Inequality in the Labor Market: Insurance, Unions, and Discrimination

Erik Höglin

EFI THE ECONOMIC RESEARCH INSTITUTE

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

Keywords: Social Insurance, Unemployment Insurance, Disability Insurance, Unions, Turbulence, Unemployment, Welfare, Discrimination.

ISBN 978-91-7258-759-5
Printed by:
Elanders, Stockholm 2008

Distributed by:
EFI, The Economic Research Institute
Stockholm School of Economics
P.O. Box 6501, SE-113 83 Stockholm
www.hhs.se/efi

to my family

Contents

Acknowledgements

Many people have influenced this thesis and deserve acknowledgement. First and foremost, I am grateful to my advisor Lars Ljungqvist for his continuous support all through the development of this thesis. Lars has devoted a lot of time to discuss the ideas, challenges and shortcomings of my work but has still given me the freedom to pursue my own ideas and make my own mistakes. Gently advising changes when my arguments were weak or flawed. Importantly, Lars has conveyed the idea that economics is fun, interesting, and important, which has been imperative in motivating me to work harder and do better research.

Magnus Johannesson has been a *de facto* second advisor, advising me and Björn Tyrefors on chapter 5 of this thesis. I am grateful to Magnus for his generosity with time but also for his commitment to research and his easy going attitude.

Other faculty member such as David Domeij, Tore Ellingsen, Martin Flodén, Richard Friberg and Jörgen Weibull have generously helped me with comments, questions, explanations and interesting teaching assignments from which I have benefited a lot. Carin Blanksvärd, Ritva Kiviharju, Anneli Sandbladh and Lillian Öberg deserve a special mentioning. They have always been enormously helpful and friendly and have made my life as a graduate student much easier. I would also like to take the opportunity to thank my undergraduate advisor Nils Gottfries for inspiration and for encouraging me to pursue graduate studies.

I entered the program with five great guys without whom these years would have been much less fun. Björn Tyrefors has been a coauthor, an office-mate, and a true friend. It has been a pleasure to work with Björn, but I have appreciated his warm personality and our discussions of the big and small parts of life even more. I have also had an enormous amount of fun with Ola Granström. No joke has ever been too silly! Without Ola, it would have been more difficult to enter the office with a smile on your face in the mornings, and I would have certainly missed the late nights. Henrik Lundvall has influenced this thesis a lot more than he would ever admit. I am particularly grateful for our endless economics discussions and for Henrik's inability to accept my sometimes flawed intuition. Per Sonnerby's mind bending analogies have also been much appreciated, as has his stand-up-guy qualities. Per has also made me think hard about optimal policy, having 'social planner' as his dream job. Milo Bianchi has taught me a lot, ranging from 'the core' to how to save on your bar bill. In doing so, he has proved to be fun, easy going, and helpful. Fredrik Wilander has been an adjunct member of our class, in the first line of the late nights team, and has become a great friend. Marieke Bos, Harald Edquist, Max Elger, Erik Grönqvist, Peter Gustafsson, Mia Hinnerich, Erik Lindqvist, Kristin Magnusson, Erik Mohlin, Conny Olovsson, Elena Paltseva, Eva Ranehill, Roine Vestman, Nina Waldenström, Sara Åhlén, and Robert Östling have also contributed to making these years more enjoyable.

I am also grateful to those that made this project possible. The generous financial support from Stockholm School of Economics and the Jan Wallander and Tom Hedelius has therefore been much appreciated.

Last but not least, I would like to thank my parents, Christina and Hans-Erik, and my brothers, Alexander and Karl. Their unconditional support and faith in me has been essential. Therefore, I dedicate this thesis to you.

Stockholm, July 2008
Erik Höglin

Introduction

CHAPTER 1

Introduction

All men are not equal in the labor market. For example, in the United States, the average income of the top 10 percent of the income distribution was more than 4 times higher than the average income among the bottom 10 percent in the 1990s (Katz and Autor, 1999). Moreover, income inequality has also been growing over the course of the last decades. The collage to high school wage ratio increased by almost 10 percent from the 1970s to the 1990s. Inequality between observationally identical workers has also increased. Juhn, Murphy and Pearce (1993) and Gottschalk and Moffit (1994) estimate a substantial increase in this so-called residual inequality, i.e. inequality among people with the same education, age, etc.

Income inequality has many sources. For example, some people are more productive than others because they are more highly educated or because they have built up their productivity through on the job training and experience. In addition, some individuals' incomes temporarily or permanently drop because of unemployment or disability. There may also be inequalities that do not reflect productivity differences, but instead are caused by discrimination.

This thesis deals with three topics, all related to inequality. First, chapters 2 and 3 ask how to insure individuals against unemployment and disability. Chapter 4 inquires if union wage setting is the reason for the lower inequality but higher unemployment in Europe relative to the United States. Finally, chapter 5 assesses if the fact that females outperform males in high school is explained by discrimination.

1.1. Insurance

With average unemployment rates exceeding 5 percent and 4 percent of the U.S. male population labeling themselves as severely and chronically disabled, unemployment and disability constitute important sources of income fluctuations. Individuals fearing risk would like to be insured against these shocks.

Chapter 2 and 3 analyses different aspects of how to insure unemployment and disability optimally. This question is interesting if at least three plausible conditions are fulfilled. First, agents have to be risk averse so that they demand insurance against unemployment and disability shocks. Second, there must be some financial markets incompleteness so that people lack access to complete contingent claims markets where they could perfectly insure each other before any uncertainty is resolved. Third, there has to be private information in the sense that agents' outcomes are at least partly determined by actions that are not commonly observed. If this third condition is not fulfilled, finding the optimal insurance system is trivial, and consists of a constant consumption sequence over time, regardless of the prevalence of unemployment and disability in an agent's history.

Chapter 2 develops a model in which these three conditions are fulfilled. In this model economy, agents are subject to the risk of becoming unemployed or disabled and wants to be insured against these shocks. However, the insurance agency, depicted as an abstract benevolent dictator, or social planner if you will, in unable to infer if an agent remains unemployed because of bad luck or because he does not exert the proper amount of effort to find a new job. Moreover, the planner cannot tell if a disability insurance claimant is in fact disabled, or merely claims to be to collect insurance benefits.

One might question if private information leads to moral hazard in this context, that is if people stay to long in unemployment or misreport disability if the insurance systems are too generous. The empirical support for moral hazard in unemployment and disability insurance is massive however. For example, Katz and Meyer (1990) show that unemployment insurance leads to longer unemployment spells. Moreover, Gruber (2000) reports that increasing the generosity of the disability insurance system leads to a increase of disability claimants from the labor force. There is also strong support for the notion that unemployment and disability insurance affect each other and therefore should be studied within the same model. For example, Autor and

Duggan (2003) attribute as much as one half percentage point of the U.S. mid 1980s decline in unemployment to the contemporaneous rise in disability benefits.

In an environment with moral hazard, providing the optimal unemployment and disability insurance is not an easy task. If optimal is to have its natural meaning, the planner has to be able to condition all parts of the insurance contract on the agents whole history of reported states, which is the only information he has access to. As the set of possible histories grows rapidly over time, problems of this type are difficult to tackle. Fortunately, Spear and Srivastava (1987), Abreu, Pearce and Stacchetti (1990) and Thomas and Worall (1990) have developed methods to circumvent this 'curse of dimensionality.' They showed that the relevant part of an agents history can be fully summarized by a 'promised future utility,' which serves as a state variable in the planner's problem. The problem can then be cast as a principal agent problem and solved by standard dynamic programming.

The goal of chapter 2 is to theoretically characterize the optimal insurance contract in terms of consumption for different histories. It turns out that benefits conditional on disability are optimally decreasing during unemployment and increased following successful job search. For low levels of effort costs of working, a no-quitting constraint becomes slack, and it is optimal to award an agent constant consumption during employment and an initial unemployment insurance replacement rate exceeding 100 percent. For high effort costs, consumption should instead increase during employment. Across all levels of the effort cost of working, consumption should decrease during unemployment, be constant during disability, and drop upon disability.

All variations in consumption across histories are due to private information. If the planner could observe the agents' states and actions, consumption would be the same for every possible history. The planner needs to vary consumption because he has to induce job search effort and abstain suboptimal quits and premature disability claims. Consumption being decreasing during an unemployment spell then becomes the optimal way to provide incentives to exert job search effort, just as in the seminal work by Shavell and Weiss (1979) and the more recent papers by Hopenhayn and Nicolini (1997, 2004). As agents are risk averse, it is cheaper for the planner to not only reduce consumption in future unemployment if the agent's unemployment spell is prolonged, but in all future states. This is the reason why future disability benefits are decreasing during unemployment. The same argument explains why future disability benefits rise when finding a new job. Upon job finding, the agent should be rewarded and risk aversion makes it cheaper to award him in all future states.

The most interesting feature of the optimal contract is thus the nontrivial way in which unemployment affects the optimal disability insurance, something that naturally did not appear in Diamond and Mirrlees' (1978) seminal paper on optimal disability insurance.

Chapter 3 takes the model from chapter 2 and solves it numerically. The purpose of that exercise is threefold. First, to study the quantitative aspects of the joint system of insurance against unemployment and disability insurance and compare it to the existing literature. Second, to compute the welfare gains associated with moving from the current system to the optimal one. Third, to provide a measure of the effectiveness of self insurance compared to the planner's insurance capabilities.

The optimal contract features a lot of consumption smoothing despite the moral hazard problems. Unemployment insurance benefits are high and decline at a moderate rate during an unemployment spell, disability benefits are higher than in the current system, and the optimal distribution of consumption has a much smaller standard deviation than the distribution implied by the current system. The cost reductions of moving to the optimal system are large, amounting to 15 percent of the costs of the current system. Despite the large costs reductions from moving to the optimal system, self insurance is rather effective compared to the optimal insurance system. An agent is willing to forgo no more than 6 weekly wages in order to join the planners optimal system.

1.2. Unions

During the last decades, Europe and the United States have taken different routes in important labor market dimensions. While U.S. wage inequality has increased substantially, unemployment rates have remained roughly constant. The increase in European inequality is much less pronounced, and unemployment rates have instead increased to levels far above those of the United States.

Chapter 3 sets out to explain these differences by union wage setting. Unions are more powerful in Europe than in the U.S., and a common feature of models with union wage setting is unemployment, even in otherwise frictionless environments. Chapter 3 uses a calibrated version of a dynamic monopoly union model to argue that a monopoly union increases the rate of unemployment as a response to increased microeconomic turbulence faced by its members. As unemployment reduces inequality, at least temporarily, the model is broadly consistent with the stylized facts described above. The union prefers higher unemployment in turbulent times both because such times implies increased opportunities for rent seeking and because unemployment leads to wage compression, which insures the workers from the increased wage volatility associated with turbulence.

A natural question becomes if the union equilibrium ('Europe') is preferred to the laissez-faire equilibrium without a union ('the U.S.'). It turns out that the welfare consequences of union wage setting is positive for workers and negative for firm owners. In total, union wage setting reduces welfare substantially in turbulent times.

1.3. Discrimination

It is well known that females have lower incomes than males. Despite being low wage earners compared to males, females outperform males before entering the labor market, that is, in school.

In chapter 5, Björn Tyrefors and I ask if females outperform males because of discrimination. To do this, we compare students' tests in Swedish under two different grading regimes. Specifically, we compare the grading outcomes generated under 'normal' circumstances, where the teachers know the identity of the student, to the grading outcomes where they do not. The difference in grading between the two regimes is then our measure of discrimination. As we grade the very same tests twice, we are confident that the differences between the grading regimes cannot be due to differences in ability between the genders.

Our analysis does not find strong support for the case that males are discriminated in school. Females outperform males in our sample, and less so when graders are unaware of the students' gender. The difference is however not large enough to pass our statistical tests on any reasonable level of significance.

References

Abreu, Dilip, David Pearce and Ennio Stacchetti. (1990), "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica* **58**, pp. 1041-1063.

Autor, David H. and Mark G. Duggan. (2003), "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics* **118**, pp. 157-202.

Diamond, Peter A. and James A. Mirrlees. (1978), "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics* **10**, pp. 295-336.

Gottschalk, Peter and Robert Moffit. (1994). "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity* **2**, pp. 217-272.

Gruber, Jonathan. (2000), "Disability Benefits and Labor Supply," *Journal of Political Economy* **108**, pp. 1162-1183.

Hopenhayn, Hugo A. and Juan Pablo Nicolini. (1997),"Optimal Unemployment Insurance," *Journal of Political Economy* **105**, pp. 413-438.

Hopenhayn, Hugo A. and Juan Pablo Nicolini. (2004), "Optimal Unemployment Insurance and Employment History," *mimeo.* Universidad Torquato Di Tella.

Juhn, Chinhui., Kevin M. Murphy and Brooks Pierce. (1993). "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* **101**, pp. 410-442.

Katz, Lawrence and David H. Autor. (1999). "Changes in the Wage Structure and Earnings Inequality," ch. 26 of *Handbook of Labor Economics* vol. **3**, ed. Orley Ashenfelter and David Card. Elsevier Science.

Katz, Lawrence and Bruce D. Meyer. (1990). "The impact of the potential duration of unemployment benefits on the duration of unemployment," *Journal of Public Economics* **41**, pp. 45-72.

Shavell, Steven. and Laurence Weiss. (1979), "The Optimal Payment of Unemployment Insurance Benefits Over Time," *Journal of Political Economy* **87**, pp. 1347-1362.

Spear, Stephen E. and Sanjay Srivastava. (1987), "On repeated Moral Hazard with Discounting," *Review of Economic Studies* **54**, pp. 599-617.

Thomas, Jonathan and Tim Worrall. (1990), "Income Fluctuations and Asymmetric Information: An Example of a Repeated Principal Agent Problem," *Journal of Economic Theory* **51**, pp. 367-390.

Insurance

CHAPTER 2

Optimal Unemployment and Disability Insurance

Erik Höglin

Abstract. This paper considers the optimal joint insurance system against unemployment and disability shocks. Benefits conditional on disability are optimally decreasing during unemployment and increased following successful job search. For low levels of effort costs of working, a no-quitting constraint becomes slack, and it is optimal to award an agent constant consumption during employment and an initial unemployment insurance replacement rate exceeding 100 percent. For high effort costs, consumption should instead increase during employment. Across all levels of the effort cost of working, consumption should decrease during unemployment, be constant during disability, and drop upon disability.

2.1. Introduction

Unemployment and disability are important sources of income fluctuations, and individuals facing such risks demand insurance to smooth consumption. Moral hazard problems stemming from private information however limit the extent to which insurance can be achieved. When an insurance agency cannot infer if an unemployed searches for a new job or whether a disability insurance claimant is in fact disabled, the optimal system of insurance against unemployment and disability risk has to acknowledge private information in order not to provide perverse incentives. With private information, unemployment and disability insurance are naturally related. For example, if disability insurance benefits are generous, unemployed might stop looking for jobs and opt for disability insurance instead. If they are not generous enough, disabled

[0] I am grateful to Lars Ljungqvist for providing continuous support, comments, and suggestions during this project. I have also benefited from comments by Milo Bianchi, Kaiji Chen, Martin Flodén, Nils Gottfries, Bertil Holmlund, Henrik Lundvall, Jörgen Weibull, and seminar participants at the Stockholm School of Economics, Uppsala University, the 1st Nordic Symposium in Macroeconomics in Smögen (Sweden), and the 2008 ENTER Jamboree in Madrid. Financial support from the Jan Wallander and Tom Hedelius foundation is gratefully acknowledged.

people might instead claim to be unemployed and be provided a consumption stream designed to provide incentives for job search, which is clearly not called upon for a truly disabled.

There is indeed ample empirical evidence that unemployment and disability insurance affect each other. Gruber (2000) studies a policy change in Canada giving rise to arguably exogenous changes in disability benefits. He finds a significant reduction in labor force participation due to increased benefits, with an implied elasticity of around .3. Gruber's paper is an argument in its own right for studying unemployment and disability insurance jointly, as one would expect some of those falsely claiming disability to otherwise be unemployed. More generous unemployment insurance would then reduce the disability claims, and vice versa. A direct test of this conjecture is performed by Autor and Duggan (2003). They study the interaction between aggregate unemployment and the generosity of the disability insurance system explicitly, and attribute as much as one half percentage point of the U.S. mid 1980's decline in unemployment to the contemporaneous rise in disability benefits.

Despite these empirical findings, previous literature on optimal social insurance under asymmetric information has mainly dealt with unemployment and disability insurance separately. Shavell and Weiss (1979) and Hopenhayn and Nicolini (1997, 2004) used search effort models to highlight the consequences of unobservable search effort for the optimal unemployment insurance system. Diamond and Mirrlees (1978, 1986), and, more recently, Golosov and Tsyvinski (2006), focused on disability insurance by assuming that it is impossible for a planner to distinguish disabled from able. I am aware of no previous paper studying optimal unemployment and disability insurance within the same model. The purpose of this paper is to do exactly that.

In order to study optimal unemployment and disability insurance within the same model, I merge Hopenhayn and Nicolini's (2004) model with a simplified version of Diamond and Mirrlees (1978), making them both special cases of my model. The agent goes through life experiencing multiple spells of employment and unemployment and eventually turn disabled. When an agent has reached disability, he can never leave that state. During unemployment, the agent has to exert search effort in order to have the chance to become employed. The planner does not know which state the agent is in or whether he exerts job search effort or not when unemployed. These are the sources of moral hazard in the model.

Following the methodological advances by Spear and Srivastava (1987), Thomas and Worall (1990), and Abreu, Pearce and Stacchetti (1990), I model the problem as a dynamic principal agent problem under asymmetric information. A social planner constructs a history dependent contract that induces the agent to reveal his true state and exert job search effort if unemployed. The optimal contract minimizes the cost of providing a given lifetime utility to the agent.

The presence of unemployment affects the optimal provision of disability benefits substantially. Specifically, benefits conditional on disability are reduced every period the agent is unemployed and increased when he finds a new job. The reason comes directly from the fact that the main result Hopenhayn and Nicolini (1997, 2004) is generalized to the environment of this paper. They showed that not only should consumption fall during unemployment to provide incentives to avoid prolonged unemployment, as in Shavell and Weiss (1979), but promised values in all future states should also be reduced during unemployment. Thus, in this paper, the longer an agent is unemployed, the lower is his promised utility should he turn disabled. Promises in all future states are reduced during unemployment because the agents are risk averse, which makes it cheaper for the planner to alter their lifetime utility in the whole future, rather than to temporarily change their consumption. By the same argument, disability benefits rise when the agent finds a new job. The planner has to award successful job search, and does so in all future states. In Diamond and Mirrlees' (1978) model that does not include unemployment, an agent is awarded higher disability benefits the later he becomes disabled. In their paper, rising disability benefits is an award for tenure. If the agent is able one more period, he is awarded in future employment as well as in future disability. When you introduce unemployment into the analysis, it is not clear if a history including many periods as unemployed reduces disability benefits or not. If the unemployment spells are short and many, benefits when turning disabled are higher compared to if they are few and long.

It also turns out that important aspects of the optimal contract hinges crucially on the level of effort costs of working. For low levels of the effort cost of working, a no-quitting constraint becomes slack. With a slack no-quitting constraint, it is optimal to award an agent constant consumption and unaltered promised values during employment, and an initial unemployment insurance replacement rate exceeding 100 percent. For effort cost levels high enough to make the no-quitting constraint binding, consumption should instead increase during employment, and the replacement rate in the unemployment insurance could be either above or below 100 percent. These features of the optimal contract are the same as those in Hopenhayn and Nicolini (2004), with the exception of the above 100 percent replacement rate. The reason why it is optimal to replace more that 100 percent of previous consumption when turning unemployed indeed stems from the introduction of disability. Specifically, it is a part of an award for truth telling. The planner optimally makes an employed agent indifferent between turning unemployed and disabled. When turning unemployed, the planner has to compensate him for the search effort and the declining consumption profile that an unemployed, but not a disabled, is facing. The planner accomplishes this by a high initial replacement rate and falling consumption throughout the unemployment spell.

The next two sections outline the economic environment and the allocations under autarky, where the agents fend for themselves without a planner and without the possibility of trading with each other. Section 2.4 then presents the planner's problem. In section 2.5, the optimal contract under the baseline assumption of no effort during employment is characterized. Section 2.6 extends the analysis to the case where effort during employment is large enough to alter the results from the case with low effort costs of working. Finally, section 2.7 concludes. Proofs of all propositions are in an appendix.

