneli Mäkinen, Björn Ohl, Karin Hederos Eriksson, Anna Sandberg, Arna Vardardottir and Simon Wehrmüller.

Last but not least I am grateful to have met my wife-to-be, Lisa Jönsson, in the PhD program. She provided me not only with food and shelter, but also helped me integrate into Swedish society. Her deep interest in policy issues and her profound knowledge of data have impressed me more than once and have given my papers a much needed connection to reality.

Contents

Introduction

This thesis consists of four independent papers covering macroeconomic topics such as unemployment insurance, retirement, business cycles and disability insurance. The unifying theme connecting all papers is the importance of health when addressing these issues. The first paper derives the optimal insurance against unemployment and disability in a setting where individuals can determine the probability of becoming disabled by exerting prevention effort. The second paper develops a life cycle model of labor supply and retirement to study the interactions between health and the labor supply behavior of older workers, in particular disability insurance and retirement. The third paper looks at the role of health in a business cycle framework. Here, health is treated as an asset that generates time and utility. The last paper derives a theoretical model of the application decision for disability benefits and proposes an empirical strategy for measuring screening stringency in the disability insurance program.

1. Optimal Social Insurance with Endogenous Health

This paper analyzes optimal insurance against unemployment and disability in a private information economy with endogenous health and search effort. The question of how a government should insure workers against unemployment and disability is a recurring and controversial theme in the public policy debate. In the last twenty years many developed countries have reformed their unemployment insurance programs in order to decrease costs for the government and make people return to employment. These changes have often included a decrease in benefits if the worker is unemployed longer than a certain amount of time. While these reforms obviously increase the incentives of unemployed workers to find new employment, it might also create situations in which the long-term unemployed would rather exit the labor force, and go on disability insurance, than continue trying to find new employment. Empirical studies, e.g. Larsson (2006) and Karlström, Palme, and Svensson (2008), have shown that substitution between social insurance programs is indeed a common phenomenon. In order to receive disability insurance benefits, an individual either has to falsely claim to be unable to work or actually become unemployable. The former is a well known problem and taken into account when designing unemployment and disability insurance systems. This issue is also addressed in this paper. The idea that strict unemployment insurance systems can create incentives for individuals to actually become disabled has, to my knowledge, not yet been formally investigated.

To study this issue, I combine unemployment and disability insurance in one framework and assume that the probability of becoming disabled is endogenous. Combining

moral hazard and adverse selection, enables me to study the optimal design of unemployment insurance when individuals have the option to leave the labor force. In this framework I am also able to analyze how unemployment and disability benefits should depend on the employment history. I assume that individuals, whether employed or unemployed, can exert so-called prevention effort. In so doing, they increase the probability of remaining healthy, and hence staying in the labor force, for one more period. Furthermore, an unemployed individual has the option of exercising search effort, which increases the probability of finding employment in the next period. Both the prevention and the search effort are costly in terms of utility. Disability is an absorbing state. The planner's goal is to minimize the expected discounted cost of providing the individual with a certain level of utility. I first consider the full information case in which the planner can observe effort levels and the individual's state of health and employment. I then relax those assumptions and assume that the planner can only observe whether an individual works or not, i.e., the planner cannot distinguish between unemployment and disability and he cannot observe effort levels. I show that the optimal sequence of consumption and promised utilities of an employed worker is increasing with tenure. Once a worker is disabled, she will receive constant benefits and her utility will remain constant. During unemployment, decreasing benefits are not necessarily optimal anymore. The prevention constraint implies increasing benefits and promised utilities during unemployment while the search constraint has the opposite effect. However,if individuals respond sufficiently much to search incentives, the latter effect dominates the former and the optimal consumption sequence is decreasing during unemployment.

2. A Life Cycle Model of Health and Retirement: The Case of Swedish Pension Reform

In this paper we develop a life cycle model of labor supply and retirement to study the interactions between health and the labor supply behavior of older workers, in particular disability insurance and pension claiming. Faced with ageing populations and the looming insolvency of social security, governments the world over are grappling with the question of how to reform retirement programs. Understanding how changes to retirement programs affect life cycle labor supply, particularly retirement behavior, is critical for assessing the effects of these changes on allocations, welfare and government finances. Accurate assessments require a model of retirement that captures the key forces underlying retirement decisions.

Various institutional features have potentially large implications for the labor supply behavior of older workers; key among them are the design of pension systems, disability insurance and healthcare. Disability insurance is particularly relevant, as in many countries a large fraction of retirement occurs before the normal retirement age. A discussion of disability insurance programs naturally leads to a discussion of health, as disability insurance programs without exception have some eligibility criteria regarding health. Health, in turn, has potential implications for labor supply outcomes both directly and through the healthcare system.

In this paper, we construct a life cycle model of labor supply and retirement, which enables us to study the interactions between health, disability insurance and old-age

retirement benefits. The key features of our framework are that individuals choose when to stop working and, given eligibility criteria, when/if to apply for disability and pension benefits. Individuals care about their health and can partially insure against health shocks by investing in health. The fact that people can impact their own health and choose whether or not to claim disability benefits, are novel features in relation to the existing literature.

While many countries have come to understand the need for social security reform, Sweden is one of the few countries to have actually undertaken a major pension reform in recent years. While there are big expectations of the reform, to the best of our knowledge no formal analysis of the expected implications of the changes to the pension system exists to date. We use our model to study the labor supply implications of the recent Swedish pension reform and to ask what particular aspects of the reform are driving the results. We find that the Swedish pension reform creates large incentives for older workers to continue working longer. Our main findings are threefold: (1) the model predicts an increase in the average retirement age of 2.3 years from 62.4 to 64.7, and (2) there is an increasing tendency for workers to continue working while collecting pension benefits, but (3) the fraction of older workers claiming disability insurance only declines by roughly one percentage point, from 18.6% to 17.7%.

3. Health and Business Cycles

This paper develops a framework to analyze the interactions between health and business cycles. The importance of health for growth and economic development has been greatly acknowledged. Most studies, however, consider health an exogenous variable rather than something the individual can influence himself. A justification for not modeling health explicitly is that, if health is endogenous, it can be assumed to be a part of human capital. However, health is more than just a part of human capital. Besides affecting labor supply and productivity, health affects survival probabilities and also has direct utility effects. The former operates through the effective discount factor which impacts the attractiveness of saving.

In this paper, I incorporate endogenous health into a business cycle model. The individual's health is determined by his stock of health capital and a stochastic component. Health increases the individual's utility as well as the total time the individual can spend on either work or leisure. The individual invests in health capital as well as in physical capital. He further decides how much of his available time to allocate to labor supply and how much to leisure.

In this setting, I analyze the effects of health and productivity shocks on the variables in the model and derive the optimal investment in health capital. I show that an unexpected decline in health causes a reduction in output, consumption and labor supply. In response to a decline in health, the individual increases investment in health capital and reduces savings accordingly. I also show that a positive shock to productivity increases both health and physical capital investment. Better health therefore leads to increased savings. Higher productivity, in turn, increases savings and improves health.

4. Screening Stringency in the Disability Insurance Program

The disability insurance program has become one of the largest income maintenance programs in modern welfare states.[1] Autor (2011) shows that, in the U.S., the share of 25–64 year olds receiving benefits from the Social Security Disability Insurance (SSDI) increased from 2.3 to 4.6 percent between 1989 and 2009. Sweden has also experienced a large growth in disability benefits recipiency. The share of 30–64 year olds receiving disability benefits increased from 8 to 12 percent between 1985 and 2008. In this period, the fluctuations in the award rate have been large.

In both the U.S. and Sweden, changes in screening stringency seem to be important for the growth of the disability insurance program. Changes in formal eligibility criteria are easily observable but the actual screening stringency also depends on the implementation of formal program rules, such as caseworker discretion and internal processes at the Social Insurance Agency. The contribution of changes in actual screening stringency to program growth has therefore been difficult to evaluate. One potential indicator of screening stringency is the denial rate. However, the denial rate depends on the composition of the applicant pool and the decision to apply for disability benefits is potentially correlated with screening stringency.

In this paper, we provide a theoretical model showing that the denial rate does not necessarily capture the screening stringency of the disability insurance program if the application decision depends on the likelihood of getting admitted. We further show that the relative health of awarded beneficiaries versus non-beneficiaries improves with a reduction in screening stringency. Based on this result, we propose an empirical strategy for assessing changes in screening stringency in the disability insurance program over time. We use the mortality rate after admittance as an objective measure of health and estimate the excess mortality of new disability beneficiaries over time. The strength of the empirical strategy is that it captures changes in screening stringency both due to changes in formal eligibility criteria and due to changes in the implementation of program rules. The latter aspect has been difficult to assess in previous studies.

Applying the empirical strategy to Sweden, we find that changes in screening stringency are an important contributing factor for the fluctuations in the disability benefit award rate over time. Screening stringency was comparatively low during the periods of large inflow to the program in the late 1980s and the early 2000s, whereas the rapid decline in disability benefit awards since 2005 is reflected in a substantial increase in screening stringency. The removal of eligibility for pure labor market reasons for workers aged 60–64 results in an increase of the relative screening stringency for older workers compared to younger. The large inflow of women compared to men during the early 2000s corresponds to a relatively lower screening stringency for women.

[1] See e.g. Wise (2012) for a more detailed description of disability insurance programs in several developed countries.

References

Autor, D. (2011), 'The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options', *Working Paper 17697*, National Bureau of Economic Research, Cambridge, MA.

Karlström, A., Palme, M., and Svensson, I. (2008), 'The Employment Effect of Stricter Rules for Eligibility for DI: Evidence from a Natural Experiment in Sweden', *Journal of Public Economics*, 92, 2071–2082.

Larsson, L. (2006), 'Sick of Being Unemployed? Interactions between Unemployment and Sickness Insurance', *Scandinavian Journal of Economics*, 108, 97–113.

Wise, D. (2012), *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, University of Chicago Press, forthcoming.

PAPER 1

Optimal Social Insurance with Endogenous Health

Tobias Laun

Abstract This paper analyzes optimal insurance against unemployment and disability in a private information economy with endogenous health and search effort. Individuals can reduce the probability of becoming disabled by exerting, so-called, prevention effort, which is costly in terms of utility. A healthy, i.e., not disabled, individual either works or is unemployed. An unemployed individual can exert search effort in order to increase the probability of finding a new job. I show that the optimal sequence of consumption is increasing for a working individual and constant for a disabled individual. During unemployment, decreasing benefits are not necessarily optimal in this setting. The prevention constraint implies increasing benefits over time while the search constraint demands decreasing benefits while being unemployed. However, if individuals respond sufficiently much to search incentives, the latter effect dominates the former and the optimal consumption sequence is decreasing during unemployment.

1. Introduction

The question of how a government should insure workers against unemployment and disability is a recurring and controversial theme in the public policy debate. In the last twenty years many developed countries have reformed their unemployment insurance programs in order to decrease costs for the government and make people return to employment. These changes have often included a decrease in benefits if the worker is unemployed longer than a certain amount of time. While these reforms obviously increase the incentives of unemployed workers to find new employment, it might also create situations in which the long-term unemployed would rather exit the labor force, and go on disability insurance, than continue trying to find new employment. Empirical studies, e.g. Larsson (2006) and Karlström, Palme, and Svensson (2008), have shown

The author thanks Lars Ljungqvist, Johanna Wallenius, David Domeij, Nils Gottfries, Sebastian Koehne, Lisa Jönsson and seminar participants at Stockholm School of Economics, Institute for International Economic Studies, Swedish Institute for Social Research and Uppsala University for helpful comments and suggestions. Financial support from the Jan Wallander and Tom Hedelius Foundation at Svenska Handelsbanken is gratefully acknowledged.

that substitution between social insurance programs is indeed a common phenomenon. In order to receive disability insurance benefits, an individual either has to falsely claim to be unable to work or actually become unemployable. The former is a well known problem and taken into account when designing unemployment and disability insurance systems. This issue is also addressed in this paper. The idea that strict unemployment insurance systems can create incentives for individuals to actually become disabled has, to my knowledge, not yet been formally investigated.

To study this issue, I combine unemployment and disability insurance in one framework and assume that the probability of becoming disabled is endogenous. Combining moral hazard and adverse selection, enables me to study the optimal design of unemployment insurance when individuals have the option to leave the labor force. In this framework I am also able to analyze how unemployment and disability benefits should depend on the employment history. I assume that individuals, whether employed or unemployed, can exert so-called prevention effort. In so doing, they increase the probability of remaining healthy, and hence staying in the labor force, for one more period. Furthermore, an unemployed individual has the option of exercising search effort, which increases the probability of finding employment in the next period. Both the prevention and the search effort are costly in terms of utility. Disability is an absorbing state. The planner's goal is to minimize the expected discounted cost of providing the individual with a certain level of utility. I first consider the full information case in which the planner can observe effort levels and the individual's state of health and employment. I then relax those assumptions and assume that the planner can only observe whether an individual works or not, i.e., the planner cannot distinguish between unemployment and disability and he cannot observe effort levels. I show that the optimal sequence of consumption and promised utilities of an employed worker is increasing with tenure. Once a worker is disabled, she will receive constant benefits and her utility will remain constant. During unemployment, decreasing benefits are not necessarily optimal anymore. The prevention constraint implies increasing benefits and promised utilities during unemployment while the search constraint has the opposite effect. However,if individuals respond sufficiently much to search incentives, the latter effect dominates the former and the optimal consumption sequence is decreasing during unemployment.

The literature on optimal unemployment insurance started with Shavell and Weiss (1979). They show that if individuals have no wealth, cannot borrow and can influence the probability of finding a job, monotonically decreasing benefits throughout unemployment are optimal.

Hopenhayn and Nicolini (1997) extend the model of Shavell and Weiss (1979) and apply the recursive solution methods for repeated games and dynamic principal-agent

problems.[1] They show that unemployment benefits are monotonically decreasing with the length of unemployment. Furthermore, after the individual finds employment there is a wage tax which depends on the unemployment history. A tax which is increasing with the length of previous unemployment is optimal, because this reduces claims to consumption in all future states (employment and unemployment), and hence, gives a stronger incentive to search for a job. In numerical simulations the authors also show that, for certain parameter values, the wage tax is negative (i.e., a subsidy) for individuals who were unemployed for five weeks or less.

In a setting in which an individual must exert effort not only to find a job but also to remain employed, Hopenhayn and Nicolini (2009) show that monotonically decreasing unemployment benefits are optimal. In other words, the presence of moral hazard while working does not alter this basic finding. The authors assume that work causes a certain level of disutility and allow for the possibility of quits which are indistinguishable from lay-offs. For a high enough level of disutility, unemployed individuals who find a new job prefer to quit after one period of employment and return to unemployment with increased benefits. In order to prevent quits, the promised utilities have to increase with the duration of employment. This increasing utility profile is achieved by decreasing the wage tax (i.e., increasing consumption) with the length of employment.

Fredriksson and Holmlund (2001) analyze a model with two states of unemployment: insured and uninsured. Here, decreasing benefits are defined as a drop in consumption when going from insured to uninsured unemployment. The optimality of this drop cannot be established analytically. Higher benefits for insured individuals increase the search incentives of the uninsured because finding employment entitles the individual to those higher benefits in the future. However, more generous benefits in the first stage of unemployment reduce the search incentives for insured individuals by making employment comparatively less attractive. The authors numerically show that this wage pressure effect is dominated by the entitlement effect and that decreasing unemployment benefits are therefore optimal.

For a more thorough review on the literature on optimal unemployment insurance see Fredriksson and Holmlund (2006).[2] An issue not addressed in the literature is the possibility for workers to leave the labor force. If individuals only have a choice between being unemployed or working, providing them with incentives to search for a job is easier than when they also have the option to go into other social insurance

[1] For more on these solution methods, see for example Spear and Srivastava (1987), Thomas and Worrall (1990), Abreu, Pearce, and Stacchetti (1990), Atkeson and Lucas (1992) and Chang (1998).

[2] A few examples include Baily (1978), Flemming (1978), Wang and Williamson (1996), Acemoglu and Shimer (1999), Boone, Fredriksson, Holmlund, and van Ours (2007) Pavoni and Violante (2007), Pavoni (2007), Hagedorn, Kaul, and Mennel (2010).

systems, e.g. disability insurance. Given the importance of disability insurance in many countries, it is, therefore, crucial to take disability insurance into account when designing optimal unemployment insurance.

Optimal disability insurance was first investigated by Diamond and Mirrlees (1978). In their model, individuals either have full or no work capacity and the government cannot distinguish between those who cannot work and those who choose not to work. They show that, in the optimum, individuals are indifferent between working and not working and that a tax on savings should be part of the optimal social insurance policy. Disability benefits are higher the longer the individual previously worked.

An example of the more recent literature on disability insurance is Golosov and Tsyvinski (2006). They define disability as a permanent negative shock to the individual's skill level. At the beginning of each period this shock occurs with an exogenous probability. Whether an individual experienced such a shock or not is private information. With full information, the optimal allocation implies full insurance against disability, i.e., the consumption of the able and the disabled worker are equal. The second-best optimum with private information is characterized by an able individual having consumption increasing in the duration of her work history. Once disability occurs, the individual's consumption drops and remains constant after that.

A common assumption in the literature on optimal disability insurance is that individuals can only be in two possible states: work or retirement. Introducing a third state, namely unemployment, implies that the age at which one stops working and the retirement age are not necessarily equivalent. The question, whether benefits are still increasing in the retirement age or only increasing in the time spent working, thus arises. Moreover, it is worth exploring how unemployment affects the level of disability benefits.

The work closest related to mine is Höglin (2008). He was the first to look at unemployment and disability insurance simultaneously. In the second chapter of his thesis he combines the models of Hopenhayn and Nicolini (1997) and Diamond and Mirrlees (1978). In his framework, there is an exogenous probability that employed individuals can become unemployed or disabled. The probability of an unemployed individual finding a job depends on her search effort. The probability of becoming disabled is exogenous and the same as for employed individuals. Disability is an absorbing state. The author shows that in the optimum employed and disabled workers have constant consumption, while unemployed individuals face decreasing benefits over time. These findings resemble the results of Hopenhayn and Nicolini (1997). Another result is that the length of employment is irrelevant for the level of benefits the individual receives when she becomes unemployed or disabled. This stands in stark contrast to the results

of Diamond and Mirrlees (1978) who find that benefits are increasing in the length of employment. The decreasing unemployment benefits and the irrelevance of employment history both stem from the assumption that disability is an exogenous state. The individual does not need to be provided with an incentive to work because job loss and disability are exogenous. While unemployed, benefits can be decreasing in order to provide search incentives without running the risk of creating incentives to leave the labor force.

The key distinction between this paper and Höglin (2008) is that here the probability of becoming disabled is endogenous. This implies that when designing optimal unemployment insurance one faces a trade-off between providing incentives to search for a job while at the same time keeping the individual in the labor force. This additional constraint makes characterizing the solution more difficult but it is necessary in order to avoid creating the wrong incentives.

The structure of the paper is as follows. In section 2, I characterize the environment of the model. Section 3 analyzes the autarky case, where there is no insurance provided by the planner. In section 4, I describe the planner's problem, and in section 5, I derive the optimal insurance for the full information case and for the case of asymmetric information. Section 6 provides a numerical illustration of the model, and section 7 concludes. All proofs can be found in the appendix.

2. The Environment

In this model an individual can be in one of three possible states; healthy and employed, healthy and unemployed, or disabled. Following Diamond and Mirrlees (1978), I assume that disability is an absorbing state. Disability should, in this context, be interpreted as a state of non-employability, i.e., a permanent loss of work capacity. The prevention effort is hence an effort to remain employable. Disability itself, however, has no direct utility effects and individuals do not necessarily want to avoid it at any cost.

An employed individual earns the constant wage $w > 0$ and maximizes her utility by choosing a level of prevention effort $a_t \geq 0$. An unemployed individual earns no income and chooses jointly the optimal levels of prevention and search effort: $a_t \geq 0$ and $e_t \geq 0$. A disabled individual also has no income and, since disability is an absorbing state, does not exert any search or prevention effort. Following the literature on repeated moral hazard, individuals have no access to credit markets or storage technology. The planner can hence directly control their consumption.

Hopenhayn and Nicolini (2009) make the simplifying assumption of discrete effort levels, i.e., $e_t = e > 0$ or $e_t = 0$. They argue that a continuum of effort levels[3] complicates the analysis significantly without providing additional insights. That is true for their case with only one incentive problem, that is, the search-incentive problem. However, making this assumption here, in the presence of two incentive problems (search and prevention), leads to multiple equilibria with different combinations of binding and non-binding constraints.

If a worker exerts prevention effort a_t, then the probability of remaining employable in the next period is $p(a_t)$. The function $p(\cdot)$ is strictly increasing, strictly concave and twice differentiable. Prevention effort can be thought of as typical prevention measures, such as exercising, a healthy diet and regular medical checkups, as well as seeking treatment for medical problems. Jönsson, Palme, and Svensson (2012) show that circulatory diseases and musculoskeletal diseases are common reasons for awarding disability benefits in Sweden. These are arguably areas in which the individual can influence the probability of becoming sick by exerting some kind of prevention effort.

The probability of remaining on the job when employed is exogenous and given by $s > 0$. The unemployed individual will remain employable in the next period with probability $p(a_t)$ and she will find a job with probability $q(e_t)$. The function $q(\cdot)$ is also strictly increasing, strictly concave and twice differentiable.

As in Hopenhayn and Nicolini (1997), effort is costly in terms of utility. An individual's expected lifetime utility is then given by

$$E \sum_{t=0}^{\infty} \beta^t \left[u(c_t) - a_t - e_t \right],$$

where $0 < \beta < 1$ is the discount factor, c_t is consumption and $u(\cdot)$ is strictly increasing, strictly concave and twice differentiable. I further assume that $u(0)$ is well defined.

The individual's only source of income is the transfer from the planner and her wage if she is working.[4] The planner has unlimited access to a perfect capital market (with a constant gross interest rate equal to $1 \,/\, \beta$) while the individual can neither lend nor borrow.

The state of an individual is private information. The planner can only observe the individual's income and from this infer whether she is employed or not. He cannot distinguish between an unemployed and a disabled individual. The effort levels are also unobservable.

---

[3] As it is the case in this model or in Hopenhayn and Nicolini (1997).

[4] Since the planner can control the individual's consumption with the transfers, I use the terms transfer and consumption interchangeably throughout this paper.

3. Autarky

In autarky there is no planner to provide insurance against unemployment and disability. An employed worker consumes her wage, w, and decides on how much prevention effort to exert. The autarky value of being employed is then

$$V^e_{aut} = \max_{a\geq 0} \left\{ u(w) - a + \beta p(a) \left[s V^e_{aut} + (1-s) V^u_{aut} \right] + \beta [1-p(a)] V^d_{aut} \right\}, \tag{1.1}$$

where V^u_{aut} and V^d_{aut} are the autarky values of being unemployed and disabled respectively. Recall that, $p(a)$ is the endogenous probability of remaining employable while s is the exogenous probability of remaining employed. Since there are no state variables in this problem, there is a time-invariant optimal prevention effort and an associated value of being healthy and employed.

The healthy but unemployed individual has no income in autarky, and hence, zero consumption. She has to determine the optimal levels of prevention and search effort. Her problem is given by

$$V^u_{aut} = \max_{a,e\geq 0} \left\{ u(0) - a - e + \beta p(a) \left[q(e) V^e_{aut} + [1-q(e)] V^u_{aut} \right] + \beta [1-p(a)] V^d_{aut} \right\}, \tag{1.2}$$

where $q(e)$ is the endogenous probability of finding a new job.

The disabled worker has no consumption in autarky. Since disability is an absorbing state, there are no decisions about effort levels. Furthermore, the value of being in this state is

$$V^d_{aut} = u(0) + \beta V^d_{aut} \quad \Leftrightarrow \quad V^d_{aut} = \frac{u(0)}{1-\beta}. \tag{1.3}$$

Equations (1.1) – (1.3) together determine the autarky levels of utility. These utility levels provide a lower bound for the planner. If he were to promise less than those values, the individual would not participate in the insurance system.

4. The Planner's Problem

The aim of the planner is to minimize the expected discounted cost of providing the individual with a certain level of utility. The planner does so by choosing the individual's consumption in the current period, and by promising her a certain utility level in the next period in an incentive compatible way. As in Spear and Srivastava (1987), the problem can be defined recursively with the individual's promised utility acting as a state variable which summarizes the employment history.

4.1. Employed Worker. An employed worker faces the problem of determining the optimal prevention effort. She receives the transfer $c^e - w$ from the planner, and hence, has a consumption of c^e. Furthermore, she is promised a utility of $V^{e,e}$ if she remains employed in the next period. If she becomes unemployed, she is promised

a utility of $V^{e,u}$. If she becomes disabled, her promised utility is $V^{e,d}$. Given these continuation values and consumption, the employed worker solves the following problem

$$\max_{a^e \geq 0} \left\{ u(c^e) - a^e + \beta p(a^e) \left[sV^{e,e} + (1-s)V^{e,u} \right] + \beta \left[1 - p(a^e) \right] V^{e,d} \right\},$$

The first-order condition is given by

$$\beta p'(a^e) \left[sV^{e,e} + (1-s)V^{e,u} - V^{e,d} \right] \leq 1, \tag{1.4}$$

with equality for $a^e > 0$. The left hand side of this inequality represents the benefit from a marginal increase in prevention effort, which consists of an increase in the probability of remaining employable multiplied by the corresponding utility benefit. The latter is the difference between the utility of being healthy, that is, the probability weighted average of the utilities in case of employment and unemployment, and the utility of being disabled. The optimal prevention effort, therefore, depends on the utility levels promised by the planner.

Since this is a repeated game, the individual working in the current period was promised a certain utility level in the previous period. This utility promise has to be kept by providing the individual with a consumption level c^e, and in turn, promising utility levels for the next period. If the planner previously promised to provide the employed worker with a utility level of V^e, the promise-keeping constraint is given by

$$V^e \leq u(c^e) - a^e + \beta p(a^e) \left[sV^{e,e} + (1-s)V^{e,u} \right] + \beta \left[1 - p(a^e) \right] V^{e,d}. \tag{1.5}$$

Since the planner cannot distinguish between an unemployed and a disabled individual, the unemployed worker has to be given an incentive to not falsely claim disability. This moral hazard can be prevented by making sure that an unemployed individual always enjoys a utility at least as high as that of a disabled individual. This truth-telling constraint is given by

$$V^{e,d} \leq V^{e,u}. \tag{1.6}$$

The problem of the planner is now to minimize the cost of providing consumption and continuation values to the employed worker such that the prevention-incentive constraint (1.4), the promise-keeping constraint (1.5) and the truth-telling constraint (1.6) are fulfilled.

$$\begin{aligned} C_e(V^e) = \min_{c^e, a^e, V^{e,e}, V^{e,u}, V^{e,d}} \Big\{ & c^e - w + \beta p(a^e) \left[sC_e(V^{e,e}) + (1-s)C_u(V^{e,u}) \right] \\ & + \beta \left[1 - p(a^e) \right] C_d(V^{e,d}) \Big\} \end{aligned}$$

subject to conditions (1.4) – (1.6). $C_e(\cdot)$, $C_u(\cdot)$ and $C_d(\cdot)$ are the minimized costs of providing utility to an employed, an unemployed and a disabled individual, respectively.

Assigning the Lagrange parameters γ^e to the promise-keeping constraint (1.5), δ^e to the prevention-incentive constraint (1.4) and $\beta\eta^e$ to the truth-telling constraint (1.6), yields the following first-order conditions

$$\gamma^e = \frac{1}{u'(c^e)}, \tag{1.7}$$

$$C'_e(V^{e,e}) = \gamma^e + \delta^e \frac{p'(a^e)}{p(a^e)}, \tag{1.8}$$

$$C'_u(V^{e,u}) = \gamma^e + \delta^e \frac{p'(a^e)}{p(a^e)} + \eta^e \frac{1}{p(a^e)(1-s)}, \tag{1.9}$$

$$C'_d(V^{e,d}) = \gamma^e - \delta^e \frac{p'(a^e)}{1-p(a^e)} - \eta^e \frac{1}{1-p(a^e)}. \tag{1.10}$$

The Envelope theorem further implies that

$$C'_e(V^e) = \gamma^e. \tag{1.11}$$

4.2. Unemployed Worker. The unemployed individual receives a transfer c^u from the planner. In addition, she is promised a utility of $V^{u,e}$ if she finds employment. If she remains employable but does not find a job, she is promised a utility level of $V^{u,u}$. If she becomes disabled, her promised utility is $V^{u,d}$. The individual takes these values as given and maximizes her utility with respect to the levels of prevention and search effort.

$$\max_{a^u, e^u \geq 0} \left\{ u(c^u) - a^u - e^u + \beta p(a^u) \left[q(e^u) V^{u,e} + \left[1 - q(e^u)\right] V^{u,u} \right] + \beta \left[1 - p(a^u)\right] V^{u,d} \right\}.$$

The first-order condition with respect to prevention effort is given by

$$\beta p'(a^u) \left[q(e^u) V^{u,e} + \left[1 - q(e^u)\right] V^{u,u} - V^{u,d} \right] \leq 1, \tag{1.12}$$

which has a similar interpretation as condition (1.4). The first-order condition with respect to search effort reads

$$\beta p(a^u) q'(e^u) \left[V^{u,e} - V^{u,u} \right] \leq 1. \tag{1.13}$$

A marginal increase in search effort yields an increase in the probability of finding employment. Multiplying this increase by the probability of remaining employable, $p(a^u)$, and the utility difference associated with finding a job, $V^{u,e} - V^{u,u}$, yields the benefit of a marginal increase in search effort.

Both first-order conditions hold with equality for $a^u, e^u > 0$, respectively.