2.2. The Economy

The economy is populated by many risk averse agents. At each point in time an agent can be employed, unemployed or disabled. If employed, the agent earns a constant wage $w > 0$, while if unemployed or disabled he earns no income. Disability is an absorbing state. In contrast, employed and unemployed can transition between employment and unemployment as well as into disability. Specifically, an unemployed becomes disabled with per period probability ξ and can gain employment by exerting search effort. He can either choose $a > 0$ or zero search effort. If search effort is a, the probability of gaining employment is p, conditional on not becoming disabled. If search effort is zero, an unemployed remain so with probability one, again conditional on not turning disabled. Employed agents face the same probability of becoming disabled as unemployed and maintain employment with exogenous probability λ, if not becoming disabled. Quits are possible. Effort is assumed to be costly in utility terms. Following Hopenhayn and Nicolini (1997), the preferences of an agent are ordered by

$$E\sum_{t=0}^{\infty}\beta^{t}\left[u\left(c_{t}\right)-a_{t}\right], \tag{2.1}$$

where c_t is consumption, $u(\cdot)$ is strictly increasing, strictly concave, twice continuously differentiable and $\beta \in (0,1)$ is the common discount factor.[1] Assume also that $u(0)$ is well defined. No effort is required to maintain employment, so a_t is zero during employment and disability. All the results remain with an effort cost of working, provided it is sufficiently small. In section 2.6, I consider the case where the cost of working is large enough to alter the qualitative results from the case with no effort costs of working.

The agents can neither save nor borrow. Lack of access to capital markets in combination with risk aversion will constitute scope for improvement upon the autarky equilibrium by a social planner that has the ability to transfer resources over time and across agents.[2] I assume that such a planner has unlimited access to a perfect capital

[1] The discount factor may include a state independent per period probability of dying. The results are invariant to such an additional assumption.

[2] Lack of access to any market for borrowing or lending is unrestrictive as long as an agents' savings can be costlessy monitored by the planner, or, equivalently, as long as the planner can control the agent's consumption. Extending the analysis to include hidden savings is *not* straightforward. Kocherlakota (2004) describes the technical difficulties in using the recursive constract methodology when a hidden state variable such as savings can take on a continuum of values. Motivated by these difficulties, I abstract from hidden savings. Werning (2002) and Mitchell and Zhang (2007) studies the role of hidden savings in optimal unemployment insurance under some functional forms assumptions.

market, where he can borrow and lend at a constant gross interest rate equal to the reciprocal of the agents' discount factor.

The current state of an agent is unobservable but income is observed by the planner. Hence, an employed cannot claim to be unemployed or disabled. An employed can quit however, rendering him in a situation identical to one that got laid off. The planner cannot distinguish quits from layoffs. Effort while unemployed is also unobservable. The optimal contract must induce an employed never to quit, unemployed and disabled not to lie, and unemployed to exert effort to become employed.

2.3. Autarky

To highlight the agents' problem we start the analysis without a planner. In autarky, the agents then lack the opportunity to trade in any way and have to fend for themselves. The only choice the agents have is whether they should exert search effort of not if unemployed, as no other transition probabilities are endogenous. Accordingly, the autarky values for employed and disabled are simply given by:

$$V_{aut}^{e} = u(w) + \beta \left\{ (1-\xi) \left[\lambda V_{aut}^{e} + (1-\lambda) V_{aut}^{u} \right] + \xi V_{aut}^{d} \right\},$$

and

$$V_{aut}^{d} = \frac{u(0)}{1-\beta}.$$

The autarky value for an unemployed is the solution to the Bellman equation:

$$V_{aut}^{u} = u(0) + \max \left\{ -a + \beta \left\{ (1-\xi) \left[pV_{aut}^{e} + (1-p) V_{aut}^{u} \right] \right\}, \beta \left[(1-\xi) V_{aut}^{u} \right] \right\} + \beta \xi V_{aut}^{d}.$$

Evidently, an unemployed exerts effort if and only if

$$\beta (1-\xi) p \left(V_{aut}^{e} - V_{aut}^{u} \right) \geq a. \tag{2.2}$$

From now on we assume that parameters are such that this condition is fulfilled. Note that then $(V_{aut}^{e} - V_{aut}^{u}) > 0$, so an employed agent will never quit. When the planner constructs the optimal contract in the proceeding sections he will, in addition to other incentive constraints, have to see to it that a condition like (2.2) is fulfilled.

2.4. The Planner's Problem

The planner's problem is to specify history dependent transfers and effort recommendations to minimize the cost of giving the agent a specified value V. Relying on the revelation principle, I restrict attention to contracts in which the planner would like the agent to tell the truth. Furthermore, I assume that the primitives of the model, including the initial promised value, are such that the planner would like to induce positive search effort. The planner collects reports on state in each period and makes history dependent transfers in order to induce no quitting, truth telling, and effort if the agent is unemployed. The timing of the contract within each period is:

(1) The agent learns his state and reports a state to the planner.
(2) Conditional on the history of reports, including the current period, the planner sets up a consumption level and promises for each possible future state.
(3) If the agent is unemployed, he chooses to exert search effort or not, which determines his probability distribution over next period's state.

If the promised value for a given state is V, the planner chooses current consumption c, and promises V^e, V^u, and V^d for each possible future state. Here, V^e is the expected utility from next period on, should the agent be employed then, and V^u and V^d are defined analogously. In order to deliver the promise, the planner has to set c, V^e, V^u, and V^d such that an agent's expected utility equals V. In addition, the contract must involve combinations of V^e, V^u, and V^d that makes it optimal for an agent not to quit, not to lie, and to search if unemployed. The value of the planner's problem when the agent has reported to be employed is $C_e(V)$, and denotes the minimal cost of delivering V to a currently employed agent with a promised value equal to V. Corresponding cost functions when an agent is unemployed or disabled, $C_u(\cdot)$ and $C_d(\cdot)$, are defined analogously.

2.4.1. Employed.

Suppose an agent is employed and has promised value equal to V. The optimal contract then solves the Bellman equation:

$$C_e(V) = \min_{c,V^e,V^u,V^d} \left\{c - w + \beta \left\{(1-\xi)\left[\lambda C_e(V^e) + (1-\lambda) C_u(V^u)\right] + \xi C_d\left(V^d\right)\right\}\right\},$$

where the minimization is subject to

$$V = u(c) + \beta \left\{ (1-\xi) \left[\lambda V^e + (1-\lambda) V^u \right] + \xi V^d \right\}, \tag{2.3}$$

$$V^e \geq V^u, \tag{2.4}$$

$$V^u \geq V^d. \tag{2.5}$$

Here, (2.3) is the promise keeping constraint, and (2.4) is the no quitting constraint that makes sure that an agent does not quit if still employed in the beginning of next period. Also, (2.5) is the truth telling constraint that assures that if unemployed next period, the agent will report so. But what if the agent becomes disabled, how do we make sure that he does not claim to be unemployed and collect (possibly) higher benefits? This can be a serious problem, but it turns out not to be. Later we will see that along the optimal path, it is never optimal to claim unemployment if disabled.

The cost function is strictly increasing and strictly convex. As promises increase, it becomes more and more expensive for the planner to deliver an additional util to the agent, a feature due to the concave utility function.[3] Attaching nonnegative Lagrange multipliers $\mu, \beta(1-\xi)\theta_e$ and $\beta(1-\xi)\theta_u$ on (2.3), (2.4) and (2.5), respectively, the first order conditions are:

$$\frac{1}{u'(c)} = \mu, \tag{2.6}$$

$$\lambda \left[C_e'(V^e) - \mu \right] - \theta_e = 0, \tag{2.7}$$

$$(1-\lambda) \left[C_u'(V^u) - \mu \right] - \theta_u + \theta_e = 0, \tag{2.8}$$

$$\xi \left[C_d'\left(V^d\right) - \mu \right] + \theta_u (1-\xi) = 0. \tag{2.9}$$

By the envelope theorem,

$$C_e'(V) = \mu. \tag{2.10}$$

[3] As a more formal argument, note that we could rewrite the problem as choosing a utility level u instead of c. The return function is then the strictly increasing, strictly convex function u^{-1} and the constraints are linear in u, V^e, V^u, and V^d. This makes the cost function become strictly increasing and strictly convex. The first order conditions with the alternative specification are identical to those presented here. These conditions are hence both necessary and sufficient.

2.4.2. Unemployed. If an agent has reported unemployment and his promised value in this state is V, the optimal contract is characterized by the solution to the dynamic program:

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

where the minimization is subject to

$$u(c) - a + \beta\left\{(1-\xi)\left[pV^e + (1-p)V^u\right] + \xi V^d\right\} = V, \tag{2.11}$$

$$\beta(1-\xi)p\left[V^e - V^u\right] \geq a, \tag{2.12}$$

$$V^u \geq V^d. \tag{2.13}$$

Here, (2.11) is the promise keeping constraint and (2.12) is the incentive compatibility constraint inducing effort. The incentive compatibility constraint makes sure that it is optimal for the unemployed to exert search effort. Finally, in order for an unemployed not to claim to be disabled, a truth telling constraint like (2.13) has to be satisfied. As we will show later, an unemployed turning disabled will not lie in this case either. The cost function is strictly increasing and strictly convex, for the same reason as in the employment problem. To proceed, attach nonnegative Lagrange multipliers μ, ϕ and $\beta(1-\xi)\theta$ on (2.11), (2.12) and (2.13) respectively. The first order necessary conditions are:

$$\mu = \frac{1}{u'(c)}, \tag{2.14}$$

$$\left[C_e'(V^e) - \mu\right]p - \phi p = 0, \tag{2.15}$$

$$\left[C_u'(V^u) - \mu\right](1-p) + \phi p - \theta = 0, \tag{2.16}$$

$$\left[C_d'\left(V^d\right) - \mu\right]\xi + \theta(1-\xi) = 0. \tag{2.17}$$

By the envelope theorem,

$$C_u'(V) = \mu. \tag{2.18}$$

2.4.3. Disabled. Disability is an absorbing state. Hence, the incentive problems between the planner and the agent are bygones once an agent has truthfully reported to be disabled. It is then optimal for the planner to award the agent a constant level of consumption that delivers the promised value. Let this transfer be c^d, which is implicitly defined by

$$V = \frac{u\left(c^d\right)}{1-\beta}.$$

The cost to the planner is then

$$C_d\left(V\right) = \frac{c^d}{1-\beta} = \frac{u^{-1}\left(V\left(1-\beta\right)\right)}{1-\beta}.$$

Note that

$$C_d'\left(V\right) = \frac{1}{u'\left(c^d\right)} \tag{2.19}$$

is positive and increasing, making $C_d\left(V\right)$ convex.

2.5. Optimal Insurance

Absent the informational asymmetries, the optimal insurance contract would award an agent the same consumption in every state and period. To see this, note that without private information the first order conditions are the same but without the multipliers on the no-quitting, truth telling, and effort incentive constraints. Then, the first order conditions imply

$$C_e'\left(V\right) = C_e'\left(V^e\right) = C_u'\left(V^u\right) = C_d'\left(V^d\right), \tag{2.20}$$

in the employment problem, and

$$C_u'\left(V\right) = C_e'\left(V^e\right) = C_u'\left(V^u\right) = C_d'\left(V^d\right), \tag{2.21}$$

in the unemployment problem. The envelope conditions in the employment and unemployment problems along with (2.19), (2.20) and (2.21) reveal that consumption is fully equalized across states and periods. The planner would simply instruct an unemployed to search, and with effort being observable the planner could punish him such that he does adhere to the instruction. Although consumption is the same, the values for different types are not. A disabled will never have to exert search effort and thus enjoys the highest level of utility. An employed faces some risk of having to search if unemployed in the future, but less so than an unemployed. Hence, an employed has higher utility than an unemployed, who is worst off in the full information contract.

Note also that in the full information contract, the cost of delivering an additional util is the same across states. With private information, this will no longer be true. Private information will induce the planner to vary consumption across states to maintain incentive compatibility, and this feature of the optimal information constrained contract implies that the costs of delivering utils differs for different types.

2.5.1. Distribution of values. Under asymmetric information, the state independent allocations from the perfect information contract are not incentive compatible. Regardless of his true state, an agent could always claim disability to enjoy the highest possible utility, or, equivalently, stop searching if unemployed. A natural first step in the inquiry of how the optimal information constrained contract differs from the unconstrained optimum is therefore to find out which information constraints are important. It turns out that the structure of binding incentive constraints is simple and intuitive. An agent is indifferent between becoming or remaining unemployed or becoming disabled and he always strictly prefers to be employed. Also, while unemployed, the agent is indifferent between exerting search effort or not.

PROPOSITION 1. *The truth telling constraints in both the employment and the unemployment problem and the effort incentive constraint in the unemployment problem are binding. The no-quitting constraint is not binding.*

All proofs are in the appendix. It might be useful to compare the results in proposition 1 to the full information benchmark to appreciate the limitations private information puts on the optimal contract. In the full information contract, the marginal costs of delivering utils to different types are equal. With private information, this is no longer true. In particular, the marginal cost of delivering utils to disabled agents is lower than the marginal cost of giving other types utils. To see this, consider the employment problem and use (2.7) and (2.9) to get

$$C_e'\left(V^e\right) - C_d'\left(V^d\right) = \frac{\theta_e}{\lambda} + \frac{\theta_u\left(1-\xi\right)}{\xi},$$

which implies

$$C_d'\left(V^d\right) \leq C_e'\left(V^e\right),$$

with strict inequality if at least one of the constraints bind. The intuition is clear. Once and agent has truthfully reported disabled, the incentive problems between him and the planner are bygones. For this reason, it is relatively cheap to deliver another util to such an agent, since variations over time is not needed to maintain incentive compatibility. Thus, the planner wants to give as much utility as possible to a disabled, given that it is not tempting to falsely claim disability. The way to accomplish this is to make the truth telling constraint binding, so that whether currently employed or unemployed, future promises between unemployment and disability are equalized, $V^u = V^d$.

A similar reasoning can be applied to explain why the no-quitting constraint is not binding. The planner has more severe moral hazard issues with an unemployed than with an employed. First, the search effort has to be compensated for so that the unemployed does not go into disability, and, second, incentives to exert search effort has to be provided. The more severe moral hazard problems makes it expensive for the planner to deliver an additional util to the unemployed, and hence relatively cheaper to deliver an additional util to an employed. This reasoning explains why the no quitting constraint is slack, $V^e > V^u$.

With a binding truth telling constraint, it is clear why a disabled would never claim to be unemployed. A truly unemployed receives exactly the same utility as a disabled claiming disability. The former's utility is based on a positive probability of finding a job. A disabled falsely claiming to be unemployed does not have this chance, and is hence worse of by claiming to be unemployed.

Note also the difference between the distribution of promised values in this paper and in Diamond and Mirrlees (1978). With only employed and disabled, as in their model, the values across employment and disability should be equalized. The reason in their paper is that disabled have a higher marginal utility of consumption. In the model of this paper it is not optimal to push the future values of employment and disability to equality, because it will by necessity push the value of unemployment into a region where the marginal cost of increasing the future value for unemployment is too high. This difference compared to Diamond and Mirrlees' model will have further implications for the optimal contract.

2.5.2. Consumption dynamics. Let us now study the consumption dynamics within a given spell. The assertion by Shavell and Weiss (1979) and Hopenhayn and Nicolini (1997, 2004) that consumption should fall during an unemployment spell remains valid also in this environment. Furthermore, consumption should not change over time during an employment spell or when disabled.

PROPOSITION 2. *Consumption is constant during an entire employment spell, decreasing during an unemployment spell, and constant during disability.*

Consumption being constant during employment is a direct consequence of the no-quitting constraint being slack. To see this, use (2.6) and (2.7) in combination with the envelope condition (2.10) to get

$$\frac{1}{u'(c^e)} - \frac{1}{u'(c)} = \frac{\theta_e}{\lambda}, \tag{2.22}$$

where c is current consumption as employed and c^e denotes consumption in the next period, should the agent still be employed. As the no-quitting constraint is slack, $\theta_e = 0$, (2.22) implies that

$$\frac{1}{u'(c^e)} = \frac{1}{u'(c)}. \tag{2.23}$$

Recursive application of (2.23) shows that consumption does not change during an employment spell, as the utility function is strictly concave. Given that the employed is not tempted to quit, there is no reason to vary consumption during a spell of employment. On the contrary, the cheapest way of delivering a given value when there are no incentive problems is a constant sequence of consumption, a feature due to the concave utility function. Constant consumption during employment is also a feature shared by Hopenhayn and Nicolini (2004), for their specification with low enough effort cost while employed.

Declining consumption during unemployment reflects the planner's need to induce search from the unemployed and his need to abstain the unemployed from reporting disabled. The first order conditions (2.14), (2.16) and (2.18) yields

$$\frac{1}{u'(c^u)} - \frac{1}{u'(c)} = \frac{\theta - \phi p}{1-p}, \tag{2.24}$$

where c is current consumption as unemployed and c^u denotes next period's consumption, should the agent remain unemployed. Whether consumption is increasing or decreasing depends on the relative size of the Largrange multipliers θ and ϕ. The higher the marginal value of relaxing the truth telling constraint (a higher θ), the more consumption tends to grow during unemployment. This illustrates the planner's need to prevent the unemployed from falsely claiming to be disabled. On the other hand, the higher the marginal value of relaxing the effort incentive constraint (a higher ϕ), the more consumption declines during unemployment. As shown in the appendix, the latter effect is stronger than the former, so consumption does decline during unemployment, $\theta - \phi p < 0$. The intuition is the same as in Shavell and Weiss (1979) and Hopenhayn and Nicolini (1997, 2004). An unemployed must be provided incentives to avoid prolonged spells of unemployment, and the planner provides those incentives by gradually lowering consumption during the whole spell.

All incentive problems between the planner and the agent are bygones once an agent has reported to be disabled. Again, concavity of the utility function prescribes constant consumption as the cost minimizing way of delivering a disabled's promise.

Within a spell there are no qualitative differences in results or intuition between this model and Hopenhayn an Nicolini (1997, 2004), but a clear difference compared to Diamond and Mirrlees (1978). In Diamond and Mirrlees' analysis, the worker consumes more and more during a spell of employment. The reason in their paper is that an agent is optimally indifferent between claiming disability or not, so rising consumption becomes an award for tenure. Such an award is not needed here, because an employed is always strictly better off than an unemployed or a disabled.

We now turn to what happens to consumption at the time of transition between states. Consumption has to jump at all transitions, partly because different states means different values and partly because of the way the planner delivers these values according to the preceding proposition.

PROPOSITION 3. *Consumption rise upon layoff and upon employment, and fall upon disability.*

The quite intuitive conclusion that consumption increases upon gaining employment follows naturally from the planner's need to deliver the high promise upon employment. Utility increases when going into employment both because of higher consumption and because effort is zero during employment. However, not having to exert search effort is not sufficient. If consumption would not increase when finding a job, the value of not searching for a job would be higher than that of searching, which is clearly not incentive compatible.

The rise in consumption upon layoff is harder to explain. It is certainly not a reward for layoff, nor is it a punishment. There is absolutely no need for the planner to punish layoffs since they are exogenous. The answer should instead be sought in the relation between unemployment and disability. When an employed agent gets laid off, the planner uses (2.16) and the envelope condition in the unemployment problem to update his consumption. Combining these two with the envelope condition in the employment problem yields,

$$\frac{1}{u'(c^u)} - \frac{1}{u'(c)} = \frac{\theta_u - \theta_e}{1-\lambda}, \tag{2.25}$$

where c is consumption prior to the layoff and c^u is consumption in the first period of the unemployment spell. Now as the truth telling constraint is binding while the no-quitting constraint is slack, (2.25) can be written as

$$\frac{1}{u'(c^u)} - \frac{1}{u'(c)} = \frac{\theta_u}{1-\lambda} > 0 \Longrightarrow c^u > c, \tag{2.26}$$

so the binding truth telling constraint pushes the replacement rate above 100 percent of previous consumption. Since the truth telling constraint while employed is binding, the agent is as well off getting unemployed as getting disabled. When unemployed we have already seen that consumption should fall, while it remains constant for a disabled. The high replacement rate is then a consumption award for truth telling, compensating for the job search effort cost. The same expression as (2.26) appears in Hopenhayn and Nicolini (2004), but without the multiplier on the truth telling constraint as there is no disability state in their model. As a consequence, their model has an initial replacement rate of exactly 100 percent. The reason is that since layoffs are exogenous, the agent is completely insured against this shock. Here, layoffs are exogenous too, but upon layoff, there is a threat that the laid off claims to be disabled.