An individual who was unemployed in the previous period was promised a utility of V^u in case of continued unemployment. The planner now has to fulfill this promise by providing the individual with consumption and by promising utility levels for all

three possible future states. The promise-keeping constraint can be written as

$$V^u \leq u(c^u) - a^u - e^u + \beta p(a^u) \left[q(e^u)V^{u,e} + (1 - q(e^u))V^{u,u}\right] + \beta \left[1 - p(a^u)\right] V^{u,d}. \quad (1.14)$$

Since the planner cannot distinguish unemployment from disability, the individual needs to be provided with an incentive to truthfully report her state of employability. The utility from being unemployed has to be at least as high as that from disability. The truth-telling constraint is given by

$$V^{u,d} \leq V^{u,u}. \quad (1.15)$$

The planner now solves the problem of providing consumption and promising utilities to an unemployed individual in a cost-minimizing and incentive compatible way

$$C_u(V^u) = \min_{c^u, a^u, e^u, V^{u,e}, V^{u,u}, V^{u,d}} \Big\{ c^u + \beta p(a^u) \left[q(e^u)C_e(V^{u,e}) + \left[1 - q(e^u)\right] C_u(V^{u,u})\right] + \beta \left[1 - p(a^u)\right] C_d(V^{u,d}) \Big\}$$

subject to the conditions (1.12) – (1.15). The first-order conditions are given by

$$\gamma^u = \frac{1}{u'(c^u)}, \quad (1.16)$$

$$C'_e(V^{u,e}) = \gamma^u + \delta^u \frac{p'(a^u)}{p(a^u)} + \mu^u \frac{q'(e^u)}{q(e^u)}, \quad (1.17)$$

$$C'_u(V^{u,u}) = \gamma^u + \delta^u \frac{p'(a^u)}{p(a^u)} - \mu^u \frac{q'(e^u)}{1 - q(e^u)} + \eta^u \frac{1}{p(a^u)[1 - q(e^u)]}, \quad (1.18)$$

$$C'_d(V^{u,d}) = \gamma^u - \delta^u \frac{p'(a^u)}{1 - p(a^u)} - \eta^u \frac{1}{1 - p(a^u)}. \quad (1.19)$$

where γ^u is the Lagrange parameter for the promise-keeping constraint (1.14), δ^u is the multiplier for the prevention-incentive constraint (1.12), μ^u is the multiplier for the search-incentive constraint (1.13), and $\beta\eta^u$ is the parameter for the truth-telling constraint (1.15).

The Envelope conditions is

$$C'_u(V^u) = \gamma^u. \quad (1.20)$$

4.3. Disabled Worker. Disability is an absorbing state, and the disabled individual cannot exert any search or prevention effort. Therefore, the planner does not have to consider any incentive constraints when providing consumption and continuation values to a disabled individual. The only constraint to consider is the promise-keeping constraint. For a disabled individual who was promised a utility of V^d, this constraint reads

$$V^d \leq u(c^d) + \beta V^{d,d}, \quad (1.21)$$

where c^d is the transfer to the disabled individual and $V^{d,d}$ is the utility promised in the next period of disability.

The planner's problem is now given by

$$C_d(V^d) = \min_{c^d, V^{d,d}} \Big\{ c^d + \beta C_d(V^{d,d}) \Big\}$$

subject to the promise-keeping constraint (1.21).

Letting γ^d be the Lagrange parameter on the promise-keeping constraint, the first-order conditions are

$$\gamma^d = \frac{1}{u'(c^d)}, \tag{1.22}$$

$$C_d'(V^{d,d}) = \gamma^d. \tag{1.23}$$

The Envelope condition reads

$$C_d'(V^d) = \gamma^d. \tag{1.24}$$

5. Optimal Insurance

Using the results from the previous section I am now able to derive the optimal insurance against unemployment and disability. In order to establish a benchmark case, I first assume that the planner has full information.

5.1. Full Information. Full information implies that the planner can distinguish between an unemployed and a disabled individual and that he is able to observe the individual's effort levels. The first assumption implies that the truth-telling constraint does not have to be considered, while the second assumption makes the prevention-incentive and the search-incentive constraints obsolete. The Lagrange parameters associated with those constraints can hence be set equal to zero in the first-order conditions.

5.1.1. *Employed Worker.* After setting the Lagrange parameters δ^e and η^e equal to zero, the first-order conditions with respect to the promised utilities together with the Envelope condition imply[5]

$$C_e'(V^e) = C_e'(V^{e,e}) = C_u'(V^{e,u}) = C_d'(V^{e,d}). \tag{1.25}$$

Because of the strict convexity of $C_e(\cdot)$, the first equality implies that the utility level of an employed worker remains constant while employed, i.e., $V^e = V^{e,e}$.

If an employed individual remains healthy and does not lose her job, she is guaranteed a utility of $V^{e,e}$ in the next period. Facing such an individual, the planner has first-order conditions similar to the ones presented above. In particular, the first-order

[5] See equations (1.8) – (1.11).

condition with respect to consumption and the Envelope condition are given by

$$\gamma^{e,e} = \frac{1}{u'(c^{e,e})} \quad \text{and} \quad C_e'(V^{e,e}) = \gamma^{e,e}.$$

This implies

$$C_e'(V^{e,e}) = \frac{1}{u'(c^{e,e})}.$$

For individuals that are employed in this period and unemployed or disabled in the next, a similar result can be derived

$$C_u'(V^{e,u}) = \frac{1}{u'(c^{e,u})} \quad \text{and} \quad C_d'(V^{e,d}) = \frac{1}{u'(c^{e,d})}.$$

Plugging these conditions, together with the Envelope condition (1.11), into equations (1.25) and using the strict concavity of the utility function yields

$$\frac{1}{u'(c^e)} = \frac{1}{u'(c^{e,e})} = \frac{1}{u'(c^{e,u})} = \frac{1}{u'(c^{e,d})} \quad \Leftrightarrow \quad c^e = c^{e,e} = c^{e,u} = c^{e,d}.$$

As expected in a case without asymmetric information and moral hazard, this implies full insurance against unemployment and disability and a constant consumption while remaining employed.

5.1.2. *Unemployed Worker.* If the Lagrange parameters δ^u, μ^u and η^u are all equal to zero, the first-order conditions with respect to the promised utilities and the Envelope condition[6] can be combined to yield

$$C_u'(V^u) = C_e'(V^{u,e}) = C_u'(V^{u,u}) = C_d'(V^{u,d}). \tag{1.26}$$

The strict convexity of $C_u(\cdot)$ implies that the utility level of an unemployed worker remains constant during unemployment, i.e., $V^u = V^{u,u}$.

As in the previous section, combining the first-order condition with respect to consumption and the Envelope condition of an individual who has been unemployed for two periods yields

$$C_u'(V^{u,u}) = \frac{1}{u'(c^{u,u})}.$$

Similarly, we have the following conditions for an individual who has been unemployed and then found a job and an individual who has been unemployed and then became disabled that

$$C_e'(V^{u,e}) = \frac{1}{u'(c^{u,e})} \quad \text{and} \quad C_d'(V^{u,d}) = \frac{1}{u'(c^{u,d})}.$$

[6] See equations (1.17) – (1.20).

Substituting theses conditions, together with the Envelope condition (1.20), into equation (1.26) and applying the strict concavity of the utility function yields

$$\frac{1}{u'(c^u)} = \frac{1}{u'(c^{u,e})} = \frac{1}{u'(c^{u,u})} = \frac{1}{u'(c^{u,d})} \quad \Leftrightarrow \quad c^u = c^{u,e} = c^{u,u} = c^{u,d}.$$

In other words, the unemployed worker is fully insured against disability and her consumption remains constant during unemployment.

5.1.3. *Disabled Worker.* The first-order condition with respect to promised utility (1.23) and the Envelope condition (1.24) can be combined to yield

$$C_d'(V^d) = C_d'(V^{d,d}). \tag{1.27}$$

Due to the strict convexity of the cost function, this implies that $V^d = V^{d,d}$, i.e., constant utility during disability.

As before, the first-order condition with respect to consumption and the Envelope condition of an individual who has been disabled for two periods together imply

$$C_d'(V^{d,d}) = \frac{1}{u'(c^{d,d})}.$$

This can again be plugged into equation (1.27), together with the Envelope condition (1.24), to yield

$$\frac{1}{u'(c^d)} = \frac{1}{u'(c^{d,d})} \quad \Leftrightarrow \quad c^d = c^{d,d}.$$

Once a worker becomes disabled his consumption and utility levels remain constant.

5.2. Asymmetric Information. After having derived the optimal insurance against unemployment and disability in the full information setting, I turn now to the case of asymmetric information. While it was optimal to have perfect insurance against unemployment and disability in the full information case, this is no longer incentive compatible in the asymmetric information setting. Here, the planner is neither able to observe effort levels nor can he distinguish between an unemployed and a disabled individual. The planner merely observes whether the individual has income or not. The individual has to be provided with incentives to prevent disability, search for employment and truthfully report her state of health. When providing consumption and continuation values, the planner has, therefore, to consider the prevention-incentive, the search-incentive and the truth-telling constraint as well as the promise-keeping constraint.

It is straightforward to show that the promise-keeping constraints are always binding in equilibrium. Also, because of the continuous effort levels, the prevention-incentive constraints and the search-incentive constraint are always binding as well.

Whether the truth-telling constraints are binding or slack cannot be determined analytically. We, therefore, know the following about the Lagrange parameters

$$\gamma^e, \delta^e, \gamma^u, \delta^u, \mu^u > 0 \quad \text{and} \quad \eta^e, \eta^u \geq 0.$$

5.2.1. *Employed Worker.* A planner facing an employed individual has the following first-order conditions with respect to the promised utilities[7]

$$\begin{aligned}
C'_e(V^{e,e}) &= \gamma^e + \delta^e \frac{p'(a^e)}{p(a^e)}, \\
C'_u(V^{e,u}) &= \gamma^e + \delta^e \frac{p'(a^e)}{p(a^e)} + \eta^e \frac{1}{p(a^e)(1-s)}, \\
C'_d(V^{e,d}) &= \gamma^e - \delta^e \frac{p'(a^e)}{1-p(a^e)} - \eta^e \frac{1}{1-p(a^e)}.
\end{aligned}$$

Since $\gamma^e, \delta^e > 0$ and $\eta^e \geq 0$ these conditions, together with the Envelope condition (1.11), imply that

$$C'_u(V^{e,u}) \geq C'_e(V^{e,e}) > C'_e(V^e) > C'_d(V^{e,d}). \tag{1.28}$$

As in the full-information case, these inequalities of marginal costs can be transformed to inequalities in transfers

$$c^{e,u} \geq c^{e,e} > c^e > c^{e,d}.$$

The consumption of an employed worker increases as long as she remains healthy and employed. Increasing consumption does not, however, necessarily imply a positive transfer from the planner. Since the worker earns a positive wage, increasing consumption can also be achieved by a decreasing wage tax, i.e., $c^e - w < c^{e,e} - w < 0$.

Moreover, consumption increases when the individual becomes unemployed and decreases at the time of disability. It can be seen below that the former does not necessarily imply a higher utility when becoming unemployed since the individual will have to exert search effort. In either case, there is no problem of moral hazard because job loss is exogenous. The drop in consumption upon disability, together with the increasing consumption while employed, creates an incentive to prevent disability.

In (1.28) it can further be seen that

$$C'_e(V^{e,e}) > C'_e(V^e) \quad \Leftrightarrow \quad V^{e,e} > V^e,$$

which implies that the continuation value of an employed worker increases with tenure. The truth-telling constraint further implies that $V^{e,u} \geq V^{e,d}$.

[7] See equations (1.8) – (1.10).

Concerning the relationship between the current level of utility V^e and the promised utility in case of disability, the following proposition holds.

> **Proposition 1:** *The utility of a worker is decreasing when she becomes disabled.*

The results can be summarized as follows

$$V^{e,e} > V^e > V^{e,d} \quad \text{and} \quad V^{e,u} \geq V^{e,d}.$$

Consumption and utility of an employed individual are increasing with tenure. This stands in contrast to the results of Hopenhayn and Nicolini (1997) where consumption and utility are constant while working. However, it resembles the conclusions from models with disutility of work, e.g. Hopenhayn and Nicolini (2009). In that context, it is necessary to provide incentives to remain employed because work decreases utility, while here it is necessary because keeping up employability requires effort. If the individual becomes disabled, her consumption and her utility decrease. This decrease together with the increasing utility while healthy and employed gives her an incentive to prevent disability. In the case of unemployment, the individual's consumption increases. This increase compensates for the search effort the individual will have to exert. Whether the utility level increases when becoming unemployed is unclear. What is certain, is that the utility when unemployed is (weakly) larger than the utility when disabled.

5.2.2. *Unemployed Worker.* The planner's first-order conditions with respect to promised utilities in the case of an unemployed individual are given by[8]

$$\begin{aligned}
C'_e(V^{u,e}) &= \gamma^u + \delta^u \frac{p'(a^u)}{p(a^u)} + \mu^u \frac{q'(e^u)}{q(e^u)}, \\
C'_u(V^{u,u}) &= \gamma^u + \delta^u \frac{p'(a^u)}{p(a^u)} - \mu^u \frac{q'(e^u)}{1-q(e^u)} + \eta^u \frac{1}{p(a^u)[1-q(e^u)]}, \\
C'_d(V^{u,d}) &= \gamma^u - \delta^u \frac{p'(a^u)}{1-p(a^u)} - \eta^u \frac{1}{1-p(a^u)}.
\end{aligned}$$

The Envelope condition (1.20) and the fact that $\gamma^u, \delta^u, \mu^u > 0$ and $\eta^u \geq 0$ together imply

$$C'_e(V^{u,e}) > C'_u(V^u) > C'_d(V^{e,d}). \tag{1.29}$$

Applying the same logic as before yields

$$c^{u,e} > c^u > c^{u,d}.$$

An unemployed worker has higher consumption in the next period if she finds employment and remains healthy, giving her an incentive to both prevent disability and exert

[8] See equations (1.17) – (1.19).

search effort. If she becomes disabled, her consumption level will be lower than under unemployment, which provides an additional incentive to exert prevention effort.

If the individual remains healthy but does not succeed in finding a job, the change in utility, and hence consumption, is ambiguous. In other words, it is not clear wether $V^{u,u}$ and $c^{u,u}$ are larger, smaller or equal to V^u and c^u, respectively. The intuition is that the incentive to search for a job has a negative effect on the promised utility while the incentive to prevent disability and the truth-telling constraint both have a positive one. Decreasing benefits and utilities as in Hopenhayn and Nicolini (1997) are therefore not necessarily optimal in this model.[9]

While it may seem that this ambiguity can only be resolved numerically, it turns out that there is a condition under which this trade-off between providing incentives for search and providing incentives for prevention disappears.

A decrease in the promised utility in case of continued unemployment, $V^{u,u}$, increases the incentives to search for a job and thereby also search effort, e^u. This can be seen in the search-incentive constraint (1.13). In the prevention-incentive constraint (1.12), a decrease in $V^{u,u}$ has a direct and an indirect effect on the value of remaining employable, which is a weighted average of the value of being employed and the value of being unemployed

$$q(e^u)V^{u,e} + [1 - q(e^u)]\, V^{u,u}. \tag{1.30}$$

The direct effect is that the value of remaining employable decreases because $V^{u,u}$ decreases. This reduces the incentives to exert prevention effort. The indirect effect comes from the fact that a decrease in $V^{u,u}$ increases search effort, and therefore the probability of reemployment, $q(e^u)$. This in turn increases the value of remaining healthy, because the value of becoming employed is larger than the value of remaining unemployed. The indirect and the direct effect are therefore of opposing signs. If the positive indirect effect is as least as large as the negative direct effect, then the value of remaining employable is not reduced when $V^{u,u}$ decreases.

In other words, if individuals react sufficiently to changed search incentives, the positive effect of decreasing utilities during unemployment on prevention effort outweighs the negative one and there is no trade-off between providing incentives for search and prevention. How much individuals react to changed search incentives depends on the concavity of $q(e)$. The less concave the function $q(e)$ is, the more individuals react to a given change in promised utility during unemployment. In order for individuals to react sufficiently, $q(e)$ can therefore not be too concave. This implies that the convex

[9] As an example, consider a case where the probability of finding employment is independent of search effort, i.e., $q(e) = \bar{q}$. Then, the negative effect of the search constraint disappears and benefits as well as utility are increasing during unemployment.

function $1-q(e)$ cannot be too convex, which is the case if $1-q(e)$ is log-concave, i.e., the log of the convex function is concave.[10] This condition is stated in the following proposition.

Proposition 2: *If the search technology satisfies the following condition*

$$-\frac{1-q(e)}{q'(e)} \geq \frac{q'(e)}{q''(e)}, \tag{1.31}$$

then decreasing utilities during unemployment lead to constant or increasing levels of prevention effort.

Examples of functions that satisfy this condition are the the logistical and the exponential distribution functions. The latter is commonly used in the literature as a search function.

Proposition 2 establishes that if individuals react sufficiently to search incentives, prevention effort does not decrease when the promised utility of continued unemployment decreases. Conversely, one could think that there is a condition for the responsiveness towards prevention incentives which makes the optimal search effort non-decreasing when $V^{u,u}$ increases.[11] However, using the same techniques as in the proof of Proposition 2, it is possible to show that there exists no such condition.

If the search function $q(e)$ fulfills condition (1.31), the positive effect of the prevention-constraint in the first-order condition with respect to the promised utility in case of continued unemployment disappears or becomes negative. Since there is also the negative effect of the search-incentive constraint as well as the (possibly) positive effect of the truth-telling constraint, it remains unclear how consumption and utility change during unemployment. To shed further light on the analysis the following lemma can be established.

Lemma 1: *For promised utility and consumption to increase or remain constant during unemployment, the truth-telling constraint has to be binding.*

[10] A function $f(x)$ is log-concave if $f(x)f''(x) \leq [f'(x)]^2$. The function $1-q(e)$ is therefore log-concave if

$$[1-q(e)][-q''(e)] \leq [-q'(e)]^2 \quad \Leftrightarrow \quad -\frac{1-q(e)}{q'(e)} \geq \frac{q'(e)}{q''(e)}.$$

[11] The intuition here would be that an increase in $V^{u,u}$ increases the incentives to prevent disability, and hence prevention effort. This can be seen in the prevention-incentive constraint (1.12). In the search-incentive constraint (1.13) there is then a negative direct effect through the increase in $V^{u,u}$ and an indirect positive effect through the increase in prevention effort. Search effort would be non-decreasing if the indirect positive effect is as least as large as the negative direct effect.

Increasing or constant utilities when remaining unemployed, together with the binding truth-telling constraint, imply that

$$V^u \leq V^{u,u} = V^{u,d} \quad \Leftrightarrow \quad V^u \leq V^{u,d}.$$

In other words, in order to have increasing or constant utilities during unemployment, the utility of an unemployed agent has to increase at the time of disability. However, the following proposition establishes that increasing utility upon disability cannot be optimal.

> **Proposition 3:** *The utility of an unemployed agent decreases when she becomes disabled.*

Since an increase in utility when going from unemployment into disability is not optimal, consumption and promised utility can neither increase nor remain constant during unemployment. Therefore, proposition 3 together with lemma 1 implies that decreasing utilities and benefits during unemployment, as in Hopenhayn and Nicolini (1997), are optimal if individuals react sufficiently to changed search incentives.

Finally, the binding search-incentive constraint demands that the promised value of employment is strictly larger than the corresponding value for unemployment, $V^{u,e} > V^{u,u}$. The truth-telling constraint, on the other hand, implies that the promised value of unemployment has to be weakly larger than the value of disability, $V^{u,u} \geq V^{u,d}$.

The results can then be summarized as follows

$$V^{u,e} > V^{u,u} \geq V^{u,d} \quad \text{and} \quad V^u > V^{u,u} \geq V^{u,d}.$$

Finding a job guarantees higher utility and higher consumption compared to further unemployment. This utility promise, together with the decreasing utility and consumption during unemployment, provides an incentive to exert search effort. However, remaining unemployed yields a weakly higher utility than becoming disabled, which generates an incentive to be truthful about the health state. This, together with the higher utility when finding a job, creates an incentive to prevent disability and remain in the labor force.

5.2.3. *Disabled Worker.* The first-order condition associated with the continuation value of a disabled individual together with the corresponding Envelope condition implies[12]

$$C_d'(V^d) = C_d'(V^{d,d}) \quad \Leftrightarrow \quad c^d = c^{d,d} \quad \text{and} \quad V^{d,d} = V^d.$$

Since disability is an absorbing state and there are no incentive problems to consider, it is optimal to provide the disabled individual with a constant utility and a constant stream of consumption.

[12] See equations (1.23) and (1.24).

6. Numerical Simulations

In order to illustrate the results derived in the previous section and to answer the questions which cannot be solved analytically, I simulate the model numerically. Since the model is computationally very intensive, I focus my analysis on the problem of the unemployed individual. I assume that employment is an absorbing state and that the utility of finding employment is therefore given by $V^e = u(w)/(1-\beta)$. Hence, the planner only decides on the utility promises for unemployment and disability.

Regarding functional forms and parameters, most of my assumptions are in accordance with the calibration of Hopenhayn and Nicolini (1997). A model period is one week. The discount factor β is equal to 0.999, which implies a yearly discount rate of 0.95. The utility function has the form

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma},$$

where $\sigma = 0.5$. The hazard functions are given by the exponential distribution function. The probability of remaining healthy depends on prevention effort, a, in the following way

$$p(a) = 1 - \exp(-\phi a).$$

The probability of finding employment is

$$q(e) = 1 - \exp(-\theta e),$$

where e is search effort. As mentioned above, this functional form is commonly used in the literature and fulfills condition (1.31).

The values of the parameters ϕ and θ are assigned in the following way. In an autarky setting, without promised utilities and with wage as the only source of income, the agent's problem is stationary. The resulting effort levels and probabilities are constant over time. The values of the parameters ϕ and θ are then chosen to yield a probability of 99.99% of remaining healthy from one week to the next and a probability of 15.12% for an unemployed individual to find new employment in autarky. This is similar to the numbers in Hopenhayn and Nicolini (1997).

Table 1.1 presents the parameter values used in the simulation. Table 1.2 shows a simulation of an individual who remains unemployed for one year. All promised utilities as well as consumption are decreasing over time. The replacement ratio decreases from 100.3% to 30.3% during the first year of unemployment. The promised disability insurance replacement ratio decreases from 23.9% to 7.5%. Search effort is increasing during unemployment, yielding an increasing probability of finding employment. Prevention effort also increases, but the probability of remaining healthy increases very little.

Parameter	Value	Description
β	0.999	Discount factor
ϕ	1.0	Probability parameter
σ	0.5	Parameter of risk aversion
θ	0.0005	Probability parameter
w	100	Wage

TABLE 1.1. Parameter Values

Weeks	UI Rep. Ratio (%)	DI Rep. Ratio (%)	Prob. Healthy	Prob. Employment
1	100.2895	23.9287	0.9999	0.0905
10	88.8153	21.3006	0.9999	0.0965
20	74.7624	18.0558	0.9999	0.1034
30	60.0244	14.6150	0.9999	0.1105
40	45.6098	11.2052	0.9999	0.1174
52	30.2666	7.5189	0.9999	0.1252

TABLE 1.2. Simulation of an Unemployed Worker's Values

In summary, consumption and all promised utilities are decreasing during unemployment in order to create search incentives. The latter means that the disability benefits are decreasing in the length of the unemployment spell. This resembles the results of Hopenhayn and Nicolini (1997) who show that it is optimal to reduce all claims to future consumption in order to create incentives to leave unemployment.

These results can now be compared to a policy in the spirit of Hopenhayn and Nicolini (1997) where the planner chooses only the utility during unemployment and the possibility of disability is not taken into account. I assume that individuals who become disabled in this setting receive the same benefit as unemployed individuals. Figure 1.1 shows a simulation of the model presented in this paper (continuous line) and a model where disability is not taken into account (dashed line).

The upper left panel shows the utility of an unemployed individual under the two policies over 52 weeks. Both policies imply decreasing utility during unemployment. As expected, the Hopenhayn-Nicolini policy is less generous. Since the planner's only problem there is to provide search incentives, the utility can decrease faster than in the model presented in this paper where the planner provides incentives for search as well as for remaining in the labor force. In the upper right panel the utility promises in case of disability are plotted over 52 weeks of unemployment. I assume that under the Hopenhayn-Nicolini policy individuals who become disabled receive unemployment

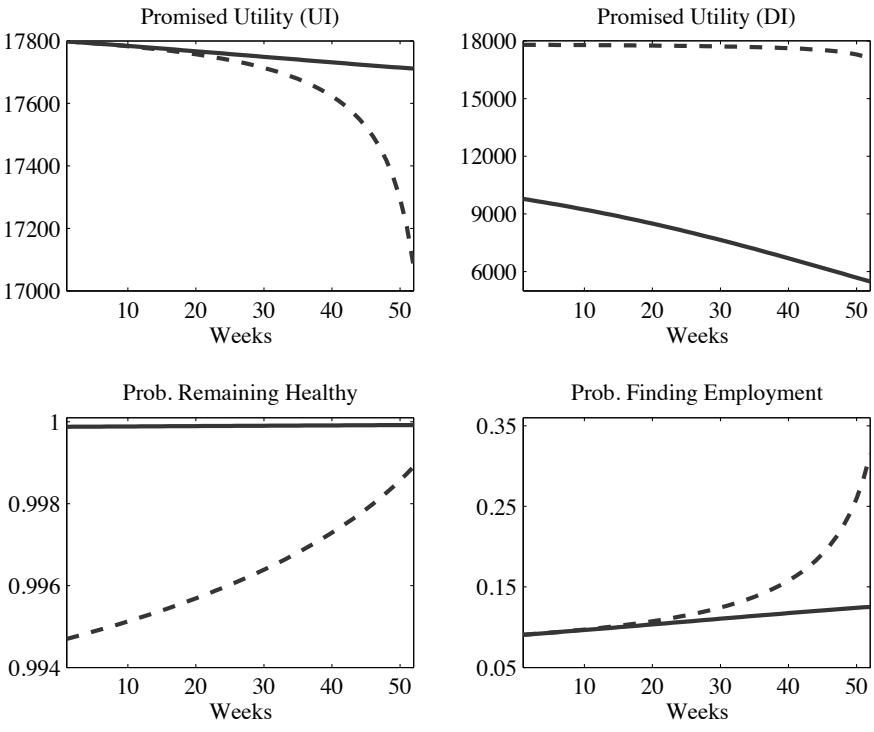

FIGURE 1.1. Simulation of an Unemployed Worker's Values. Optimal Insurance against Unemployment and Disability (continuous line), Hopenhayn-Nicolini Insurance (dashed line).

benefits. This makes the disability insurance more generous compared to the optimal insurance derived in this paper. This generosity, together with the less generous unemployment insurance, generates weaker incentives to prevent disability. The latter can be seen in the lower left panel where the probability or remaining healthy for another period is shown. Finally, in the lower right panel it can be seen that the probability of finding employment increases during unemployment under both policies, but the increase is steeper under the Hopenhayn-Nicolini policy.

7. Conclusion

In this paper, I derive the optimal insurance against unemployment and disability in a private information economy with endogenous health and search effort. Introducing endogenous health makes the analysis significantly more difficult but it addresses a

crucial shortcoming in the literature, namely the possibility of individuals to leave the labor force and go into other insurance systems such as disability insurance. If individuals only have a choice between being unemployed or working, providing them with incentives to search for a job is easier than when they also have the option to exit the labor force.

I demonstrate that the optimal sequence of consumption and promised utilities of an employed worker is increasing with tenure. This implies among other things that the promised disability benefits are increasing in the time spent working. Once a worker is disabled she will receive constant benefits and her utility will remain constant. Both findings are in line with the results of Diamond and Mirrlees (1978). For an unemployed worker I show that the prevention constraint implies increasing benefits and utility levels during unemployment while the search constraint has the opposite effect. Decreasing benefits and utilities as in Hopenhayn and Nicolini (1997) are therefore not necessarily optimal in this model. I show, however, that the search effect dominates the prevention effect if individuals respond sufficiently much to a change in search incentives. The sequence of consumption and utility levels is then decreasing during unemployment. Moreover, I show numerically that the promised utility of disability is decreasing as well. This means that disability benefits are no longer monotonically increasing in the retirement age as in Diamond and Mirrlees (1978), instead they are lower the longer the individual was previously unemployed.

References

Abreu, D., Pearce, D., and Stacchetti, E. (1990), 'Toward a Theory of Discounted Repeated Games with Imperfect Monitoring', *Econometrica*, 58, 1041–1063.

Acemoglu, D. and Shimer, R. (1999), 'Efficient Unemployment Insurance', *Journal of Political Economy*, 107, 893–928.

Atkeson, A. and Lucas, R. (1992), 'On Efficient Distribution with Private Information', *The Review of Economic Studies*, 59, 427–453.

Baily, M. (1978), 'Some Aspects of Optimal Unemployment Insurance', *Journal of Public Economics*, 10, 379–402.

Boone, J., Fredriksson, P., Holmlund, B., and van Ours, J. (2007), 'Optimal Unemployment Insurance with Monitoring and Sanctions', *The Economic Journal*, 117, 399–421.

Chang, R. (1998), 'Credible Monetary Policy in an Infinite Horizon Model: Recursive Approaches', *Journal of Economic Theory*, 81, 431–461.

Diamond, P. and Mirrlees, J. (1978), 'A Model of Social Insurance with Variable Retirement', *Journal of Public Economics*, 10, 295–336.

Flemming, J. (1978), 'Aspects of Optimal Unemployment Insurance', *Journal of Public Economics*, 10, 403–425.

Fredriksson, P. and Holmlund, B. (2001), 'Optimal Unemployment Insurance in Search Equilibrium', *Journal of Labor Economics*, 2, 370–399.

Fredriksson, P. and Holmlund, B. (2006), 'Improving Incentives in Unemployment Insurance: A Review of Recent Research', *Journal of Economic Surveys*, 20, 357–386.

Golosov, M. and Tsyvinski, A. (2006), 'Designing Optimal Disability Insurance: A Case for Asset Testing', *Journal of Political Economy*, 114, 257–279.

Hagedorn, M., Kaul, A., and Mennel, T. (2010), 'An Adverse Selection Model of Optimal Unemployment Insurance', *Journal of Economic Dynamics and Control*, 34, 490–502.

Höglin, E. (2008), 'Inequality in the Labor Market: Insurance, Unions, and Discrimination', *PhD Thesis*, Stockholm School of Economics.

Hopenhayn, H. and Nicolini, J. (1997), 'Optimal Unemployment Insurance', *Journal of Political Economy*, 105, 412–438.

Hopenhayn, H. and Nicolini, J. (2009), 'Optimal Unemployment Insurance and Employment History', *The Review of Economic Studies*, 76, 1049–1070.

Jönsson, L., Palme, M., and Svensson, I. (2012), 'Disability Insurance, Population Health and Employment in Sweden', *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, ed. by Wise, D., University of Chicago Press, forthcoming.