The reason that consumption falls upon disability is easy to see if we keep in mind that consumption is constant during employment. If consumption did not fall upon disability, a disabled would be as well of as an employed with zero unemployment risk. This is clearly not incentive compatible, as everyone would then state disability.

The drop in consumption upon disability also appears in Diamond and Mirrlees (1978). The intuition is also similar. The drop in consumption is a tool for avoiding that an agent goes into disability too early. In Diamond and Mirrlees (1978), this is achieved by delivering the same value to a disabled as to an employed. In this paper, premature claims of disability is avoided by delivering the same value to a disabled as to an unemployed.

2.5.3. Dynamics of promised values. It is also interesting to study what happens to agents with different histories. The evolution of promised values is informative in this respect. Naturally, promised values move in the same direction as does consumption.

PROPOSITION 4. *During employment, future promises are constant with tenure. During unemployment, all promises fall. Promises in future unemployment and disability rise when finding a job.*

The longer an agent is unemployed, the lower is his utility in the future, not just in unemployment but also in future spells of employment, and when he eventually becomes disabled as well. The intuition is that since agents are risk averse, it is cheaper to maintain incentive compatibility by permanent alterations of values in all future states.

As promises are unaltered during employment, the contract exhibits limited memory. It simply does not matter if an agent's employment spell was long or short, just the number of spells. For example, the planner awards exactly the same consumption and future promises to an agent that has been employed 100 periods, unemployed for 10 and then gains employment as to one who was employed for one period, unemployed for 10 and then gained employment.[4] Limited memory in this respect means that the common feature of real world social insurance systems that a number of periods of employment is needed to qualify for unemployment insurance does not find support in this model. Here, finding a job is sufficient.

Even though there is some limited memory in the optimal contract, the evolution of promised values has important implications for the difference across histories. The falling promises during unemployment together with the constant promises during employment suggests that inequality between lucky agents, with few and short unemployment spells, and unlucky agents, with the opposite experience, gets larger over time. There are also some non-trivial dynamics of the conditional value of disability benefits that did not appear in Diamond and Mirrlees (1978). For every period that the agent is unemployed, the value should he become disabled falls. If he finds a new job, the disability benefits increase however, so whether the prevalence of unemployment in an agents history increases or reduces his benefits if becoming disabled thus depends not only on the number of periods as unemployed, but also on the distribution of spells and durations. An agent with many short spells of unemployment clearly receives higher disability benefits than an agent that has experienced few but long spells. This stands in sharp contrast to Diamond and Mirrlees (1978), where benefits are higher the later an agent becomes disabled. Here, the unemployment/employment experience becomes important.

The dynamics of promised values also has implications for the long run evolution of consumption. In fact, an agent will typically consume less and less as he gets older. Falling expected consumption is reminiscent the immiseration property in Thomas and Worall (1990).[5] Behind these properties of the optimal contract lurk an inverse Euler equation, which provides intuition for consumption being front loaded (see Rogerson,

[4] Other than in these very simple cases it is difficult or impossible to compare exact histories analytically.

[5] Thomas and Worall (1990) study a planner providing optimal insurance to an agent hit by identically independently distributed income shocks and show that the agents promised value reaches an arbitrarily low level with probability one.

1985). To see this, consider an employed agent. Combine the first order conditions in the employment problem with the envelope conditions for all problems to get

$$\frac{1}{u'(c)} = (1-\xi)\lambda\frac{1}{u'(c^e)} + (1-\xi)(1-\lambda)\frac{1}{u'(c^u)} + \xi\frac{1}{u'(c^d)}. \tag{2.27}$$

The inverse Euler equation says that the inverse of marginal utility today should equal the expected value of the inverse of marginal utility tomorrow. A similar equation, with p replacing λ, holds for an unemployed. Applying Jensen's inequality to (2.27) yields,

$$u'(c) < (1-\xi)\lambda u'(c^e) + (1-\xi)(1-\lambda)u'(c^u) + \xi u'\left(c^d\right), \tag{2.28}$$

so consumption has a tendency to be front loaded.[6] The reason for this intertemporal wedge is that it is more costly to give consumption tomorrow as it will have to be varied across states to maintain incentive compatibility. In other words, the planner's marginal rate of transformation exceeds the agent's marginal rate of substitution.

2.6. High Effort Cost of Working

So far, I have assumed that the effort cost of employment is zero. In this section, I relax that assumption and provide some qualitative differences and similarities as compared to the preceding section.

Let e be the effort cost while employed. Algebraically, the planner's problem is only changed in one way, namely that the promise keeping constraint (2.3) in the employment problem is now replaced by

$$V = u(c) - e + \beta\left\{(1-\xi)\left[\lambda V^e + (1-\lambda)V^u\right] + \xi V^d\right\}.$$

It should be made clear however that although this change looks minimal it generally is not, as the cost functions are not the same as before. I keep the notation from before and trust the reader to recall that these functions are not the same as in the preceding section.

---

[6] Rogerson (1985) proves that for a large class of HARA (including Quadratic, CARA and CRRA) utility functions (2.28) implies that expected consumption is falling over time. As the utility function in this paper is not necessarily in the HARA class, Rogerson's proof does not go through generically. However the tendecy for front loading is clear from (2.28).

The first step is again to establish which incentive constrains are binding. When effort during employment is present, it is not as easy to establish unambiguous results as when it is not. For low levels of effort, the results from the preceding section holds, while for high levels some stark differences compared to the low effort case emerge. Specifically, the no-quitting constraint becomes binding for high enough levels of the effort cost. A binding no-quitting constraint implies different consumption dynamics and different dynamics of promised values as compared to the case with low effort. Introducing effort costs of employment does not affect the qualitative results in the states of unemployment and disability.

Let us first establish that the no-quitting constraint is slack for low levels of the effort cost of working.

PROPOSITION 5. *There exists an effort level* $e > 0$ *such that Proposition 1 holds.*

To develop an understanding for the agents' situation when employed, consider an agent who contemplates to quit. Suppose he finds out that he is employed at the beginning of period t. If he quits directly, his utility from now on is V_t^u. Alternatively, he can truthfully report to be employed this period, quit in the beginning of next period, and earn a utility from now on equal to $\widehat{V} = u(c_t^e) - e + \beta V_{t+1}^u$. Evidently, he postpones his quit if $\widehat{V} > V_t^u$. Whether the no-quitting constraint is binding or not, it turns out that the agent's future value of turning unemployed never decreases if he is employed another period. In other words, $V_{t+1}^u \geq V_t^u$. Using the fact that the truth telling constraint is always binding ($V^u = V^d$ for any history), this implies that a sufficient condition for $\widehat{V} > V_t^u$, is

$$u(c_t^e) - e + \beta V_t^u \geq V_t^u \Longleftrightarrow u(c_t^e) - e \geq V_t^d (1 - \beta) = u\left(c_t^d\right).$$

Now as consumption always drops upon disability, i.e. $u(c_t^e) > u\left(c_t^d\right)$, there always exists an $e > 0$ such that it is optimal to postpone the quit. Applying this principle recursively, there exists an $e > 0$ such that the agent never quits. If e is high however, the no-quitting constraint becomes binding.

Proposition 6. *For high enough effort cost of employment, the no-quitting constraint is binding. When this happens, the truth telling constraint is also binding.*

Recall the intuition for why the no-quitting constraint is slack when e is low. An inversion of that reasoning explains why this constraint binds for high levels of effort during employment. When the effort cost of working is high, the planner has to compensate the employed in order to abstain him from quitting. The compensation pushes consumption upwards more for higher e, and because decreasing marginal utility implies increasing marginal cost of delivering a util, there has to be a high enough e such that, while employed, it is no longer optimal to promise strictly more for employment than for unemployment.

Note that the binding no-quitting constraint implies that during employment, the agent is indifferent between all future states. This does not, however, imply that there is complete consumption smoothing across histories, as will become clear below. The reason for the incomplete consumption smoothing is of course the different, and, in the case of unemployment, unobserved, effort costs. For example, consider two employed agents with the same history, and suppose that one of them immediately becomes disabled and the other remains employed for a long time before becoming disabled. The first agent then receives lower lifetime consumption, but has to exert less effort.

2.6.1. Consumption dynamics. A natural consequence of a different structure of binding constraints is different consumption dynamics. Also unsurprisingly, the difference pertains to the state of employment, where the only difference arise. During employment, consumption has to rise when the no-quitting constraint is binding.

Proposition 7. *Consumption is increasing during employment, decreasing during unemployment, and constant during disability.*

Consumption rises during employment when the no-quitting constraint is binding because it is optimal to award the agent for staying employed. This is done by a rising profile of consumption during employment. In the other states, nothing is changed and the intuition from before remains. The falling consumption during unemployment thus seems like a very robust feature of this class of models.

The introduction of costly effort of employment has one implication for consumption at the transitions. It is not always optimal to award truth telling to the extent that the replacement rate exceeds 100 percent.

PROPOSITION 8. *Consumption rise upon employment and fall upon disability. Consumption could either rise or fall upon layoff.*

The intuition for why the replacement rate is not always above 100 percent if the no-quitting constraint is binding is pretty straightforward. When the no-quitting constraint is binding, it is tempting to quit, and an award for truth telling that implies a high replacement rate means that, in order to avoid quits, the award for tenure has to be large. The planner always want to smooth consumption subject to incentive compatibility, so a lower award for both tenure and truth telling becomes optimal.

2.6.2. Dynamics of promised values. One of the major results in the preceding section was that future disability benefits should be decreasing during unemployment. Moreover, when effort costs during employment was low, promised values remained constant during an entire employment spell, making the length of the employment spell irrelevant. When effort costs of employment is high, the first result is still valid whereas the second is not.

PROPOSITION 9. *Promised values in all future states are increasing during employment and decreasing during unemployment. Promises in future unemployment and disability rise when finding a job.*

As above, promised values fall during unemployment to provide incentives to avoid long lasting unemployment spells, i.e. to provide incentives to exert search effort. Disability benefits being decreasing during unemployment is thus a very robust feature of this environment.

As in the case with low effort costs of working, benefits conditional on disability rise upon job finding. In addition, as all promises increase during employment, benefits conditional on disability move more like in Diamond and Mirrlees (1978) with high than with low effort costs of working.

During employment, the mirror image of the reasoning used during unemployment could be used. When the no-quitting constraint is binding it is optimal to award tenure, and thus avoid suboptimal quits, by increasing the promises in all future states. The intuition is again that it is cheaper for the planner to alter the risk averse agents' values in all future states.

Note that the increasing promises during employment could be broadly interpreted as an argument for benefit entitlements being conditional on history involving a certain number of periods as employed. Indeed, the contract when the no-quitting constraint is binding has no loss of memory, the planner updates the promised values in a non-trivial way in every period until an agent reaches disability.

Finally, it is important to note that although promised values rise during employment, the inverse Euler equation (2.28) still holds. Hence, expected consumption is still falling for many utility functions and Thomas and Worall's (1990) immiseration result is still relevant.

2.7. Conclusions

There are good reasons to study how to design unemployment and disability insurance in the same model as they are important for each other empirically. When doing so a number of qualitative differences emerge, compared to models in which only one of them can be analyzed.

The dynamics of the conditional value of disability benefits did not appear in Diamond and Mirrlees (1978), as unemployment does not exist in their model. For every period that the agent is unemployed, the value should be become disabled falls. If he finds a new job, the disability benefits increase however. Whether the prevalence of unemployment in an agent's history increases or reduces his benefits if becoming disabled thus depends not only on the number of periods as unemployed, but also on the distribution of spells and durations. An agent with many short spells of unemployment clearly receives higher disability benefits than an agent that has experienced few but long spells. Moreover, when the effort costs of working is high, benefits conditional on disability rise during employment. These dynamics are more involved than in Diamond and Mirrlees (1978), where benefits are higher the later an agent becomes disabled. Here, the unemployment/employment experience becomes important.

The qualitative features of the optimal contract hinges crucially on the level of effort cost while employed. For low levels of the effort cost while working, a no-quitting constraint becomes slack. With a slack no-quitting constraint, it is optimal to award an agent constant consumption during employment and an initial unemployment insurance replacement rate exceeding 100 percent. For effort cost levels high enough to make the no-quitting constraint binding, consumption should instead increase during employment, and the replacement rate could be either above or below 100 percent. Across all levels of the effort cost during employment, consumption should decrease during unemployment, be constant during disability, and drop upon disability.

Natural extensions of this paper would include variable job search effort and moral hazard on the job, as in Wang and Williamson (1996) and Zhao (2001). When search effort is variable, the planner has an additional challenge: to provide the incentives to exert the right amount of effort without luring the unemployed to quit. With moral hazard on the job in the form of wages being conditional on the level of effort exerted, additional incentive constraints come into play.

Finally, numerical simulations of the model would be an interesting extension. With a numerical solution to the planner's problem it would be straightforward to simulate examples of complex histories, which would further reveal the properties of the optimal contract. Numerical simulations would also be appropriate to make quantitative comparisons with Hopenhayn and Nicolini (1997, 2004) and Diamond and Mirrlees (1978). In a companion paper, Höglin (2008), I perform such simulations and show examples where not only the qualitative differences shown in this paper are present, but also how large they are quantitatively.

References

Abreu, Dilip, David Pearce and Ennio Stacchetti. (1990), "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica* **58**, pp. 1041-1063.

Autor, David H. and Mark G. Duggan. (2003), "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics* **118**, pp. 157-202.

Diamond, Peter A. and James A. Mirrlees. (1978), "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics* **10**, pp. 295-336.

Diamond, Peter A. and James A. Mirrlees. (1986), "Payroll-Tax Financed Social Insurance with Variable Retirement," *Scandinavian Journal of Economics* **88**, pp. 25-50.

Golosov, Mikhail and Aleh Tsyvinski. (2006), "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy* **114**, pp. 257-269.

Gruber, Jonathan. (2000), "Disability Benefits and Labor Supply," *Journal of Political Economy* **108**, pp. 1162-1183.

Hopenhayn, Hugo A., and Juan Pablo Nicolini. (1997),"Optimal Unemployment Insurance," *Journal of Political Economy* **105**, pp. 413-438.

Hopenhayn, Hugo A., and Juan Pablo Nicolini. (2004), "Optimal Unemployment Insurance and Employment History," *mimeo.* Universidad Torquato Di Tella.

Höglin, Erik. (2008) "Optimal Unemployment and Disability Insurance: Allocations, Welfare Gains, and the Effectiveness of Self Insurance," chapter 3 of this volume.

Kocherlakota, Narayana R. (2004), "Figuring out the Impact of Hidden Savings on Optimal Unemployment Insurance," *Review of Economic Dynamics* **7**, pp. 541-554.

Mitchell, Matthew and Yuzhe Zhang (2007). "Unemployment Insurance with Hidden Savings," mimeo. University of Toronto.

Rogerson, William P. (1985), "Repeated Moral Hazard," *Econometrica* **53**, pp. 69-76.

Shavell, Steven. and Laurence Weiss. (1979), "The Optimal Payment of Unemployment Insurance Benefits Over Time," *Journal of Political Economy* **87**, pp. 1347-1362.

Spear, Stephen E. and Sanjay Srivastava. (1987), "On repeated Moral Hazard with Discounting," *Review of Economic Studies* **54**, pp. 599-617.

Thomas, Jonathan and Tim Worrall. (1990), "Income Fluctuations and Asymmetric Information: An Example of a Repeated Principal Agent Problem," *Journal of Economic Theory* **51**, pp. 367-390.

Wang, Cheng. and Stephen D. Williamson. (1996). "Unemployment Insurance with Moral Hazard in a Dynamic Economy." *Carnegie Rochester Conference Series on Public Policy* **44**, pp. 1-41.

Werning, Ivan. (2002), "Optimal Unemployment with Unobservable Savings," *mimeo.* University of Chicago and MIT.

Zhao, Rui. (2001), "The Optimal Unemployment Insurance Contract: Why a Replacement Ratio?," *mimeo.* University of Illinois, Champagne-Urbana.

Appendix - Proofs

Proof of proposition 1. I prove the proposition by a sequence of Lemmas and a corollary. First, I prove that (a) exactly one of either the truth telling constraint or the no quitting constraint is binding in the employment problem. Then, I prove that (b), given (a), both constraints in the unemployment problem are binding. Finally, I prove that given (a) and (b), the truth telling constraint is the only binding constraint in the employment problem.

Lemma A1 *In the employment problem, either the no quitting constraint or the truth telling constraint is binding.*

Proof. This is a proof by contradiction in two steps. First, I show that neither of the constraints being binding cannot be optimal and then that both being binding cannot be optimal. Suppose neither are binding. Then, $V^e > V^u > V^d$, $\theta_e = \theta_u = 0$, and by the first order conditions (2.7) - (2.10),

$$C_e'(V^e) = C_u'(V^u) = C_d'\left(V^d\right) = C_e'(V) \Longrightarrow V^e = V, \tag{2.A.1}$$

where the implication follows from the strict convexity of $C_e(\cdot)$. Using (2.6), (2.9), and (2.19),

$$V^d = \frac{u(c)}{1-\beta}. \tag{2.A.2}$$

The promise keeping constraint together with (2.A.1) and (2.A.2) implies

$$\begin{aligned} V &= V^d(1-\beta) + \beta\left\{(1-\xi)\left[\lambda V + (1-\lambda)V^u\right] + \xi V^d\right\} \\ &< V^u(1-\beta) + \beta\left\{(1-\xi)\left[\lambda V + (1-\lambda)V^u\right] + \xi V^u\right\} \\ &= V^u + \beta\lambda(1-\xi)(V - V^u) \Longrightarrow V^u > V = V^e, \text{ a contradiction.} \end{aligned}$$

Next, suppose both are binding. Then $V^e = V^u = V^d$. The first order conditions (2.7) - (2.10) with θ_e and θ_u being strictly positive now imply that

$$C_e'(V^e) > C_e'(V) > C_d'\left(V^d\right).$$

This implies $V^e = V^u = V^d > V$ and, by (2.6), (2.9), and (2.19),

$$V^e = V^u = V^d < \frac{u(c)}{(1-\beta)}.$$

The promise keeping constraint is then $V = u(c) + \beta V^d > u(c) + \beta V \implies V > \frac{u(c)}{1-\beta} \implies V > V^d = V^e$, a contradiction. ■

Lemma A2 *If the truth telling constraint in the employment problem is binding when employed in period* t, *it is also binding if still employed in period* $t+1$.

Proof. If the truth telling constraint is binding, the no-quitting constraint is not, so $\theta_e = 0$. Then $C_e'(V^e) = C_e'(V) \implies V^e = V$, so the contract is self generating. ■

Lemma A3 *If the no quitting constraint in the employment problem is binding when employed in period* t, *it is also binding if still employed in period* $t+1$.