Karlström, A., Palme, M., and Svensson, I. (2008), 'The Employment Effect of Stricter Rules for Eligibility for DI: Evidence from a Natural Experiment in Sweden', *Journal of Public Economics*, 92, 2071–2082.

Larsson, L. (2006), 'Sick of Being Unemployed? Interactions between Unemployment and Sickness Insurance', *Scandinavian Journal of Economics*, 108, 97–113.

Pavoni, N. (2007), 'On Optimal Unemployment Compensation', *Journal of Monetary Economics*, 54, 1612–1630.

Pavoni, N. and Violante, G. (2007), 'Optimal Welfare-to-Work Programs', *The Review of Economic Studies*, 74, 283–318.

Shavell, S. and Weiss, L. (1979), 'The Optimal Payment of Unemployment Insurance Benefits over Time', *Journal of Political Economy*, 87, 1347–1362.

Spear, S. and Srivastava, S. (1987), 'On Repeated Moral Hazard with Discounting', *The Review of Economic Studies*, 54, 599–617.

Thomas, J. and Worrall, T. (1990), 'Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem', *Journal of Economic Theory*, 51, 367–390.

Wang, C. and Williamson, S. (1996), 'Unemployment Insurance with Moral Hazard in a Dynamic Economy', *Carnegie-Rochester Conference Series on Public Policy*, 44, 1–41.

Appendix

A1. Proof of Proposition 1. Since the incentive constraint (1.4) is fulfilled, the equilibrium effort level will give the individual the highest possible utility. This utility level is, therefore, weakly larger than the utility implied by zero effort. The promise-keeping constraint (1.5), which is binding in equilibrium, can then be rearranged to yield

$$\begin{aligned} V^e &= u(c^e) - a^e + \beta p(a^e)\left[sV^{e,e} + (1-s)V^{e,u}\right] + \beta\left[1 - p(a^e)\right]V^{e,d} \\ &\geq u(c^e) - 0 + \beta p(0)\left[sV^{e,e} + (1-s)V^{e,u}\right] + \beta\left[1 - p(0)\right]V^{e,d} \\ &= u(c^e) + \beta V^{e,d}, \end{aligned} \tag{1.32}$$

where the last equality comes from the fact that $p(0) = 0$.

The first-order condition (1.28) implies that

$$C_e'(V^e) > C_d'(V^{e,d}),$$

which can be transformed to yield

$$\frac{1}{u'(c^e)} > \frac{1}{u'(c^{e,d})} \quad \Leftrightarrow \quad u(c^e) > u(c^{e,d}) \quad \Leftrightarrow \quad \frac{u(c^e)}{1-\beta} > \frac{u(c^{e,d})}{1-\beta}.$$

Since disability is an absorbing state, consumption will be constant at the optimum and the latter inequality can be rewritten as

$$\frac{u(c^e)}{1-\beta} > \frac{u(c^{e,d})}{1-\beta} = V^{e,d} \quad \Leftrightarrow \quad u(c^e) > (1-\beta)V^{e,d}. \tag{1.33}$$

(1.32) and (1.33) can then be combined to yield

$$V^e \geq u(c^e) + \beta V^{e,d} > (1-\beta)V^{e,d} + \beta V^{e,d} = V^{e,d} \quad \Leftrightarrow \quad V^e > V^{e,d} \quad \square$$

A2. Proof of Proposition 2. The value of being employable (1.30) is affected by the promised utility in case of continued unemployment, $V^{u,u}$, in two ways: directly through the promised utility and indirectly through search effort, e^u, which depends on $V^{u,u}$. The value of being employable can therefore be expressed as a function of $V^{u,u}$

$$F(V^{u,u}) \equiv q\left(e^u\left(V^{u,u}\right)\right)V^{u,e} + \left[1 - q\left(e^u\left(V^{u,u}\right)\right)\right]V^{u,u}.$$

If $V^{u,u}$ decreases, the direct effect reduces the value of being employable while the indirect effect increases the value. If the positive effect is as least as large as the negative one, the value of F does not increase with $V^{u,u}$, i.e.,

$$\frac{\mathrm{d}F(V^{u,u})}{\mathrm{d}V^{u,u}} = q'\left(e^u\left(V^{u,u}\right)\right)\frac{\partial e^u\left(V^{u,u}\right)}{\partial V^{u,u}}\left[V^{u,e} - V^{u,u}\right] + 1 - q\left(e^u\left(V^{u,u}\right)\right) \leq 0,$$

which implies for the change in effort

$$\frac{\partial e^u\left(V^{u,u}\right)}{\partial V^{u,u}} \leq -\frac{1-q\left(e^u\left(V^{u,u}\right)\right)}{q'\left(e^u\left(V^{u,u}\right)\right)}\frac{1}{\left[V^{u,e}-V^{u,u}\right]}. \tag{1.34}$$

Equation (1.34) describes how equilibrium effort, e^u, should react to a change in the promised utility in case of continued unemployment, $V^{u,u}$, so that the value of being employable is not reduced when $V^{u,u}$ is decreasing.

How much equilibrium effort actually changes can be derived from the search-incentive constraint (1.13)

$$\beta p(a^u)q'(e^u)\left[V^{u,e}-V^{u,u}\right]=1 \;\Leftrightarrow\; q'(e^u)=\frac{1}{\beta p(a^u)\left[V^{u,e}-V^{u,u}\right]}\equiv G(V^{u,u}). \tag{1.35}$$

The equilibrium effort is then given by

$$e^u\left(V^{u,u}\right)=\left(q'\right)^{-1}\left[G(V^{u,u})\right],$$

and the derivative with respect to $V^{u,u}$ is

$$\frac{\partial e^u\left(V^{u,u}\right)}{\partial V^{u,u}}=\frac{\partial\left(q'\right)^{-1}\left[G(V^{u,u})\right]}{\partial G(V^{u,u})}\frac{\partial G(V^{u,u})}{\partial V^{u,u}}, \tag{1.36}$$

where $\left(q'\right)^{-1}$ is the inverse of the first derivative of the search function $q(\cdot)$. The first part of this derivative can be transformed using the inverse function theorem[13]

$$\frac{\partial\left(q'\right)^{-1}\left[G(V^{u,u})\right]}{\partial G(V^{u,u})}=\frac{1}{q''\left(\left(q'\right)^{-1}\left[G(V^{u,u})\right]\right)}=\frac{1}{q''\left(e^u(V^{u,u})\right)}. \tag{1.37}$$

The second part of the derivative in (1.36) is given by

$$\frac{\partial G(V^{u,u})}{\partial V^{u,u}}=\frac{\partial}{\partial V^{u,u}}\left(\frac{1}{\beta p(a^u)\left[V^{u,e}-V^{u,u}\right]}\right)=\frac{1}{\beta p(a^u)\left[V^{u,e}-V^{u,u}\right]^2}.$$

Substituting the second equality in (1.35) gives

$$\frac{\partial G(V^{u,u})}{\partial V^{u,u}}=q'\left(e^u(V^{u,u})\right)\frac{1}{\left[V^{u,e}-V^{u,u}\right]}. \tag{1.38}$$

Plugging in the two expressions from (1.37) and (1.38) into equation (1.36) yields

$$\frac{\partial e^u\left(V^{u,u}\right)}{\partial V^{u,u}}=\frac{q'\left(e^u(V^{u,u})\right)}{q''\left(e^u(V^{u,u})\right)}\frac{1}{\left[V^{u,e}-V^{u,u}\right]} \tag{1.39}$$

[13] The theorem states that

$$\left(f^{-1}\right)'(b)=\frac{1}{f'(f^{-1}(b))}=\frac{1}{f'(a)},$$

where $b=f(a)$.

The expression in (1.39) shows how much the optimal search effort will react to a change in the promised utility of continued unemployment. Equation (1.34) demonstrates how much search effort should change in order for a decrease in promised utility of unemployment not to reduce the value of remaining employable. Combining equations (1.34) and (1.39) yields

$$-\frac{1-q(e^u)}{q'(e^u)} \geq \frac{q'(e^u)}{q''(e^u)}.$$

A3. Proof of Lemma 1. If condition (1.31) is satisfied, the first-order condition with respect to $V^{u,u}$ reads

$$C'_u(V^{u,u}) = \gamma^u - \delta^u \frac{p'(a^u)}{p(a^u)} - \mu^u \frac{q'(e^u)}{1-q(e^u)} + \eta^u \frac{1}{p(a^u)[1-q(e^u)]}.$$

Since the Envelope condition is given by $C'_u(V^u) = \gamma^u$, increasing or constant utilities (and hence consumption) during unemployment are optimal if

$$\gamma^u - \delta^u \frac{p'(a^u)}{p(a^u)} - \mu^u \frac{q'(e^u)}{1-q(e^u)} + \eta^u \frac{1}{p(a^u)[1-q(e^u)]} \geq \gamma^u,$$

or

$$\eta^u \geq \delta^u p'(a^u)[1-q(e^u)] + \mu^u p(a^u) q'(e^u) > 0,$$

i.e., the Lagrange parameter associated with the truth-telling constraint has to be positive, and hence the truth-telling constraint has to be binding.

A4. Proof of Proposition 3. Since the incentive constraints (1.12) and (1.13) are fulfilled, the equilibrium effort levels will give the individual a weakly higher utility than zero search and prevention effort. The binding promise-keeping constraint (1.14) can then be rearranged to yield

$$\begin{aligned} V^u &= u(c^u) - a^u - e^u + \beta p(a^u)\left[q(e^u)V^{u,e} + (1-q(e^u))V^{u,u}\right] + \beta\left[1-p(a^u)\right]V^{u,d} \\ &\geq u(c^u) - 0 - 0 + \beta p(0)\left[q(0)V^{u,e} + (1-q(0))V^{u,u}\right] + \beta\left[1-p(0)\right]V^{u,d} \\ &= u(c^u) + \beta V^{u,d}, \end{aligned} \tag{1.40}$$

where the last equality comes from the fact that $p(0) = q(0) = 0$.

The first-order condition (1.29) implies that

$$C'_u(V^u) > C'_d(V^{u,d})$$

which can be transformed to yield

$$\frac{1}{u'(c^u)} > \frac{1}{u'(c^{u,d})} \quad \Leftrightarrow \quad u(c^u) > u(c^{u,d}) \quad \Leftrightarrow \quad \frac{u(c^u)}{1-\beta} > \frac{u(c^{u,d})}{1-\beta}.$$

Since consumption is constant once the individual is disabled, this condition can be rewritten as

$$\frac{u(c^u)}{1-\beta} > \frac{u(c^{u,d})}{1-\beta} = V^{u,d} \quad \Leftrightarrow \quad u(c^u) > (1-\beta)V^{u,d}. \tag{1.41}$$

(1.40) and (1.41) can then be combined to yield

$$V^u \geq u(c^u) + \beta V^{u,d} > (1-\beta)V^{u,d} + \beta V^{u,d} = V^{u,d} \quad \Leftrightarrow \quad V^u > V^{u,d} \quad \square$$

PAPER 2

A Life Cycle Model of Health and Retirement: The Case of Swedish Pension Reform

Tobias Laun Johanna Wallenius

Abstract In this paper we develop a life cycle model of labor supply and retirement to study the interactions between health and the labor supply behavior of older workers, in particular disability insurance and pension claiming. In our framework, individuals choose when to stop working and, given eligibility criteria, when/if to apply for disability and pension benefits. Individuals care about their health and can partially insure against health shocks by investing in health. Sweden is one of the few Western economies to have undertaken a large pension reform in recent years. We use our framework to study the labor supply implications of this reform and find that the new pension system creates big incentives for the continued employment of older workers.

1. Introduction

Faced with ageing populations and the looming insolvency of social security, governments the world over are grappling with the question of how to reform retirement programs. Understanding how changes to retirement programs affect life cycle labor supply, particularly retirement behavior, is critical for assessing the effects of these changes on allocations, welfare and government finances. Accurate assessments require a model of retirement that captures the key forces underlying retirement decisions.

Various institutional features have potentially large implications for the labor supply behavior of older workers; key among them are the design of pension systems, disability insurance and healthcare. Disability insurance is particularly relevant, as in many countries a large fraction of retirement occurs before the normal retirement age. A discussion of disability insurance programs naturally leads to a discussion of health, as disability insurance programs without exception have some eligibility criteria

The authors thank Lars Ljungqvist and David Domeij for their valuable suggestions. We also thank seminar participants at Uppsala University for their comments. Laun gratefully acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation at Svenska Handelsbanken.

regarding health. Health, in turn, has potential implications for labor supply outcomes both directly and through the healthcare system.

In this paper, we construct a life cycle model of labor supply and retirement, which enables us to study the interactions between health, disability insurance and old-age retirement benefits. The key features of our framework are that individuals choose when to stop working and, given eligibility criteria, when/if to apply for disability and pension benefits. Individuals care about their health and can partially insure against health shocks by investing in health. The fact that people can impact their own health and choose whether or not to claim disability benefits, are novel features in relation to the existing literature.

While many countries have come to understand the need for social security reform, Sweden is one of the few countries to have actually undertaken a major pension reform in recent years. While there are big expectations of the reform, to the best of our knowledge no formal analysis of the expected implications of the changes to the pension system exists to date. We use our model to study the labor supply implications of the recent Swedish pension reform and to ask what particular aspects of the reform are driving the results. Our interest in Sweden is spurred by the unique nature of the large changes to social security, but also by some of the distinguishing country characteristics. Much of the existing literature on social security, particularly disability insurance claiming, has focused on the United States. Many of the institutional features in much of Europe, including Sweden, differ drastically from those in the United States. This is of course true of social insurance programs, but also of healthcare. Unlike the United States where many people receive health insurance through their employer, Sweden has a public healthcare system. The fact that Medicare eligibility in the United States starts at age 65 creates a potentially large incentive for people to continue working until then. This motive is absent in Sweden, and more broadly most of Europe.

Sweden is in the process of switching from a pay-as-you-go (PAYG), defined benefit program to a notional, pay-as-you-go, defined contribution plan. The first benefits from the new system were paid out in 2001. But due to the gradual phasing-in of the reform, not until 2040 will all benefits be paid from the new system. There are many issues inherent with the old Swedish pension system, key among them the fact that the pension benefit is based on earnings from only the 15 highest years and only income up to a relatively low ceiling counts toward the benefit. Not only does this have the potential to treat workers with equivalent lifetime earnings very unequally, it does not provide incentives for older individuals to remain employed. In fact, given that wages tend to level off in the 40s or 50s, there is no expected increase in pension benefits from continued employment for the majority of older workers. Furthermore,

the system is sensitive to demographic change. One should also note that under the old system disability insurance is very generous. The new pension scheme hopes to address these issues.

We find that the Swedish pension reform creates large incentives for older workers to continue working longer. Our main findings are threefold: (1) the model predicts an increase in the average retirement age of 2.3 years from 62.4 to 64.7, and (2) there is an increasing tendency for workers to continue working while collecting pension benefits, but (3) the fraction of older workers claiming disability insurance only declines by roughly one percentage point, from 18.6% to 17.7%. To understand the results, consider the incentives for continued employment faced by someone of, say, age 65. Under the old pension system, the net present value of lifetime pension benefits was only marginally higher for someone who continued working past 65 than for someone who stopped at 65. Under the new system, the net present value of lifetime pension benefits for someone that stops working at age 65 is lower than in the old system, but the net present value of lifetime pension benefits increases rather steeply from continued employment. An approximate calculation reveals that roughly 40% of the increase in the length of the average working life is due to the reduction in the generosity of pension benefits, while the remaining 60% are due to the increase in the present value of benefits from continued employment. Only a negligible share of the increase in aggregate labor supply is coming from a decline in disability insurance incidence. While the computation of the disability insurance benefit changes as part of the reform, we find that the net present value of lifetime benefits for someone that goes on disability insurance at, for example, age 50 is only slightly lower following the reform. This explains why the model does not predict a large change in disability insurance incidence.

While reform is often times a slow and painful process, our results regarding the Swedish pension reform suggest that with an appropriately designed carrot-and-stick approach it is indeed possible to reform an ailing pension system. This is good news for the world community at large and indicates that there are significant lessons to be learned from the case of Sweden.

While the focus in the quantitative exercise has been on Sweden, our framework is general in nature and can be used to address a host of policy questions pertaining to retirement. In fact, the development of a model to study the interactions between health, disability insurance and old-age pension benefits is an important contribution of the paper.

There is a vast literature on retirement, pertaining to both the claiming of old-age pension benefits and disability insurance. Most of it is centered on the United States.[1] Methodologically, the paper most similar to ours is French (2005). There are several notable differences, however. In our framework individuals can impact the evolution of their health, whereas in French (2005) they cannot. Furthermore, our five tiers allow for a finer distinction between health categories than French's assumption of two health states, good and bad. Additionally, we allow for the possibility of individuals in poor health to go on disability insurance.

Our paper also contributes to the literature on the impact of tax and transfer programs on life cycle labor supply. See, e.g., Rogerson and Wallenius (2009), Wallenius (2009). The key distinction between our paper and the aforementioned ones is that we are explicitly interested in disability insurance and as such also include endogenous health in our framework.

Jönsson, Palme, and Svensson (2012) document the prevalence of disability insurance incidence in Sweden, whereas Sundén (2006) and Palmer (2003) document the Swedish pension reform and its intended consequences. These papers are, however, descriptive in nature and do not provide any quantitative analysis of the policy reform. Sundén (2002) studies the ability of the post-reform Swedish pension system to adjust to demographic change. His analysis, however, assumes that retirement is exogenous.

An outline of the paper follows. Section 2 presents the model and the solution method, while Section 3 describes the calibration procedure. Section 4 outlines the quantitative exercise that is carried out in the paper and Section 5 describes the results from this exercise. Section 6 discusses the robustness of the results. Section 7 concludes.

2. Model

We consider a discrete time overlapping generations framework, in which a measure one of identical, finitely lived individuals is born every period. A model period is a year, and individuals live for 61 periods with certainty. Model age zero corresponds to age 20 in the data. Furthermore, individuals are endowed with one unit of time each period.

[1] See for example Gustman and Steinmeier (1986), Pozzebon and Mitchell (1989), Stock and Wise (1990), Berkovec and Stern (1991), Rust and Phelan (1997), French (2005), Gruber and Wise (2004), Gruber and Wise (2009), Coile and Gruber (2007), Coile and Levine (2007), Low, Meghir, and Pistaferri (2010), Laun (2012), French and Song (2009).

Letting a denote model age, individuals have preferences over sequences of consumption (c), labor supply (l) and health (h) given by:

$$\sum_{a=0}^{60} \beta^a \left[\ln(c_a) - b(h_a)l_a + h_a\right],$$

where β is the discount factor. Preferences are assumed to be separable and consistent with balanced growth, thereby dictating the $\ln(c)$ choice. We assume that the disutility from working is health dependent.[2] Specifically, working is more unpleasant the worse the health of an individual. Additionally, the health of an individual enters directly in the utility function.

Each period there are markets for consumption, labor, capital and health investment. Let w_a denote the exogenous age-varying wage profile, r the interest rate and $p(h)$ the cost of health investment as a function of health. The individual faces a sequence of budget constraints given by:

$$c_a + k_{a+1} - (1+r)k_a + (1-s)p(h_a)i_a^h = (1-\tau)w_a l_a + I_a^{DI} DI_a + I_a^{PB} PB_a + T.$$

The agent's capital stock at age a is denoted by k_a. We impose a no-borrowing constraint, $k_a \geq 0$. This is one way of ensuring that people work when young, even at a low wage.[3] Furthermore, we assume that individuals must have non-negative assets at death.

Health investments are denoted by i_a^h, and take the value of zero or one. Health investments are subsidized at the rate s. Following the OECD self-assessed health measure, health is discretized into five states: very good, good, fair, bad and very bad. All individuals start out in very good health. Health evolves according to the following law of motion:

$$h_{a+1} = h_a + I_a^{HI} i_a^h + \varepsilon_a^h.$$

I_a^{HI} is an indicator function, which takes the value one if the health investment is effective and zero otherwise. The probability that the health investment is effective is dependent on both the age and the health of the individual. ε_a^h is the exogenous health shock. The probability of the health shock is also age- and health-dependent.

We assume a discrete labor supply choice where the individual either works full-time or not at all, $l_a \in \{0, \bar{l}\}$. While for some individuals retirement is a gradual

[2] This is an alternative to assuming that productivity, or the wage, is health dependent. Both result in a distribution of retirement ages. French (2005) finds that there is surprisingly little difference in the wages of healthy and unhealthy individuals. This observation appears to hold for Sweden as well.

[3] In the absence of a borrowing constraint, and with exogenous wages and individuals choosing the timing of work, people would choose not to work when young but rather at a higher wage later on. This is not what we observe in the data.

transition from full-time work to no work, for most this transition occurs abruptly.[4] Assuming the presence of some non-convexity in the individual's choice problem is one way of generating movements from full time work to no work. Examples of non-convexities are fixed costs associated with work and non-linear wage schedules. Instead of modeling these underlying details, here we simply assume that individuals are faced with a discrete choice problem of full time work or no work. Labor income is the product of the exogenous, age-dependent wage and labor supply.[5] The government levies a proportional tax, τ, on labor income.

I_a^{DI} is an indicator function, which takes the value one if the individual claims disability benefits and zero otherwise. Similarly, I_a^{PB} is the indicator function associated with pension benefits. DI_a denotes the disability benefits and PB_a the pension benefits. Both benefits depend on the age and past earnings of the claimant, and in the case of disability benefits, on health. The benefits will be discussed in more detail in the calibration section. T denotes a lump-sum transfer, which is the same for all individuals.

The government uses the proceeds from the proportional tax levied on labor income, τ, to finance the subsidy on health investment, pension and disability insurance benefits, as well as the lump-sum transfer. We assume a balanced budget in equilibrium.[6]

2.1. Solving the Model. Each period an individual must choose: how much to consume, how much to invest in physical capital, whether or not to invest in health, whether or not to work, whether or not to apply for disability insurance and whether or not to apply for pension benefits. The large number of combinations implies a large state space. This in turn yields a computationally intensive problem.

As labor supply and health investment are discrete choices by construction, we only need to discretize physical capital. We assume a capital grid with 31 grid points, ranging from 0 to 1 500 000 SEK (roughly 227 000 USD).

We solve for decision rules via backward induction. Assuming zero utility when dead, we know the value function at age 81. This allows us to solve the agent's problem at age 80 for each possible employment history, consisting of disability, pension and retirement decisions, and for each state of health and physical capital. We then know the value function at age 80 and can solve the agent's problem at age 79 and so on.

Having solved for the decision rules, we simulate the model 61 000 times. For aggregation purposes we assume that at any given point in time, the economy consists

[4] See Rogerson and Wallenius (2011) for a discussion of the United States. The same observation is true for Sweden. We return to this point in Section 3.

[5] We assume that the price per efficiency unit of labor has been normalized to one.

[6] We consider alternatives to the lump-sum transfer in Section 6.

of 10 000 20 year olds, 10 000 21 year olds, 10 000 22 year olds, and so forth. All agents start out with zero capital and in very good health.

3. Calibration

In this section we discuss the approach for assigning parameter values. Recall that a model period is a year, and that agents enter the model at age 20. We assume that the initial capital stock of an individual is zero.

The policy parameters are chosen to match the Swedish pre-reform social security system. The remaining parameters are chosen to match various moments of the Swedish data. We now describe this process in greater detail.

The preference parameters needing to be assigned a value are the discount factor, β, and the disutility from working parameter, $b(h)$. We target an annual interest rate of 3%, and simply assume that $\beta = 1/(1+r)$.[7] The disutility from working is larger, the worse the health status of an individual. We assume a linear relationship between the disutility levels associated with the five discrete health states and parameterize it so as to target the retirement age distribution. See Figure 2.1 for an illustration of the target distribution.[8]

The exogenous age-varying wage profile is constructed from Eurostat data for the year 2009. The data reports average labor income for five-year age bins, 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, and 60-64. We fit a quadratic function to the data and use that to interpolate values for individual ages. Figure 2.2 plots this function. It exhibits the typical hump-shaped profile, with income leveling out in the 40s and 50s and declining slightly in the 60s.[9]

We assume an indivisible labor supply choice, where people either work a fixed workweek or not at all. The length of the workweek is set to 1/3 of the time endowment. Given our emphasis on retirement, the nature of the transition from full-time work to little or no work is particularly relevant. As depicted in Figure 2.3, at age 55, more than 60% of Swedes are working full-time (35 hours or more per week), whereas 20% are not working at all. At age 70, virtually no one is working. The SHARE[10] dataset is not

[7] Relaxing this assumption would introduce life-cycle effects in the consumption profile. While there is some empirical evidence of this, these effects are not of first-order importance for the questions addressed here. We have, therefore, chosen to abstract from them.

[8] Due to classification issues by the Swedish pension authority based on pension collection these numbers may exclude some people who are working. Therefore the values constitute a lower bound of employed people.

[9] The results are robust to adjusting the labor earnings profile by average hours for each particular age group. At older ages, there are issues associated with selection. We return to this point in Section 6.

[10] This paper uses data from SHARELIFE release 1, as of November 24th 2010. The SHARE data collection has been primarily funded by the European Commission through the 5th framework

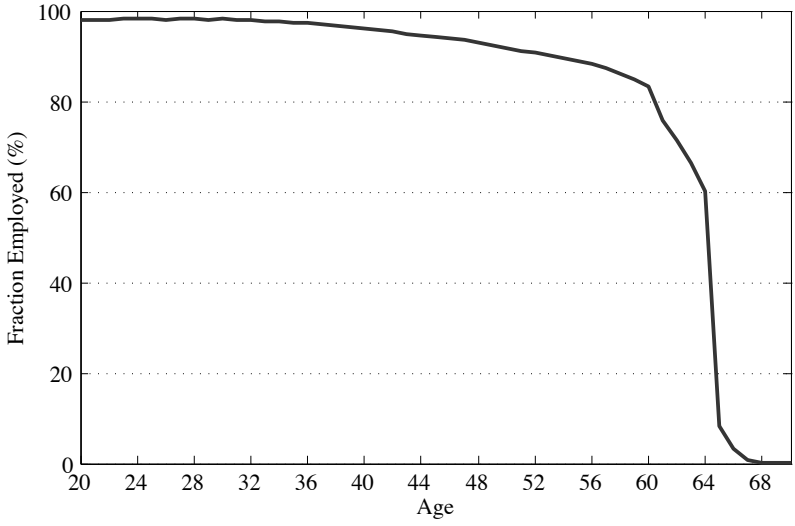

FIGURE 2.1. Fraction Employed by Age. Data source: Swedish Pension Authority, 2009.

a panel, and therefore does not preclude the fact that some people could be transiting sequentially between hours bins. The fact that the fraction of the age group working part-time (between 10 and 25 hours per week) stays roughly constant at 10% until it drops to essentially zero at age 65, however, suggests that that this is very limited in scope. Moreover, the low incidence of part-time work leads one to conjecture that part-time work is not very desirable in Sweden, be it due to a limited availability of part-time jobs or a wage penalty associated with them. We therefore feel that the inclusion of a part-time work option is not of first-order importance for our analysis.

Pertaining to health, three objects need to be parameterized: the cost function for health investments, the process governing the effectiveness of health investments and the process governing shock to health. With the cost function for health investment we wish to capture both the overall level of health expenditures and the differences in health expenditure by health status. Our measure for health expenditures is from the

program (project QLK6-CT-2001-00360 in the thematic program Quality of Life), through the 6th framework program (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th framework program (SHARE-PREP, 211909 and SHARE-LEAP, 227822). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064, IAG BSR06-11, R21 AG025169) as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).

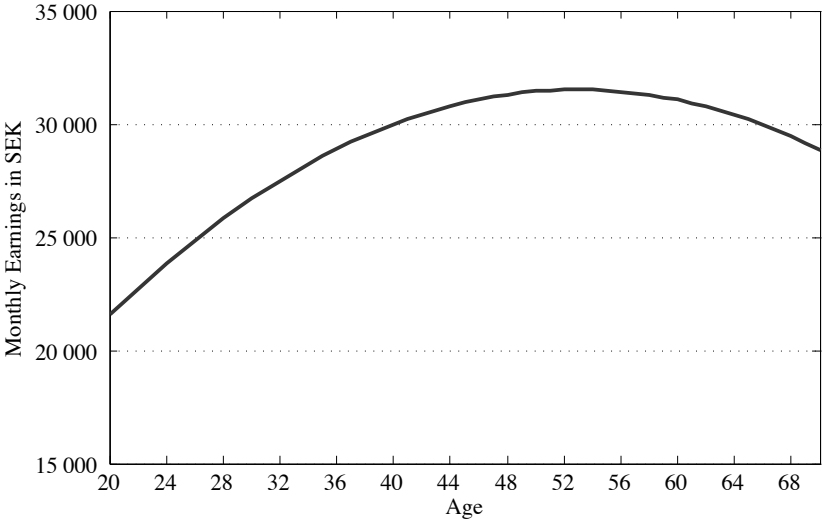

FIGURE 2.2. Average Earnings Profile. Data source: Eurostat, 2009.

Survey on Health, Aging and Retirement in Europe (SHARE) and is the sum of expenditures on inpatient care, outpatient care, prescription drugs and health insurance. The appealing feature of the SHARE data is that it is possible to tabulate average health expenditures by health status. The health expenditure measure in this dataset, however, is an incomplete measure of health expenditures and captures just over 50% of all health expenditures in Sweden. Our approach, therefore, is to use the SHARE data to capture the differences in health expenditure by health status and to set the level of health costs so as to match total health expenditures as a fraction of total tax revenue. In Sweden, total health expenditures constitute roughly 18% of total tax revenue.