Proof. Suppose towards a contradiction that no quitting binds at t but not at $t+1$. Then, $\theta_{e,t} > 0$, $\theta_{u,t} = 0$ and $\theta_{e,t+1} = 0, \theta_{u,t+1} > 0$, and by (2.7) and (2.8)

$$C_u'(V_t^u) < C_e'(V_t^e) = C_e'\left(V_{t+1}^e\right) < C_u'\left(V_{t+1}^u\right).$$

This implies, by convexity of the cost functions, that

$$V_{t+1}^u > V_t^u \text{ and } V_{t+1}^e = V_t^e.$$

Since the no quitting constraint binds at t, we have that

$$V_{t+1}^e = V_t^e = V_t^u.$$

The no quitting constraint does not bind at $t+1$, so the truth telling constraint binds at $t+1$. Hence, $V_{t+1}^u < V_{t+1}^e = V_t^u$, a contradiction. ■

Corollary A1. *Either the no quitting or the truth telling constraint is binding the entire employment spell.*

Proof. Follows immediately from the three Lemmas above. ■

Lemma A4 *In the unemployment problem, if either the effort constraint or the truth telling constraint is binding, then both are.*

Proof. Suppose that only the truth telling constraint binds. Then $\theta > 0$ and $\phi = 0$ and by (2.17)

$$C_d'\left(V^d\right) < C_u'(V). \tag{2.A.3}$$

Let c^d be consumption if disabled in the beginning of next period. Using (2.14) and (2.19), (2.A.3) implies that

$$c^d < c \Longrightarrow V^d = \frac{u\left(c^d\right)}{1-\beta} < \frac{u\left(c\right)}{1-\beta}$$

By (2.16) and (2.18), with $\theta > 0$

$$C_u'\left(V^u\right) > C_u'\left(V\right) \Longrightarrow V^u > V,$$

where the implication follows from strict convexity of $C_u\left(\cdot\right)$. As the truth telling constraint is binding, $V^u = V^d > V$. Since $\phi = 0$ it follows that

$$\beta\left(1-\xi\right)p\left(V^e - V^u\right) > a,$$

and then

$$\begin{aligned} V &= u\left(c\right) - a + \beta\left\{\left(1-\xi\right)\left[pV^e + \left(1-p\right)V^u\right] + \xi V^d\right\} > u\left(c\right) + \beta\left\{\left(1-\xi\right)V^u + \xi V^d\right\} \\ &= u\left(c\right) + \beta V^d > \left(1-\beta\right)V^d + V^d = V^d \end{aligned}$$

But then $V > V^d$, a contradiction. Hence, if $\theta > 0$ then $\phi > 0$. Suppose next that $\phi > 0$ and $\theta = 0$. Then, by (2.17)

$$C_d'\left(V^d\right) = C_u'\left(V\right). \tag{2.A.4}$$

Again, using (2.14) and (2.19), (2.A.4) implies that

$$c^d = c \Longrightarrow V^d = \frac{u\left(c\right)}{1-\beta}$$

Moreover,

$$\begin{aligned} V &= u\left(c\right) - a + \beta\left\{\left(1-\xi\right)\left[pV^e + \left(1-p\right)V^u\right] + \xi V^d\right\} = u\left(c\right) + \beta\left\{\left(1-\xi\right)V^u + \xi V^d\right\} \\ &= V^d\left(1-\beta\right) + \beta\left\{\left(1-\xi\right)V^u + \xi V^d\right\} > V^u. \end{aligned}$$

The last equation implies $V^d > V^u$, violating the truth telling constraint. Hence, if the effort incentive constraint is binding, so is the truth telling constraint. ■

Lemma A5 *In the unemployment problem, if the constraints bind in one period, they bind in all subsequent periods during the unemployment spell.*

Proof. Suppose the constraints bind when unemployed in period t with promise equal to V but not if still unemployed in $t+1$. The constraints being binding at t but not in $t+1$ implies, by (2.15), (2.16), and (2.18)

$$C_e'(V_t^e) > C_u'(V) > C_u'(V_t^u) = C_u'\left(V_{t+1}^u\right) = C_e'\left(V_{t+1}^e\right).$$

Hence, by strict convexity of the cost functions

$$V_t^e > V_{t+1}^e \text{ and } V_t^u = V_{t+1}^u.$$

The effort constraint is binding in t but not in $t+1$. Hence,

$$\beta(1-\xi)\,p\,(V_t^e - V_t^u) = a$$

$$\beta(1-\xi)\,p\left(V_{t+1}^e - V_{t+1}^u\right) > a$$

But then $V_{t+1}^e > V_t^e$, a contradiction. ■

Lemma A6 *Consider an agent employed at time t with a value V such that the truth telling constraint is binding in the employment problem. Then, if unemployed at $t+1$, the constraints in the unemployment problem are binding.*

Proof. Suppose towards a contradiction that the truth telling constraint is binding when employed at t but the constraints in the unemployment problem are not when unemployed in $t+1$. Then, $V_t^u = V_t^d$, and, by (2.6), (2.7) and (2.8)

$$C_d'\left(V_t^d\right) < C_e'(V) = 1/u\left(c_t^e\right).$$

Using (2.19) and the fact that consumption is constant after turning disabled,

$$V_t^u = V_t^d < u\left(c_t^e\right)/(1-\beta).$$

Moreover with $\theta_{e,t} = 0$ and $\theta_{u,t} > 0$, by (2.7) and (2.8),

$$C_u'(V_t^u) > C_e(V_t^e) \Longrightarrow 1/u'\left(c_{t+1}^u\right) > 1/u'\left(c_t^e\right).$$

where the implication follows from the envelope conditions in the employment and unemployment problems. Since the constraints are not binding in $t+1$, $C_u'(V_t^u) = C_u'\left(V_{t+1}^u\right) = C_d'\left(V_{t+1}^d\right) \Longrightarrow V_{t+1}^d = u\left(c_{t+1}^u\right)/(1-\beta)$. Combining these three relations we have that $V^d_{t+1} > V^d_t = V^u_t$. The effort constraint is slack in $t+1$ and $V^u_{t+1} > V^d_{t+1}$ since the truth telling constraint when unemployed in $t+1$ is also slack. Therefore,

$$\begin{aligned} V_t^u &= u\left(c_{t+1}^u\right) - a + \beta\left\{(1-\xi)\left[pV_{t+1}^e + (1-p)V_{t+1}^u\right] + \xi V_{t+1}^d\right\} \\ &> u\left(c_{t+1}^u\right) + \beta\left[(1-\xi)V_{t+1}^u + \xi V_{t+1}^d\right] > u\left(c_{t+1}^u\right) + \beta V_{t+1}^d = V_{t+1}^d > V_t^u, \end{aligned}$$

a contradiction. ■

Lemma A7 *Consider an agent employed at time t with a value V such that the no quitting constraint is binding in the employment problem. Then, if unemployed at $t+1$, the constraints in the unemployment problem are binding.*

Proof. Suppose the agent is employed at time t with the no quitting constraint binding and then unemployed in $t+1$ with no constraint binding. Then, by (2.7), (2.8) and (2.10),

$$C_e'\left(V_t^e\right) > C_e'\left(V\right) > C_u'\left(V_t^u\right).$$

This implies $V_t^e = V_t^u > V$, by convexity of $C_e'(\cdot)$ and the fact that the no-quitting constraint binds. Since $C_e'(V) > C_u'(V_t^u)$ and $V_t^u > V$,

$$C_e'\left(V_t^u\right) > C_u'\left(V_t^u\right).$$

No constraints being binding when unemployed at $t+1$ implies that

$$C_e'\left(V_{t+1}^e\right) = C_u'\left(V_{t+1}^u\right) = C_u'\left(V_t^u\right) \Longrightarrow V_{t+1}^u = V_t^u.$$

Hence $C_e'\left(V_{t+1}^e\right) < C_e'\left(V_t^u\right) = C_e'\left(V_{t+1}^u\right) \Longrightarrow V_{t+1}^u > V_{t+1}^e$, violating the effort incentive constraint while unemployed. ■

Lemma A8 *Only the truth telling constraint is binding in the employment problem.*

Proof. Suppose towards a contradiction that the no-quitting constraint binds while employed. Then $V_t^e = V_t^u$. Let the promises set in $t+1$ be without hats if unemployed then and with hats if employed then. Then, if unemployed in $t+1$, the truth telling and effort constraints are binding. Hence,

$$V_t^u = u\left(c^u\right) + \beta V_{t+1}^u = u\left(c^u\right) + \beta V_{t+1}^d.$$

If still employed in $t+1$, the no quitting constraint is still binding, $\widehat{V}^e_{t+1} = \widehat{V}^u_{t+1}$. Hence,

$$V^e_t = u(c^e) + \beta\left[(1-\xi)\widehat{V}^u_{t+1} + \xi\widehat{V}^d_{t+1}\right].$$

With a binding no-quitting constraint and a slack truth telling constraint, (2.7) (2.8) (2.10) and (2.18) imply that

$$C'_e(V^e_t) > C'_u(V^u_t) \Longrightarrow \frac{1}{u'(c^e)} > \frac{1}{u'(c^u)}$$

So $c^e > c^u$. Then

$$u(c^e) - u(c^u) = \beta V^d_{t+1} - \beta\left[(1-\xi)\widehat{V}^u_{t+1} + \xi\widehat{V}^d_{t+1}\right] > 0$$

Since the no quitting constraint is the binding one while employed it follows that

$$\left[(1-\xi)\widehat{V}^u_{t+1} + \xi\widehat{V}^d_{t+1}\right] > \widehat{V}^d_{t+1}.$$

The last two equations imply

$$V^d_{t+1} > \widehat{V}^d_{t+1}.$$

The first order conditions however state that

$$C'_d\left(\widehat{V}^d_{t+1}\right) = C'_e(V^e_t) > C'_u(V^u_t) > C'_d\left(V^d_{t+1}\right).$$

Convexity of the cost functions then imply $V^d_{t+1} < \widehat{V}^d_{t+1}$, a contradiction. ■

Proof of proposition 2. The truth telling constraint during employment is binding. Thus

$$C'_e(V^e) = C'_e(V) \Longrightarrow V^e = V,$$

which by using the envelope condition with (2.6), means that

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

so, by strict concavity of $u(\cdot)$, consumption is constant during the entire employment spell. During unemployment, the truth telling constraint being binding implies that $\theta > 0$. By (2.14), (2.17), and (2.18),

$$C'_d\left(V^d\right) < C'_u(V) \Longrightarrow \frac{1}{u'(c^d)} < \frac{1}{u'(c)} \Longrightarrow V^d < \frac{u(c)}{1-\beta}, \tag{2.A.5}$$

where c^d denotes consumption in the next period if disabled then and where the the last implication follows since

$$\frac{u\left(c^d\right)}{1-\beta} = V^d.$$

The incentive constraint for effort implies that

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

where the last equality follows since the truth telling constraint is binding, $V^u = V^d$. Note that by (2.A.5)

$$V \geq u\left(c\right) + \beta V^d \Longrightarrow V > \left(1-\beta\right)V^d + \beta V^d = V^d.$$

Hence

$$V > V^d = V^u.$$

Combining (2.14) and (2.18) repeatedly it follows that $V^u < V$ implies that consumption is falling during the entire unemployment spell. Consumption being constant during disability is trivial. ■

Proof of proposition 3. During unemployment, since the effort incentive constraint is binding and $V^u < V$ (see the proof of proposition 2) by (2.15) and (2.16)

$$C_u'\left(V\right) < C_e'\left(V^e\right),$$

and this implies, by the envelope conditions in the two problems and the first order conditions with respect to consumption, that $c < c^e$. The truth telling constraint during employment is binding. Then by (2.6), (2.8), (2.10) and (2.18),

$$\frac{1}{u'\left(c^u\right)} > \frac{1}{u'\left(c\right)},$$

which means that consumption rise upon layoff. If employed last period before turning disabled, by (2.6), (2.9), (2.10), and (2.19),

$$\frac{1}{u'\left(c^d\right)} < \frac{1}{u'\left(c\right)}.$$

If unemployed prior to turning disabled the same result is produced using (2.14), (2.17), (2.18), and (2.19). ■

Proof of proposition 4. During employment, the no quitting constraint is not binding, so $C_e'(V^e) = C_e'(V) \Longrightarrow V^e = V$. This means that the problem is self generating, so all promised values remain the same during an employment spell. During unemployment, proposition 2 showed that $V > V^u$, so V^u is decreasing during an unemployment spell. By recursively applying the incentive constraint, which binds at the optimum, as given by

$$V^e = V^u + \frac{a}{\beta(1-\xi)p},$$

it can be inferred that V^e has to decrease during the unemployment spell, as V^u decrease during an unemployment spell. As the truth telling constraint is binding, it follows that V^d has the same dynamics as V^u. Finally, to prove that benefits conditional on disability rise upon employment, consider an unemployed agent with promise V. The planner awards him new promises $\{V^e, V^u, V^d\}$, and it suffices to prove that the agents disability benefits are larger if he finds a job and then becomes disabled than if he becomes disabled immediately. Let V_e^u be the promise to this agent if he finds a job and then becomes unemployed and V_e^d.be the promise to the agent if he finds a job and then becomes disabled. By (2.7), (2.8), and (2.10),

$$C_u'(V_e^u) > C_e'(V^e).$$

Moreover, by (2.15), (2.16), and (2.18)

$$C_e'(V^e) > C_u'(V^u).$$

Hence, by convexity of $C_u(\cdot)$, we have $V_e^u > V^u$. The truth telling constraint being binding completes the proof, as $V_e^u = V_e^d$. ■

Proof of proposition 5. This proposition is proved in several steps. Fist, I prove that the truth telling constraint is always binding by showing that no constraint being binding cannot be optimal and then showing that if the no-quitting constraint binds, so does the truth telling constraint. I then show that promises for future unemployment is never decreased during employment. Given these proofs, I then show that there exists an $e>0$ such that it is not optimal to quit.

Lemma A9 *In the employment problem, there exists an $e>0$ such that at least one of the no quitting constraint or the truth telling constraint is binding.*

Proof. Suppose neither the no-quitting not the truth telling constraints are binding. Then, $V^e > V^u > V^d$, and by the first order conditions (2.7) - (2.10),

$$C_e'(V^e) = C_u'(V^u) = C_d'\left(V^d\right) = C_e'(V).$$

This implies $V^e = V$. Using (2.19) and (2.6),

$$V^d = \frac{u(c)}{1-\beta}.$$

The promise keeping constraint implies

$$\begin{aligned} V &= u(c) - e + \beta\left\{(1-\xi)\left[\lambda V^e + (1-\lambda)V^u\right] + \xi V^d\right\} < u(c) - e + \beta V \Longrightarrow \\ V &< (u(c)-e)/(1-\beta) \Longrightarrow V < V^d, \end{aligned}$$

a contradiction.

Lemma A10 *If the no-quitting constraint binds, so does the truth telling constraint.*

Proof. Suppose that the no-quitting constraint is the only binding constraint in the employment problem in period t. Then $V_t^e = V_t^u > V_t^d$, and

$$C_e'(V_t^e) > C_d'\left(V_t^d\right) = C_e'(V_t) > C_u'(V_t^u).$$

This implies $V^d = u(c_t)/(1-\beta)$. If turning unemployed in the next period the truth telling constraint is binding, and the value should the agent remain unemployed in $t+2$ is reduced. In other words $V_{t+1}^u = V_{t+1}^d < V_t^u$. Moreover,

$$V_{t+1}^d \leq \frac{u\left(c_{t+1}^u\right)}{1-\beta} < \frac{u(c_t)}{1-\beta},$$

because consumption upon layoff drop if the no-quitting constraint is the only binding constraint, as seen from (2.6), (2.8) and the envelope condition in the unemployment problem. Now, this implies $V_{t+1}^u \leq V_t^d$. The promise keeping constraint in period $t+1$ for an unemployed reads

$$V_{t+1} = V_t^u = u\left(c_{t+1}^u\right) + \beta V_{t+1}^u < u(c_t) + \beta V_{t+1}^u = (1-\beta)V_t^d + \beta V_{t+1}^u \leq V_t^d,$$

violating the period t truth telling constraint. ■

Lemma A11 *Promises for future unemployment are never reduced during employment.*

Proof. This is trivial since if the no-quitting constraint is slack, then $V^e = V$ and hence V^u remains constant during the whole spell while if the no-quitting constraint binds $V^e = V^u > V$. ■

Proposition 5 *There exists an $e > 0$ such that the no-quitting constraint is not binding.*

Proof. Suppose he finds out that he is employed at the beginning of period t. If he quits directly, his utility from now on is V_t^u. Alternatively, he can truthfully report to be employed this period, quit in the beginning of next period, and earn a utility from now on equal to $\widehat{V} = u(c_t^e) - e + \beta V_{t+1}^u$. Evidently, he postpones his quit if $\widehat{V} > V_t^u$. Whether the no-quitting constraint is binding or not, it turns out that thee agent's future value of turning unemployed never decreases if he is employed another period. In other words, $V_{t+1}^u \geq V_t^u$. If we use the fact that the truth telling constraint is always binding ($V^u = V^d$ for any history), this implies that a sufficient condition for $\widehat{V} > V_t^u$, is

$$u(c_t^e) - e + \beta V_t^u \geq V_t^u \Longleftrightarrow u(c_t^e) - e \geq V_t^d (1 - \beta) = u\left(c_t^d\right).$$

Now as consumption always drops upon disability, i.e. $u(c_t^e) > u\left(c_t^d\right)$, there always exists an $e > 0$ such that it is optimal to postpone the quit. Applying this principle recursively, there exists an $e > 0$ such that the agent never quits. ■

Proof of proposition 6. When the effort cost of working is high, the planner has to compensate the employed in order to abstain him from quitting. The compensation pushes consumption upwards more for higher e, and because decreasing marginal utility implies increasing marginal cost of delivering a util, there has to be a high enough e such that it is no longer optimal to give strictly more to an employed than to an unemployed. The proof that no constraint can be binding is the same as before, so if the truth telling constraint is slack, the no quitting constraint is binding. Suppose that the no-quitting constraint is the only binding constraint in the employment problem in period t. Then $V_t^e = V_t^u > V_t^d$, and

$$C_e'(V_t^e) > C_d'\left(V_t^d\right) = C_e'(V_t) > C_u'(V_t^u).$$

This implies $V^d = u(c_t)/(1-\beta)$. If turning unemployed in the next period the truth telling constraint is binding, and the value should the agent remain unemployed in $t+2$ is reduced. In other words $V_{t+1}^u = V_{t+1}^d < V_t^u$. Moreover,

$$V_{t+1}^d \leq \frac{u\left(c_{t+1}^u\right)}{1-\beta} < \frac{u(c_t)}{1-\beta},$$

because consumption upon layoff drop if the no-quitting constraint is the only binding constraint. Now, this implies $V_{t+1}^u \leq V_t^d$. The promise keeping constraint in period $t+1$ for an unemployed reads

$$V_{t+1} = V_t^u = u\left(c_{t+1}^u\right) + \beta V_{t+1}^u < u(c_t) + \beta V_{t+1}^u = (1-\beta) V_t^d + \beta V_{t+1}^u \leq V_t^d,$$

violating the period t truth telling constraint. ■

Proof of proposition 7. When the no-quitting constraint is binding, (2.6), (2.7), and the envelope condition in the employment problem (2.10) imply that.

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

The rest of the proposition is the same as in the proof of proposition 2.

Proof of proposition 8. As the truth telling constraint is still binding in both problems, the fact consumption upon disability drops can be proved in the same way as in proposition 3. Also, as there is no change in the structure of binding constraints in the unemployment problem, the fact that consumption rises upon employment is also proved as in proposition 3. It is not possible to show either that consumption rises or falls upon layoff, so either could be true.

Proof of proposition 9. When the no-quitting constraint is binding, (2.7) implies that $C_e'(V^e) > C_e'(V)$. Convexity of $C_e(\cdot)$ then assures that $V^e > V$. the binding no quitting constraint and truth telling constraints implies $V^e = V^u = V^d$ so all values move in tandem. During unemployment, the results remain from the preceding section, an the proof is identical to that in proposition 4. To prove that benefits conditional on disability rise upon job finding is easy since both the truth telling and the no-quitting constraint is binding, in combination with the binding incentive constraint for effort. Consider an unemployed agent with promise V. The incentive constraint assures that $V^e > V^u$, and when having found a job $V_e^e = V_e^u = V_e^d > V^e > V^u = V^d$, where V_e^e is the promise for employment if the agent found a job, and V_e^u and V_e^d are the promises conditional on unemployment and disability for the agent that has found a job. As $V_e^d > V^d$ the agent who finds a job and then becomes disabled then gets higher benefits than if he got disabled immediately. ■

CHAPTER 3

Optimal Unemployment and Disability Insurance: Allocations, Welfare Gains, and the Effectiveness of Self Insurance

Erik Höglin

Abstract. This paper asks what are the quantitative implications of the optimal unemployment and disability insurance system. Unemployment insurance benefits are slowly decreasing during an unemployment spell. Disability insurance benefits increases in the number of periods as unemployed if the unemployment spells are shorter than of average length, and decreases if unemployment spells are longer. A stationary distribution of consumption exhibits much less variation in the optimal system than in the current one. Moreover, the optimal contract is not well replicated by the current system, making the potential cost savings from moving to the optimal system large, amounting to about 15 percent. Finally, self insurance through a riskless bond is relatively effective: an agent is willing to give away no more than the equivalent of 6 weekly wages in order to join the planner's system.

3.1. Introduction

Unemployment and disability are important sources of income fluctuations. Individuals facing such risks demand insurance to smooth consumption. However, private insurance and capital markets have proved to provide insufficient unemployment and disability insurance. For example, Gruber (1997) reports a 20 percent average drop in consumption upon unemployment for workers not eligible for governmental unemployment insurance. The failure of private markets may be explained by informational asymmetries leading to adverse selection problems. Governments are able to overcome adverse selection by simply developing mandatory insurance systems. Indeed, governmental social insurance programs are important in most industrialized countries. In the United States, the cost of the unemployment insurance system was $39 billion in

[0] I am grateful to Martin Flodén, Lars Ljungqvist, and Henrik Lundvall for useful comments and suggestions. Financial support from the Jan Wallander and Tom Hedelius foundation is gratefully acknowledged.

2003 (Feldstein, 2005). The outlays on disability insurance even surpasses that amount, being $61 billion already in 2001 (Golosov and Tsyvinski, 2006).

Providing governmental insurance is by no means unproblematic. At the core are moral hazard problems: with job search effort being unobservable, unemployed might search less than what is desirable. Unemployment insurance is therefore believed to increase unemployment (Katz and Meyer, 1990; Carling *et al*, 2001). Disability insurance has similar problems. Determining if a claimant is really disabled is an almost impossible task, especially since many applying for disability insurance have difficult-to-verify conditions, such as mental disorders and back pain. Although hard to measure, previous studies such as Gruber (2000) and Author and Duggan (2003) have indeed found excessive claims in the disability insurance system.

Motivated by these private information considerations, Shavell and Weiss (1979) and Diamond and Mirrlees (1978) provided the seminal contributions to the field of optimal social insurance theory under asymmetric information. Shavell and Weiss (1979) used a one spell model with unobservable job search effort and unemployment insurance benefits as the only policy instrument to ask how unemployment insurance benefits should evolve over time. They concluded that in order to provide the proper incentives to exert search effort, consumption should decrease during an unemployment spell. Hopenhayn and Nicolini (1997) extended the analysis to include a second policy instrument, a tax upon employment, and showed that not only should consumption decrease with the length of the unemployment spell, but so should consumption in future employment. They also generalized the proposition of declining consumption during unemployment to a model with multiple spells in Hopenhayn and Nicolini (2004). The intuition is that to minimize consumption variance and still provide incentives to search, the planner gradually lowers the consumption in all future states during an unemployment spell.

In their seminal paper on optimal disability insurance, Diamond and Mirrlees (1978) assumed that whether an agent is disabled or not is private information. Under this assumption, they showed that consumption should increase up to the point of disability, decrease upon disability, and then remain constant. They also proved an important proposition of history dependence: the later an individual becomes disabled, the higher his disability benefits should be. The drop in consumption when turning disabled is needed to abstain able workers from prematurely claiming to be disabled. The reason why future disability benefits rise with tenure is more complex, but similar to why consumption upon employment fall with increased unemployment spell length in Hopenhayn and Nicolini (1997, 2004). The planner optimally makes an agent indifferent between becoming disabled and remaining able. To ensure that the agent does not claim to be disabled, conditional on future ability, an award for tenure becomes necessary. As agents are risk averse, this reward is optimally delivered in all future states. Hence, disability benefits are higher for agents becoming disabled after a longer history of job tenure.

None of the papers above analyzed the interdependence between unemployment and disability insurance. Empirically, important interactions between the systems are well documented. Gruber (2000) finds that disability insurance reduces labor supply; workers unable to find work opt for disability insurance rather than unemployment insurance, and hence leave the labor force. Moreover, Autor and Duggan (2003) report that mid 1980's U.S. unemployment was significantly reduced by the presence of disability insurance. The mechanism suggested in these papers is straightforward. When it is difficult to infer the actual state of an agent having access to a veritable menu of social insurance programs, the agent does not necessarily chose the program that reflects his true state, but rather the program maximizing his well being.

In Höglin (2008), I merge Hopenhayn and Nicolini's (2004) multiple spells model with a simplified version of Diamond and Mirrlees (1978) in order to address the interactions between the systems. The falling consumption during unemployment is robust to the introduction of disability in the model, and the intuition is the same as in Shavell and Weiss (1979) and Hopenhayn and Nicolini (1997, 2004). The results concerning optimal unemployment insurance however differs qualitatively from Shavell and Weiss (1997) and Hopenhayn and Nicolini (1997,2004) in one respect because the planner has to induce an agent to tell the truth. Specifically, to avoid unemployed to prematurely claim disability, the planner offers a reward for truth telling by replacing more than 100 percent of the past consumption during the first period of the unemployment spell. Without disability, i.e. in Hopenhayn and Nicolini (2004), the replacement rate is exactly 100 percent.

The optimal disability insurance is also altered in a non trivial way by the prevalence of unemployment. Every period an agent is unemployed, his promised value in all future states should be reduced, for the same reason as the increasing employment tax in Hopenhayn and Nicolini (1997, 2004). Therefore, disability insurance benefits decrease during an unemployment spell. If an agent finds a job however, he is rewarded in all future states, which implies higher future disability benefits. These two opposing forces makes it impossible to analytically prove if disability benefits increase or decrease with periods of past unemployment. It suggests however that an agent's disability benefits are higher if he has many short spells compared to few long spells.

The purpose of this paper is threefold. First, I study the quantitative aspects of the joint system of insurance against unemployment and disability insurance and compare it to the existing literature by numerically calculating the optimal contract in Höglin (2008). Second, I compute the welfare gains associated with moving from the current system to the optimal one. Finally, I provide a measure of the effectiveness of self insurance compared to the planner's insurance capabilities.

To inquire the quantitative aspects of the joint system of insurance against unemployment and disability insurance, I numerically calculate the optimal contract in Höglin (2008). In this exercise it is possible to address several questions that Höglin (2008) could not answer. How fast should unemployment benefits fall during a spell? Should disability benefits be higher for an agent becoming disabled early than for one becoming disabled late? Does repeated spells of unemployment deteriorate future disability benefits? I also perform quantitative comparisons with Hopenhayn and Nicolini (1997, 2004) that shed further light on how the joint system differs from one considering only unemployment insurance.

The optimal insurance system against unemployment and disability features a high degree of consumption smoothing over time and across states, as compared to the current U.S. system. The initial unemployment insurance replacement rate is only marginally above 100 percent, dropping below 100 percent within a month, but one year into a spell of unemployment, benefits are still so high that the replacement rate is above 96 percent. In other words, the fall in the unemployment insurance benefits is quantitatively small. In the optimal contract, disability benefits are higher if the agent's history includes many short spells of unemployment than if it includes few long spells, for a given number of periods unemployed prior to becoming disabled. Moreover, if the agents' spells are around or below average length, unemployment tends to increase future disability insurance benefits while if they are of above average length, the opposite is true. The quantitative effects are substantial; an agent with many short spells of unemployment is awarded more than 10 percent higher disability benefits than an agent with the same number of total periods as unemployed, but with fewer and longer spells.

In addition to these questions concerning a single agent, I also simulate a large population of individuals to obtain a stationary distribution of consumption. I can then calculate moments of that stationary distributions and compare them to moments reflecting the current system as well as to the distribution implied by autarky. By doing this, it is possible to find out, for example, if consumption inequality is higher or lower in the optimal system than in the current one. Indeed, the standard deviation of consumption is much smaller in the optimal system compared to the current U.S. system, being twice as large in the current system as in the optimal one.

I measure the welfare gains associated with moving from the current system to the optimal one by comparing the costs of delivering the value of the current system optimally to the cost of the current system. The costs of delivering the value of the current system optimally is 15 percent lower than the cost of the current system. The source of this increased efficiency in the optimal system comes entirely from increased consumption smoothing. The model features a binary choice of searching if unemployed. Agents find it optimal to search in the optimal contract, by construction, but also in the current system. Moreover, it is not optimal to claim a false state, neither in the current nor in the optimal system. Hence, the aggregate and individual dynamics over states are the same in the current and the optimal system. This implies that the cost reductions stem only from increased consumption smoothing. To the extent that variations along an intensive margin of job search is important, the calculated welfare gains of the optimal system could be even higher in such an environment. Whether this is actually the case is left for future research however.

Following Hopenhayn and Nicolini (1997) and, more recently, Pavoni and Violante (2007), cost reductions are used to measure the welfare gains of moving to the optimal system. A natural alternative would be consumption equivalents, as is standard when measuring the welfare costs of business cycles (e.g. Lucas, 1987). However, private information makes the consumption equivalents method difficult to implement. Specifically, if consumption is increased in every state, the agents' behavior might change. Therefore, measuring welfare gains by cost reductions is a more straightforward method.

The final purpose of this paper is to provide a measure of the effectiveness of self insurance. Self insurance is a natural alternative to the planner's system. Borrowing constrained agents are not able to achieve the planner's consumption allocation as the planner has the ability to borrow. To measure the planner's ability and thus the effectiveness of self insurance, I compute the wealth level a newborn has to have in order to be indifferent between the planner's system and self insurance. It turns out that for initial wealth levels below 6 weekly wages, a newborn agent prefers to forgo his wealth and join the planner's balanced budget system.

The paper is organized as follows. First, the next three sections describe the environment, the planner's problem, and the calibration. Section 3.5 calculates the value of living in the current U.S. system. Section 3.6 then shows example consumption dynamics, explains how the prevalence of disability affects the optimal unemployment insurance and how unemployment affects the optimal disability insurance. In section 3.7, I simulate a large population over a long time horizon and analyze a stationary distribution over consumption. Section 3.8 calculates the cost of the current system as well as the cost reductions associated with moving to the optimal system and section 3.9 discusses the sources of these cost savings. Finally, section 3.10 studies the effectiveness of self insurance and section 3.11 concludes.

3.2. The Economy

The economy is populated by a measure one of risk averse agents. In each period, an agent can be employed, unemployed or disabled. All agents are born employed and face a constant per period probability of dying equal to ψ, and a measure ψ of agents are born each period, so that the population is constant. If employed, an agent earns a constant wage $w > 0$, while if unemployed or disabled he earns no income. Disability is an absorbing state, but disabled face death risk too. In contrast, employed and unemployed can transition between employment and unemployment as well as into disability and death. Specifically, an unemployed becomes disabled with per period probability ξ and can gain employment by exerting search effort. He can either choose $a > 0$ or zero search effort. If search effort is a, the probability of gaining employment is p, conditional on not dying or becoming disabled. If search effort is zero, an unemployed remain so with probability one, again conditional on not dying or turning disabled. Employed agents face the same probability of becoming disabled as unemployed and maintain employment with exogenous probability λ, if not dying or becoming disabled. Quits are possible. Effort is assumed to be costly in utility terms. Following Hopenhayn and Nicolini (1997), the preferences of an agent are ordered by

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

where c_t is consumption, $u(\cdot)$ is strictly increasing, strictly concave, twice continuously differentiable and $\beta \equiv \widehat{\beta}\left(1-\psi\right) \in (0,1)$ is the effective discount factor, including the pure time discount factor $\widehat{\beta}$ as well as the survival probability $(1-\psi)$. Assume also that $u(0)$ is well defined. No effort is required to maintain employment, so a_t is zero during employment and disability. All the results remain if we introduce an effort cost of employment, provided it is sufficiently small.

The agents can neither save nor borrow. Lack of access to capital markets in combination with risk aversion will constitute scope for improvement upon the autarky equilibrium by a social planner that has the ability to transfer resources over time and across agents.[1] I assume that such a planner has unlimited access to a perfect capital market, where he can borrow and lend at a constant gross interest rate equal to β^{-1}.

The current state of an agent is unobservable but income is observed by the planner. Hence, an employed cannot state unemployment or disability. An employed can quit however, rendering him in a situation identical to one that got laid off. The planner cannot distinguish quits from layoffs. Effort while unemployed is also unobservable. The optimal contract must induce an employed never to quit, unemployed and disabled not to lie, and unemployed to exert effort to become employed.

The evolution of the measure of employed, unemployed and disabled and the corresponding steady state values of these statistics will be important for the calibration and when computing the stationary distribution in section 3.7. Let E_t be the time t measure of employed agents and define U_t and D_t analogously. The measure of each type of agent evolve according to the difference equations

$$E_t = \psi + (1-\psi)(1-\xi)\lambda E_{t-1} + (1-\psi)(1-\xi)pU_{t-1}, \tag{3.1}$$

$$U_t = (1-\psi)(1-\xi)(1-\lambda)E_{t-1} + (1-\psi)(1-\xi)(1-p)U_{t-1}, \tag{3.2}$$

$$D_t = (1-\psi)\xi E_{t-1} + (1-\psi)\xi U_{t-1} + (1-\psi)D_{t-1}. \tag{3.3}$$

I will use the steady state version of these equations to calibrate transition probabilities consistent with observed long run averages of, for example, the unemployment rate.

3.3. The Planner's Problem

The planner's problem is to specify history dependent transfers and effort recommendations to minimize the cost of giving the agent a specified value V. Relying on the revelation principle, I restrict attention to contracts where the planner would like the

[1] Lack of access to any market for borrowing or lending is unrestrictive as long as an agent's savings can be costlessy monitored by the planner, or, equivalently, as long as the planner can control the agent's consumption. Extending the analysis to include hidden savings is *not* straightforward. Kocherlakota (2004) describes the technical difficulties in using the recursive constract methodology when a hidden state variable such as savings can take on a continuum of values. Motivated by these difficulties, I abstract from hidden savings. Werning (2002) and Mitchell and Zhang (2007) studies the role of hidden savings in optimal unemployment insurance under some functional forms assumptions. In section 10, I analyze self insurance by allowing for observable savings.

agents to tell the truth. Furthermore, I assume that the primitives of the model, including the initial promised value, are such that the planner would like to induce positive search effort. The planner collects reports on state in each period and makes history dependent transfers in order to induce no quitting, truth telling, and effort if the agent is unemployed. The timing of the contract within each period is:

(1) The agent learns his state and reports a state to the planner.
(2) Conditional on the history of reports, including the current period, the planner sets up a consumption level and promises for each possible future state.
(3) If the agent is unemployed, he chooses to exert search effort or not, which determines his probability distribution over next period's state.

If the promised value for a given state is V, the planner chooses current consumption c, and promises V^e, V^u, and V^d for each possible future state. Here, V^e is the expected utility from next period on, should the agent be employed then, and V^u and V^d are defined analogously. In order to deliver the promise, the planner has to set c, V^e, V^u, and V^d such that an agent's expected utility equals V. In addition, the contract must involve combinations of V^e, V^u, and V^d that makes it optimal for an agent not to quit, not to lie, and to search if unemployed. The value of the planner's problem when the agent has reported to be employed is $C_e(V)$, and denotes the minimal cost of delivering the value V to an agent currently employed. Corresponding cost functions when an agent is unemployed or disabled, $C_u(\cdot)$ and $C_d(\cdot)$, are defined analogously.

3.3.1. Employed. Suppose an agent is employed and has a promised value equal to V. The optimal contract then solves the Bellman equation:

$$C_e(V) = \min_{c,V^e,V^u,V^d} \left\{c - w + \beta\left\{(1-\xi)\left[\lambda C_e(V^e) + (1-\lambda)C_u(V^u)\right] + \xi C_d\left(V^d\right)\right\}\right\},$$

where the minimization is subject to

$$V = u(c) + \beta\left\{(1-\xi)\left[\lambda V^e + (1-\lambda)V^u\right] + \xi V^d\right\}, \tag{3.4}$$

$$V^e \geq V^u, \tag{3.5}$$

$$V^u \geq V^d. \tag{3.6}$$

Here, (3.4) is the promise keeping constraint, and (3.5) is the no quitting constraint that makes sure that an agent does not quit if still employed in the beginning of next period. Also, (3.6) is the truth telling constraint that assures that if unemployed next period, the agent will report so.

Höglin (2008) proves that in this environment, the no-quitting constraint (3.5) is slack whereas the truth-telling constraint (3.6) is binding. The reason is that since the incentive problems are bygones once an agent has truthfully claimed disability, it is cheap for the planner to give such an agent as much utils as possible, provided unemployed or employed are not given incentives to falsely claim disability. This is achieved by letting the truth telling constraint be binding. A similar reasoning applies for the no-quitting constraint. The planner has many incentive issues with an unemployed agent; he is to be given incentives to search and not to claim disability. This calls for a lot of variation in consumption and promised values, an admittedly expensive policy for the planner. The planner has less incentive problems with a currently employed agent, so the marginal cost of delivering utils to such an agent is lower than to an unemployed. This makes the no-quitting constraint become slack, since it calls for giving more utils to the employed than to the unemployed.

With a slack no-quitting constraint and a binding truth telling constraint, promised values satisfy

$$V^e = V > V^u = V^d \equiv \widehat{V}.$$

So the problem may be stated as

$$C_e(V) = \min_{\widehat{V}<V} \left\{ c - w + \beta \left\{ (1-\xi) \left[\lambda C_e(V) + (1-\lambda) C_u\left(\widehat{V}\right) \right] + \xi C_d\left(\widehat{V}\right) \right\} \right\}, \quad (3.7)$$

where

$$c = u^{-1}\left(V - \beta \left\{ (1-\xi) \left[\lambda V + (1-\lambda) \widehat{V} \right] + \xi \widehat{V} \right\} \right).$$

3.3.2. Unemployed. If an agent has reported unemployment and his promised value in this state is V, the optimal contract is characterized by the solution to the dynamic program:

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

where the minimization is subject to

$$u(c) - a + \beta\left\{(1-\xi)\left[pV^e + (1-p)V^u\right] + \xi V^d\right\} = V, \tag{3.8}$$

$$\beta(1-\xi)p\left[V^e - V^u\right] \geq a, \tag{3.9}$$

$$V^u \geq V^d. \tag{3.10}$$

Here, (3.8) is the promise keeping constraint and (3.9) is the incentive compatibility constraint inducing effort. The incentive compatibility constraint makes sure that it is optimal for the unemployed to exert search effort. Finally, in order for an unemployed not to claim to be disabled, a truth telling constraint like (3.10) has to be satisfied.

The structure of binding constraints is similar to that in the employment problem. The truth telling constraint is binding for the same reason as in the employment problem. The incentive constraint is also binding; forces that calls for consumption smoothing dictates that it is not optimal to vary values across future state to a greater extent than what makes it worthwhile to exert search effort. Promised values then satisfy

$$V^e = V^u + \frac{a}{\beta(1-\xi)p} > V^u = V^d \equiv \widehat{V}.$$

Moreover, Höglin (2008) proves that $\widehat{V} < V$. Hence, the problem can be rewritten to include a single control variable. The Bellman equation becomes,

$$C_u(V) = \min_{\widehat{V}<V}\left\{c + \beta\left[(1-\xi)\left[pC_e(V^e) + (1-p)C_u\left(\widehat{V}\right)\right] + \xi C_d\left(\widehat{V}\right)\right]\right\}, \tag{3.11}$$

where

$$c = u^{-1}\left(V + a - \beta\left\{(1-\xi)\left[pV^e + (1-p)\widehat{V}\right] + \xi\widehat{V}\right\}\right),$$

and

$$V^e = \widehat{V} + \frac{a}{\beta(1-\xi)p}.$$

3.3.3. Disabled. Disability is an absorbing state. The incentive problem between the planner and the agent are then bygones once an agent has truthfully reported to be disabled. It is then optimal for the planner to award the agent a constant level of consumption that delivers the promised value. Let this transfer be c^d, which is implicitly defined by

$$V = \frac{u\left(c^d\right)}{1-\beta}.$$

The cost to the planner is then

$$C_d\left(V\right) = \frac{c^d}{1-\beta} = \frac{u^{-1}\left(V\left(1-\beta\right)\right)}{1-\beta}.$$

Before proceeding to the calibration and solution of the model, we note that since in all states it is equally good for an agent to become disabled as to become unemployed, there is no reason for a disabled to falsely claim to be unemployed. If a disabled did do so, he would be worse off than by telling the truth, since he could never reach employment.

3.4. Calibration

The model period is assumed to be one week, and the pure time discount factor is set to .999. Following Ljungqvist and Sargent (1998), I assume that the individual spends an average of 42.7 years in the model before retiring and leaving the model, that is, before dying. This makes the weakly survival probability equal $1 - .00045$ and $\beta = .999*(1 - .00045) = .99855$. Based on Meyer (1990), I set the probability of gaining employment, conditional on job search, to .1, which makes the average unemployment spell last for 10 weeks. This is also the value that Hopenhayn and Nicolini (1997) target in their calibration.

Using the system (3.1) - (3.3), the values of ξ and ψ yields the steady state measure of disabled. Meyer and Mok (2006) use the Panel Studies of Income Dynamics to calculate that 20 percent of U.S. male household heads are disabled and that one fifth of those are severely and chronically disabled. Based on their estimate, I calibrate ξ so that, given $\psi = .00045$, the steady state measure of disabled equals 4 per cent of the population. The final transition parameter is λ, the probability of keeping the job if employed. Using (3.1) - (3.3), I set λ so that the steady state unemployment rate $\widetilde{u} \equiv U/(E+U)$ equals 5.5 per cent. This implies that the probability of lay off $(1-\lambda)$ equals .00585, so that an agent on average keeps a job for 3.3 years.