We assume two possible shocks to health, a small shock and a big shock. The small shock constitutes a one-unit drop in health, whereas the big shock constitutes a three-unit drop in health. Given that health investment is at most one, and not always effective, agents can only partially insure against health shocks. The probability of health shocks is increasing in age. Furthermore, the probability of being hit by a shock is bigger, the worse ones health. This feature is included to mimic persistence in health. The probability that the health investment is effective is decreasing in age. Also, the probability that the investment is effective is lower, the worse the health of the individual. The probabilities of the health shocks and the probabilities governing

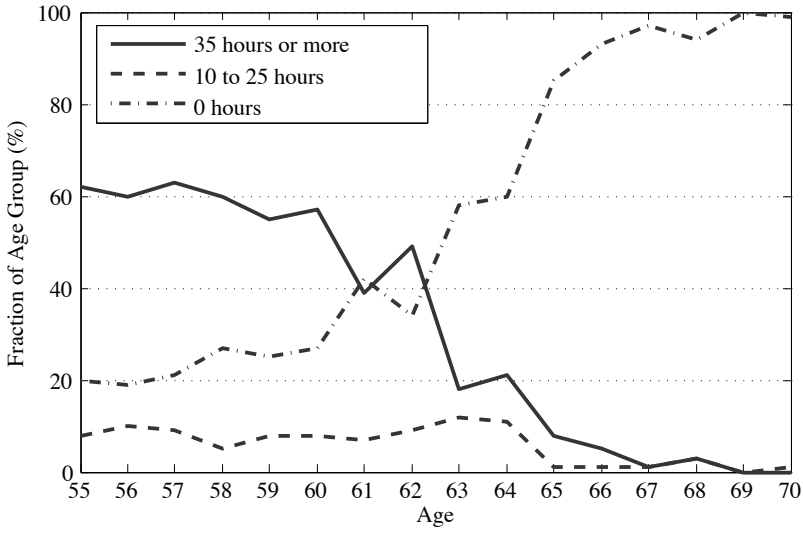

FIGURE 2.3. Incidence of Full-Time vs Part-Time Work by Age. Data source: SHARE.

the effectiveness of health investments are chosen to target the fraction of people on disability insurance, the timing of disability incidence and the health distribution at older ages. Table 2.1 reports the fraction of each age group on disability insurance. Disability insurance incidence is quite high in Sweden, with 18.6% of the population going on disability insurance at some point during their life. The majority of disability insurance incidence occurs after the age of 50.

Age	Fraction on DI
20-24	0.02
25-29	0.02
30-34	0.02
35-39	0.03
40-44	0.04
45-49	0.06
50-54	0.09
55-59	0.12
60-64	0.18

TABLE 2.1. Fraction on Disability Insurance by Age. Data source: Swedish pension authority, 2009.

Table 2.2 reports self-assessed health states for older individuals. The reported values are expressed as a fraction of the relevant age group. In the model we assume that an individual must be in either bad or very bad health to be eligible for disability insurance.

Age	Very good	Good	Fair	Bad	Very bad
55 to 64	29.9	35.1	24.5	7.7	2.9
65 to 74	25.0	40.4	27.0	5.9	1.7
75 to 84	15.7	33.5	35.4	10.6	4.8

TABLE 2.2. Self-Assessed Health by Age (reported as percentage of age group). Data source: OECD, 2009.

Sweden has a public healthcare system. We capture this in an extremely stylized way, by assuming a subsidy on all health investments. The rate of the subsidy is set to match public spending as a fraction of total spending on healthcare. This share is 81.9% for Sweden.

The tax on labor income is set to equal the average effective labor tax burden in Sweden in 2009. The tax rate of 0.438 includes income taxes, payroll taxes and consumption taxes.[11] Recall that the labor tax is used to fund the subsidy on health investment, pensions, disability insurance and the lump-sum transfer. We assume the government balances its budget in equilibrium; the lump-sum transfer is set so as to accomplish this.

The model is calibrated to the pre-reform Swedish pension system. It is a PAYG, defined benefit plan financed through a payroll tax. The pension benefit is comprised of two parts, a basic allowance and an earnings dependent supplement. Both are tied to the so-called basic amount (BA), which equaled 43 600 SEK (roughly 6 600 USD) in 2009. The basic allowance is the same for everyone and equal to 0.96BA. The earnings dependent supplement is given by:

$$0.6AP_a \min(a/30, 1)\text{BA},$$

where AP_a is average pension points at age a. One accrues pension points from earned income in the 15 highest years of earnings. They are computed by taking income in excess of the BA up to 7.5BA and dividing by the BA. Furthermore, there is an adjustment when there are less than 30 years of work.

[11] The method for computing tax rates is outlined in McDaniel (2007). The actual tax series can be found at http://www.caramcdaniel.com/researchpapers.

When mapping this to the model, average pension points are a state variable. The state variable is updated according to the following rule:

$$AP_{a+1} = AP_a + \frac{1}{15}\frac{\min\{w_a\bar{l}, 7.5\text{BA}\} - \text{BA}}{\text{BA}} \qquad \text{if} \quad a < 15,$$

$$AP_{a+1} = AP_a + \frac{1}{15}\max\left\{0, \frac{\min\{w_a\bar{l}, 7.5\text{BA}\} - \text{BA}}{\text{BA}} - AP_a\right\} \qquad \text{if} \quad a \geq 15.$$

In other words, if the individual has worked for less than 15 years, an additional year of work always increases average pension points. If the individual has worked for 15 or more years, average pension points only increase if earnings exceed average earnings to date. We make the simplifying assumption that a high income year replaces an average income year, instead of the lowest income year. This is the same as in French (2005).

The first age at which the pension benefit can be claimed is 61. The full retirement age is 65. The actuarial adjustment for early claiming is 0.5%-points for every month up to age 65. The actuarial adjustment for delayed claiming is 0.7% for every month up to age 70.

The disability insurance benefit is computed in much the same way as the pension benefit. The notable exceptions are: (1) there is no actuarial reduction for early claiming, (2) assumed pension points are computed up to age 65 based on average income from the last three years prior to disability. These features make disability insurance under the old rules very generous. People are automatically transferred from disability insurance to regular pension at the age of 65. The benefit stays the same throughout.

If one experiences only a partial loss in earnings capacity, it is possible to claim partial disability insurance in Sweden. As roughly three-quarters of all disability claimants are on full-disability insurance, we abstract from partial disability insurance in our model. It is not possible to continue working while on full-disability insurance. In the model, we assume that the individual must be in bad or very bad health to qualify for disability insurance. We arrived at this cut-off after examining data on the self-reported heath status of older workers, as well as the fraction of the age group on disability insurance. According to the Swedish pension authority (Pensionsmyndigheten), the fraction of people aged 55-64 on disability insurance in Sweden in 2009 was roughly 16%. According to the OECD self-assessed health questionnaire, of this same age group roughly 11% reported being in bad or very bad health. We assume that if the individual satisfies this health criterion, and applies for disability, he/she receives it.[12] Furthermore, we assume that disability is an absorbing state.

[12] Alternatively we could assume that a person applying for disability insurance receives the benefit with some positive probability, and that this probability is bigger, the worse the health of the applicant.

Table 2.3 summarizes the benchmark calibrated parameter values for the model.

Parameter	Value	Explanation
Policy Parameters		
τ	0.438	Tax on labor income
s	0.82	Subsidy on health expenditure
Utility Parameters		
β	0.97	Discount factor
b_5	2.5	Disutility from work when health very good
b_4	3.0	Disutility from work when health good
b_3	3.5	Disutility from work when health fair
b_2	4.0	Disutility from work when health bad
b_1	4.5	Disutility from work when health very bad
$\bar{l}$	$\frac{1}{3}$	Labor supply
Health Parameters		
e^l	1	Decrease in health from low shock
e^h	3	Decrease in health from high shock
p^l	$0.1 \to 0.5$	Probability of low shock
p_1^l	0.5	Probability of low shock when health very bad
p^h	0.01	Probability of high shock
p_1^h	0.1	Probability of high shock when health very bad
q_5	$1 \to 0.5$	Probability health investment effective when health very good
q_4	$0.9 \to 0.5$	Probability health investment effective when health good
q_3	$0.8 \to 0.5$	Probability health investment effective when health fair
q_2	$0.4 \to 0.1$	Probability health investment effective when health bad
q_1	$0.2 \to 0.1$	Probability health investment effective when health very bad

TABLE 2.3. Calibrated Parameter Values

A brief explanation of a few of the listed parameters is in order. As previously noted, the disutility from working is greater the worse the health of the individual. We assume a linear relationship. The table reports the two boundary points.

The probability of being hit by the small health shock increases linearly with age from 0.1 to 0.5. However, if the individual is in the worst health state, the probability of being hit by the small shock is 0.5, regardless of age. The probability of the big health shock is constant over age at 0.01, unless the individual is in the worst health state, in which case the probability is 0.1. As noted previously, the dependency of the shock probability on health status mimics persistence.

The reason we decided not to pursue this option is that it would require the determination of several parameter values, of which we have little way of knowing how to discipline.

Recall that the probability that the health investment is effective is dependent on both age and health. Given a particular level of health, the probability that health investment is effective decreases linearly with age. A decline in health, however, shifts the probabilities to a lower trajectory. The table reports the boundary values for each health state.

3.1. Calibrated Economy. We now outline the calibrated economy and discuss how well we are able to match the data on Sweden prior to the pension reform.

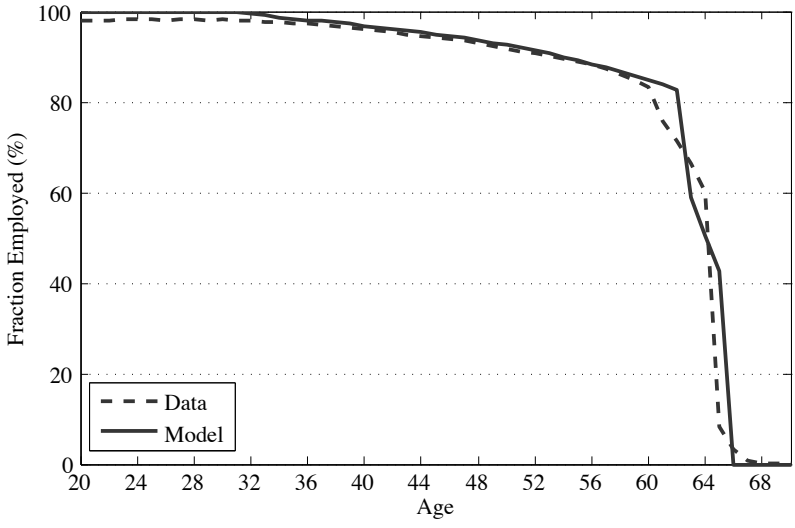

FIGURE 2.4. Fraction Employed by Age

Figure 2.4 plots the distribution of retirement ages for the benchmark economy relative to the data. From the figure one notes that the timing of retirement predicted by the model mimics that in the data quite closely. In particular, the model predicts first a gradual decline in employment rates in the early 60s followed by a sharper decline in the mid 60s. The average age at which people stop working in the benchmark economy is 62.4. Given that pension claiming does not require that one stop working, there is no reason to expect that the age at which people stop working would coincide with the age at which they start collecting pension benefits. Moreover, the adjustments for early and delayed claiming are roughly actuarially fair. With certain lifetimes, the agents are then rather indifferent about when to start taking out benefits. In the model, everyone (who does not go on disability insurance) starts collecting pension benefits at

age 63. For the purposes of any policy analysis, the age at which people stop working is of more interest than the age at which they start collecting pension benefits. In contrast to pension benefits, to claim disability benefits one must stop working.

Figure 2.5 plots the fraction of a particular age group that is on disability insurance. The solid line denotes the model predicted values, while the dashed line sketches the data. The model does a relatively good job of matching both the incidence and timing of disability insurance. The model predicts that 18.6% of people go on disability insurance during their lifetime. The average age at which people claim disability benefits in the model is 51.3. Note that the last age at which people are eligible for disability is 64; at age 65 disabled individuals are automatically transferred to pension.

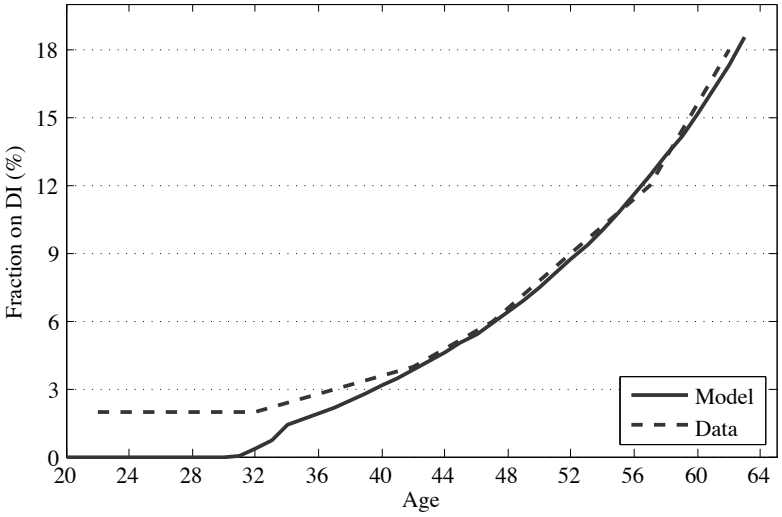

FIGURE 2.5. Fraction on Disability Insurance by Age.

One aspect of the data that the model struggles to match is the health distribution at older ages. To illustrate, the model predicts that at age 70: 47% of people are in very good health, 21% in good health, 8% in fair health, 5% in bad health and 19% in very bad health. In contrast, according to the data on people aged 65-74, 25% are in very good health, 40.4% in good health, 27% in fair health, 5.9% in bad health and 1.7% in very bad health.[13] The health distribution predicted by the model places too little mass at intermediate health states. There is tension in the model between matching the fraction of people on disability insurance and matching the health distribution at

[13] The OECD self-assessed health data is only available for 10-year age bins.

older ages. We feel that matching the fraction of people on disability insurance is more important for the policy analysis to follow.

Note also that in our model the majority of people who go on disability insurance, and thereby satisfy the poor health criterion, stay in poor health throughout the remainder of their life.

The tax rate on labor income is 43.8% and the budget balancing lump-sum transfer is 28 000 SEK (roughly 4 200 USD) per person annually. The ratio of total health expenditures to total tax revenue in the calibrated economy is 18.6%. This is very close to the 18% reported in the data.

To summarize, the model does a good job of replicating the salient features of the pre-pension reform Swedish economy, particularly as it pertains to the labor supply behavior of older workers.

4. Quantitative Exercise: Pension Reform

We now analyze the implications of the recent Swedish pension reform. This entails modifying the pension and disability insurance schemes to reflect the new policies. All other parameters are as in the benchmark calibration. Specifically, we assume a small open economy with a fixed interest rate and price per efficiency unit of labor. The effective labor tax burden has stayed roughly constant in recent years. We therefore keep τ fixed at 0.438, but compute a new budget balancing lump-sum transfer.[14]

The new pension scheme is comprised of two parts, a notional defined contribution component and a funded individual account. The contribution rate is 18.5% on all earnings, of which 16% are credited to the defined contribution part and 2.5% to the individual accounts. Pension rights are accrued on earnings up to a ceiling, which equaled 50 900 SEK in 2009 (roughly 7 700 USD). The annuity is then computed by taking total pension capital and dividing by life expectancy.

The system is still a PAYG system, with current contributions used to fund the benefits of the current old. The size of the benefit is dependent on current economic conditions. This is why the benefit scheme is classified as a *notional* defined contribution plan.

The first age at which one can collect pension benefits is unchanged at 61. If one continues to work while collecting pension benefits, one continues to accrue pension capital. The benefit is then recalculated when the individual stops working.

The computation of the disability insurance benefit has also changed as part of the reform. Under the new system, the disability benefit is equal to 64% of the average income from the three years prior to disability. One accrues pension benefits while

[14] We discuss the implications of keeping the transfer fixed at the pre-reform level in Section 6.

on disability. Also, as was the case prior to the reform, people are automatically transferred from disability insurance to pension at age 65.

When modeling the pension reform, we don't explicitly model the funded accounts component and instead treat all contributions as if they were part of the defined contribution part. In Section 6 we consider an alternative to this, namely treating the funded accounts component as purely forced savings. This entails lowering the labor tax rate by 2.5 percentage points.

5. Results

There are several issues inherent with the old Swedish pension system, in particular the fact that the pension benefit is based on earnings from only the 15 highest years and only income up to a relatively low ceiling counts toward the benefit. This has the potential to treat workers with equivalent lifetime earnings very unequally, as someone with low income in many years would earn significantly less than someone with the same lifetime income concentrated in 15 years. Furthermore, it does not provide incentives for older individuals to remain employed. In fact, given that wages tend to level off in the 40s or 50s, there is no expected increase in pension benefits from continued employment for most older workers. The new pension scheme hopes to address this issue.

In this section we discuss the implications of the pension reform as predicted by the model. We are particularly interested in whether the reform creates incentives for people to continue working longer.

5.1. Timing of Retirement. We find that the Swedish pension reform does indeed create large incentives for workers to remain employed longer. In fact, the model predicts an increase in the average retirement age of 2.3 years from 62.4 to 64.7. Figure 2.6 illustrates the shift in the overall retirement age distribution. Following the reform, a notable number of people are predicted to still be working at ages 66-68.

Following the changes to the Swedish pension system, there is an increasing tendency for workers to continue working while collecting pension benefits. According to the model people take out pension benefits earlier than before, with the majority of people now taking out pension benefits starting at age 61, compared with age 63 in the old system.

The Swedish pension reform changes the computation of pension benefits along several dimensions. In terms of understanding the results, two features are of paramount importance. They are the reduction in the generosity of benefits (holding the stop working age constant) and the increase in benefits from deferred retirement. In the old system the net present value of lifetime pension benefits as a function of the

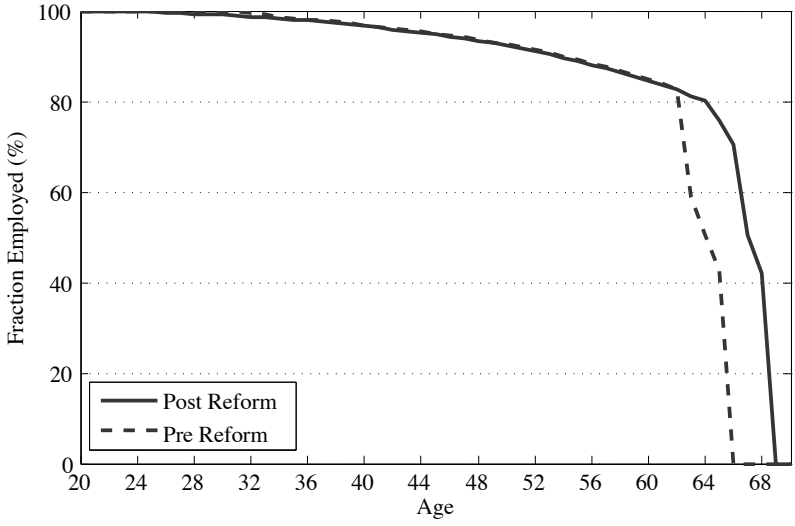

FIGURE 2.6. Fraction Employed by Age

age at which one stops working is very flat. In the new system this schedule rises much more steeply as a function of the age at which one stops working. However, were one to stop working at the same age in the new system as in old system, the present value of lifetime benefits would be lower. To disentangle these two effects, consider the following simple exercise. Compute the sum of the present value of lifetime pension benefits for everyone in the old system, given their optimal retirement choices. Then compute the hypothetical benefits from the new system if people were to stop working at the same age as in the old system. This calculation indicates that total lifetime benefits in net present value terms are about 90% of their previous level in the new system. We then uniformly scale down the pre-reform pension scheme by a factor of 0.903.[15] Following this reduction in the scale of pension benefits, the average retirement age rises by roughly one year. This approximate exercise indicates that roughly 40% of the predicted increase of 2.3 years in the average retirement age resulting from the pension reform is due to the reduction in the generosity of benefits, whereas the remaining 60% is due to the fact the the present value of lifetime pension benefits increases when one defers retirement.

[15] The calculation of the scale factor assumes that the collection of pension benefits in each system starts when it is optimal. This is independent of the age at which one stops working.

5.2. Disability Insurance Incidence. In response to the changes to social security, the model predicts that the fraction of older workers claiming disability insurance falls by roughly 1 percentage point, from 18.6% to 17.7%. This implies that only a small fraction of the increase in employment is coming from a decline in the incidence of disability insurance.

All disability claimants are automatically transferred into pension at age 65. This is true of both the old and the new system. In the old system, however, this distinction was irrelevant from the individual's perspective, as the benefit was constant. This is no longer the case in the new system. The disability insurance benefit is actually somewhat higher in the new system, but the expected pension benefit is lower for someone who has been on disability for an extended amount of time, compared with that in the old system. The reason for this is that one continues to contribute to pension capital when on disability insurance, but at a significantly lower rate than when working. These two opposing effects roughly offset for someone who claims disability insurance at the average age for disability incidence. In other words, the net present value of lifetime benefits for someone that goes on disability insurance under the new system at, say age 50, is only slightly lower than for someone who did the same under the old system. Given that the economic incentives for disability insurance claiming change very little as a result of the pension reform, it is not surprising that the model does not predict a more significant decline in disability insurance incidence following the reform.

The average age for going on disability insurance declines by roughly one year following the reform, from 51.30 to 49.98. This is explained by the fact that the present value of lifetime disability and pension benefits (for someone who goes on disability insurance) is higher in the new system than in the old system when one claims benefits before age 50, but lower when one claims later.

5.3. Other Implications. One of the additional concerns with the old Swedish social security system was the heavy financial burden the funding of the system placed on taxpayers. In our model, the decline in the share of tax revenue going to fund pensions following the reform is exemplified by the increase in the lump-sum transfer. All else equal, our model predicts an increase in the lump-sum transfer from 28 000 SEK to 36 000 SEK (from 4 200 USD to 5 400 USD) needed to balance the budget subsequent to the changes in social security programs. As previously mentioned, we explore alternatives to this model assumption in the following section.

Health expenditures as a fraction of tax revenue are virtually unchanged at 17.6%, previously 18.6%. Similarly, the health distribution is also unchanged by the reform.

6. Sensitivity Analysis

In this section we discuss the robustness of the results to various features of the model and the data.

One key issue in matching a wage or labor income profile is that we only observe wages for those who work. The problem of selection is particularly relevant at older ages. As a robustness check we re-calibrate the model to a wage profile where wages after age 62 are kept constant at the age 62 level. Comparing the model predicted post-pension reform retirement age distributions for the two specifications, one notes that there is slightly more mass retiring at ages 68 and 69 when wages are assumed constant after age 62 than when they were allowed to decline. Overall, however, the effect is negligible, with the model now predicting an average retirement age of 64.8 for the post-reform pension system, compared with 64.7 in the baseline calibration.

As previously noted, the Swedish pension reform implies a decline in the tax revenue needed to fund social security. Given that the labor tax rate has not declined following the reform, the assumption of budget balancedness in the model implies an increase in the lump-sum transfer. Alternatively, one could ignore this general equilibrium aspect and only consider the partial equilibrium decision problem of agents. This would entail keeping the lump-sum transfer fixed at the pre-reform level. This results in a small shift in the post-reform retirement age distribution, with some of the people previously retiring at age 67 choosing to defer retirement until age 68 or 69. The aggregate effect is to raise the average retirement age in the post-reform economy from 64.7 to 65. This in turn implies that the partial equilibrium version of the model implies an increase in the average retirement age of 2.6 years following the Swedish pension reform, compared with the 2.3 years predicted by the general equilibrium version.

Recall that when modeling the Swedish pension reform we abstract from the fully funded component and treat all contributions as part of the defined benefit component. The funded component is a form of forced savings. Given that one could reduce other savings by a corresponding amount, one could in fact argue that it isn't really different from regular savings. Assuming agents are cognizant of this, this suggests lowering the labor tax rate by the size of the funded component, i.e., 2.5 percentage points, when going from the old pension scheme to the new one. When modeled this way, the labor supply implications of the Swedish pension reform are even larger than when treating all contributions as part of the defined benefit portion. Figure 2.7 depicts the retirement age distribution for this case, and contrasts it with the retirement age distributions for the old pension system and the new system when treating all contributions as part of the defined benefit component. The average retirement age rises to 65.5. This

corresponds to an increase in the average retirement age of 3.1 years relative to the old pension system, compared with our baseline prediction of an increase of 2.3 years.

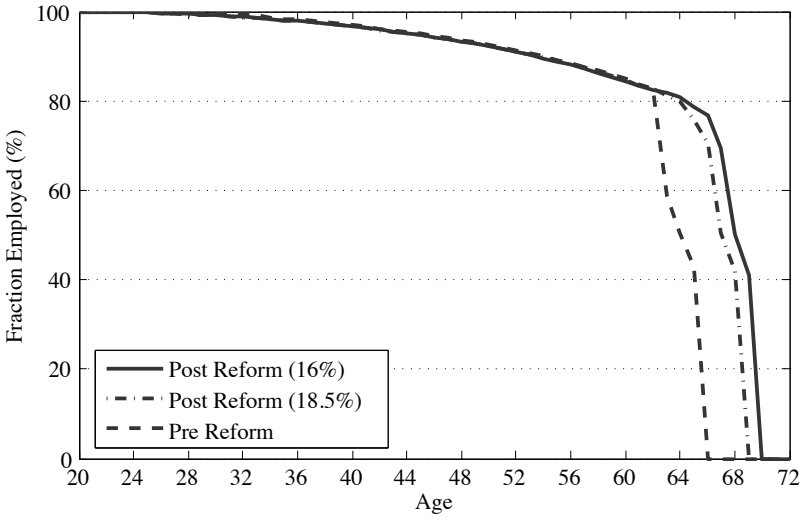

FIGURE 2.7. Fraction Employed by Age

In our analysis we set the labor tax rate to match the average effective labor tax burden in Sweden. Although pension and disability insurance benefits and healthcare expenditures constitute a large share of government expenditures, we are required to take a stand on what to do with the additional tax revenue in the model. The results presented in this section and the previous one have assumed a lump-sum transfer of equal size to all individuals. There are many alternatives to the assumption of a lump-sum transfer, and following Rogerson (2007) and Ragan (2005) we know that the labor supply implications of labor taxes can be quite different depending on what the government does with the tax revenue. As a robustness check, we set the lump-sum transfer to zero and instead assume that the tax revenue that is left over after pension and disability insurance payments and healthcare subsidies is spent on government consumption. We assume that the individual values government consumption but that it enters separately in the utility function. In other words, the marginal utility of private consumption is unaffected by government consumption. We recalibrate the model to the old Swedish pension system with this assumption and study the labor supply implications of the pension reform. The setting with government consumption

implies a slightly larger increase in the average retirement age following the pension reform than the setting with the lump-sum transfer, 2.6 years compared with 2.3 years.

To summarize, our results are robust to various model features. In fact, the sensitivity analysis presented in this section indicates that our baseline results are a conservative estimate of the labor supply effects of the recent Swedish pension reform.

7. Conclusions

Governments around the world are grappling with the question of how to reform ailing pension systems. Sweden is one of the few countries to have recently undertaken a large-scale reform. In this paper we study what potential lessons there are to be learned from this reform for the international community. This objective requires the development of a model capturing the key driving forces underlying retirement behavior, particularly the interactions between health, disability insurance and old-age pension benefits. To this end, we develop an overlapping generations model of life cycle labor supply and retirement. In our framework, individuals care about their health, and can partially insure against negative health shocks. Agents choose when to stop working and, given eligibility criteria, if/when to claim disability insurance and pension benefits. The endogeneity of health and disability incidence are novel features of the framework in relation to the previous literature.

We use the model to study the labor supply implications of the recent Swedish pension reform. Our interest in Sweden naturally stems from the unique nature of the recent large reform, but also from some of the inherent country characteristics. Much of the literature on disability insurance and pension claiming has centered on the United States. Many institutional features, such as healthcare, have potentially large implications for the labor supply outcomes of older workers. These institutions differ markedly between the United States and Sweden, or more generally most of Western Europe.

The Swedish reform entails a switch from a defined benefit to a defined contribution scheme. Under the old system, pension benefits were based only on income from the 15 highest years and only income up to a relatively low ceiling contributed to benefits. This had the potential of treating individuals with equivalent lifetime earnings very unequally. Moreover, under the old system disability insurance was extremely generous, treating the disabled individual as if he/she had earned the pre-disability income until age 65. The pension reform seeks to address these issues.

We find that the new Swedish pension system creates large incentives for the continued employment of older individuals. In fact, the model predicts an increase in the average retirement age of 2.3 years, from 62.4 to 64.7. We find that the incentives for

working longer following the reform are two-fold. First, were one to retire at the same age in the new system as in the old system, the implied pension benefit would be lower. Second, unlike in the old system, in the new system the present value of lifetime pension benefits increases if one continues to work longer. Both effects are quantitatively important in accounting for the increase in the employment rates of older workers. Only a small share of the increase in aggregate labor supply comes from a decline in the fraction of people on disability insurance. In fact, the fraction of people to go on disability insurance during their lifetime only drops by roughly 1 percentage point, from 18.6% to 17.7%. This is unsurprising, as it turns out that the present value of lifetime disability insurance and pension benefits for someone that went on disability insurance at, for example, age 50 declines only marginally following the reform.

Our focus in this paper has been on the recent pension reform in Sweden. However, a key virtue of the framework developed in this paper is that it is quite general in nature and can therefore be used to study a host of interesting policy questions related to the labor supply behavior of older individuals.

References

Berkovec, J. and Stern, S. (1991), 'Job Exit Behavior of Older Men', *Econometrica*, 59, 189–210.

Coile, C. and Gruber, J. (2007), 'Future Social Security Entitlements and the Retirement Decision', *The Review of Economics and Statistics*, 89, 234–246.

Coile, C. and Levine, P. (2007), 'Labor Market Shocks and Retirement: Do Government Programs Matter?', *Journal of Public Economics*, 91, 1902–1919.

French, E. (2005), 'The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior', *The Review of Economic Studies*, 72, 395–427.

French, E. and Song, J. (2009), 'The Effect of Disability Insurance Receipt on Labor Supply', *Working Paper 2009–05*, Federal Reserve Bank of Chicago.

Gruber, J. and Wise, D. (2004), 'Social Security Programs and Retirement Around the World: Micro Estimation', *University of Chicago Press*.

Gruber, J. and Wise, D. (2009), 'Social Security Programs and Retirement Around the World: The Relationship to Youth Employment', *University of Chicago Press*.

Gustman, A. and Steinmeier, T. (1986), 'A Structural Retirement Model', *Econometrica*, 54, 555–584.

Jönsson, L., Palme, M., and Svensson, I. (2012), 'Disability Insurance, Population Health and Employment in Sweden', *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, ed. by Wise, D., University of Chicago Press, forthcoming.

Laun, T. (2012), 'Optimal Social Insurance with Endogenous Health', *Working Paper 742*, SSE/EFI Working Paper Series in Economics and Finance.

Low, H., Meghir, C., and Pistaferri, L. (2010), 'Wage Risk and Employment Risk over the Life Cycle', *American Economic Review*, 100, 1432–1467.