In order for the analysis to be comparable to Hopenhayn and Nicolini (1997), I adopt their constant relative risk aversion utility function, $u(c) = c^{1-\sigma}/(1-\sigma)$, with $\sigma = .5$. Furthermore, w is normalized to 100. The last parameter is a, the effort cost of searching for a job. As a baseline, I set a such that a newly unemployed in the current system (described in the next section) is indifferent between exerting search effort or not. This turns out to imply $a = 65$, and constitutes an upper bound on what search effort costs can be if search is to be optimal within the current system.

3.5. The Value of the Current System

The agent's initial value is primitive to the model. Given the parameters above, it is possible to calculate the value to the agent of the current system. This value will then be delivered in the optimal way, and comparisons to the actual system in terms of consumption profiles and cost reductions are possible.

Following Hopenhayn and Nicolini (1997), I assume that an unemployed gets 66 percent of his pre unemployment *wage* for 26 weeks of unemployment and nothing thereafter. If becoming disabled, the agent is awarded 40 percent of his wage when lastly employed. This figure is used in Golosov and Tsyvinski (2006). I assume that no history dependent taxes upon employment are used in the current system. Finally, once an agent is employed, his value if becoming laid off is the same regardless of his history.

The value of a disabled agent is then

$$\widetilde{V}^d = \frac{u(40)}{(1-\beta)}.$$

Let $\widetilde{V}^u(t)$ be the value of an unemployed agent with t weeks of unemployment insurance benefits entitlements, living in the current system and behaving optimally. The value of an agent with exhausted unemployment benefits is then

$$\widetilde{V}^u(0) = \max\left\{u(0) - a + \beta\left\{(1-\xi)\left[p\widetilde{V}^e + (1-p)\widetilde{V}(0)\right] + \xi\widetilde{V}^d\right\}, \widetilde{V}^d\right\},$$

where the max operator reflects the possibility to claim to be disabled.[2] For $t \in \{1, 2, ...26\}$,

$$\widetilde{V}^u(t) = \max\left\{u(66) - a + \beta\left\{(1-\xi)\left[p\widetilde{V}^e + (1-p)\widetilde{V}^u(t-1)\right] + \xi\widetilde{V}^d\right\}, \widetilde{V}^d\right\}.$$

Finally, the value of an employed is given by

$$\widetilde{V}^e = \max\left\{u(100) + \beta\left\{(1-\xi)\left[\lambda\widetilde{V}^e + (1-\lambda)\widetilde{V}^u(26)\right] + \xi\widetilde{V}^d\right\}, \widetilde{V}^u(26), \widetilde{V}^d\right\}.$$

The current system promises $\widetilde{V}^e$ to a newborn. In the proceeding analyses, the planner will deliver $\widetilde{V}^e$, but within the optimal system. The agents will thus be given the same expected lifetime utility as in the current system, but the planner will save costs. The costs reductions can then serve as a measure of the efficiency loss from the current system and will be calculated in section 3.8. Note that the current system does not have a balanced budget. On the contrary, a positive budget is needed in order for the current system to be able to deliver $\widetilde{V}^e$. Since I will ask the optimal system to deliver $\widetilde{V}^e$ later, the optimal system is, *a priori*, not a balanced budget system either.

3.6. Single Agents

Within the environment of Hopenhayn and Nicolini's (1997) one spell model, single agent analyses is the only viable option for numerical examples. In this section, I study single agents' outcomes within the optimal contract of this paper and compare them to Hopenhayn and Nicolini (1997). In addition, I numerically solve Hopenhayn and Nicolini's (2004) model of multiple spells to be able to single out if the differences depend on the one spell nature of the contract or on the introduction of disability in the model. I also compare the difference in unemployment insurance benefits between the optimal system delivering the value of the current system and the optimal balanced budget system.

The purpose of this section is not only comparisons to the previous literature; the single agent analyses shed light on the contract the agent is faced with and questions important in their own right, such as the effect of unemployment on future disability insurance benefits, can be inquired for different example histories.

---

[2] Given how a is calibrated, this option is never exerted however.

3.6.1. Consumption profiles. From Höglin (2008), consumption should be constant during an employment spell, decreasing during unemployment, and constant after an agent has turned disabled. With the no-quitting constraint being slack in the employment problem, there is no way in which the planner has to vary consumption during an employment spell in order for the employed agent not to quit. Therefore, the agents' strictly concave utility function prescribes that the cheapest way for the planner to deliver utils is a constant sequence of consumption during an employment spell. During unemployment, the planner faces a completely different situation. It is crucial to provide the right incentives so that an unemployed agent searches for new jobs. The efficient way to provide such incentives is a falling sequence of consumption during unemployment, a well known result already from Shavell and Weiss (1979). Moreover, consumption should go up when an agent is laid off and when he finds a new job, but drop if he turns disabled. The drop upon disability is a natural way to make sure that no agent falsely claims disability. The upward jump upon layoff is a reward for telling the truth, abstaining the recently unemployed from claiming disability prematurely.

Figure 3.1 shows examples of consumption profiles for two agents experiencing different employment histories. The solid line represents the more unlucky of the two. He is employed for some 40 periods, gets laid off, and finds a new job almost immediately. The planner then awards him for the job finding, so his consumption is higher during his second employment spell than in his first. However, come another 50 periods of job tenure, he is laid off to experience a relatively long spell of unemployment. During this unemployment spell, consumption falls and when he finally finds a new job, unemployment has deteriorated his promised value for employment to the extent that consumption as employed is lower than during his last employment spell. During the almost 10 years we observe in this figure, he further experiences one more long spell of unemployment and three shorter spells.

The agent represented by the dashed line is more lucky. He experiences no spell of unemployment long enough to depress consumption enough so that it is lower when he finds a new job as compared to his last employment spell. Instead he has several short spells that increase his consumption as employed when he returns to employment.

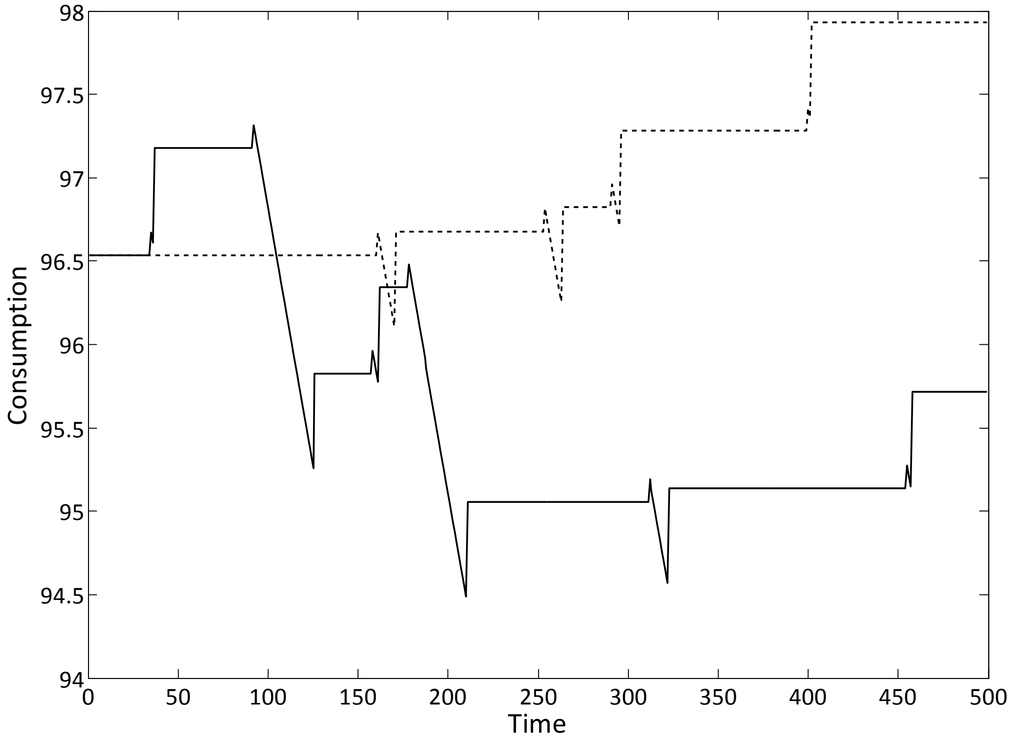

Figure 3.1. Examples of Consumption Dynamics.

Figure 3.1 starkly illustrates history dependence in consumption. As these agents experience different histories, consumption is affected in all future states. As a consequence, they consume differently in both employment and unemployment depending on their history. Moreover, as history evolves, the difference between the lucky and the unlucky agent's consumption grows. The growing inequality over time illustrates the planner's way of providing incentives to an unemployed. For every period an agent is unemployed, consumption in all future states is reduced to provide incentives to avoid prolonged unemployment.

Figure 3.2 depicts the consumption profiles for two agents who eventually become disabled. The solid line shows consumption for an agent who experiences one long and several short spells of unemployment before turning disabled after slightly less than 4 years. Upon becoming disabled, consumption drops substantially, from 97 to around 60 units of the single good, or by almost 40 percent. The dashed line depicts

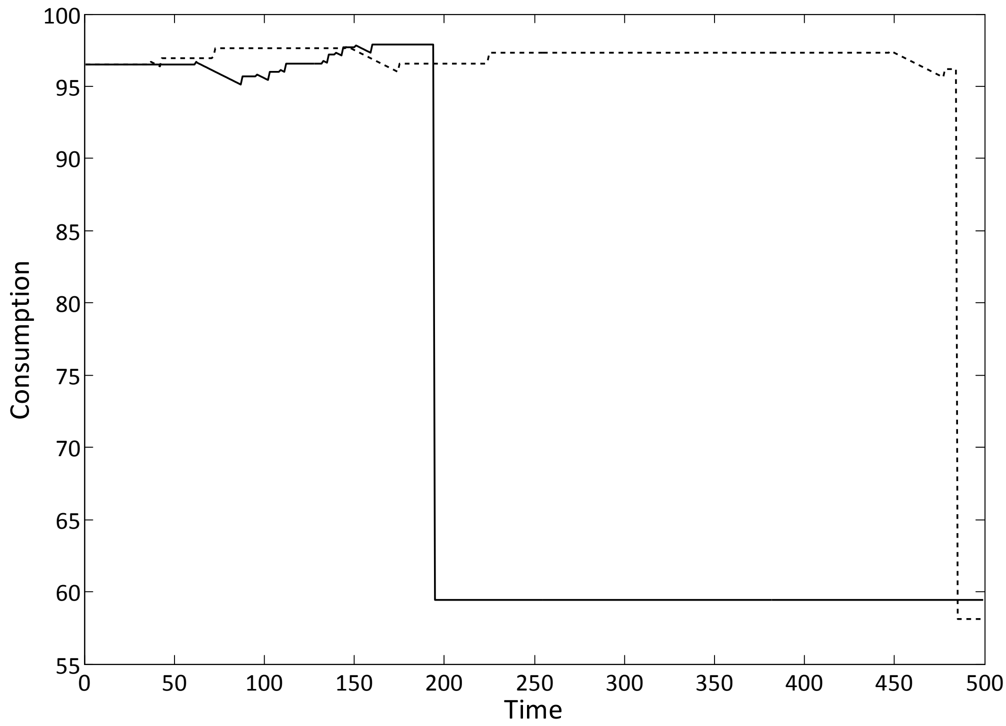

FIGURE 3.2. CONSUMPTION FOR THE DISABLED.

consumption for an agent who has a longer history of employment and unemployment before becoming disabled. Despite becoming disabled much later than the other agent, and having more periods of employment in his history prior to disability, he is awarded lower benefits upon turning disabled. This is in contrast to Diamond and Mirrlees's (1978) proposition that disability insurance benefits should be higher the later an agent becomes disabled. In Höglin (2008), there are two opposing forces that makes it analytically unclear if disability benefits should be higher or lower for an agent becoming disabled early as opposed to late. In this particular example, the agent becoming disabled early is awarded higher benefits than the one getting disabled late. A more general analysis of this property follows in subsection 3.6.3.

3.6.2. Disability's effect on the optimal unemployment insurance. Comparing the qualitative results in Höglin (2008) to those in Hopenhayn and Nicolini (1997, 2004) reveals only one difference: the above 100 percent replacement rate found optimal in Höglin's environment with risk of premature disability claims. This is not to say that the systems are identical. There might be quantitative differences that are substantial. To figure out if this is the case, table 3.1 compares this model's unemployment insurance to Hopenhayn and Nicolini (1997) as well as to Hopenhayn and Nicolini (2004). Unemployment insurance benefits rise upon layoff, becoming only slightly above 100 percent of consumption prior to layoff. During the unemployment spell, consumption decreases at a slow and relatively constant rate. Unemployment benefits becomes lower than consumption prior to layoff after about one month. As far as a year into the unemployment spell, benefits are almost 97 percent of consumption prior to layoff.

The optimal unemployment benefits during a spell are not that different from neither Hopenhayn and Nicolini (1997) nor Hopenhayn and Nicolini (2004). The difference between Hopenhayn and Nicolini (2004), that is, the third column of table 3.1, and this model, as represented in the first column of table 3.1, is that the probability of becoming disabled is zero in the former. Without disability, the initial unemployment insurance benefit level is exactly equal to consumption prior to layoff, as proved in Hopenhayn and Nicolini (2004) and discussed in Höglin (2008). The upward jump upon layoff that appears in this paper is a reward for telling the truth, abstaining the recently unemployed from claiming disability prematurely. As is obvious from inspection of column 1 of table 3.1, the truth telling award is not quantitatively substantial, being only .14 percent higher than the model without disability. Other than this small level effect, unemployment benefits within a spell follows the same pattern as a model without disability closely. Benefits decline at approximately the same rate, and the level of the benefits is higher than in the current system, even for long lasting unemployment spells.

There are also relatively small differences compared to Hopenhayn and Nicolini's (1997) one spell model. In column 2 of table 3.1, Hopenhayn and Nicolini's (1997) simulation results are reproduced. The level is lower than in Hopenhayn and Nicolini (2004), but quantitatively, the differences are small. Also in this model, benefits are high and decline slowly.

TABLE 3.1. OPTIMAL UNEMPLOYMENT INSURANCE

	Replacement rate			
Weeks of unemployment	Model	H&N (1997)	H&N(2004)	Balanced Budget
1	100.14	99.0	100.00	100.14
2	100.08	98.9	99.93	100.08
3	100.01	98.8	99.87	100.01
4	99.94	98.7	99.80	99.94
5	99.88	98.6	99.74	99.88
6	99.81	98.5	99.67	99.81
7	99.75	98.4	99.61	99.74
8	99.68	98.3	99.54	99.67
12	99.42	97.9	99.28	99.41
16	99.16	97.5	99.02	99.14
26	98.51	96.5	98.37	98.47
52	96.82	94.0	96.68	96.72

Finally, the 4th column of table 3.1 depicts the optimal unemployment insurance benefits for a balanced budget system.[3] These results suggest that the evolution of unemployment benefits is not sensitive to the way the current system is calculated: until the 8th week of unemployment, the replacement rate is smaller than visible in the second decimal place. For longer spells, the systems differ, but the differences are minuscule.

The main message from this section is thus that unemployment insurance is not quantitatively affected by the risk of premature disability claims to a large extent.

3.6.3. Unemployment's effect on optimal disability benefits. An important implication of Höglin (2008) is that unemployment and employment spells affect disability benefits in a non-trivial way. An agent's disability benefits are reduced for each period he is unemployed and increased for each time he finds a new job. Every period an agent is unemployed, his promised value in all future states should be reduced to provide incentives for job search. Therefore, disability insurance benefits decrease during an unemployment spell. If an agent finds a job however, he should be

[3] A balanced budget system delivers V^e_{BB} to a newborn agent, where V^e_{BB} is the solution to $C_e\left(V^e_{BB}\right) = 0$.

rewarded in all future states, which implies higher future disability benefits. These two opposing forces makes it impossible to analytically prove if disability benefits increase or decrease with periods of past unemployment. It suggests however, that an agent's disability benefits are higher if he has many short spells compared to few long spells.

Figure 3.3 reveals that this conjecture is true, by plotting multiples of the total number of weeks as unemployed before turning disabled and the length of each unemployment spell. For a given number of periods as unemployed before turning disabled, disability benefits fall with increased length of the unemployment spells. An agent with 80 weeks as unemployed before turning disabled receives 5.5 percent higher benefits if his spell length is 10 weeks compared to 20 weeks. Moreover, an agent with the same 80 weeks of unemployment but with shorter than average spell length of 5 weeks get more than 12 percent higher benefits than the agent whose spells last for 20 weeks. The total number of weeks as unemployed is thus a poor proxy for the relevant part of an agents unemployment history that should determine the optimal disability benefits. For a given spell length, agents with more periods as unemployed get lower benefits only if the spells are relatively long lasting. For spells shorter than 12 weeks, an agent gets higher disability benefits the more periods as unemployed he has experienced before turning disabled. Note that 12 weeks is the breaking point for all levels of total weeks in unemployment, and that is constitutes a spell length that is only slightly above the average of 10 weeks.

3.7. The Optimal Distribution of Consumption

In order for a law of large numbers to apply approximately, I simulate a population of 40,000 individuals for 30,000 periods. All agents enter the model as employed and holding the value of an employed agent, as calculated using the current system in section 3.5. When an agent dies between two periods, he in reborn as employed with the value of an employed in the current system.

Let us first study the distribution of consumption and compare it to the consumption distributions implied by the current system and by autarky. Table 3.2 depicts

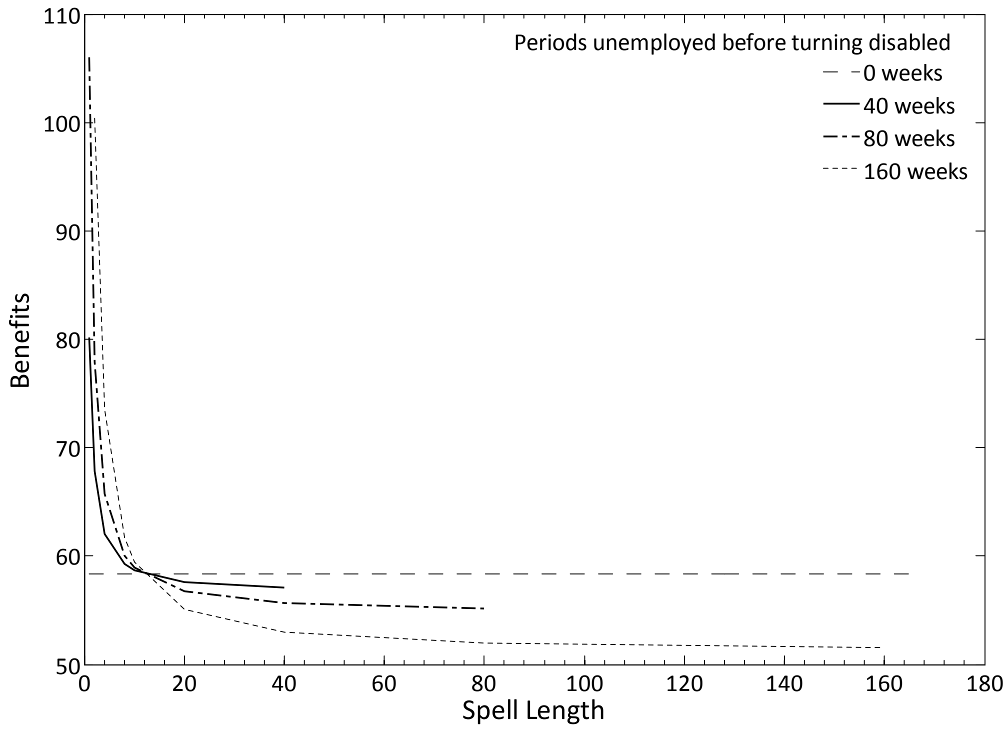

FIGURE 3.3. DISABILITY BENEFITS FOR DIFFERENT UNEMPLOYMENT HISTORIES.

moments from these three distributions.[4] The average consumption is higher in the optimal system than in the current one and in autarky.[5] Average consumption being higher in the optimal system than in autarky is unsurprising, as the optimal system delivers the value of the current system which has a positive budget (see the next section). The reason why the optimal system manages to deliver higher average consumption than the current system is that when an agent enters the economy as employed, the planner collects some of his earnings and stores it at the gross interest rate β^{-1}, while the current system implies that all earnings are consumed when employed. This implies that in the steady state, the planner has built up a stock of resources to spend on the

[4] The distribution under the current system is calculated taking into account that some unemployed workers are not eligible for unemployment insurance. Let $U(n)$ be the measure of unemployed with current unemployement duration of n weeks and E be the steady state measure of employed agents. Then, $U(n) = \frac{1-\lambda}{1-p} E\left[(1-\xi)(1-\psi)(1-p)\right]^n$, and the measure of non-eligible unemployed is given by $\sum_{n=26}^{\infty} U(n)$, which amounts to 149 workers, or .037 percent of the population.