McDaniel, C. (2007), 'Average Tax Rates on Consumption, Investment, Labor and Capital in the OECD 1950-2003', *Working Paper*.

Palmer, E. (2003), 'The New Swedish Pension System', *Working Paper*.

Pozzebon, S. and Mitchell, O. (1989), 'Married Womens Retirement Behavior', *Journal of Population Economics*, 2, 39–53.

Ragan, K. (2005), 'Taxes, Transfers and Time Use: Fiscal Policy in a Household Production Model', *Working Paper*, University of Chicago.

Rogerson, R. (2007), 'Taxation and Market Work: Is Scandinavia an Outlier?', *Economic Theory*, 32, 59-85.

Rogerson, R. and Wallenius, J. (2009), 'Micro and Macro Elasticities in a Life Cycle Model with Taxes', *Journal of Economic Theory*, 144, 2277–2292.

Rogerson, R. and Wallenius, J. (2011), 'Fixed Costs, Retirement and the Elasticity of Labor Supply', *Working Paper*.

Rust, J. and Phelan, C. (1997), 'How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets', *Econometrica*, 65, 781–832.

Stock, J. and Wise, D. (1990), 'Pensions, the Option Value of Work and Retirement', *Econometrica*, 58, 1151–1180.

Sundén, A. (2006), 'The Swedish Experience with Pension Reform', *Oxford Review of Economic Policy*, 22, 133–148.

Sundén, D. (2002), 'The Dynamics of Pension Reform', *Ph.D. Thesis*, Stockholm School of Economics.

Wallenius, J. (2009), 'Social Security and Cross-Country Differences in Hours: A General Equilibrium Analysis', *Working Paper*.

PAPER 3

Health and Business Cycles

Tobias Laun

Abstract This paper develops a framework to analyze the interactions between health and business cycles. The individual's health is determined by his stock of health capital and a stochastic component. Health increases the individual's utility as well as the total time the individual can spend on either work or leisure. The individual invests in health capital as well as in physical capital. He further decides how much of his available time to allocate to labor supply and how much to leisure. In this setting, I show that an unexpected decline in health causes a reduction in output, consumption and labor supply. In response to a decline in health, the individual increases investment in health capital and reduces savings accordingly. I also show that a positive shock to productivity increases both health and physical capital investment. Better health therefore leads to increased savings. Higher productivity, in turn, increases savings and improves health.

1. Introduction

The importance of health for growth and economic development has been greatly acknowledged. Most studies, however, consider health an exogenous variable rather than something the individual can influence himself. A justification for not modeling health explicitly is that, if health is endogenous, it can be assumed to be a part of human capital. However, health is more than just a part of human capital. Besides affecting labor supply and productivity, health affects survival probabilities and also has direct utility effects. The former operates through the effective discount factor which impacts the attractiveness of saving.

In this paper, I incorporate endogenous health into a business cycle model. The individual's health is determined by his stock of health capital and a stochastic component. Health increases the individual's utility as well as the total time the individual

The author thanks Lars Ljungqvist, Johanna Wallenius, David Domeij, Henrik Lundvall and seminar participants at Sveriges Riksbank and the Stockholm School of Economics for helpful comments and suggestions. Financial support from the Jan Wallander and Tom Hedelius Foundation at Svenska Handelsbanken is gratefully acknowledged.

can spend on either work or leisure. The individual invests in health capital as well as in physical capital. He further decides how much of his available time to allocate to labor supply and how much to leisure. In this setting, I analyze the effects of health and productivity shocks on the variables in the model and derive the optimal investment in health capital.

Mushkin (1962) was the first to discuss the role of health as an investment good by highlighting the similarities and differences between health and other forms of human capital, such as education. Grossman (1972) argued that health is not just a form of human capital, since it not only affects productivity but also determines the time an individual can allocate to earning income and producing goods. The author then develops a formal model in which he derives the optimal demand for the commodity "good health". The Grossman model with the notion of health being both an investment good and a consumption good, inspired an extensive literature; some examples are Ehrlich and Chuma (1990) and Ried (1998).[1]

In the macroeconomic literature, the first to recognize the importance of considering health in the analysis of human capital was Mankiw, Romer, and Weil (1992). Subsequently, Fogel (1994) and Barro (1996) investigated the relationship between health and economic growth. This development gave rise to a extensive empirical literature in which different health measures were used to evaluate the effect of health on growth and economic development.[2] Those papers find that health generally has a positive effect on growth.

There are only a few papers which consider health an endogenous factor. Chakraborty (2004) assumes that the probability of individuals surviving from the first to the second period in a two period OLG model increases in public health expenditures. Here, a poverty trap can arise in which low health expenditures, i.e. high mortality, lead to a low rate of saving, which in turn implies low tax revenues and hence low health expenditures. In the model of Johansson and Löfgren (1995) individuals invest in health capital which generates utility and determines the individuals' expected length of life. Furthermore, the overall health of the labor force is an argument in the production function. Because of this positive externality, individuals invest less in health than is socially optimal. In the setting of van Zon and Muysken (2001) health generates utility and, together with human capital, determines labor supply. Here, health and growth are both compliments (through labor supply) and substitutes (through utility).

[1] For a deeper discussion of the Grossman model and the literature that followed see Grossman (2000).

[2] Examples are Knowles and Owen (1995), Arora (2001), Bhargava, Jamison, Lau, and Murray (2001), Acemoglu and Johnson (2007), Lorentzen, McMillan, and Wacziarg (2008), Aghion, Howitt, and Murtin (2010).

The growth maximizing and welfare maximizing solutions are therefore not necessarily identical.

The paper is organized as follows. Section 2 describes the model environment. In section 3 I derive the equilibrium conditions. In section 4 the model is solved and in section 5 I discuss the results.

2. The Environment

I consider a standard business cycles model with an infinitely lived representative household and a continuum of identical firms. The key distinction to the standard literature is the introduction of health as an endogenous variable. Health is produced with the help of health capital and a stochastic component. Households have then two investment options: health capital and physical capital. Health increases utility as well as the time endowment which can be spent on either work or leisure.

Each firm has access to a Cobb-Douglas production technology. Output in period t is therefore given by

$$Y_t = F(K_t, N_t) = A_t K_t^{\alpha} N_t^{1-\alpha}, \tag{3.1}$$

where K_t is physical capital, N_t is labor input and A_t is production technology. The latter is an exogenous stochastic process with the following properties

$$\ln A_{t+1} = \rho \ln A_t + \varepsilon_{t+1}, \tag{3.2}$$

where the productivity shock $\{\varepsilon_{t+1}\}$ is i.i.d. with mean zero and a finite variance σ_Y^2.

The period t net investment in physical capital equals gross investment I_t^K minus depreciation

$$K_{t+1} - K_t = I_t^K - \delta_K K_t, \tag{3.3}$$

where $\delta_K \in (0, 1)$ is the exogenous depreciation rate.

Following Grossman (1972) health enters the model in the form of a stock of so-called health capital. This stock depreciates at rate $\delta_H \in (0, 1)$ and increases with gross investment I_t^H

$$H_{t+1} - H_t = I_t^H - \delta_H H_t. \tag{3.4}$$

In this setting, investment in health capital can be thought of as the costs of preventive measures like exercise or a healthy diet but also medical expenses which are necessary once health deteriorates.

To capture the fact that an individual's health is not a purely deterministic variable, I assume that there is an exogenous stochastic health component Z_t, which has the following properties

$$\ln Z_{t+1} = \psi \ln Z_t + \nu_{t+1}, \tag{3.5}$$

where $\{\nu_{t+1}\}$ is an i.i.d. health shock with mean zero and finite variance σ_H^2.

Similarly to output, which is produced using endogenous physical capital as well as exogenous and stochastic production technology, the individual's effective health is determined by the endogenous stock of health capital and by the exogenous and stochastic health component. Health affects the individual in the model in two ways. First, health increases the individual's utility. The utility generated by consumption and leisure at time t is multiplied by a function $P(\cdot)$ which is twice continuously differentiable, strictly concave and increasing in its arguments: health capital and the stochastic health technology. The individual's utility at time t is hence given by

$$U_t = P(H_t, Z_t)[u(C_t) + v(L_t)], \tag{3.6}$$

where both u and v are twice continuously differentiable, strictly concave and increasing in their respective arguments. C_t denotes consumption and L_t leisure at time t.

The second channel through which health affects the individual is a constraint on the individual's time endowment. The gross time endowment is given by 1. The individual, however, can only use a share $0 \leq T_t \leq 1$ of this endowment for leisure and labor supply. The individual's time constraint then reads

$$T_t = T(H_t, Z_t) = L_t + N_t, \tag{3.7}$$

where $T(\cdot)$ is twice continuously differentiable, strictly concave and increasing in its arguments. The share $1 - T_t$ can be interpreted as the time spent being sick.

The individual receives income from renting out physical capital and from working. He spends this income on consumption and investment in physical and health capital. Furthermore, there is a convex cost of investing in physical and health capital intended to capture the diminishing returns inherent in such investments. The budget constraint then reads

$$W_t N_t + R_t K_t = C_t + I_t^K + I_t^H + \frac{\mu_K}{2}\left(I_t^K\right)^2 + \frac{\mu_H}{2}\left(I_t^H\right)^2, \tag{3.8}$$

where W_t is the wage and R_t is the rental rate of physical capital.

3. The Decentralized Problem

3.1. The Household's Problem. In every period $i = 0, \ldots, \infty$ the individual maximizes utility (3.6) such that the budget constraint (3.8) is fulfilled. After substituting in the law of motion of physical and health capital, (3.3) and (3.4), as well as the time constraint (3.7), the first order conditions are given by

$$\frac{\partial L}{\partial C_{t+i}} = E_t\left\{\beta^i P_{t+i} u'(C_{t+i}) - \lambda_{t+i}\right\} = 0, \tag{3.9}$$

$$\begin{aligned}
\frac{\partial L}{\partial N_{t+i}} =& E_t\Big\{-\beta^i P_{t+i} v'\big(T_{t+i} - N_{t+i}\big) + \lambda_{t+i} W_{t+i}\Big\} = 0, && (3.10)\\
\frac{\partial L}{\partial K_{t+i+1}} =& E_t\Big\{\lambda_{t+i}\Big[-1 - \mu_K\big(K_{t+i+1} - (1-\delta_K)K_{t+i}\big)\Big]\\
&+\lambda_{t+i+1}\Big[R_{t+i+1} + 1 - \delta_K + \mu_K(1-\delta_K)\big(K_{t+i+2} - (1-\delta_K)K_{t+i+1}\big)\Big]\Big\}\\
=& 0, && (3.11)\\
\frac{\partial L}{\partial H_{t+i+1}} =& E_t\Big\{\beta^{i+1} P_{t+i+1} v'\big(T_{t+i+1} - N_{t+i+1}\big) T_{H_{t+i+1}}\\
&+\beta^{i+1} P_{H_{t+i+1}}\Big[u(C_{t+i+1}) + v\big(T_{t+i+1} - N_{t+i+1}\big)\Big]\\
&+\lambda_{t+i}\Big[-1 - \mu_H\big(H_{t+i+1} - (1-\delta_H)H_{t+i}\big)\Big]\\
&+\lambda_{t+i+1}\Big[1 - \delta_H + \mu_H(1-\delta_H)\big(H_{t+i+2} - (1-\delta_H)H_{t+i+1}\big)\Big]\Big\}\\
=& 0, && (3.12)
\end{aligned}$$

for all $i = 0, \ldots, \infty$.

(3.9) and (3.10) evaluated at $i = 0$ yield together the intratemporal first order condition representing the tradeoff between consumption and leisure

$$v'(T_t - N_t) = W_t u'(C_t). \tag{3.13}$$

The left hand side of this equation shows the benefit of having an additional marginal unit of leisure while the right hand side stands for the opportunity cost of enjoying leisure, i.e. the wage rate times the marginal utility of consumption.

Substituting (3.9) into (3.11) gives the intertemporal first order condition with respect to physical capital[3]

$$P_t u'(C_t)\left(1 + \mu_K I_t^K\right) = E_t \beta P_{t+1} u'(C_{t+1})\Big[R_{t+1} + (1-\delta_K)\left(1 + \mu_K I_{t+1}^K\right)\Big]. \tag{3.14}$$

The left hand side of equation (3.14) represents the utility cost of investing a marginal unit in physical capital while the right hand side stands for the benefit following this investment.

[3] For simplicity, I reintroduce the expression for gross investment in physical and health capital I_t^K and I_t^h.

The intertemporal health condition comes from combining (3.9) and (3.12)

$$\begin{aligned} P_t u'(C_t)\left(1+\mu_H I_t^H\right) =& E_t\Big\{\beta P_{t+1} v'\big(T_{t+1}-N_{t+1}\big) T_{H_{t+1}} \\ &+\beta P_{H_{t+1}}\Big[u(C_{t+1})+v\big(T_{t+1}-N_{t+1}\big)\Big] \\ &+\beta P_{t+1} u'(C_{t+1})(1-\delta_H)\left(1+\mu_H I_{t+1}^H\right)\Big\}. \end{aligned} \tag{3.15}$$

This equation can be interpreted in a similar way as the intertemporal first order condition for physical capital. The left hand side equals the cost of investing in health capital. The right hand side of the equation gives the benefit from investing in health arising from (1) an increase in next period's net time endowment, evaluated at the marginal utility of leisure, (2) an increase in next period's utility and (3) the benefit of having a larger stock of health capital in the next period.

3.2. The Firm's Problem. In each period t the firm maximizes profits

$$\max_{N_t,K_t} \quad \Pi_t = A_t K_t^{\alpha} N_t^{1-\alpha} - W_t N_t - R_t K_t.$$

The standard result holds that firms pay the marginal product to the input factors and have zero profits in equilibrium

$$(1-\alpha)\frac{Y_t}{N_t} = W_t \tag{3.16}$$

and

$$\alpha\frac{Y_t}{K_t} = R_t. \tag{3.17}$$

3.3. Equilibrium. Combining equations (3.13) and (3.16) to eliminate wages gives the intratemporal equilibrium condition

$$\frac{v'(T_t-N_t)}{u'(C_t)} = (1-\alpha)\frac{Y_t}{N_t}. \tag{3.18}$$

Substituting (3.17) into (3.14) yields the intertemporal equilibrium condition for physical capital

$$P_t u'(C_t)\left(1+\mu_K I_t^K\right) = E_t \beta P_{t+1} u'(C_{t+1})\left[\alpha\frac{Y_{t+1}}{K_{t+1}} + (1-\delta_K)\left(1+\mu_K I_{t+1}^K\right)\right]. \tag{3.19}$$

Plugging (3.16) and (3.17) into the household's budget constraint (3.8) gives the market clearing constraint

$$Y_t + (1-\delta_K)K_t + (1-\delta_H)H_t = C_t + K_{t+1} + H_{t+1} + \frac{\mu_K}{2}\left(I_t^K\right)^2 + \frac{\mu_H}{2}\left(I_t^H\right)^2. \tag{3.20}$$

4. Solution

Equations (3.18) – (3.20) together with equations (3.1) – (3.5) and (3.15) constitute a system of nine non-linear difference equations. In order to solve this system, I log-linearize the nine equations around the non-stochastic steady state.

Approximating[4] the production function (3.1) around its steady state $Y = AK^{\alpha}N^{1-\alpha}$ yields[5]

$$y_t = a_t + \alpha k_t + (1-\alpha)n_t. \tag{3.21}$$

The steady state levels of production and health technology are given by $A = Z = 1$. Log-linearizing around these steady states gives

$$a_{t+1} = \rho a_t + \varepsilon_{t+1} \tag{3.22}$$

and

$$z_{t+1} = \psi z_t + \nu_{t+1}. \tag{3.23}$$

The log-linear version of the intratemporal equilibrium condition (3.18) is given by

$$\theta_1 c_t + \theta_2 h_t + \theta_3 z_t + \theta_4 n_t = y_t - n_t, \tag{3.24}$$

where

$$\theta_1 \equiv -\frac{u''(C)C}{u'(C)}, \quad \theta_2 \equiv \frac{v''(T-N)T_H H}{v'(T-N)}, \quad \theta_3 \equiv \frac{v''(T-N)T_Z Z}{v'(T-N)}, \quad \theta_4 \equiv -\frac{v''(T-N)N}{v'(T-N)}.$$

The intertemporal equilibrium condition for physical capital can be expressed as

$$\begin{aligned}&\theta_5 h_t + \theta_6 z_t + \theta_7 c_t + \theta_8 i_t^K \\ =&\theta_5 h_{t+1} + \theta_6 E_t z_{t+1} + \theta_7 E_t c_{t+1} + \theta_9 E_t y_{t+1} - \theta_9 k_{t+1} + \theta_{10} E_t i_{t+1}^K.\end{aligned}$$

Substituting $E_t z_{t+1} = \psi z_t$ yields

$$\begin{aligned}&\theta_5 h_t + \theta_6(1-\psi) z_t + \theta_7 c_t + \theta_8 i_t^K \\ =&\theta_5 h_{t+1} + \theta_7 E_t c_{t+1} + \theta_9 E_t y_{t+1} - \theta_9 k_{t+1} + \theta_{10} E_t i_{t+1}^K,\end{aligned} \tag{3.25}$$

where

$$\theta_5 \equiv P_H u'(C)\left(1+\mu_K I^K\right)H, \quad \theta_6 \equiv P_Z u'(C)\left(1+\mu_K I^K\right)Z, \quad \theta_7 \equiv Pu''(C)\left(1+\mu_K I^K\right)C,$$

$$\theta_8 \equiv Pu'(C)\mu_K I^K, \quad \theta_9 \equiv \beta P u'(C)\alpha\frac{Y}{K}, \quad \theta_{10} \equiv \beta P u'(C)(1-\delta_K)\mu_K I^K.$$

---

[4] Details can be found in section A2 in the appendix.

[5] Let the variable x_t be defined as the log-deviation of X_t from its steady state X

$$x_t \equiv \ln X_t - \ln X = \ln\left(\frac{X_t}{X}\right).$$

Log-linearizing the intertemporal equilibrium condition for health and substituting $E_t z_{t+1} = \psi z_t$ yields

$$\begin{aligned}&\theta_{11}h_t + (\theta_{12} - \theta_{16}\psi)z_t + \theta_{13}c_t + \theta_{14}i_t^H \\ =&\theta_{15}h_{t+1} + \theta_{17}E_t n_{t+1} + \theta_{18}E_t c_{t+1} + \theta_{19}E_t i_{t+1}^H,\end{aligned} \tag{3.26}$$

where

$$\begin{aligned}
\theta_{11} \equiv& P_H u'(C)\left(1+\mu_H I^H\right)H \\
\theta_{12} \equiv& P_Z u'(C)\left(1+\mu_H I^H\right)Z \\
\theta_{13} \equiv& Pu''(C)\left(1+\mu_H I^H\right)C \\
\theta_{14} \equiv& Pu'(C)\mu_H I^H \\
\theta_{15} \equiv& \beta\Big\{2P_H v'(T-N)T_H + Pv''(T-N)T_H^2 + Pv'(T-N)T_{HH} \\
&+ P_{HH}\Big[u(C) + v\big(T-N\big)\Big] + P_H u'(C)(1-\delta_H)\left(1+\mu_H I^H\right)\Big\}H \\
\theta_{16} \equiv& \beta\Big\{P_Z v'(T-N)T_H + Pv''(T-N)T_Z T_H + Pv'(T-N)T_{HZ} + P_{HZ}\Big[u(C) + v\big(T-N\big)\Big] \\
&+ P_H v'(T-N)T_Z + P_Z u'(C)(1-\delta_H)\left(1+\mu_H I^H\right)\Big\}ZE_t z_{t+1} \\
\theta_{17} \equiv& -\beta\Big\{Pv''(T-N)T_H + P_H v'\big(T-N\big)\Big\}N, \\
\theta_{18} \equiv& \beta\Big\{P_H u'(C) + Pu''(C)(1-\delta_H)\left(1+\mu_H I^H\right)\Big\}C, \\
\theta_{19} \equiv& \beta P u'(C)(1-\delta_H)\mu_H I^H
\end{aligned}$$

The log-linearized versions of the laws of motion of physical and health capital are given by

$$k_{t+1} = \delta_K i_t^k + (1-\delta_K)k_t \tag{3.27}$$

and

$$h_{t+1} = \delta_H i_t^h + (1-\delta_H)h_t. \tag{3.28}$$

Finally, the log-linearization of the market clearing condition (3.20) yields

$$\begin{aligned}&Yy_t + (1-\delta_K)Kk_t + (1-\delta_H)Hh_t = \\ &Cc_t + Kk_{t+1} + Hh_{t+1} + \mu_K\left(I^K\right)^2 i_t^K + \mu_H\left(I^H\right)^2 i_t^H.\end{aligned} \tag{3.29}$$

After the log-linearization, the system consists of nine linear difference equations: (3.21) – (3.29). These equations can be expressed in matrix notation as

$$\Phi\begin{pmatrix} x_{t+1}^s \\ E_t x_{t+1}^u \end{pmatrix} = \Xi\begin{pmatrix} x_t^s \\ x_t^u \end{pmatrix} + \begin{pmatrix} q_{t+1}^s \\ 0 \end{pmatrix}, \tag{3.30}$$

where $x_t^s \equiv [k_t \ h_t \ a_t \ z_t]'$ is a $(n_s \times 1)$ vector of state variables at time t, $x_t^u = \left[y_t \ c_t \ n_t \ i_t^K \ i_t^H\right]'$ is a $(n_u \times 1)$ vector of control variables and $q_{t+1}^s = [\varepsilon_{t+1} \ \nu_{t+1} \ 0 \ 0]'$ is a $(n_s \times 1)$ vector of shocks. n_s and n_u stand here for the number of state and control variables, respectively. The coefficient matrices are given by

$$\Phi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K & H & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\theta_9 & \theta_5 & \theta_9 & \theta_7 & 0 & \theta_{10} & 0 \\ 0 & 0 & 0 & \theta_{15} & 0 & \theta_{18} & \theta_{17} & 0 & \theta_{19} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\Xi = \begin{pmatrix} \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1-\delta_K)K & (1-\delta_H)H & Y & -C & 0 & -\mu_K \left(I^K\right)^2 & -\mu_H \left(I^H\right)^2 \\ 0 & 0 & 1-\delta_K & 0 & 0 & 0 & 0 & \delta_K & 0 \\ 0 & 0 & 0 & 1-\delta_H & 0 & 0 & 0 & 0 & \delta_H \\ 0 & \theta_6(1-\psi) & 0 & \theta_5 & 0 & \theta_7 & 0 & \theta_8 & 0 \\ 0 & \theta_{12}-\theta_{16}\psi & 0 & \theta_{11} & 0 & \theta_{13} & 0 & 0 & \theta_{14} \\ -1 & 0 & -\alpha & 0 & 1 & 0 & -(1-\alpha) & 0 & 0 \\ 0 & \theta_3 & 0 & \theta_2 & -1 & \theta_1 & 1+\theta_4 & 0 & 0 \end{pmatrix}$$

Applying expectations to all variables in (3.30) yields

$$\Phi E_t \begin{pmatrix} x_{t+1}^s \\ x_{t+1}^u \end{pmatrix} = \Xi \begin{pmatrix} x_t^s \\ x_t^u \end{pmatrix}. \tag{3.31}$$

In order to solve equation (3.31) for decision rules and laws of motion, I use the generalized Schur decomposition.[6] The generalized Schur decomposition of the two square matrices Φ and Ξ yields four matrices Q, Z, S and T with the following properties

(1) Q and Z are Hermitian, i.e. $Q^H Q = QQ^H = I$ and similarly for Z, where Q^H is the conjugate or Hermitian transpose of Q.
(2) T and S are upper triangular.
(3) $Q\Phi = SZ^H$ and $Q\Xi = TZ^H$.

[6] See for example Blanchard and Kahn (1980), Gomme and Klein (2010), King and Watson (2002), Klein (2000), Sims (2002) and Uhlig (1999).

(4) There is no i such that the diagonal elements $s_{ii} = t_{ii} = 0$.[7] Furthermore, s_{ii} and t_{ii} can appear in any order.[8]

Premultiplying equation (3.31) with Q and substituting in the expressions from above gives

$$SZ^H E_t \begin{pmatrix} x^s_{t+1} \\ x^u_{t+1} \end{pmatrix} = TZ^H \begin{pmatrix} x^s_t \\ x^u_t \end{pmatrix}. \tag{3.32}$$

Define then

$$\begin{pmatrix} s_t \\ u_t \end{pmatrix} \equiv Z^H \begin{pmatrix} x^s_t \\ x^u_t \end{pmatrix}, \tag{3.33}$$

where s_t has the same length as x^s_t and u_t has the same length as x^u_t. Plugging this into (3.32) yields

$$SE_t \begin{pmatrix} s_{t+1} \\ u_{t+1} \end{pmatrix} = T \begin{pmatrix} s_t \\ u_t \end{pmatrix}. \tag{3.34}$$

The matrices S and T can then be partitioned in the following way

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} E_t \begin{pmatrix} s_{t+1} \\ u_{t+1} \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} s_t \\ u_t \end{pmatrix},$$

where S_{11} has the dimension $n_s \times n_s$, S_{12} is $n_s \times n_u$, S_{21} is $n_u \times n_s$ and S_{22} is $n_u \times n_u$. T is partitioned similarly.

Upper triangularity of S and T implies that $S_{21} = T_{21} = 0$. The last n_u equations of this system can then be written as

$$S_{22} E_t u_{t+1} = T_{22} u_t. \tag{3.35}$$

Since S_{22} and T_{22} constitute an unstable matrix pair[9] any solution to equation (3.35) with bounded mean must satisfy

$$u_t = 0 \quad \forall t. \tag{3.36}$$

Substituting this into (3.33) and rearranging yields

$$\begin{pmatrix} x^s_t \\ x^u_t \end{pmatrix} = \begin{pmatrix} Z_{11} \\ Z_{21} \end{pmatrix} s_t,$$

or

$$x^u_t = Z_{21} Z_{11}^{-1} x^s_t. \tag{3.37}$$

This equation gives the decision rules for control variables x^u_t, given the state variables x^s_t.

---

[7] m_{ij} denotes the row i, column j element of any matrix M.

[8] The matrices S and T are arranged such that $|s_{ii}| > |t_{ii}|$ for $i = 1, \ldots, n_s$ and $|s_{ii}| < |t_{ii}|$ for $i = n_s + 1, \ldots, n_s + n_u$.

[9] The generalized eigenvalue pairs of these matrices satisfy $|s_{ii}| < |t_{ii}|$.

The result from equation (3.36) that $u_t = 0$ for all t implies that for the first n_s equations of the system (3.34)

$$S_{11}E_t s_{t+1} = T_{11}s_t.$$

Substituting $s_t = Z_{11}^{-1}x_t^s$ into this equation gives

$$S_{11}Z_{11}^{-1}E_t x_{t+1}^s = T_{11}Z_{11}^{-1}x_t^s.$$

Since S_{11} and T_{11} constitute a stable matrix pair,[10] S_{11} is invertible and it therefore holds that

$$E_t x_{t+1}^s = Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}x_t^s.$$

Dropping the expectations operator yields

$$x_{t+1}^s = Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}x_t^s + q_{t+1}^s. \tag{3.38}$$

This equation gives the law of motion for the state variables x^s.

5. Numerical Results

In this section I proceed with a numerical analysis of the model. In order to do so, some assumptions on functional forms have to be made. Let the utility function be additively separable and given by

$$u(C_t) + v(L_t) = \ln C_t + B \ln L_t.$$

The health multiplier is assumed to be between zero and one and has the form

$$P(H_t, Z_t) = 1 - e^{-H_t Z_t}.$$

The net time endowment is also between zero and one and given by

$$T(H_t, Z_t) = \frac{H_t Z_t}{H_t Z_t + 1}.$$

I calibrate the model to match moments from the Swedish data.[11] A model period corresponds to 3 months. The discount rate $\beta = 0.995$ implies a steady-state annualized interest rate of 2 percent. The utility parameter $B = 2$ is set so that in steady state labor supply comprises about one-third of the household's gross time endowment. The depreciation rate of both health and physical capital are set at 0.025 which implies an annual depreciation rate of 10 percent. The other parameters are set to standard levels and presented in Table 3.1.

With the assumed parameter values and functional forms, the steady state of the economy can be determined. As can be seen from Table 3.2, the individual spends

[10] The generalized eigenvalue pairs of these matrices satisfy $|s_{ii}| > |t_{ii}|$.

[11] All data referred to in this paper comes from Statistics Sweden.

Parameter	Value	Description
α	0.333	Physical capital share
β	0.995	Discount factor
B	2.0	Utility parameter
δ_H	0.025	Depreciation rate of health capital
δ_K	0.025	Depreciation rate of physical capital
μ_H	0.1	Health capital investment cost parameter
μ_K	0.1	Physical capital investment cost parameter
ψ	0.9	Persistence of shocks to health technology
ρ	0.9	Persistence of shocks to production technology
σ_H	0.01	Standard deviation of health technology shock
σ_K	0.01	Standard deviation of production technology shock

TABLE 3.1. Parameter Values

roughly 33 percent of her gross time endowment working and about 55 percent on leisure. The missing 12 percent of the time the individual is sick and can neither work nor enjoy leisure. This number is chosen to match Swedish data on sick days.

Variable	SS Value	Description
A_t	1.00	Production technology
Z_t	1.00	Health technology
K_t	11.71	Physical capital
H_t	7.37	Health capital
Y_t	1.09	Output
C_t	0.60	Consumption
N_t	0.33	Labor supply
L_t	0.55	Leisure
T_t	0.88	Time budget
U_t	-1.70	Utility
I_t^K	0.29	Physical capital investment
I_t^H	0.18	Health capital investment
W_t	2.19	Wage
R_t	0.03	Rental rate of capital

TABLE 3.2. Steady State of the Model

The decision rules for the control variables presented in equation (3.37) now have the following form

$$\begin{aligned} y_t =& 1.49a_t + 0.08z_t + 0.26k_t - 0.07h_t, \\ c_t =& 0.32a_t + 0.08z_t + 0.44k_t + 0.29h_t, \\ n_t =& 0.73a_t + 0.12z_t - 0.11k_t - 0.11h_t, \\ i_t^k =& 3.91a_t + 1.14z_t - 2.11k_t + 2.40h_t, \\ i_t^h =& 1.29a_t - 1.62z_t + 3.49k_t - 5.22h_t. \end{aligned}$$

This shows how output, consumption, labor supply, physical capital investment and health capital investment vary depending on the production technology, the health technology, physical capital and health capital. Since the variables are measured in log-deviations from steady state, the coefficients can be interpreted as elasticities. For example, a one percent increase in production technology A yields a 0.32% increase in consumption. It can be seen that positive health shock has a positive effect on output, consumption and labor supply. Health capital h_t has a negative effect on labor supply, and hence on output. A one percent increase in health capital increases the net time endowment by 0.12%. As a result, leisure increases by 0.26% implying a decline in labor supply. An improvement in the health technology causes a reduction in health investment and an increase, of roughly the same relative size, in physical capital investment. A one percent increase in the production technology leads to an increase in health and physical capital investment by more than one percent. Hence, better health hence leads to increased savings. Higher productivity, in turn, increases savings and improves health.