[5] Autarky refers to a situation in which the agents cannot trade with each other and in which it does not exist any social insurance system at all. For details, see Höglin (2008).

TABLE 3.2. CONSUMPTION.

	Optimal	Current	Autarky
Mean	96.65	95.40	90.32
Std. Dev.	8.08	14.94	29.57
Min	53.76	0	0
Max	120.32	100	100
Min/Max	.45	0	0

unemployed and disabled, states which are more prevalent in the steady state than when he system starts with only employed. To appreciate this, consider the following example of an 'average' agent. He is born employed and remain so for approximately 3.3 years, or 172 weeks. During these 172 weeks, he consumes 96.18 units of consumption per period in the optimal system, but 100 units per period in the current system. Thus, when turning unemployed in his 173rd week of life, the planner has saved $\sum_{j=0}^{171} \beta^{-j} (100 - 96.18) = 745.76$ units of consumption, or approximately 7.5 weekly wages, on him. The virtue of the optimal system is that it accomplishes more consumption smoothing than the current system. In other words, consumption is higher during unemployment and disability in the optimal system compared to the current system. As unemployment and disability are more prevalent in the steady state distribution over states than in the initial distribution (in which everyone is employed), the planner has more resources available than the current system in the steady state.

As indicated above, the optimal system indeed features much more consumption smoothing across histories than the current and the autarky system. The cross sectional standard deviation in the optimal system is almost half of that in the current system and less than a third of the autarky standard deviation.

Although the optimal system features lower cross sectional inequality than the current system, moral hazard still limits consumption smoothing significantly. A few unlucky workers have markedly lower consumption than others. To see the details, figure 3.4 depicts the consumption distribution for the optimal system. A relatively small mass of agents consume around 60 units of the single good whereas the bulk of the mass lies around 100. The distribution is also relatively unskewed. Finally, the

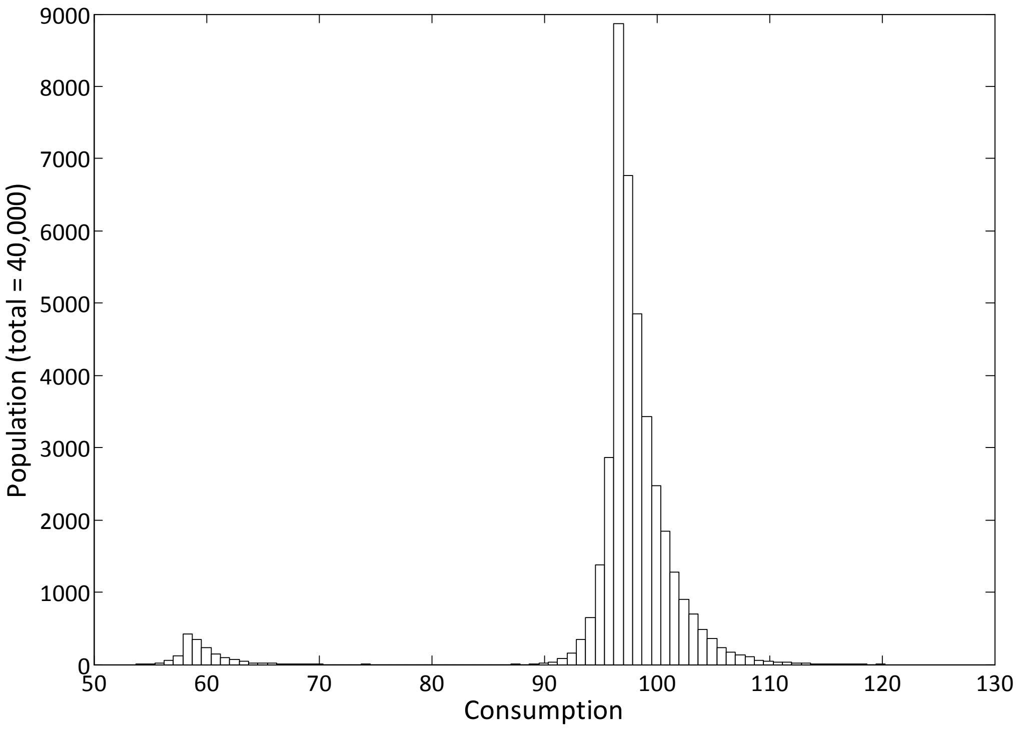

FIGURE 3.4. THE DISTRIBUTION OF CONSUMPTION.

spike slightly below 100 represents the consumption of newborn and those that has never experienced unemployment or disability since their birth.

To figure out the differences between agents in different states, I produce separate distributions for employed, unemployed and disabled. The moments of these distributions are depicted in table 3.3. The average consumption is similar between employed and unemployed. Moreover, the variation of these distributions are very similar, showing no divergence even in minimum or maximum values. In this sample, disabled consume less than employed or unemployed, irrespective of their history before turning disabled. The cross sectional maximum consumption for a disabled is more than 15 percent lower than the minimum among the unemployed.

Table 3.3 Consumption by State.

	Employed	Unemployed	Disabled
Mean	98.25	97.85	59.67
Std. Dev.	2.72	2.82	2.04
Min	87.27	89.49	53.76
Max	120.32	119.98	73.75
Min/Max	.73	.75	.73

Indeed, the disabled agents' consumption distribution is disjoint from the other types, as shown in figure 3.5. Figure 3.5 also displays the distributions for all different types. The distribution for employed and unemployed are almost completely overlapping. Both are only sightly skewed to the left and has the bulk of the mass in a small range, as suggested by the low standard deviation. The disabled agents' consumption distribution is also concentrated at a fairly small interval.

3.8. Cost Reductions from the Optimal System

The features of the optimal contract displayed so far does not pin down the welfare gains from moving to the optimal system from the current one. The reason is of course that the agents experience the same expected utility in both systems. The welfare gains are instead seen when computing the costs of delivering the value of the current system within the current system as compared to delivering the same value optimally.

3.8.1. Cost of the current system. The costs of the current system can be calculated in a similar way as the value of the current system was calculated. Assume that the agent gets 66 percent of his pre unemployment *wage* for 26 weeks of unemployment and nothing thereafter. If becoming disabled, the agent is awarded 40 percent of his wage when employed. In addition, assume that no history dependent taxes upon employment are used in the current system. Finally, once an agent is employed, his value if becoming laid off is the same regardless of his history.

It is convenient to start with a disabled agent. The cost to the current system of a disabled agent is

$$\widetilde{C}_d = \frac{40}{1-\beta}$$

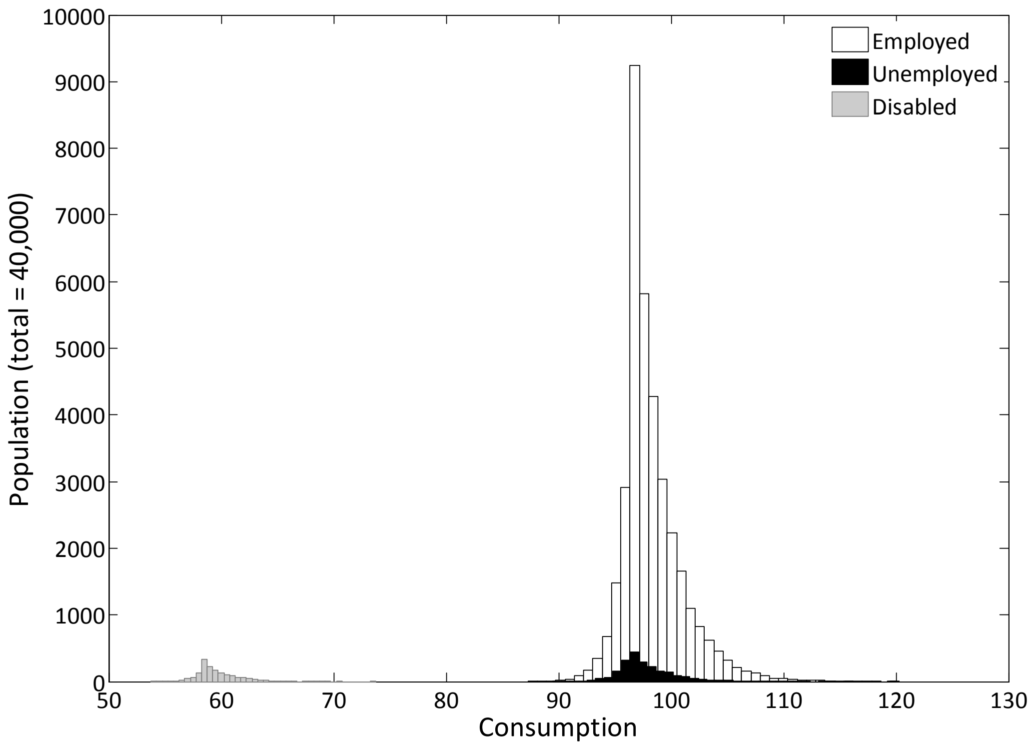

FIGURE 3.5. CONSUMPTION DISTRIBUTION BY STATE.

Consider next an unemployed agent and denote by $\widetilde{C}_u(n)$ the cost to the current system of an unemployed agent with n periods of unemployment insurance benefits left. An unemployed agent that has exhausted his unemployment insurance benefits costs

$$\widetilde{C}_u(0) = 0 + \beta \left\{ (1-\xi) \left[p\widetilde{C}_e + (1-p)\widetilde{C}_u(0) \right] + \xi \widetilde{C}_d \right\},$$

where $\widetilde{C}_e$ is the cost of an employed agent. For $n \in \{1, 2, ...26\}$,

$$\widetilde{C}_u(n) = 66 + \beta \left\{ (1-\xi) \left[p\widetilde{C}_e + (1-p)\widetilde{C}_u(n-1) \right] + \xi \widetilde{C}_d \right\}.$$

Finally, the cost of an employed agent is given by

$$\widetilde{C}_e = 100 - w + \beta \left\{ (1-\xi) \left[\lambda\widetilde{C}_e + (1-\lambda)\widetilde{C}_u(26) \right] + \xi \widetilde{C}_d \right\}.$$

3.8.2. Quantitative cost reductions. To be able to calculate the cost reductions, we need to obtain the cost in the optimal system of delivering the value of the current system. This number is simply given by evaluating the minimal cost function at the value of the current system, $C_e\left(\widetilde{V}^e\right)$. The cost reductions from moving to the optimal system is then $\widetilde{C}_e - C_e\left(\widetilde{V}^e\right)$, and measures the planners excess ability of delivering utils as compared to the current system.

The costs of the current and the optimal system are depicted in table 3.4. The cost of delivering $\widetilde{V}^e$ within the current system equals approximately one half of a yearly wage. The cost of delivering $\widetilde{V}^e$ within the optimal system is substantially lower. The cost reductions from moving to the optimal system amounts to approximately one monthly wage, or 15 percent of the cost of the current system.

3.9. What Enhances Efficiency?

There are, in principle, two ways in which the planner might improve upon the current system. First, the optimal system may provide better incentives to exert search effort and tell the truth, so that unemployment or disability is too high in the current system. The other way is that the planner might achieve better consumption smoothing without ruining the incentives to act in a socially optimal way. In this environment with a calibration that implies that unemployed search and tell the truth even within the current system, increased consumption smoothing is responsible for all of the welfare gain. Optimality of search and truth telling within the current system implies that the aggregate and individual dynamics over states are the same in the current and the optimal system, so no improvement in behavior is achieved.

The calculated welfare gains from moving to the optimal system thus constitute a lower bound on the welfare gains. To the extent that variations along an intensive margin of job search is important, the calculated welfare gains of the optimal system could be even higher in such an environment. Previous research suggest however that variations in job search activity is not as important as one may think. Hopenhayn and Nicolini (1997) report that even in their model with a continuum of effort choices, there are very small differences in job search activity between the current and the optimal system. Thus, even in such an environment, the bulk of the welfare gains comes from increased consumption smoothing.

Table 3.4. Cost Reductions.

	Optimal	Current
Cost	2244	2650
Savings	15%	-

3.10. Optimal Insurance versus Self Insurance

As is standard in this literature, I have assumed that the agents lack access to any market for borrowing and lending. This is not as restrictive an assumption as one may think. As long as savings are observable, i.e. as long as the planner can control the agents' consumption, the consumption allocations chosen by the planner is invariant to the alternative assumption that the agents can save. If agents can save, consumption and benefits are not the same. Shimer and Werning (2008) has recently highlighted the distinction between providing liquidity and providing insurance that arises when agents can save.[6] I make no attempt of distinguishing between insurance and liquidity in this paper, but focus on the effectiveness of self insurance.

If agents are allowed to save but not borrow, they can build up buffer stocks during employment and dissave when unemployed and disabled. This buffer stock is a substitute for the planner. Self insurance through non state contingent claims has an advantage over the planner as there is obviously no moral hazard between the agent and himself. The planner has one advantage over self insurance though: the ability to lend. A natural question becomes how effective self insurance through a riskless bond is relative to the planner.

In this section, I ask how effective self insurance is relative to optimal insurance by solving the agents' problem when they can save, but not borrow, at a gross rate equal to β^{-1} (the same rate at which the planner borrows and saves at) and comparing that solution to the planner's problem when his budget is balanced. An indirect measure of

[6] Shimer and Werning (2008) proves that if agents have constant absolute risk aversion there is an equivalence between constant benefits and optimal benefits. If the agents' have constant relative risk aversion utility functions, as in this paper, there is 'almost' an equivalence in their numerical examples. Furthermore, Rendahl (2007) shows that in an environment like Hopenhayn and Nicolini (1997) with observed savings, an agent's assets encodes all the relevant parts of his history so that the optimal unemployment insurance lacks explicit duration dependence and is recursive in the agent's wealth.

the planners ability is then the answer to the question: what is the wealth level such that a newborn agent is indifferent between living in the planner's system and living with self insurance through a riskless bond only?

Consider an employed agent with bond wealth $b \geq 0$. The Bellman equation for such an agent is given by

$$V_{SI}^{e}(b) = \max_{c,b'} \left\{ u(c) + \beta \left\{ (1-\xi) \left[\lambda V_{SI}^{e}(b') + (1-\lambda) V_{SI}^{u}(b') \right] + \xi V_{SI}^{d}(b') \right\} \right\}, \quad (3.12)$$

where $V_{SI}^{u}(\cdot)$ and $V_{SI}^{d}(\cdot)$ are value functions if unemployed or disabled and where the maximization is subject to a budget constraint and a borrowing constraint,

$$c + b' = w + \beta^{-1} b, \quad (3.13)$$

$$b' \geq 0. \quad (3.14)$$

Similarly, the value function of an unemployed obeys the Bellman equation

$$V_{SI}^{u}(b) = \max_{c,b'} \left\{ u(c) + \beta \left\{ (1-\xi) \left[p V_{SI}^{e}(b') + (1-p) V_{SI}^{u}(b') \right] + \xi V_{SI}^{d}(b') \right\} \right\}. \quad (3.15)$$

Subject to

$$c + b' = \beta^{-1} b, \quad (3.16)$$

$$b' \geq 0. \quad (3.17)$$

In (3.15), I have implicitly assumed that the agent always exerts effort if unemployed. This is not always true. There are wealth levels high enough so that an unemployed would choose not to search, as the cost of searching relative to not searching increases in the level of wealth.[7] However, for the present calibration, this situation never occurs for the levels of wealth I am interested in. Finally, if disabled, the Bellman equation becomes

$$V_{SI}^{d}(b) = \max_{c,b'} \left\{ u(c) + \beta V_{SI}^{d}(b') \right\}, \quad (3.18)$$

where, again, the maximization is subject to

$$c + b' = \beta^{-1} b, \quad (3.19)$$

[7] Recall that the necessary condition for exerting search effort in this case would be $\beta(1-\xi) p \left[V_{SI}^{e}(b') - V_{SI}^{u}(b') \right] \geq a$. Naturally, when b' becomes very large, labor status is less important and the agent is well insured against any shock, so that $\left[V_{SI}^{e}(b') - V_{SI}^{u}(b') \right]$ is small.

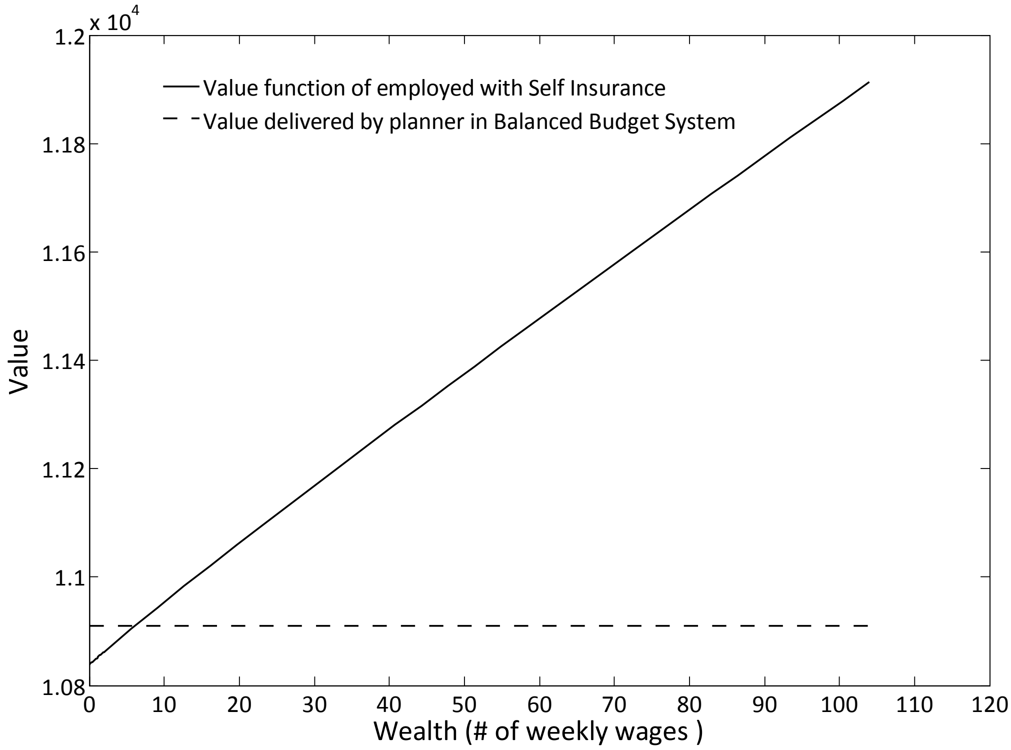

Figure 3.6. The Effectiveness of Self Insurance.

$$b' \geq 0. \tag{3.20}$$

To obtain a measure of the effectiveness of self insurance, I will solve the problems above and calculate the wealth $\bar{b}$ that satisfies

$$V_{SI}^{e}\left(\bar{b}\right) = V_{BB}^{e},$$

where V_{BB}^{e} is the solution to $C_e\left(V_{BB}^{e}\right) = 0$, and indicates the value the planner is able to deliver under the restriction of a balanced budget.

Figure 3.6 shows the value function for a self insuring employed agent, $V_{SI}^{e}\left(\cdot\right)$, along with the value of living in the planner's balanced budget system. Evidently, self insurance is rather effective. For wealth levels as low as 6 weekly wages, the planner does no better in a balanced budget system than the agent does on his own with access to a riskless bond.

3.11. Conclusions

Unemployment insurance and disability insurance are naturally related to each other and intricately so in the optimal joint system of unemployment and disability insurance. The quantitative impact of disability on the optimal unemployment insurance is not substantial. Benefit profiles resemble those in the literature that handles unemployment insurance in isolation, notably Hopenhayn and Nicolini (1997, 2004).

The agents' unemployment history is however an important part of the history when determining the optimal disability benefits. Agents who have suffered long lasting spells of unemployment receive lower disability benefits than those with no or short spells prior to becoming disabled. These effects are substantial, and may constitute consumption differences in excess of 10 percent for plausible differences in the unemployment history.