The laws of motion for the state variables were presented in equation (3.38). The production and health technology behave as follows

$$\begin{aligned} a_{t+1} =& 0.90a_t + \varepsilon_{t+1}, \\ z_{t+1} =& 0.90z_t + \nu_{t+1} \end{aligned}$$

while physical and health capital develop according to

$$\begin{aligned} k_{t+1} =& 0.10a_t + 0.03z_t + 0.92k_t + 0.06h_t, \\ h_{t+1} =& 0.03a_t - 0.04z_t + 0.09k_t + 0.84h_t. \end{aligned}$$

Figure 3.1 shows[12] the dynamic response to a one standard deviation decline in the health technology. In order to offset this decline, the individual increases his investment in health capital. This increase in health capital investment reduces savings. The

[12] The figures are presented in section A1 in the appendix.

resulting decrease in physical capital, together with the reduced labor supply (caused in turn by a tighter time budget), causes a decrease in output. Since consumption and leisure are reduced as well, utility decreases when health is hit by a negative shock.

The dynamic response to a one standard deviation decrease in the production technology can be seen in Figure 3.2. It shows the usual consequences of a negative shock to productivity, i.e., reduced savings (and hence less physical capital), a decline in labor supply and a reduction in output. Health capital investment is also negatively affected by this shock. Health capital therefore decreases. Since the stochastic health component does not change here, the time budget is reduced as well. Although the individual has less time available, he will spend more time on leisure. However, since consumption and health are reduced, total utility is also negatively affected by the drop in productivity.

Simulating the model for 1000 periods, taking logs and applying the Hodrick-Prescott filter to remove the trend yields the moments presented in Tables 3.3 and 3.4.

	Standard Deviation (%)	
Variable	Model	Data
Output (Y_t)	1.92	1.64
Consumption (C_t)	0.47	0.93
Hours Worked (N_t)	0.95	1.19
Capital Investment (I_t^K)	5.17	4.04

TABLE 3.3. Standard Deviations (Model and Data)

The data used in Tables 3.3 and 3.4 comes from Statistics Sweden. It can be seen that the model does a good job replicating the standard business cycle facts. To illustrate another feature of the model, I simulate output and health investment and compare it with data from Sweden and Germany. The values from the simulation are aggregated to yearly levels and the first observation of each series is normalized to one. The Hodrick-Prescott filter is applied to remove the trend. The resulting series are plotted in panel 1 of Figure 3.3. In the other two panels time series for GDP and health expenditures from Sweden and Germany are shown. Those series are also first normalized and then filtered to make them comparable with the series in panel 1. It turns out that the model is quite capable of reproducing the cyclical behavior of GDP and health expenditures.

Model

	Cross-Correlation of Output with:				
Variable	$x(-2)$	$x(-1)$	x	$x(+1)$	$x(+2)$
Output (Y_t)	0.44	0.69	1.00	0.69	0.44
Consumption (C_t)	0.23	0.52	0.88	0.74	0.60
Hours Worked (N_t)	0.48	0.71	0.99	0.65	0.38
Capital Investment (I_t^K)	0.47	0.70	0.97	0.62	0.34

Data

	Cross-Correlation of Output with:				
Variable	$x(-2)$	$x(-1)$	x	$x(+1)$	$x(+2)$
Output (Y_t)	0.67	0.86	1.00	0.86	0.67
Consumption (C_t)	0.71	0.79	0.75	0.65	0.48
Hours Worked (N_t)	0.24	0.50	0.69	0.78	0.80
Capital Investment (I_t^K)	0.60	0.79	0.90	0.91	0.81

TABLE 3.4. Cyclical Behavior (Model and Data)

6. Conclusion

In this paper, I develop a framework to study the interactions between health and business cycles. The individual's health is determined by his stock of health capital and a stochastic component. In my framework, health increases the individual's utility as well as the total time the individual can spend on either work or leisure. The individual invests in health capital as well as in physical capital. He further decides how much of his available time to allocate to labor supply and how much to leisure. The endogeneity of health is a novel feature of the framework in relation to the previous literature.

In this setting, I show that an unexpected decline in health causes a reduction in output, consumption and labor supply. In response to a decline in health, the individual increases investment in health capital and reduces savings accordingly. I also show that a positive shock to productivity increases both health and physical capital investment. Better health therefore leads to increased savings. Higher productivity, in turn, increases savings and improves health.

References

Acemoglu, D. and Johnson, S. (2007), 'Disease and Development: The Effect of Life Expectancy on Economic Growth', *Journal of Political Economy*, 115, 925–985.

Aghion, P., Howitt, P., and Murtin, F. (2010), 'The Relationship between Health and Growth: When Lucas Meets Nelson-Phelps', *Working Paper 15813*, National Bureau of Economic Research, Cambridge, MA.

Arora, S. (2001), 'Health, Human Productivity, and Long-Term Economic Growth', *The Journal of Economic History*, 61, 699–749.

Barro, R. (1996), 'Health and Economic Growth', *Technical report*, World Health Organization, Washington D.C.

Bhargava, A., Jamison, D., Lau, L., and Murray, C. (2001), 'Modeling the Effects of Health on Economic Growth', *Journal of Health Economics*, 20, 423– 440.

Blanchard, J. and Kahn, C. (1980), 'The Solution of Linear Difference Models under Rational Expectations', *Econometrica*, 48, 1305–1311.

Chakraborty, S. (2004), 'Endogenous Lifetime and Economic Growth', *Journal of Economic Theory*, 116, 119–137.

Ehrlich, I. and Chuma, H. (1990), 'A Model of the Demand for Longevity and the Value of Life Extension', *Journal of Political Economy*, 98, 761–782.

Fogel, R. (1994), 'Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy', *American Economic Review*, 84, 369–395.

Gomme, P. and Klein, P. (2010), 'Second-Order Approximation of Dynamic Models without the Use of Tensors', *Working Paper*, Concordia University, Montreal.

Grossman, M. (1972), 'On the Concept of Health Capital and the Demand for Health', *Journal of Political Economy*, 80, 223–250.

Grossman, M. (2000), 'The Human Capital Model', in Culyer, A. and Newhouse, J., editors, *Handbook of Health Economics*, pages 347–408. North Holland.

Johansson, P. and Löfgren, K. (1995), 'Wealth from Optimal Health', *Journal of Health Economics*, 14, 65–79.

King, R. and Watson, M. (2002), 'System Reduction and Solution Algorithms for Singular Linear Difference Systems under Rational Expectations', *Computational Economics*, 20, 57–86.

Klein, P. (2000), 'Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model', *Journal of Economic Dynamics and Control*, 24, 1405–1423.

Knowles, S. and Owen, D. (1995), 'Health Capital and Crosscountry Variation in Income per Capita in the Mankiw-Romer-Weil Model', *Economic Letters*, 48, 99–106.

Lorentzen, P., McMillan, J., and Wacziarg, R. (2008), 'Death and Development', *Journal of Economic Growth*, 13, 81–124.

Mankiw, G., Romer, D., and Weil, D. (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, 107, 407–437.

Mushkin, S. (1962), 'Health as an Investment', *Journal of Political Economy*, 70, 129–157.

Ried,W. (1998), 'Comparative Dynamic Analysis of the Full Grossman Model', *Journal of Health Economics*, 17, 383–425.

Sims, C. (2002), 'Solving Linear Rational Expectations Models', *Computational Economics*, 20, 1–20.

Uhlig, H. (1999), 'A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily', in Marimon, R. and Scott, A., editors, *Computational Methods for the Study of Dynamic Economies*, pages 30–61. Oxford University Press, New York.

van Zon, A. and Muysken, J. (2001), 'Health and Endogenous Growth', *Journal of Health Economics*, 20, 169–185.

Appendix

A1. Figures.

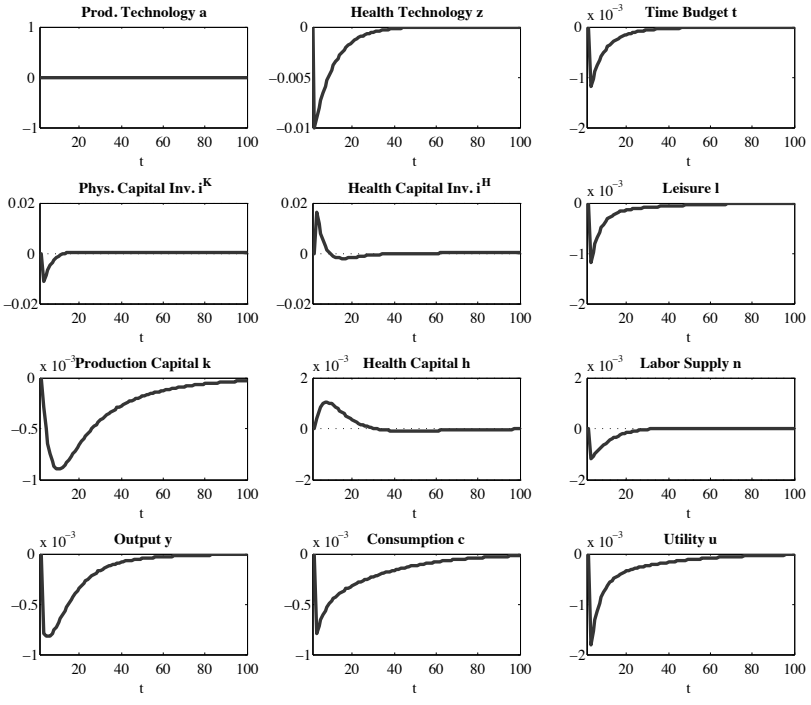

FIGURE 3.1. Response to a one standard deviation drop in health technology. All variables are expressed in log-deviations from steady state.

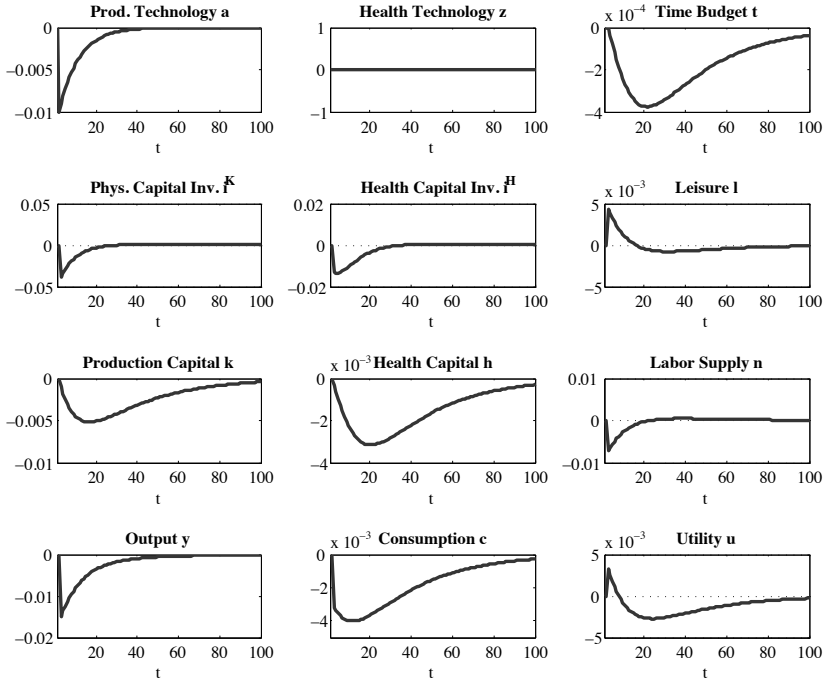

FIGURE 3.2. Response to a one standard deviation drop in production technology. All variables are expressed in log-deviations from steady state.

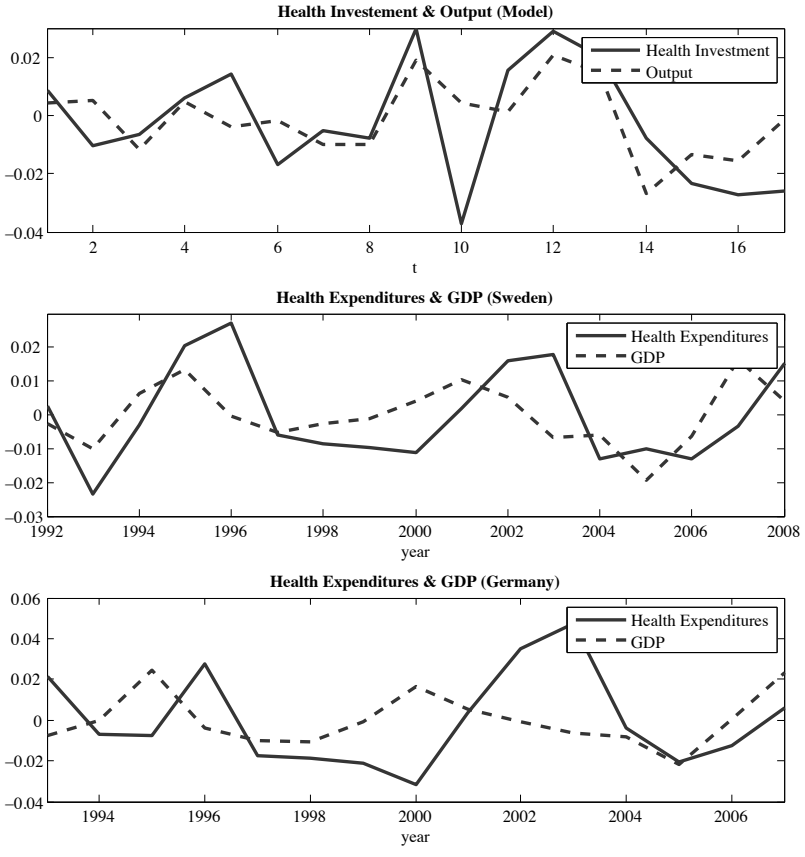

FIGURE 3.3. Output and health expenditure. The first panel shows a simulation of the model economy for 68 quarters, aggregated to 17 years. The second panel shows the corresponding data from Sweden for the years 1992 to 2008. The third panel illustrates the data from Germany from 1993 to 2007. The first observation of each time series has been normalized to one and the Hodrick-Prescott filter has been applied to remove the trend.

A2. Log-Linearization.

Production Function: The log-linearization of the production function (3.1) around its steady state $Y = AK^{\alpha}N^{1-\alpha}$ is given by

$$\begin{aligned} Y_t - Y &= K^{\alpha}N^{1-\alpha}(A_t - A) + \alpha AK^{\alpha-1}N^{1-\alpha}(K_t - K) + (1-\alpha)AK^{\alpha}N^{-\alpha}(N_t - N) \\ &= Y\frac{A_t - A}{A} + \alpha Y\frac{K_t - K}{K} + (1-\alpha)Y\frac{N_t - N}{N}. \end{aligned}$$

Dividing by Y and using

$$x_t \equiv \ln X_t - \ln X = \ln\left(\frac{X_t}{X}\right) \approx \frac{X_t - X}{X}$$

yields

$$\frac{Y_t - Y}{Y} = \frac{A_t - A}{A} + \alpha\frac{K_t - K}{K} + (1-\alpha)\frac{N_t - N}{N}$$

or

$$y_t = a_t + \alpha k_t + (1-\alpha)n_t.$$

Intratemporal Equilibrium Condition: Approximating equation (3.18) around the steady state $\frac{v'(T-N)}{u'(C)} = (1-\alpha)\frac{Y}{N}$ yields

$$\begin{aligned} &-\frac{v'(T-N)u''(C)}{\left[u'(C)\right]^2}(C_t - C) + \frac{v''(T-N)T_H}{u'(C)}(H_t - H) + \frac{v''(T-N)T_Z}{u'(C)}(Z_t - Z) \\ &-\frac{v''(T-N)}{u'(C)}(N_t - N) = (1-\alpha)\frac{1}{N}(Y_t - Y) - (1-\alpha)\frac{Y}{N^2}(N_t - N). \end{aligned}$$

Dividing both sides of the equation by the steady state gives

$$\begin{aligned} &-\frac{u''(C)}{u'(C)}(C_t - C) + \frac{v''(T-N)T_H}{v'(T-N)}(H_t - H) + \frac{v''(T-N)T_Z}{v'(T-N)}(Z_t - Z) \\ &-\frac{v''(T-N)}{v'(T-N)}(N_t - N) = \frac{Y_t - Y}{Y} - \frac{N_t - N}{N} \end{aligned}$$

or

$$-\frac{u''(C)C}{u'(C)}c_t + \frac{v''(T-N)T_H H}{v'(T-N)}h_t + \frac{v''(T-N)T_Z Z}{v'(T-N)}z_t - \frac{v''(T-N)N}{v'(T-N)}n_t = y_t - n_t$$

Intertemporal Equilibrium Condition for Physical Capital: The steady state here is defined by

$$1 + \mu_K I^K = \beta\left[\alpha\frac{Y}{K} + (1-\delta_K)\left(1 + \mu_K I^K\right)\right].$$

The first-order approximation is then given by

$$\begin{aligned}
&P_H u'(C)\left(1+\mu_K I^K\right)(H_t-H)+P_Z u'(C)\left(1+\mu_K I^K\right)(Z_t-Z)\\
&+Pu''(C)\left(1+\mu_K I^K\right)(C_t-C)+Pu'(C)\mu_K(I_t^K-I^K)\\
&=\beta P_H u'(C)\left[\alpha\frac{Y}{K}+(1-\delta_K)\left(1+\mu_K I^K\right)\right](H_{t+1}-H)\\
&+E_t\beta P_Z u'(C)\left[\alpha\frac{Y}{K}+(1-\delta_K)\left(1+\mu_K I^K\right)\right](Z_{t+1}-Z)\\
&+E_t\beta Pu''(C)\left[\alpha\frac{Y}{K}+(1-\delta_K)\left(1+\mu_K I^K\right)\right](C_{t+1}-C)\\
&+E_t\beta Pu'(C)\left[\alpha\frac{1}{K}\right](Y_{t+1}-Y)-\beta Pu'(C)\left[\alpha\frac{Y}{K^2}\right](K_{t+1}-K)\\
&+E_t\beta Pu'(C)(1-\delta_K)\mu_K(I_{t+1}^K-I^K).
\end{aligned}$$

Substituting in the steady state condition and expressing the variables in log-deviations from steady state gives

$$\begin{aligned}
&P_H u'(C)\left(1+\mu_K I^K\right)Hh_t+P_Z u'(C)\left(1+\mu_K I^K\right)Zz_t\\
&+Pu''(C)\left(1+\mu_K I^K\right)Cc_t+Pu'(C)\mu_K I^K i_t^K\\
&=P_H u'(C)\left(1+\mu_K I^K\right)Hh_{t+1}+P_Z u'(C)\left(1+\mu_K I^K\right)ZE_t z_{t+1}\\
&+Pu''(C)\left(1+\mu_K I^K\right)CE_t c_{t+1}+\beta Pu'(C)\alpha\frac{Y}{K}E_t y_{t+1}\\
&-\beta Pu'(C)\alpha\frac{Y}{K}k_{t+1}+\beta Pu'(C)(1-\delta_K)\mu_K I^K E_t i_{t+1}^K.
\end{aligned}$$

Intertemporal Equilibrium Condition for Health Capital: In the steady state this condition is given by

$$\begin{aligned}
Pu'(C)\left(1+\mu_H I^H\right)=&\beta Pv'\big(T-N\big)T_H+\beta P_H\Big[u(C)+v\big(T-N\big)\Big]\\
&+\beta Pu'(C)(1-\delta_H)\left(1+\mu_H I^H\right).
\end{aligned}$$

Approximating around the steady state yields

$$
\begin{aligned}
&P_H u'(C)\left(1+\mu_H I^H\right)(H_t-H)+P_Z u'(C)\left(1+\mu_H I^H\right)(Z_t-Z)\\
&+Pu''(C)\left(1+\mu_H I^H\right)(C_t-C)+Pu'(C)\mu_H(I_t^H-I^H)\\
&=\beta\Big[P_H v'(T-N)T_H+Pv''(T-N)T_H^2+Pv'(T-N)T_{HH}\Big](H_{t+1}-H)\\
&+E_t\beta\Big[P_Z v'(T-N)T_H+Pv''(T-N)T_Z T_H+Pv'(T-N)T_{HZ}\Big](Z_{t+1}-Z)\\
&-E_t\beta Pv''(T-N)T_H(N_{t+1}-N)\\
&+\beta\Big\{P_{HH}\Big[u(C)+v\big(T-N\big)\Big]+P_H v'(T-N)T_H\Big\}(H_{t+1}-H)\\
&+E_t\beta\Big\{P_{HZ}\Big[u(C)+v\big(T-N\big)\Big]+P_H v'(T-N)T_Z\Big\}(Z_{t+1}-Z)\\
&+E_t\beta P_H u'(C)(C_{t+1}-C)-E_t\beta P_H v'\big(T-N\big)(N_{t+1}-N)\\
&+\beta P_H u'(C)(1-\delta_H)\left(1+\mu_H I^H\right)(H_{t+1}-H)\\
&+E_t\beta P_Z u'(C)(1-\delta_H)\left(1+\mu_H I^H\right)(Z_{t+1}-Z)\\
&+E_t\beta Pu''(C)(1-\delta_H)\left(1+\mu_H I^H\right)(C_{t+1}-C)\\
&+E_t\beta Pu'(C)(1-\delta_H)\mu_H(I_{t+1}^H-I^H).
\end{aligned}
$$

Expressed in log-deviations from steady state

$$
\begin{aligned}
&P_H u'(C)\left(1+\mu_H I^H\right)Hh_t+P_Z u'(C)\left(1+\mu_H I^H\right)Zz_t\\
&+Pu''(C)\left(1+\mu_H I^H\right)Cc_t+Pu'(C)\mu_H I^H i_t^H\\
&=\beta\Big\{2P_H v'(T-N)T_H+Pv''(T-N)T_H^2+Pv'(T-N)T_{HH}\\
&\quad+P_{HH}\Big[u(C)+v\big(T-N\big)\Big]+P_H u'(C)(1-\delta_H)\left(1+\mu_H I^H\right)\Big\}Hh_{t+1}\\
&+\beta\Big\{P_Z v'(T-N)T_H+Pv''(T-N)T_Z T_H+Pv'(T-N)T_{HZ}+P_{HZ}\Big[u(C)+v\big(T-N\big)\Big]\\
&\quad+P_H v'(T-N)T_Z+P_Z u'(C)(1-\delta_H)\left(1+\mu_H I^H\right)\Big\}ZE_t z_{t+1}\\
&-\beta\Big\{Pv''(T-N)T_H+P_H v'\big(T-N\big)\Big\}NE_t n_{t+1}\\
&+\beta\Big\{P_H u'(C)+Pu''(C)(1-\delta_H)\left(1+\mu_H I^H\right)\Big\}CE_t c_{t+1}\\
&+\beta Pu'(C)(1-\delta_H)\mu_H I^H E_t i_{t+1}^H.
\end{aligned}
$$

PAPER 4

Screening Stringency in the Disability Insurance Program

Per Johansson Lisa Jönsson Tobias Laun

Abstract Changes in the screening stringency of applications to the disability insurance program are potentially important for explaining program growth. Screening stringency is, however, inherently difficult to observe since it depends on the implementation of program rules as well as formal eligibility criteria. We derive a theoretical model showing that the application decision, and with that, the average health of applicants are endogenous to screening stringency. Based on these results, we provide an empirical strategy for assessing changes in screening stringency over time, using the mortality rate after admittance to the disability insurance program. The strength of the empirical strategy is that it captures both formal and informal changes in screening stringency. Applying the empirical strategy to Sweden, we find that changes in screening stringency are an important contributing factor for fluctuations in the disability benefit award rate over time.

1. Introduction

The disability insurance program has become one of the largest income maintenance programs in modern welfare states.[1] Autor (2011) shows that, in the U.S., the share of 25–64 year olds receiving benefits from the Social Security Disability Insurance (SSDI) increased from 2.3 to 4.6 percent between 1989 and 2009, and that the SSDI Trust Fund is projected to be exhausted between 2015 and 2018. Duggan and Imberman (2009) review the literature and distinguish the relative contribution of population factors, economic conditions and program design for the expansion of the program. They argue, based on the work by Autor and Duggan (2003) and Black, Daniel and Sanders (2002), that program growth can be explained by a decline in labor market opportunities of low-skilled workers, an increase in the effective replacement rate and, most importantly, the 1984 liberalization of the program's eligibility criteria.

The authors thank Lars Ljungqvist and Mårten Palme for valuable comments. Jönsson and Laun gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation at Svenska Handelsbanken.

[1] See e.g. Wise (2012) for a more detailed description of disability insurance programs in several developed countries.

Sweden has also experienced a large growth in disability benefits recipiency. The share of 30–64 year olds receiving disability benefits increased from 8 to 12 percent between 1985 and 2008. In this period, the fluctuations in the award rate have been large. Jönsson, Palme and Svensson (2012) show that changes in the relative health of women compared to men and of young compared to old is to some degree reflected in relative award rates, but that the underlying population health cannot account for the large variations in inflow over time. The authors further demonstrate that some of the changes in the award rate can be related to changes in formal eligibility criteria. Karlström, Palme and Svensson (2008) show that the removal of the special eligibility rules for workers aged 60 to 64 in 1997 led to a decrease in disability benefits recipiency in this age group. However, the employment effect was crowded out by an increased utilization of the sickness and unemployment insurances.

In both the U.S. and Sweden, changes in screening stringency seem to be important for the growth of the disability insurance program. Changes in formal eligibility criteria are easily observable but the actual screening stringency also depends on the implementation of formal program rules, such as caseworker discretion and internal processes at the Social Insurance Agency. The contribution of changes in actual screening stringency to program growth has therefore been difficult to evaluate.

One potential indicator of screening stringency is the denial rate. However, the denial rate depends on the composition of the applicant pool and the decision to apply for disability benefits is potentially correlated with screening stringency. Halpern and Hausman (1986) estimate a model where individuals choose whether to apply for disability benefits with uncertainty about the approval of their application. The authors show that the probability of acceptance has a significant effect on the probability of applying, although, the benefit level seems more important. Parsons (1991) and Gruber and Kubik (1997) use the disability funding crisis of the late 1970s which induced a sharp increase in initial denial rates as an exogenous change in screening stringency. Parsons (1991) shows that initial eligibility determination indeed is an important self-screening mechanism. His results suggest that a 10 percent increase in the initial denial rate induced a 4 percent decrease in application rates after 2 years. de Jong, Lindeboom and van der Klaauw (2011) show for the Netherlands that an exogenous increase in stringency has led to a drop in applications for disability benefits.

In this paper, we provide a theoretical model showing that the denial rate does not necessarily capture the screening stringency of the disability insurance program if the application decision depends on the likelihood of getting admitted. We further show that the relative health of awarded beneficiaries versus non-beneficiaries improves with a reduction in screening stringency. Based on this result, we propose an empirical

strategy for assessing changes in screening stringency in the disability insurance program over time. We use the mortality rate after admittance as an objective measure of health and estimate the excess mortality of new disability beneficiaries over time. The strength of the empirical strategy is that it captures changes in screening stringency both due to changes in formal eligibility criteria and due to changes in the implementation of program rules. The latter aspect has been difficult to assess in previous studies.

Applying the empirical strategy to Sweden, we find that changes in screening stringency are an important contributing factor for the fluctuations in the disability benefit award rate over time. Screening stringency was comparatively low during the periods of large inflow to the program in the late 1980s and the early 2000s, whereas the rapid decline in disability benefit awards since 2005 is reflected in a substantial increase in screening stringency. The removal of eligibility for pure labor market reasons for workers aged 60–64 results in an increase of the relative screening stringency for older workers compared to younger. The large inflow of women compared to men during the early 2000s corresponds to a relatively lower screening stringency for women. Finally, the screening stringency has been highest in Stockholm compared to other regions during the period under study, but the relative stringency across regions has converged after the possibilities of taking labor market reasons into account have been removed.

The paper proceeds as follows. Section 2 presents the theoretical model of the application decision to the disability insurance program. Section 3 presents a numerical simulation of the model. Section 4 outlines the empirical strategy for assessing changes in screening stringency. Section 5 describes the Swedish disability insurance program and the data. Section 6 presents the estimation results and Section 7 concludes the paper.

2. The Model

In this section we describe a stylized model of an individual's application decision to the disability insurance program. The aim of the model is to show that the application decision, and with that, the average health of applicants are endogenous to screening stringency. In the model, an individual decides whether to apply for disability benefits or whether to continue working. When applying, the individual has to determine which health to signal to the caseworker. If the signaled health is lower than the disability benefit threshold, the individual receives disability benefits. If this is not the case, the individual is assumed to return to the labor force.

Let m_{it} be the disability benefit threshold that individual i faces at calendar time t. This threshold can be expressed as

$$m_{it} = m_t + \mu_{it},$$

where m_t is the average disability benefit threshold at time t and μ_{it} is an idiosyncratic error term, with zero mean and a cumulative distribution function $F(\cdot)$. The average disability benefit threshold, m_t, is known by the individual.

The individual applies for disability benefits by signaling health to the caseworker. Let h_{it} be the objective health, which is unobserved by the caseworker, and let θ_{it} be the health signaled by the individual. The deviation between signaled and objective health allows for moral hazard in the model. If signaled health, θ_{it}, is less than the disability benefit threshold the individual faces, m_{it}, then the individual is awarded disability benefits. Hence,

$$D_{it}^{AC} = \begin{cases} 1 & \text{if} \quad \theta_{it} - m_t < \mu_{it} \\ 0 & \text{if} \quad \theta_{it} - m_t \geq \mu_{it}, \end{cases}$$

where D_{it}^{AC} is an indicator of individual i's application for disability benefits at time t being accepted.

To determine the optimal value of signaled health, θ_{it}^*, the individual maximizes the expected value from applying for disability benefits

$$\begin{aligned} V^{AP}(h_{it}, m_t, w_{it}) = \max_{\theta_{it}} \big\{ & p(\theta_{it}, m_t) V^{AC}(h_{it}, w_{it}) \\ & + [1 - p(\theta_{it}, m_t)] \, V^{RJ}(h_{it}, w_{it}) \\ & - C(\theta_{it}, h_{it}, w_{it}) \big\}, \end{aligned} \tag{4.1}$$

where $V^{AC}(\cdot)$ is the value of the application for disability benefits being accepted, $V^{RJ}(\cdot)$ is the value of the application being rejected, $C(\cdot)$ is the application cost and

$$p(\theta_{it}, m_t) = \Pr(\theta_{it} - m_t < \mu_{it}) = 1 - F(\theta_{it} - m_t),$$

is the probability of the application being accepted. It follows that

$$p_\theta(\theta_{it}, m_t) \equiv \frac{\partial p(\theta_{it}, m_t)}{\partial \theta_{it}} = -f(\theta_{it} - m_t) < 0 \tag{4.2}$$

and

$$p_m(\theta_{it}, m_t) \equiv \frac{\partial p(\theta_{it}, m_t)}{\partial m_t} = f(\theta_{it} - m_t) > 0. \tag{4.3}$$

That is, the probability of obtaining disability benefits is decreasing in signaled health and increasing in the average disability benefit threshold.

The value of an accepted application for disability benefits, V^{AC}, is a function of the objective health, h_{it}, and the wage, w_{it}. The individual's consumption, when

receiving disability benefits, is given by δw_{it}, where δ is the compensation rate. The other source of utility is the individual's objective health, h_{it}. The value of receiving disability benefits is therefore increasing in all arguments.

The value of a rejected application for disability benefits, V^{RJ}, is a function of the objective health, h_{it}, and the wage, w_{it}. Health generates utility and V^{RJ} is therefore increasing in h_{it}. We assume that the individual returns to the labor force when being denied disability benefits. The value of being denied disability benefits is therefore increasing in the wage.

The application cost, C, is increasing in the deviation between signaled and objective health, which implies that it is costly to lie about the state of health

$$C_\theta(\theta_{it}, h_{it}, w_{it}) \equiv \frac{\partial C(\theta_{it}, h_{it}, w_{it})}{\partial \theta_{it}} \begin{cases} \leq 0 & \text{if} \quad \theta_{it} \leq h_{it} \\ > 0 & \text{if} \quad \theta_{it} > h_{it}. \end{cases}$$

The application cost is also increasing in the wage. This reflects the fact that in order to apply for disability benefits, for example in the U.S., the applicant has to leave his or her current employment and forego wage earnings. In Sweden, a common way to signal bad health is to go on sickness absence, which is more costly for high income earners. Hence,

$$C_w(\theta_{it}, h_{it}, w_{it}) \equiv \frac{\partial C(\theta_{it}, h_{it}, w_{it})}{\partial w_{it}} > 0.$$

The individual maximizes the expected value of applying for disability benefits, (4.1), by choosing the optimal health signal. The first order condition is given by

$$p_\theta(\theta_{it}, m_t) \left[V^{AC}(h_{it}, w_{it}) - V^{RJ}(h_{it}, w_{it}) \right] = C_\theta(\theta_{it}, h_{it}, w_{it}). \tag{4.4}$$

This equation can be solved to yield the optimal health signal as a function of the individual's health and wage as well as the disability benefit threshold

$$\theta^*_{it} = g(h_{it}, m_t, w_{it}).$$

This signal in turn determines the value of applying for disability benefits $V^{AP}(h_{it}, m_t, w_{it})$.

The sufficient conditions for a unique solution are

$$p_{\theta\theta}(\theta^*_{it}, m_t) = -f'(\theta^*_{it} - m_t) < 0 \tag{4.5}$$

and

$$C_{\theta\theta}(\theta^*_{it}, m_t) > 0. \tag{4.6}$$

Condition (4.5) is fulfilled for a wide range of symmetric distribution functions as long as the optimal signal is below the disability benefit threshold, i.e., as long as $\theta^*_{it} < m_t$.

Now that the optimal health signal and the value of applying for disability benefits have been determined, we can see when it is optimal for the individual to apply for disability benefits. The individual applies for disability benefits if the value of applying

is larger than the value of working, i.e., if

$$V^{AP}(h_{it}, m_t, w_{it}) > V^{WO}(h_{it}, w_{it}),$$

where the latter is increasing in both health and wage.

We assume that in the worst health state it is optimal for the individual to apply for disability benefits while in the best health state it is optimal to work. This assumption is not very restrictive and can be motivated by a disutility of working which is decreasing in health. In other words, a healthy individual has a low disutility of working as well as a high cost of signaling a health low enough to receive disability benefits. This makes working more attractive than applying for disability benefits. An individual in bad health, on the other hand, faces a large disutility of working while at the same time having little or no cost of signaling health below the disability benefit threshold, which makes applying for disability benefits more attractive than working. Defining $\underline{h}$ as the lowest and $\overline{h}$ as the highest possible health state, we can express these assumptions as

$$V^{AP}(\underline{h}, m_t, w_{it}) > V^{WO}(\underline{h}, w_{it})$$

and

$$V^{AP}(\overline{h}, m_t, w_{it}) < V^{WO}(\overline{h}, w_{it}).$$

Since the value of working is increasing in health, these assumptions guarantee the existence of a health level below which it is optimal to apply for disability benefits and above which it is optimal to continue working. Let us define this health cut-off level as $h^{CO}(m_t, w_{it})$. The application decision can then be formalized as follows

$$D_{it}^{AP} = \begin{cases} 1 & \text{if} \quad h_{it} < h^{CO}(m_t, w_{it}) \\ 0 & \text{if} \quad h_{it} \geq h^{CO}(m_t, w_{it}), \end{cases}$$

where D_{it}^{AP} is an indicator of individual i applying for disability benefits at time t.

We are interested in how this cut-off level changes with the average disability benefit threshold, m_t. First, we note that the value of working does not depend on screening stringency. Then, we apply the envelope theorem and show how the value of applying for disability benefits changes with m_t in equilibrium

$$\frac{\partial V^{AP}(h_{it}, m_t, w_{it})}{\partial m_t} = p_m(\theta_{it}^*, m_t) \left[V^{AC}(h_{it}, w_{it}) - V^{RJ}(h_{it}, w_{it}) \right].$$

From condition (4.3) we know that $p_m > 0$. If the value of an accepted application for disability benefits is strictly larger than the utility from being denied benefits, we know that

$$\frac{\partial V^{AP}(h_{it}, m_t, w_{it})}{\partial m_t} > 0.$$

In other words, the value of applying for disability benefits is increasing in m_t in equilibrium.

As illustrated in Figure 4.1, a reduction in stringency, i.e., an increase from m_t to m_t', pushes the value of applying for disability benefits upwards, from $V^{AP}(h_{it}, m_t)$ to $V^{AP}(h_{it}, m_t')$. Since the value of working, $V^{WO}(h_{it})$, does not change, this implies that the health cut-off level increases from $h^{CO}(m_t)$ to $h^{CO}(m_t')$. We therefore know that

$$\frac{\partial h^{CO}(m_t, w_{it})}{\partial m_t} > 0. \tag{4.7}$$

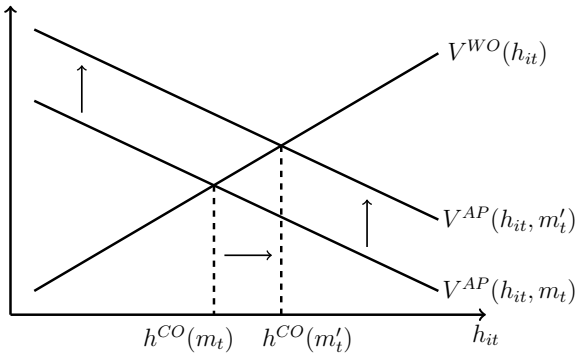

FIGURE 4.1. Change of the Health Cut-Off as Stringency Decreases.

As can be seen in Figure 4.1, the pool of disability benefit applicants increases from all individuals whose health falls in the interval $[\underline{h}, h^{CO}(m_t)]$ to all individuals with objective health in $[\underline{h}, h^{CO}(m_t')]$. The average health among applicants therefore increases when stringency decreases.

However, to address the question of how the probability of receiving disability benefits, and hence the acceptance rate to the disability insurance program, changes, we have to see how the signaled health, θ_{it}^*, is affected by a lower stringency. Implicitly differentiating the first order condition (4.4) yields

$$\frac{d\theta_{it}^*}{dm_t} = -\frac{p_{\theta m}(\cdot)\left[V^{AC}(\cdot) - V^{RJ}(\cdot)\right]}{p_{\theta\theta}(\cdot)\left[V^{AC}(\cdot) - V^{RJ}(\cdot)\right] - C_{\theta\theta}(\cdot)}, \tag{4.8}$$

where

$$p_{\theta m}(\theta_{it}^*, m_t) = f'(\theta_{it}^* - m_t) > 0$$

is the cross derivative of p with respect to θ and m. The numerator is then strictly positive if the value of an accepted application for disability benefits is strictly larger than the utility from being denied benefits. Using this assumptions as well as conditions

(4.5) and (4.6) we can show that the denominator of (4.8) is strictly negative. We can therefore conclude that

$$\frac{d\theta_{it}^*}{dm_t} > 0. \tag{4.9}$$

This implies that not only does the average objective health among applicants increase when stringency decreases, also the health signal of each applicant increases. The intuition behind this result is that a reduction in stringency increases the probability of receiving disability benefits which, in turn, allows the individual to signal health closer to the true state, and hence reducing application costs.

A change in stringency therefore has two effects. The first is a macro or population effect that describes the change in average objective health among applicants due to the change in the composition of the applicant pool. The second is a micro or behavioral effect that captures the fact that applicants signal a different health when stringency changes.

Recall that we are interested in how the probability of receiving disability benefits, and hence the acceptance rate, changes when stringency decreases. The total differential of the probability of receiving benefits, $p(\cdot)$, is given by

$$dp(\theta_{it}^*, m_t) = p_m(\theta_{it}^*, m_t)dm_t + p_\theta(\theta_{it}^*, m_t)d\theta_{it}^*.$$

Rearranging yields

$$\frac{dp(\theta_{it}^*, m_t)}{dm_t} = p_m(\theta_{it}^*, m_t) + p_\theta(\theta_{it}^*, m_t)\frac{d\theta_{it}^*}{dm_t}. \tag{4.10}$$

Since $p_m > 0$, the first part of this expression is strictly positive, while the second part is strictly negative because of $p_\theta < 0$ as well as $d\theta_{it}^*/dm_t > 0$. The first part captures the fact that a lower stringency, ceteris paribus, increases the acceptance rate to the disability insurance program. However, the behavioral effect described above leads to individuals signaling better health when stringency decreases, which in turn reduces the acceptance rate to the disability insurance program. The effect of stringency on the probability of receiving disability benefits is therefore undetermined. In other words, it is not clear whether the acceptance rate of disability applicants increases or decreases or remains unchanged when stringency changes.

3. Numerical Simulation

In this section we simulate the model numerically in order to illustrate the results from the previous section. To do so, we make the following assumptions on functional forms

$$V^{AC}(h_{it}, w_{it}) = \delta w_{it} + h_{it},$$

$$V^{RJ}(h_{it}, w_{it}) = \alpha w_{it} + h_{it} - \phi(\overline{h} - h_{it}),$$
$$V^{WO}(h_{it}, w_{it}) = w_{it} + h_{it} - \phi(\overline{h} - h_{it}).$$

In other words, utility is linear in consumption as well as in health. When working, the individual incurs disutility of labor which is decreasing in health. Working causes no disutility when the individual is in perfect health. An individual on disability benefits receives the compensation rate times the wage as a benefit. When being denied disability benefits, the individual returns to the labor force and earns the wage. However, there is a wage penalty, $\alpha < 1$, associated with having applied for disability benefits. Applying for disability benefits might signal low productivity to the employer. Also, there could be stigma associated with application to the disability insurance program.

The cost of applying for disability benefits is assumed to be quadratic in the deviation of signaled health from objective health and linear in the wage

$$C(\theta_{it}, h_{it}, w_{it}) = \rho(h_{it} - \theta_{it})^2 + \psi w_{it}.$$

In particular, we see that assumption (4.6) is fulfilled

$$C_{\theta\theta}(\theta_{it}, h_{it}, w_{it}) = 2\rho > 0.$$

The idiosyncratic error term, μ_{it}, is logistically distributed and therefore

$$p(\theta_{it}, m_t) = \frac{1}{1 + \exp(\theta_{it} - m_t)}.$$

With this functional form we can show that the first derivative with respect to θ is negative

$$p_\theta(\theta_{it}, m_t) = -p(\theta_{it}, m_t)[1 - p(\theta_{it}, m_t)] < 0.$$

The second derivative with respect to θ as well as the cross derivative with respect to θ and m are given by

$$p_{\theta\theta}(\theta_{it}, m_t) = p_\theta(\theta_{it}, m_t)\,[2p(\theta_{it}, m_t) - 1] < 0 \;\Leftrightarrow p(\theta_{it}, m_t) > \frac{1}{2} \;\Leftrightarrow\; \theta_{it} < m_t.$$
$$p_{\theta m}(\theta_{it}, m_t) = p_m(\theta_{it}, m_t)\,[2p(\theta_{it}, m_t) - 1] > 0 \;\Leftrightarrow p(\theta_{it}, m_t) > \frac{1}{2} \;\Leftrightarrow\; \theta_{it} < m_t.$$

In other words, all assumptions made in the previous section are fulfilled as long as individuals signal health below the disability benefit threshold.

The health of individual i at age a is given by

$$h_{ia} = \exp(-\kappa(a - 30) + \varepsilon_i) \quad \text{for} \quad a = 30, \ldots, 100,$$

where

$$\varepsilon_i \sim N(0, \sigma^2).$$

An individual in this model survives from one period to the next as long as her health is above a threshold, i.e., as long as $h_{ia} \geq q_1$. However, there is also a small probability of dying, $q_2(a)$, which is independent of health and increasing in age. All parameters associated with the health process are calibrated to replicate the survival probabilities that we see in the data. Figure 4.2 compares the probability of surviving from one period to the next, conditional on living to this period, in our model with the corresponding numbers in the data.

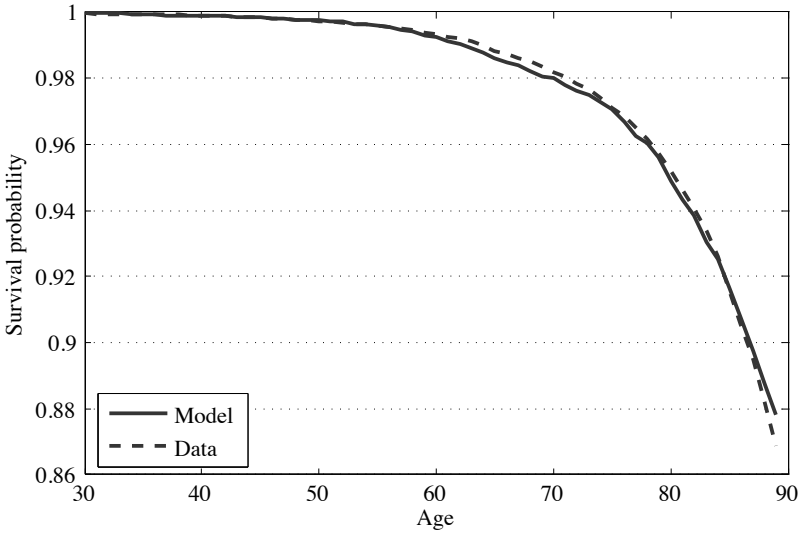

FIGURE 4.2. Conditional survival probabilities.

All parameter values used to simulate the model are presented in Table 4.1. For the purposes of this analysis we keep the wage constant over time and identical for all individuals, i.e., $w_{it} = w$.

Each period $n = 1000$ individuals aged 30 enter the model. They live at most to the age of 100. We draw a health shock, ε_i, for each individual as well as an error term, μ_{it}, for each individual in each period. We then simulate the model for each level of stringency $m = 0, 0.05, \ldots, 1$.

In Figure 4.3 we show how the health cut-off, the acceptance rate to the disability insurance program, the objective health and the signaled health of individuals applying for disability benefits as well as of individuals accepted to the disability insurance program vary with the stringency level m.

Parameter	Value	Description
α	0.5	Wage penalty
δ	0.8	Disability compensation rate
$\overline{h}$	1	Perfect health
$\underline{h}$	0	Worst health
κ	0.04	Health parameter
ρ	1	Application cost multiplier (signal)
ϕ	2	Disutility of labor multiplier
ψ	1	Application cost multiplier (wage)
q_1	0.1	Survival cut-off
$q_2(30)$	0	Probability of dying at age 30
$q_2(50)$	0.002	Probability of dying at age 50
$q_2(70)$	0.007	Probability of dying at age 70
$q_2(80)$	0.01	Probability of dying at age 80
$q_2(90)$	0.06	Probability of dying at age 90
σ	0.5	Standard deviation of health shock
w	1	Wage

TABLE 4.1. Parameter values

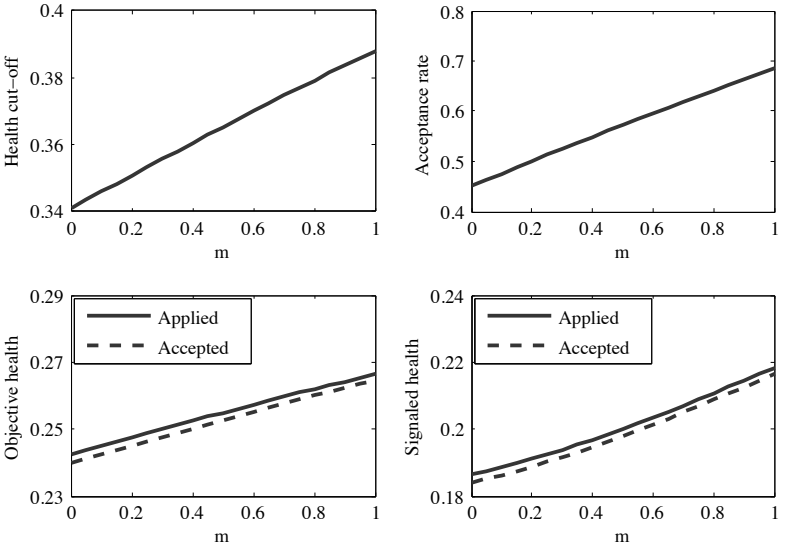

FIGURE 4.3. Simulation of the health cut-off, acceptance rate, objective and signaled health for different levels of stringency.

Recall that an increase in m corresponds to a decrease in stringency. In the upper left panel of Figure 4.3 we see the increase in the health cut-off described in condition (4.7). As stringency decreases, the health level below which it is optimal to apply to disability benefits increases. As a consequence we see that the average objective health

among disability benefits applicants and recipients increases with lower stringency. In condition (4.9), we see that not only objective health but also signaled health among applicants and recipients increases when stringency decreases. This is illustrated in the lower right panel of Figure 4.3. This increase in signaled health implies, ceteris paribus, a lower acceptance rate to the disability insurance program as stringency decreases. However, as discussed above, a lower stringency by itself leads to a higher acceptance rate and which of those two effects prevails is analytically unclear.[2] In our simulation, the latter effect dominates the former and the acceptance rate increases with falling stringency.

In Figure 4.4 we compare objective health and mortality of disability benefits applicants to non-applicants and of recipients to non-recipients. Here, mortality is measured by using longevity in the form of the difference between the maximum number of years alive and the actual number of years alive.[3]

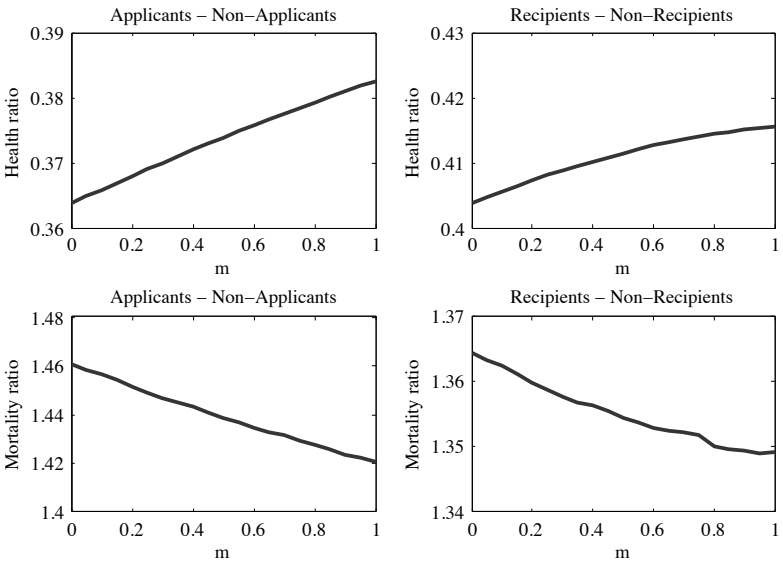

FIGURE 4.4. Comparison of the health and mortality of disability benefit applicants to non-applicants and of disability benefit recipients to non-recipients.

[2] See equation (4.10).

[3] Recall that agents enter the model at age 30 and die with certainty at age 100. The maximum number of years alive is therefore 71.

In the upper left panel of Figure 4.4 we see the average objective health of disability insurance applicants relative to the average objective health of non-applicants. We see that, as stringency decreases, the disability insurance applicants become relatively healthier. In the upper right panel we see that the same holds for disability benefits recipients relative to non-recipients.

The two lower graphs in Figure 4.4 show the average mortality of disability insurance applicants relative to non-applicants on the left side and disability insurance recipients relative to non-recipients on the right side. We can see that mortality of applicants and recipients relative to non-applicants and non-recipients decreases when stringency decreases.

Consider the null hypothesis that there are no changes in screening stringency over time. In that case, we get from the theoretical model, that the relative health of disability benefit recipients compared to non-recipients is constant. If we could observe the health of individuals receiving disability benefits, we could therefore test the null hypothesis of constant screening stringency against the alternative hypothesis that screening stringency changes over time.

It is, however, difficult to observe the objective health of individuals. One common measure of objective health is mortality. Under the assumption that mortality, at a given age, is inversely proportional to the health of the individual, the relative mortality of disability benefit recipients compared to non-recipients is constant if there are no changes in screening stringency. Thus, by observing the mortality of disability beneficiaries after admittance we can identify changes in screening stringency.

4. Empirical Strategy

In this section, we propose an empirical strategy for testing whether screening stringency changes over time. We use the result from the previous section that the relative mortality of awarded disability beneficiaries compared to non-beneficiaries should be constant over time, for a given level of screening stringency.

First, we need to make an assumption about the distribution of longevity. We could assume a specific form of the underlying distribution of the time to death and perform the analysis within a full maximum likelihood framework. Instead we assume that, under the null hypothesis of constant screening stringency, the mortality of awarded disability beneficiaries at each age is proportionally related to the mortality of non-beneficiaries. This assumption holds for a number of known distributions, such as the Weibull and the exponential distribution, and allows for a semi-parametric estimation of the parameters of interest. The mortality hazard of awarded disability beneficiaries,

$\lambda(a)$, is then given by

$$\lambda(a) = g(a)\exp(\beta),$$

where $g(a)$ is the baseline hazard at age a and $\exp(\beta)$ is the mortality ratio of awarded beneficiaries compared to non-beneficiaries. We know that $\beta > 0$ since the mortality of disability insurance beneficiaries is larger than the general mortality at each age, as is shown in Figure 4.4.

One concern when bringing the model to the data is that the baseline hazard is likely to vary across cohorts due to, for example, the expansion of education or changes in nutrition. This can be accounted for within a duration analysis framework by stratifying the baseline mortality risk by cohort. Another reason to stratify the baseline hazard by cohort is to solve the problem of length-bias sampling, that comes from the fact that we observe different cohorts during different ages. This issue is discussed further in the data section.

Under the null hypothesis that screening stringency is constant over time, β can be estimated using the stratified Cox proportional hazard model

$$\lambda_{ic}(a) = g_c(a)\exp(\beta D_i(a)), \quad i = 1, ..., n,$$

where $\lambda_{ic}(a)$ is the mortality hazard of individual i in cohort c at age a and $g_c(a)$ is the baseline hazard of cohort c. $D_i(a)$ is a step function that takes the value one after individual i is awarded disability benefits at age a. For an individual who is not awarded disability benefits during the studied time period, $D_i(a)$ is zero for all ages.

The alternative hypothesis, that the disability benefit threshold changes over time, can be analyzed by testing if $\beta_t = \beta$ for all t in the model

$$\lambda_{ic}(a) = g_c(a)\exp(\beta_t D_{it}(a)), \quad i = 1, ..., n, \quad t = 1, ...T, \tag{4.11}$$

where t is the year in which the individual was awarded disability benefits. An equivalent model would be to estimate an average mortality ratio of awarded disability beneficiaries compared to non-beneficiaries, and separate effects of the deviation from the average over time. Identification of β_t comes from the within cohort comparison of mortality hazards at age a for those who are awarded disability benefits at calendar time t compared to those who have *not yet* obtained disability benefits. The individuals in the comparison group may, however, begin receiving disability benefits later if eligible then. Since individuals are no longer eligible for disability benefits above age 64, we censor the observations above this age.[4]

[4] The reason is that we do not know if an individual aged, say, 66 would still have remained in the control group. Had the eligible age been 66 years, his or her health could have been such that he/she instead would have received disability benefits if applying. Keeping individuals in the control group

The model also allows us to test for differences in screening stringency across different groups of the population. We can test for differences across age groups by estimating the model

$$\lambda_{ic}(a) = g_c(a)\exp(\beta_t D_{it}(a) + \delta_t^a D_{it}(a) I(a_i = a)), \quad i = 1, ..., n, \quad t = 1, ...T, \quad (4.12)$$

where the coefficients δ_t^a capture differences in screening stringency across age groups over time.

We can also test for differences in screening stringency across other types of populations defined by, e.g., gender or region of residence. We could use the same model as in equation (4.12). This test, however, is based on the assumption that the productivity and health of the two populations at each age are the same, which might be unlikely for men and women or for individuals living in different regions of Sweden. However, under the assumption of a proportional health and productivity difference at each age a in the two populations, it is still possible to test for differences in screening stringency across the two populations. We then get that $g_c(a, p = 2) = g_c(a, p = 1)\exp(\delta_0)$, which gives us the proportional hazard model

$$\lambda_{ic}(a) = g_c(a)\exp(\delta_0 I(p_i = 2) + \beta_t D_{it}(a) + \delta_t^p D_{it}(a) I(p_i = 2)), \quad (4.13)$$
$$i = 1, ..., n, \quad t = 1, ...T,$$

where the coefficients δ_t^p describe the changes in screening stringency across the two populations over time.

5. An Application to Sweden

We apply the empirical strategy, outlined in the previous section, to Sweden. In this section, we describe the Swedish disability insurance program, discuss the data used for estimation and present the development of program participation over time and across groups. As we will see, Sweden is an interesting application since the disability benefit award rate has varied considerably during the time period under study, in a fashion that cannot always be attributed to formal program changes. The strength of the proposed empirical strategy is that also informal changes in the implementation of program rules can be detected.

5.1. The Swedish Disability Insurance Program. The Swedish disability insurance program replaces foregone earnings due to a lasting reduced working capacity for workers aged 19–64. Benefits can be granted part-time or full-time, depending on

after the age of 64 could lead to an attenuation of the estimated screening effect. However, when we include data above age 64 in the analysis, the patterns of the results are similar.

the extent of the work impairment, and the compensation level is 64 percent of the assumed income, up to a ceiling. Before 2003, the disability insurance program was part of the public pension system instead of the social insurance system, and benefits were calculated according to the formula that applied to old-age pension. The average level of compensation was, however, similar before and after 2003. Disability benefits can be supplemented with payments from occupational insurances, covering the majority of Swedish workers.[5] The total compensation level, including occupational insurances, is about 80 percent for workers in the private and local government sector and about 85 percent for workers in the central government sector, for earnings below the ceiling.

The basis in the eligibility determination process is that a clear relationship between medical causes and the reduction in working capacity is required in order to qualify for disability benefits. During certain time periods and for certain groups of workers, however, consideration could also be given to the individual's labor market situation. From 1972 to October 1991, individuals aged 60 to 64 could be granted disability benefits for pure labor market reasons, without a health assessment, if they were still unemployed when reaching the time limit in the unemployment insurance. From 1970 to January 1997, favorable rules for workers aged 60 to 64 further implied less strict health assessments, no test of employability and lower requirements for changing occupation or area of residence in order to find work. During the same time period, workers of all ages could be granted disability benefits for labor market reasons and health reasons combined, if they suffered from a reduced working capacity for medical reasons and had been unemployed for 1–2 years.

Since the reform in 1997, individual job opportunities should not be taken into account in the eligibility determination for disability benefits. The explicit focus on medical conditions when assessing the individual's working capacity makes the Swedish disability insurance program somewhat different from the U.S. counterpart. The Social Security definition of disability in the U.S. is "the inability to engage in a substantial gainful activity in the U.S. economy" and, in 1984, the program's eligibility criteria were liberalized from mainly considering diagnostic and medical factors to placing more weight on the applicant's ability to function in a work-like setting.[6] This implies an inherent feature in the U.S. disability insurance program of responding to changes in individual job opportunities.

[5] Sjögren Lindquist and Wadensjö (2007) estimate that 96 percent of Swedish workers are covered by a collective agreement allowing for occupational insurance, and that almost all of these workers fulfill the criteria for receiving occupational insurance when sick. Between 60 and 80 percent of workers claim occupational insurance when receiving disability benefits.

[6] See e.g. Autor (2011).

5.2. Data. To study the Swedish disability insurance program, we use data covering all individuals in Sweden aged 30–64 from 1985 to 2008. The data originates from administrative records, collected and maintained by Statistics Sweden. For each individual, we observe birth year, gender and the county of residence. We also have access to detailed spell data on the collection of disability benefits, provided by the Swedish Social Insurance Agency. For each year, we record if the individual received disability benefits and whether it was the entry year to the program. Since the data is based on disability benefits payments, there might be a slight lag of the first payment compared to the decision to award benefits. Finally, we add information about mortality, provided by the National Board of Health and Welfare. Mortality is measured until 2010 or until the year the individual turns 64.

Once awarded benefits, most individuals remain in the disability insurance program until retirement. For the individuals who are awarded disability benefits at several occasions, the first year of award is regarded as the entry year to the program.[7] We define the county of residence as the county in which the individual resides when first awarded disability benefits. For individuals never awarded benefits, the county of residence is defined as the county in which the individual lives for the most years during the period under study.

The sample used for estimation consists of individuals aged 30–64 during 1985–2008 who are not already receiving disability benefits during the first year in which they are observed. This includes individuals who are observed from the beginning of the sample period and are not receiving disability benefits in the first year, 1985, as well as individuals entering the sample by turning 30 during 1986–2008 and not already receiving disability benefits at that age. The sampling of individuals aged 30–64 from 1985 to 2008 implies that different cohorts will be observed for different periods of time and at different ages. This problem of length bias sampling is solved by stratifying the baseline hazard by cohort in the Cox regression model.

5.3. The Development of Program Participation. After we apply the proposed estimation strategy to the Swedish disability insurance program, we relate the empirical results for the changes in screening stringency over time to the development of disability benefit awards. Figure 4.5 presents the development of program participation. The solid line, plotted against the left axis, shows the share of individuals aged 30–64 receiving disability benefits. The dashed line, plotted against the right axis, shows the disability benefit award rate among those not already receiving benefits. From 1985 to 2008, the share of disability beneficiaries increased from just over

[7] This simplification leads to only 2.4 percent of the registrations of program participation not corresponding to the data.

8 percent to about 12 percent, but the fluctuations in the award rate were large. The disability benefit award rate was high during the late 1980s and early 1990s, declined during the mid 1990s and increased rapidly during the late 1990s and the early 2000s. Since the mid 2000s, there has been a remarkable drop in new disability benefit awards, which has even resulted in a decline in the number of beneficiaries in recent years.

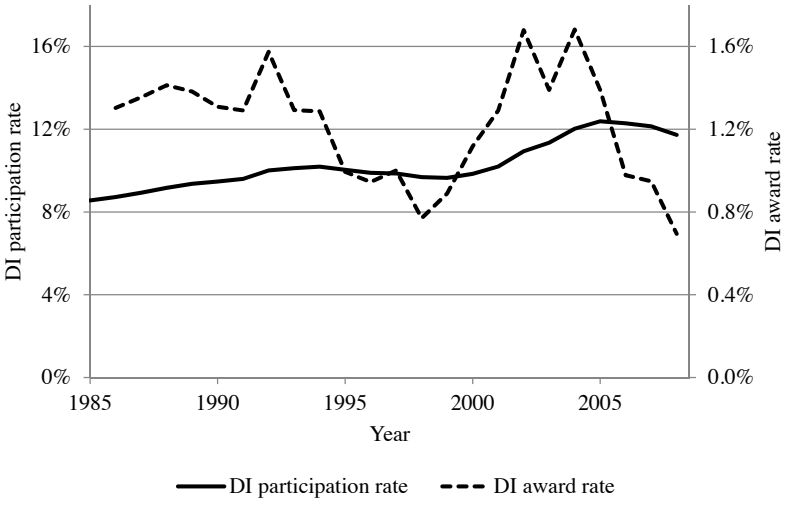

FIGURE 4.5. The share of individuals receiving disability benefits and the disability benefit award rate in ages 30–64 in Sweden, 1985–2008

The reasons for the fluctuations in the disability benefit award rate over time are not fully understood. The decline in awards in the early 1990s coincided with the removal of the pure labor market reasons for workers aged 60 to 64 in 1991, but the adaption did not appear immediately. There were also fluctuations in the award rate around the reform in 1997, when the favorable rules for workers aged 60 to 64 and the eligibility rules for health and labor market reasons combined were removed. The large increase during the early 2000s and the subsequent decline since the mid 2000s, however, cannot be attributed to any formal program changes. In 2005, the 21 regional offices of the Social Insurance Agency were integrated into one central authority. This might have affected the internal processes within the agency and be a reason for the decline in awards after the re-organization. The formal eligibility criteria were substantially

tightened in July 2008, but the large drop in disability benefit awards appeared well before this. Whether or not an informal change in the assessment of new applicants preceded the change in formal program rules will be investigated in the empirical analysis.

We use the empirical strategy to study changes in the relative screening stringency across age groups, gender and regions of residence. As discussed above, several of the changes in formal eligibility criteria have concerned the age group 60 to 64. Figure 4.6 shows the disability benefit award rate in the age groups 30–59 and 60–64. Following the removal of the pure labor market reasons for workers aged 60 to 64 in October 1991, the award rate in this age group fell from between 6 and 7 percent to just above 2 percent in 1998. Interestingly, the fluctuations in the award rate around the removal of the favorable elderly rules in 1997 also appeared exclusively for the older age group. The pattern in award rates during the 2000s was similar across the two age groups. In the mid 2000s, the award rate in the younger age group was higher than ever before. In the empirical analysis, we study whether this is solely due to changes in the underlying health, or whether it can also be related to a looser screening of younger applicants.

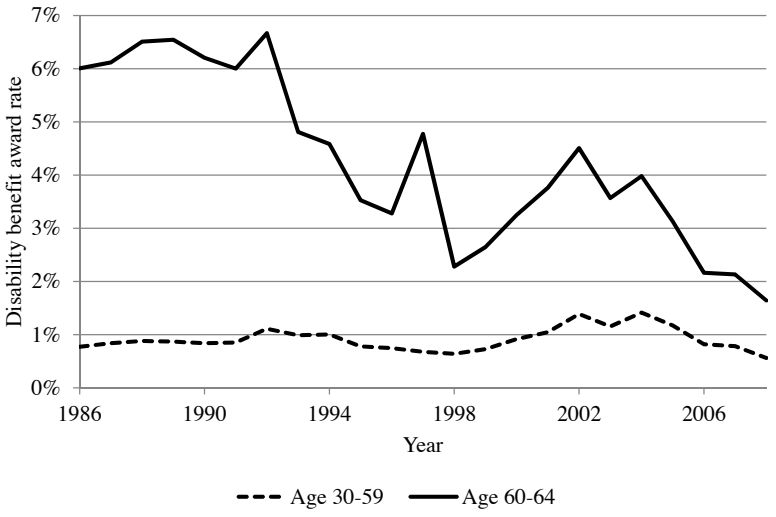

FIGURE 4.6. The disability benefit award rate in the age groups 30–59 and 60–64 in Sweden, 1986–2008

Figure 4.7 shows the disability benefit award rate during 1986–2008 for men and women. Women have been more likely to be awarded disability benefits throughout the period. Between 1986 and 1998, the award rates of women and men developed similarly. The increase in awards between 1998 and 2005, however, was much larger for women. The decrease in awards since 2005 has again led to a convergence in award rates of women and men. The changes in relative award rates across gender could be motivated by changes in the underlying health. Jönsson, Palme and Svensson (2012) show that the health of older men has improved over the last decade, while the health of older women has remained fairly constant. The changes could, however, also be due to differential assessments of eligibility for men and women over time. This is investigated in the empirical analysis.

FIGURE 4.7. The disability benefit award rate among men and women in Sweden, 1986–2008

The inflow to the disability insurance program has varied considerably across regions. Figure 4.8 presents the development of the disability benefit award rates in four large regions in Sweden. The inflow has been lowest in the Stockholm region and highest in the northern part of Sweden, Norrland, throughout the period, whereas the middle part of Sweden, Svealand (excluding Stockholm), and the southern part of

Sweden, Götaland, have had intermediate award rates. We expect differences in award rates, even after controlling for health, before 1997 when labor market opportunities could be taken into account in the eligibility determination. The unemployment rate has typically been highest in Norrland and lowest in the Stockholm region, and we see that the Stockholm region contributed less to the large inflow during the late 1980s. In the empirical analysis, we study whether the differences in award rates across regions can be explained by varying degrees of screening stringency.

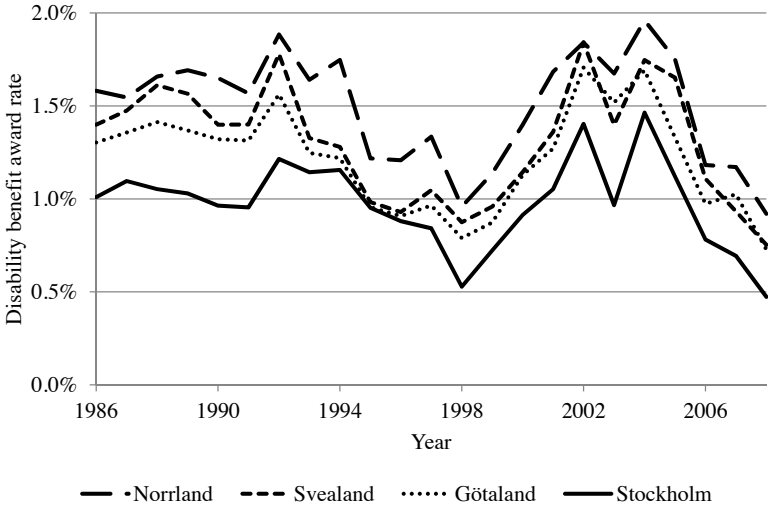

FIGURE 4.8. The disability benefit award rate across Swedish regions, 1986–2008

6. Results

This section presents the results from the empirical analysis. First, we provide the main estimation results of the overall changes in screening stringency in the Swedish disability insurance program from 1986 to 2008. Next, we analyze how the screening stringency has changed for different groups of the population, defined by age, gender and region of residence.

6.1. Main Results. Figure 4.9 presents the main results from the estimation of equation (4.11). The estimated coefficients are presented in terms of hazard ratios, $\exp(\beta_t)$, plotted by the solid line. A hazard ratio of one implies that there is no difference in mortality between new disability beneficiaries and non-beneficiaries, whereas a hazard ratio of three implies a three times higher mortality hazard of new beneficiaries. Changes in the excess mortality of new disability beneficiaries over time indicate that the screening stringency in the disability insurance program has changed. The dotted lines plot the 95 percent confidence interval, with standard errors clustered by birth cohort.

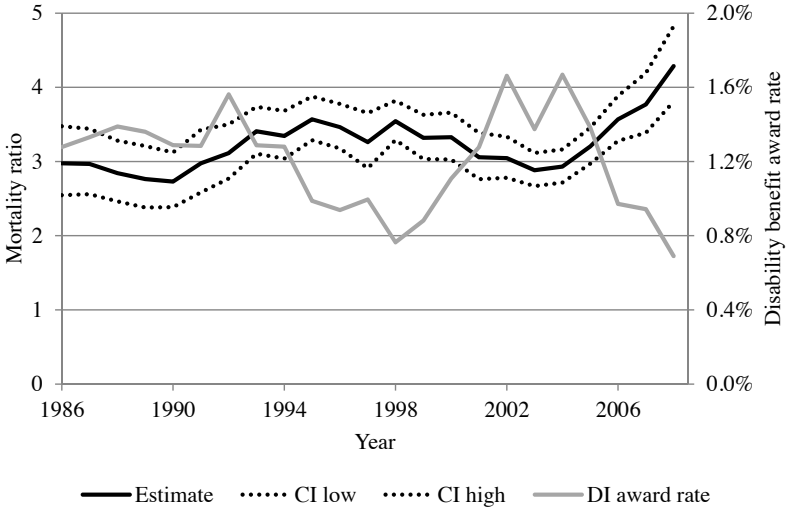

FIGURE 4.9. Estimated mortality hazard ratio of new disability beneficiaries compared to non-beneficiaries and the disability benefit award rate, 1986–2008

The estimates in Figure 4.9 suggest that the screening stringency was relatively low during the 1980s. The mortality hazard of new disability beneficiaries was less than three times as large as for non-beneficiaries during this period. In the early 1990s, screening stringency increased and then remained at the new level during the mid 1990s. Between 1998 and 2004, screening stringency steadily declined. Since 2004, however,

screening stringency has rapidly increased. The mortality hazard of new disability beneficiaries in 2008 is more than four times higher than for non-beneficiaries.

Figure 4.9 also presents the disability benefit award rate among individuals who are not already receiving benefits. The changes in screening stringency correspond well to the changes in the disability award rate over time. The award rate was high during the late 1980s, when screening stringency was low. The increasing screening stringency during the early 1990s is also reflected in a decreasing disability benefit award rate. The spike in awards in 1992 does not appear to be caused by slackening stringency. Anecdotal evidence suggests that a shortage of caseworkers led to a queue of cases during the preceding years, and that the spike in 1992 is due to a shifting of assessments across years. This would explain why the screening stringency is unaffected. The increase in awards in 1997, on the other hand, is reflected in a reduced screening stringency during that particular year, but the development is counteracted in 1998. This might be a transitory response to the removal of the elderly rules in 1997. From 1998 to 2004, when screening stringency steadily declined, the disability benefit award rate rapidly increased. Finally, the increase in screening stringency from 2004 onwards is reflected in a marked decrease in disability benefit awards.

6.2. Heterogeneity Analysis. In this section, we study the relative changes in screening stringency across age groups, gender and regions of residence. Since several changes in formal eligibility criteria have concerned the age group 60–64, we begin by studying the relative difference in screening stringency over time for workers aged 30–59 and 60–64. To do this, we estimate equation (4.12) and let $a_i = a$ if the individual was awarded benefits at age 60–64. Figure 4.10 presents the hazard ratios of the estimated coefficients, $\exp(\delta_t^a)$, representing the relative mortality hazard of individuals awarded benefits at age 60–64 compared to 30–59. The dotted lines plot the 95 percent confidence interval, with standard errors clustered by birth cohort.

The hazard ratio in Figure 4.10 is below one throughout the period, which suggests that the screening stringency of older beneficiaries has been lower than that of younger beneficiaries. The relative screening stringency for older compared to younger workers was constant during the 1980s but increased rapidly after the removal of the pure labor market reasons for the age group 60–64 in 1991. During the mid 1990s, the relative screening stringency for individuals awarded disability benefits above and below age 60 was again fairly constant, except around the time of the removal of the favorable elderly rules in 1997. This change led to a temporarily reduced screening stringency for older workers in 1997, followed by a spike in screening stringency in 1998. After that, the relative screening stringency returned to the previous level, increased again in the early 2000s and decreased slightly during the last few years of observation.

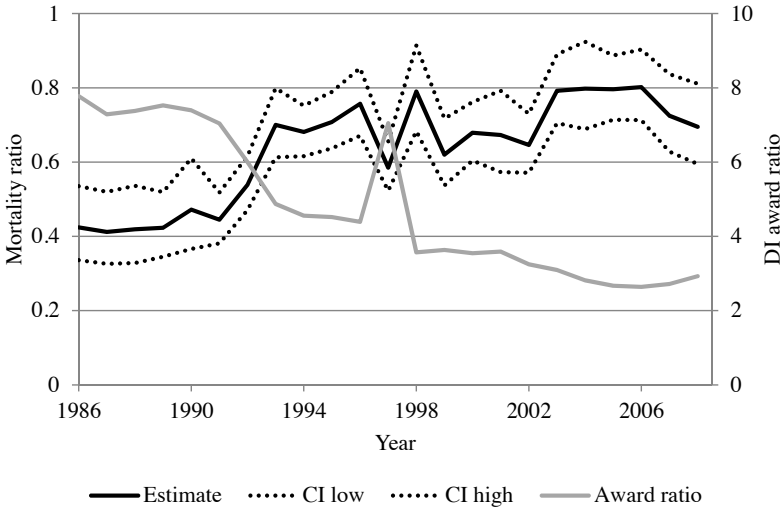

FIGURE 4.10. Estimated mortality hazard ratio of disability beneficiaries aged 60–64 compared to beneficiaries aged 30–59 and the disability benefit award ratio of the same age groups, 1986–2008

The grey line in Figure 4.10 shows the relative disability benefit award rate of the age group 60–64 compared to the age group 30–59. The development of screening stringency corresponds well to the relative award rate of older compared to younger workers. Older workers were awarded disability benefits almost eight times as much as younger workers during the 1980s, when the pure labor market reasons for older workers were in place. After these rules were removed in 1991, and the relative screening stringency for older compared to younger workers increased, the relative award rate rapidly decreased. The transitory decrease in relative screening stringency in 1997 can also be depicted in a sudden spike in relative disability benefit awards of older compared to younger workers. During the early 2000s, the relative award rate of older compared to younger workers declined and the relative screening stringency increased. This suggests that the large increase of younger disability beneficiaries during this period is not only driven by relatively worse health in this group, but also by reduced screening stringency for younger workers.

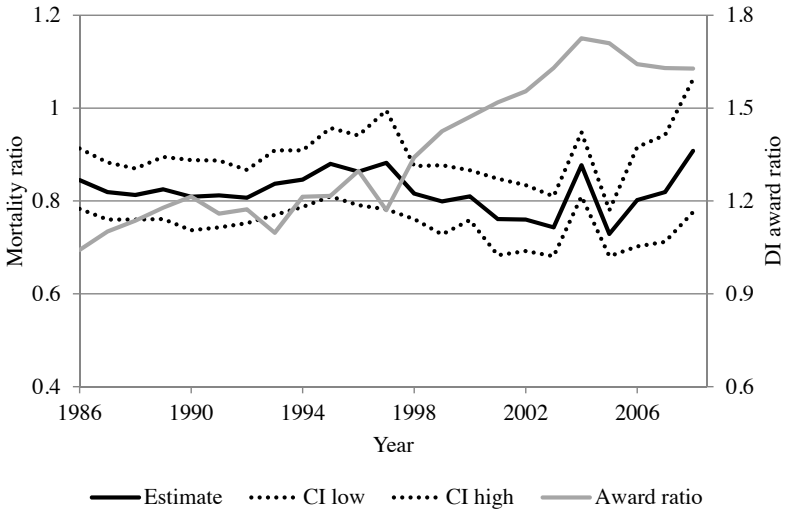

FIGURE 4.11. Estimated mortality hazard ratio of female compared to male disability beneficiaries and the disability benefit award ratio of women compared to men, 1986–2008

To study relative changes in screening stringency across gender, we use the model presented by equation (4.13) and let $p_i = 2$ indicate being a woman. The solid line in Figure 4.11 plots the mortality hazard ratio of new female beneficiaries compared to their male counterparts, $\exp(\delta_t^p)$, and the dotted lines plot the 95 percent confidence intervals with standard errors clustered by birth cohort. The mortality ratio of female compared to male beneficiaries is below one throughout the period, which implies that screening stringency has been lower for women than for men. Recall that we control for the relative mortality of women and men in the population by estimating the parameter δ_0. During the 1980s, the relative screening stringency between men and women was constant, and during the mid 1990s, the stringency across gender converged. From 1997 to 2003, however, the mortality hazard ratio of female compared to male beneficiaries decreased, but has increased during the end of the period. In 2008, there is no longer a significant difference in screening stringency across gender.

Figure 4.11 also presents the relative disability benefit award rate of women compared to men. When the relative screening stringency decreased during the early 2000s

there was a corresponding increase in the relative award rate of women compared to men. This suggests that the large increase in disability benefit awards for women during this period can not be explained solely by changes in health, but is also due to a slackening in screening stringency for women compared to men.

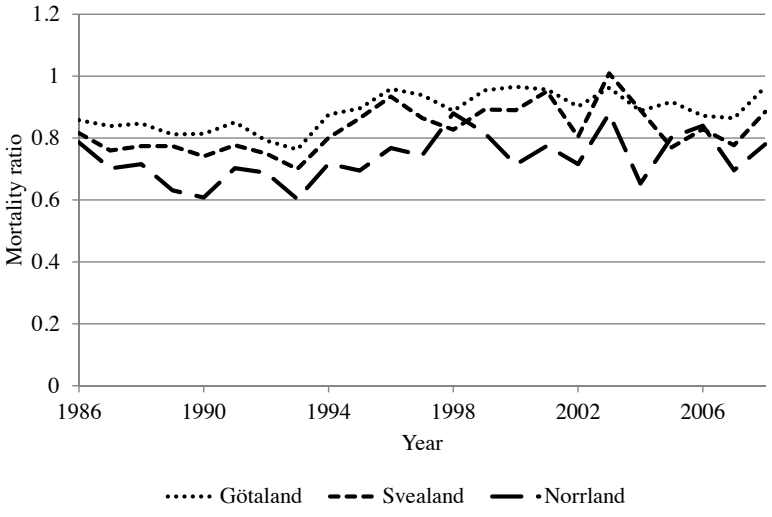

FIGURE 4.12. Estimated mortality hazard ratio of new disability beneficiaries in Swedish regions compared to beneficiaries in the Stockholm region, 1986–2008

To study the relative screening stringency across regions over time, we estimate equation (4.13), but with four groups instead of two: Götaland, Svealand (excluding Stockholm) and Norrland are compared to the Stockholm region. Figure 4.12 shows the mortality hazard ratios of new beneficiaries in each region, compared to beneficiaries in the Stockholm region, during 1986–2008. The relative screening stringency in the other regions compared to Stockholm was lowest during the late 1980s and the early 1990s, but has approached one since the possibilities of taking labor market reasons into account were removed. Figure 4.13 shows the relative disability benefit award rate in the regions relative to the award rate in Stockholm. The relative screening stringency across regions correspond well to the relative disability benefit award rates.

This suggests that part of the explanation for the differential inflow to the disability insurance program across regions is differences in screening stringency.

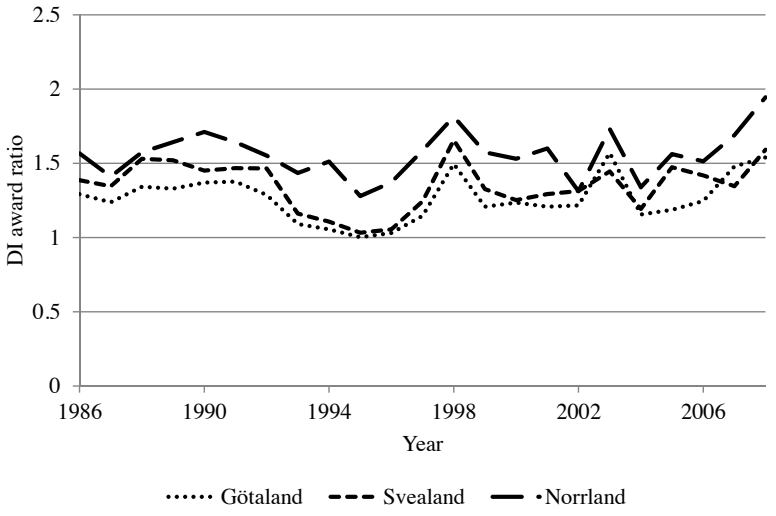

FIGURE 4.13. The disability benefit award ratio of Swedish regions compared to the Stockholm region, 1986–2008

7. Conclusion

Changes in the screening stringency of applications to the disability insurance program are potentially important for explaining program growth. Screening stringency is, however, inherently difficult to observe since it depends on the implementation of program rules as well as formal eligibility criteria.

In this paper, we derive a theoretical model of the application decision to the disability insurance program. Individuals decide whether to apply for disability benefits and, if so, which health to signal to the caseworker. The application cost is increasing in the deviation of signaled health from objective health. The probability of a successful application is increasing in the average disability benefit threshold, which is known to the individual, and decreasing in the signaled health. We show that it is optimal for individuals to apply for benefits if their health is below a certain cut-off, which

is increasing in the average disability benefit threshold. When screening stringency is reduced, i.e., when the average threshold increases, the health cut-off increases and the average health among applicants improves. Since the signaled health among applicants also increases when stringency decreases, the change in the acceptance rate to the disability insurance program is undetermined. The laxer stringency has a positive effect on the acceptance rate while the improved health signal has a negative one.

We then show that the health of disability benefits recipients relative to non-recipients improves when stringency is reduced. This can be used to detect changes in screening stringency in the disability insurance program. We propose an empirical strategy for assessing changes in screening stringency over time, using the mortality of awarded disability beneficiaries compared to non-beneficiaries over time. The strength of the empirical strategy is that it captures both formal and informal changes in screening stringency. Applying the strategy to Sweden, we find that much of the fluctuations in disability benefit award rates over time can be related to changes in screening stringency. For example, the rapid decline in the disability benefit award rate since 2004, which does not coincide with any formal changes to the program, can be attributed to a substantial increase in the stringency of screening.

References

Autor, D. (2011), 'The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options', *Working Paper 17697*, National Bureau of Economic Research, Cambridge, MA.

Autor, D. and Duggan, M. (2003), 'The Rise in Disability Rolls and the Decline in Unemployment', *Quarterly Journal Economics*, 118, 157–206.

Black, D., Daniel, K. and Sanders, S. (2002), 'The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust', *American Economic Review*, 92, 27–50.

de Jong, P., Lindeboom, M. and van der Klaauw, B. (2011), 'Screening Disability Insurance Applications', *Journal of European Economic Association*, 9, 106–129.

Duggan, M. and Imberman, S. (2009), 'Why Are the Disability Rolls Skyrocketing? The Contribution of Population Characteristics, Economic Conditions, and Program Generosity', *Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly*, ed. by Cutler, D. and Wise, D., University of Chicago Press.

Gruber, J. and Kubik, J. (1997), 'Disability Insurance Rejection Rates and the Labor Supply of Older Workers', *Journal of Public Economics*, 64, 1–23.

Halpern, J. and Hausman, J. (1986), 'Choice under Uncertainty: A Model of Applications for the Social Security Disability Insurance Program', *Journal of Public Economics*, 31, 131–161.

Jönsson, L., Palme, M., and Svensson, I. (2012), 'Disability Insurance, Population Health and Employment in Sweden', *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, ed. by Wise, D., University of Chicago Press, forthcoming.

Karlström, A., Palme, M., and Svensson, I. (2008), 'The Employment Effect of Stricter Rules for Eligibility for DI: Evidence from a Natural Experiment in Sweden', *Journal of Public Economics*, 92, 2071–2082.

Parsons, D. (1991), 'Self-Screening in Targeted Public Transfer Programs', *Journal of Political Economy*, 99, 859–876.

Sjögren Lindquist, G. and Wadensjö, E. (2007), 'Ett svårlagt pussel - kompletterande ersttningar vid inkomstbortfall' *Report for Expert Group on Economic Studies*, 2007:1.

Wise, D. (2012), *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, University of Chicago Press, forthcoming.

The Stockholm School of Economics

A complete publication list can be found at *www.hhs.se/research/publications*.
Books and dissertations are published in the language indicated by the title and can be ordered via e-mail: *publications@hhs.se*.

A selection of our recent publications

Books

Alexandersson, Gunnar. (2011). *Den svenska buss- och tågtrafiken: 20 år av avregleringar.* Forskning i fickformat.
Barinaga, Ester (2010). *Powerful dichotomies.*
Benson, Ilinca Lind, Johnny, Sjögren, Ebba & Wijkström, Filip (red.) (2011). *Morgondagens industri: att sätta spelregler och flytta gränser.*
Einarsson, Torbjörm (2011). *Medlemsorganisationen: Individen, organisationen och samhället.*
Engwall, Lars (2009). *Mercury meets Minerva: business studies and higher education: the Swedish case.*
Ericsson, Daniel (2010). *Den odöda musiken.*
Ericsson, Daniel (2010). *Scripting Creativity.*
Holmquist, Carin (2011). *Kvinnors företagande – kan och bör det öka?*
Lundeberg, Mats (2011). *Improving business performance: a first introduction.*
Melén, Sara (2010). *Globala från start: småföretag med världen som marknad.* Forskning i Fickformat.
Modig, Niklas & Åhlström, Pär. (2011). *Vad är lean?: en guide till kundfokus och flödeseffektivitet.*
Mårtensson, Pär & Mähring, Magnus (2010). *Mönster som ger avtryck: perspektiv på verksamhetsutveckling.*
Ottosson, Mikael (2011). *Skogsindustrin och energiomställningen.* Forskning i fickformat.
Sjöström, Emma (2010). *Ansiktslösa men ansvarsfulla.* Forskning i fickformat.
Wijkström, Filip (2010). *Civilsamhällets många ansikten.*

Dissertations

Bjerhammar, Lena (2011). *Produktutvecklingssamarbete mellan detaljhandelsföretag och deras varuleverantörer.*
Bohman, Claes (2010). *Attraction: a new driver of learning and innovation.*
Bottero, Margherita (2011). *Slave trades, credit records and strategic reasoning: four essays in microeconomics.*
Bång, Joakim (2011). *T Essays in empirical corporate finance.*
Carlsson-Wall, Martin (2011). *Targeting target costing: cost management and inter-organizational product development of multi-technology products.*
Formai, Sara (2012). *Heterogeneous firms, international trade and institutions.*
Freier, Ronny (2011). *Incumbency, divided government, partisan politics and council size: essays in local political economics.*
Fries, Liv (2011). *Att organisera tjänstesektorns röst: en teori om hur organisationer formar aktörer.*
Hemrit, Maetinee (2011). *Beyond the Bamboo network: the internationalization process of Thai family business groups.*
Kaisajuntti, Linus (2011). *Multidimensional Markov-functional and stochastic volatility interest rate modelling.*
Khachatryan, Karen (2011). *Biased beliefs and heterogeneous preferences: essays in behavioral economics.*
Lundmark, Martin (2011). *Transatlantic defence industry integration: discourse and action in the organizational field of the defence market.*
Lundvall, Henrik (2010). *Poverty and the dynamics of equilibrium unemployment: [essays on the economics of job search, skills, and savings].*
Migueles Chazarreta, Damián (2011). *The implications of trade and offshoring for growth and wages.*

Monsenego, Jérôme (2011). *Taxation of foreign business income within the European internal market: an analysis of the conflict between the objective of achievement of the European internal market and the principles of territoriality and worldwide taxation*.

Ranehill, Eva (2011). *Essays on gender, competition and status*.

Runsten, Philip (2011). *Kollektiv förmåga: en avhandling om grupper och kunskapsintegration*.

Setterberg, Hanna (2011). *The pricing of earnings: essays on the post-earnings announcement drift and earnings quality risk*.

Stepanok, Ignat (2011). *Essays on international trade and foreign direct investment*.

Söderblom, Anna (2011). *Private equity fund investing: investment strategies, entry order and performance*.

Wallace, Björn (2011). *Genes, history and economics*.