In general, the optimal system of insurance of unemployment and disability shocks achieves a lot more consumption smoothing than the current system. Due to this fact, the cost reductions from moving to the optimal system are large, amounting to 15 percent of the cost of the current system.

Still, self insurance is not a bad substitute for the planner, as the value of being in the optimal system is not larger than the value under self insurance with a wealth level equal to 6 weekly wages.

The welfare gains and the effectiveness of self insurance calculated in this paper hinge on the discrete effort choice. A natural extension would be to allow for a continuum of effort levels with a function mapping efforts to probabilities of successful job search, as in Hopenhayn and Nicolini (1997). Whether the insights of this paper are altered by this generalization is an open question. Regarding the cost reductions, a conjecture is that the differences would not be large as more efficient job search is only a minor part of the calculated cost reductions in Hopenhayn and Nicolini (1997).

References

Autor, David H. and Mark G. Duggan. (2003), "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics* **118**, pp. 157-202.

Carling, Kenneth, Bertil Holmlund and Altin Vesiju. (2001). "Do benefit cuts boost job-findings? Swedish evidence from the 1990s", *Economic Journal* **111**, pp. 766-790.

Diamond, Peter A. and James A. Mirrlees. (1978), "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics* **10**, pp. 295-336.

Feldstein, Martin. (2005). "Rethinking Social Insurance," *American Economic Review* **95**, pp 1-24.

Golosov, Mikhail and Aleh Tsyvinski. (2006), "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy* **114**, pp. 257-269.

Gruber, Jonathan. (1997) "The Consumption Smoothing Benefits of Unemployment Insurance," *American Economic Review* **87**, pp.192-205.

Gruber, Jonathan. (2000), "Disability Benefits and Labor Supply," *Journal of Political Economy* **108**, pp. 1162-1183.

Hopenhayn, Hugo A. and Juan Pablo Nicolini. (1997),"Optimal Unemployment Insurance," *Journal of Political Economy* **105**, pp. 413-438.

Hopenhayn, Hugo A. and Juan Pablo Nicolini. (2004), "Optimal Unemployment Insurance and Employment History," *mimeo.* Universidad Torquato Di Tella.

Höglin, Erik. (2008). “Optimal Unemployment and Disability Insurance,” chapter 2 of this volume.

Katz and Meyer. (1990). “The impact of the potential duration of unemployment benefits on the duration of unemployment,” *Journal of Public Economics* **41**, pp. 45-72.

Kocherlakota, Narayana R. (2004). “Figuring out the Impact of Hidden Savings on Optimal Unemployment Insurance,” *Review of Economic Dynamics* **7**, pp. 541-554.

Ljungqvist, Lars and Thomas J. Sargent. (1998) “The European Unemployment Dilemma,” *Journal of Political Economy* **106**, pp. 514-550.

Lucas, Robert E. Jr. (1987). *Models of Business Cycles*, Oxford: Basil Blackwell.

Meyer, Bruce D. (1990). “Unemployment Insurance and Unemployment Spells,” *Econometrica* **58,** pp. 757-782.

Meyer, Bruce D. and Wallace K. C. Mok (2006). “Disability, Earnings, Income and Consumption” *mimeo.* University of Chicago.

Mitchell, Matthew and Yuzhe Zhang (2007). “Unemployment Insurance with Hidden Savings,” *mimeo.* University of Toronto.

Pavoni, Nicola and Giovanni L. Violante. (2007). “Optimal Welfare-to-Work Programs,” *Review of Economic Studies* **74**, pp. 283-318.

Rendahl, Pontus. (2007). “Unemployment Insurance and Credit Frictions,” *mimeo.* University of Amsterdam.

Shavell, Steven. and Laurence Weiss. (1979), “The Optimal Payment of Unemployment Insurance Benefits Over Time,” *Journal of Political Economy* **87**, pp. 1347-1362.

Shimer, Robert and Ivan Werning. (2008). "Liquidity and Insurance for the Unemployed," forthcoming in *the American Economic Review*.

Werning, Ivan. (2002), "Optimal Unemployment with Unobservable Savings," *mimeo*. University of Chicago and MIT.

Unions

CHAPTER 4

Unions, Turbulence, and Unemployment

Erik Höglin

Abstract. This paper uses a calibrated version of a dynamic monopoly union model to argue that a monopoly union increases the rate of unemployment as a response to increased microeconomic turbulence faced by its members. The quantitative implications of the model matches the European aggregate time series of unemployment, both during the transition and in the steady states. The union prefers higher unemployment in turbulent times both because such times implies increased opportunities for rent seeking and as unemployment leads to wage compression which insures the workers from the increased wage volatility associated with turbulence. The welfare consequences of union wage setting is positive for workers and negative for firm owners. In total, union wage setting reduces welfare substantially in turbulent times.

4.1. Introduction

During the last decades, Europe and the United States have taken different routes in important labor market dimensions. While U.S. wage inequality has increased substantially, unemployment rates have remained roughly constant. The increase in European inequality is much less pronounced, and unemployment rates have instead increased to levels far above those of the United States.

Krugman (1994) hypothesized that the different labor market experiences between the two continents can be explained by institutional differences. The flexible U.S. labor market simply absorbed the pressures from technical change and globalization by increased inequality. In Europe, less flexible labor markets handled the pressure towards increased inequality by letting a large fraction of its workers be unemployed, thereby maintaining wage equality.

[0] I have benefited from discussions with Tore Ellingsen, Martin Flodén, Lars Ljungqvist, and Henrik Lundvall. Financial support from Jan Wallander and Tom Hedelius foundation is gratefully acknowledged.

One important institutional difference between the United States and Europe is the prevalence of collective bargaining. In this paper we study if Krugman's hypothesis is valid when the inflexibility in the European labor markets is represented by union wage setting. To this end, we study a monopoly union in a dynamic environment where its members are confronted with increased microeconomic volatility, or turbulence.

We focus on turbulence affecting *ex ante* identical workers as inequality within groups has increased substantially in the last decades. For example, Juhn, Murphy and Pierce (1993) show that the more than half of the increase in U.S. wage inequality pertains to an increase in the residual inequality, i.e. to inequality among observationally identical workers. Gottschalk and Moffit (1994) decompose the earnings inequality further to conclude that a sizable fraction of its rise is transitory, reflecting increased time series variance of individual workers' wages. Looking even deeper into the data, Violante (2002) documents that wage loss from displacement as well as the yearly return to tenure increased from the 1970s to the 1980s. We take the results from these studies and feed them into a monopoly union model, solve the union's wage setting problem and simulate a synthetic 'Europe' that we can compare to aggregate time series data.

The first contribution of this paper is to show that a monopoly union indeed increases the rate of unemployment as a response to increased microeconomic turbulence faced by its members. The quantitative implications of the model also matches the European aggregate time series of unemployment reasonably well, both during the transition and in the steady states.

The second contribution is to show that this effect has two different channels, and that one of these channels is more important than the other. The union creates unemployment either for rent seeking or insurance reasons. With a fixed factor of production which implies decreasing returns to scale, withdrawing workers from employment raises the wages of employed workers. The incidence of the reduced production associated with the reduced supply of labor falls partly on the firms. The union effectively consumes a part of the firms' profits, why we call it the rent seeking effect. This type of union rent seeking has been studied extensively before, see, for example, McDonald and Solow (1981).

Agell and Lommerud (1991) showed that unemployment also can arise as an insurance device in a world with uneven income and imperfect capital markets. Risk averse workers achieve low variance of wages over time and across states by unemployment, which implies wage compression. When turbulence increases, this insurance effect naturally gets more important. However, the rent seeking effect also gets more important, and contributes more to the increased aggregate unemployment than the insurance effect. Indeed, the rent seeking effect is imperative for the model's general performance. Without the possibility to extract rents from the firms, the union's incentives to create unemployment is dampened, and its possibility to affect the wages vanishes in the long run. As a result of this, unemployment does not exist in the steady state without rent seeking. This result is reminiscent the analysis of hours worked restrictions by Marimon and Zilibotti (2000). They showed that with a fixed factor of production, which implies decreasing returns to scale (or rent seeking in the language of this paper), an hours worked restriction can be welfare improving whereas if factors are flexible, and the production function exhibits constant returns to scale, it cannot.

Krugman (1994) emphasized the importance of the welfare states being larger in Europe. The welfare state is indeed an important factor in our model. If replacement rates during unemployment are low, unions create very little unemployment. The interaction between the welfare state and union wage setting is thus an important one in this model. This is not to say that there is no role for the union absent the welfare state, but that, in the long run, those effects vanishes. During a transition to a steady state, the union could still optimally compress wages at the benefit of their members. The transition would then be postponed, but convergence to the laissez-faire steady state would still occur.

The conclusion is that the story of unions being the cause of the rise in the European unemployment is consistent with a world where firms and unions bargain over rents in welfare states where turbulence have increased.

We are not the first to argue that unions caused the rise in and persistence of European unemployment. Blanchard and Summers (1986), Gottfries and Horn (1986), and Lindbeck and Snower (1988) develop insider-outsider models of unions to explain the persistence in European unemployment. Confronted with a negative macroeconomic shock, insiders prefer to maintain high wages and increase unemployment, because unemployment primarily hurts outsiders. As unemployment reduces the number of future insiders, this effect becomes persistent, causing 'hysteresis' in the European labor market. These papers focus on a single macroeconomic shock, such as the 1970's oil crises, while in this paper we study a permanent change in the economic environment. Moreover, in this paper all workers belong to the union and there is no conflict of interest between workers with identical skills.

We also extend the analysis by studying the welfare consequences of union wage setting. This is motivated by Agell (1999), who argues that unionization may enhance welfare in a world of incomplete markets, and Lindquist (2005), who found large welfare losses from union wage setting in a slightly different environment. In the environment of this paper, workers benefit from union wage setting, but firm owners suffer significant reductions in profits. In the aggregate, the welfare costs of union wage setting are substantial in turbulent times but negligible in tranquil times.

Technical change has also been argued to cause deunionization. Acemuglu, Aghion and Violante (2001) ask if skill biased technical change caused the deunionization that occurred the early 1980s. In their model, increased inequality causes deunionization because the outside option of high skilled workers increase, luring them to leave the union. Although we do not directly address deunionization, this is arguably nonrestrictive in the environment of this paper. As each type of worker benefits from the presence of union wage setting, deunionization would not occur in this environment.

Alesina, Glaeser and Sacerdote (2005) studies differences in hours worked between the United States and Europe using a union wage setting model like the one of this paper, but with a special twist including an externality in the form of a social multiplier that makes people enjoy leisure more if their neighbor reduces his labor supply.

Their static model does not, however, address the question of what role the increased inequality has played for the intercontinental differences in labor market experience.

Finally, this paper is also related to the series of papers on the European unemployment problem by Ljungqvist and Sargent (e.g. 1998, 2004, 2007a, 2007b, 2008). Ljungqvist and Sargent argues that Europe's welfare states are more vulnerable to increasing microeconomic turbulence than the Unites States. When turbulence increases, European workers optimally choose longer spells of unemployment as their *past* human capital is protected by generous unemployment benefits. The aggregate consequence is higher unemployment rates in Europe than in the United States. Ljungqvist and Sargent use McCall (1970) search models, search island models in the spirit of Lucas and Prescott (1974), matching models along the lines of Diamond (1982), Mortensen (1982) and Pissarides (1990), and representative family models inspired by the aggregation theories of Hansen (1985) and Rogerson (1988), but have no unions in their model economies. Their argument goes through also in the environment of this paper in the following sense: unions choose higher unemployment rates as a response to turbulence only if the value of unemployment is not too low. In other words, unions increasing unemployment in response to turbulence is more likely in welfare states.

The next section presents the model, the calibration, and some steady state properties of the model. In section 4.3, we perform an experiment where we compare the union's strategies in tranquil and turbulent times. Section 4.4 explains the economic forces at work, contrasting a rent seeking union with a model in which the union's motives are purely to insure its members. We then study what role the welfare state plays in section 4.5. In section 4.6, we perform welfare analysis. Section 4.7 concludes.

4.2. The Model

4.2.1. The economy. There are many identical price taking firms producing a single consumption good using capital and labor as their inputs. The representative firm operates a production function given by

$$F(K,H,L) = K^{\alpha}\left[\psi H^{\rho} + (1-\psi)L^{\rho}\right]^{\frac{1-\alpha}{\rho}}, \tag{4.1}$$

where K, H and L are the amounts of capital and high and low skilled workers employed, respectively, α is capital's income share of output, and $1/(1-\rho)$, with $\rho < 1$, is the constant elasticity of substitution between high and low skilled labor. For most of the analysis we will assume that capital is fixed, and normalize $K \equiv 1$. The representative firm then takes wages as given and chooses H and L to maximize profits, as given by

$$\widehat{F}(H, L) - w_h H - w_l L,$$

where

$$\widehat{F}(H, L) = [\psi H^\rho + (1-\psi) L^\rho]^{\frac{1-\alpha}{\rho}}.$$

The solution to this program is fully characterized by the first order conditions,

$$(1-\alpha)[\psi H^\rho + (1-\psi) L^\rho]^{\frac{1-\alpha-\rho}{\rho}} \psi H^{\rho-1} = w_h, \tag{4.2}$$

and

$$(1-\alpha)[\psi H^\rho + (1-\psi) L^\rho]^{\frac{1-\alpha-\rho}{\rho}} (1-\psi) L^{\rho-1} = w_l. \tag{4.3}$$

There is a continuum of workers of unit mass. A worker can be either high or low skilled at each point in time, and workers gain and lose skills stochastically depending on their labor market experience. An employed worker of skill type $j \in \{h, l\}$ remains in his skill group with probability π_j. Skill gain is impossible during unemployment, and high skilled workers become low skilled if laid off. We thus interpret $1 - \pi_h$ as the exogenous probability that a high skilled worker is laid off, and assume that skill loss occurs with probability one in that case. The workers have preferences ordered by a constant relative risk aversion utility function

$$\sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma}}{1-\sigma}. \tag{4.4}$$

There are no capital or insurance markets, so consumption equals income in every period. If unemployed, such a worker consumes R, an exogenous income equivalent of the momentary value of unemployment.

All workers belong to a union that has monopoly power in the labor market. The union sets the wages for both skill groups, taking the solution to the firms problem as given. The union is utilitarian, so it wants to maximize the weighted average of the different types of workers utilities. Let χ_t and u_t be the measure of high skilled workers in the beginning of time t and the time t unemployment rate among the low skilled, respectively. The assumption that only low skilled workers can be unemployed is not restrictive, as long as parameters are such that the high skilled earn more than the low skilled. Under such circumstances, the union would not set wages such that high skilled workers become unemployed as it (i) reduces the wage bill more than low skilled unemployment, and (ii) increases wage inequality which is disliked since workers are risk averse. The next subsection describes the parameter restrictions needed for this statement to be true.

In equilibrium, $H_t = \chi_t$ and $L_t = (1-\chi_t)(1-u_t)$. The unions momentary utility function becomes

$$U(\chi, u) = \chi \frac{w_h^{1-\sigma}}{1-\sigma} + (1-\chi)\left[(1-u)\frac{w_l^{1-\sigma}}{1-\sigma} + u\frac{R^{1-\sigma}}{1-\sigma}\right],$$

where w_h and w_l are functions of χ and u according to (4.2), (4.3), and the equilibrium conditions above. Looking at (4.2), (4.3) and the equilibrium conditions, we see that we can formulate the union's problem such that it chooses the rate of unemployment among the low skilled rather than the wages.

The measure of high skilled is an important variable for the union as it determines the underlying relative supply of high and low skilled workers which affects production possibilities and underlying inequality. The measure of high skilled workers obeys the law of motion

$$\chi' = \pi_h \chi + (1-u)(1-\pi_l)(1-\chi). \tag{4.5}$$

The union chooses a sequence of unemployment rates to maximize $\sum_{t=0}^{\infty} \beta^t U(\chi_t, u_t)$, where χ_t is a state variable. Let $J(\chi)$ be the maximum value of the union's problem when the current measure of high skilled workers is χ. The solution of the union's problem then satisfies the Bellman equation

$$J(\chi) = \max_u \left\{ \chi \frac{w_h^{1-\sigma}}{1-\sigma} + (1-\chi)\left[(1-u)\frac{w_l^{1-\sigma}}{1-\sigma} + u\frac{R^{1-\sigma}}{1-\sigma}\right] + \beta J(\chi') \right\}, \tag{4.6}$$

where the maximization is subject to (4.2), (4.3), and (4.5).

4.2.2. Steady state characterization. Before moving to the calibration, some properties of the model's steady state can be characterized analytically. In a steady state where $\chi = \chi' \equiv \overline{\chi}$ and $u = u' \equiv \overline{u}$, the measure of high skilled obeys

$$\frac{\overline{\chi}}{1-\overline{\chi}} = \frac{(1-\overline{u})(1-\pi_l)}{(1-\pi_h)}. \tag{4.7}$$

Hence

$$\overline{\chi} = \frac{(1-\overline{u})(1-\pi_l)}{(1-\pi_h)+(1-\overline{u})(1-\pi_l)}, \tag{4.8}$$

which implies that the higher steady state unemployment, the lower is, endogenously, the steady state measure of high skilled. High unemployment contributes to wage compression and a low share of high skilled to the opposite. Specifically, using (4.2), (4.3) and the equilibrium conditions $H_t = \chi_t$ and $L_t = (1-\chi_t)(1-u_t)$, the wage inequality is given by

$$\phi(\chi, u) = \frac{\psi}{1-\psi}\left(\frac{H}{L}\right)^{\rho-1} = \frac{\psi}{1-\psi}\left(\frac{\chi}{(1-\chi)(1-u)}\right)^{\rho-1}. \tag{4.9}$$

From (4.9), inequality is evidently decreasing in u and χ, as $\rho < 1$. The intuition is straightforward; the more high skilled, the lower are their marginal productivity and the fewer low skilled employed (high u), the higher is the marginal productivity of a low skilled worker. In the steady state, after substituting (4.7), the wage ratio is given by

$$\phi_{ss} = \frac{\psi}{1-\psi}\left(\frac{1-\pi_l}{1-\pi_h}\right)^{\rho-1}. \tag{4.10}$$

Inspection of the expression in (4.10) reveals that, in the steady state, the wage inequality is independent of the unions choice of the rate of unemployment, but instead given by parameters of the production function and the law of motion for the fraction of high skilled. This is intuitive, as our preferred production function implies that the ratio of employed high skilled to employed low skilled determines the current inequality. Moreover, from (4.7) one can easily tell that the ratio of employed high skilled to employed low skilled in the steady state is completely determined by the transition probabilities, and given by

$$\frac{\overline{\chi}}{(1-\overline{\chi})(1-\overline{u})} = \frac{(1-\pi_l)}{(1-\pi_h)}.$$

Regardless of how many workers the union puts into unemployment in the steady state, the transition probabilities assure that the same fraction of employed high skilled to low skilled employed is reached in the steady state. The conclusion is that the union cannot affect the long run inequality of wages. However, during any transition from a steady state to another the union can, and will, reduce inequality by putting low skilled workers into unemployment. Moreover, the union can affect the levels of the wages, even in the steady state.

4.2.3. Parameter restrictions. We want parameters such that the high skilled wage is above the low skilled wage in the laissez-faire. In order to accomplish this, we restrict parameters such that $\phi_{ss} > 1$. We also want the high skilled wage to be higher than the low skilled wage out of the steady state. The condition for $\phi(\chi, 0) > 1$ is

$$\phi(\chi, 0) > 1 \Longleftrightarrow \chi < \frac{1}{1 + \left(\frac{1-\psi}{\psi}\right)^{\frac{1}{1-\rho}}} \equiv \chi_{\max}.$$

Using (4.5), it is possible to show that for any $\chi < \chi_{\max}$ we have $\chi' < \chi_{\max}$ as long as $\phi_{ss} > 1$. So, for any initial value $\chi_0 \in (0, \chi_{\max})$, $\chi \in (0, \chi_{\max})$ in all periods. Hence, we assume that parameters are such that $\phi_{ss} > 1$ and that $\chi_0 \in (0, \chi_{\max})$.

4.2.4. Calibration. We calibrate the model so that the version without any union influence, i.e. without unemployment, matches observations from the U.S. economy. We set the coefficient of relative risk aversion $\sigma = 2$, the model period to one year, and fix the discount factor $\beta = .96$, to be consistent with an annual real interest rate of 4.0 percent. We further set $\alpha = .32$, a value commonly used as capital's income share in the Real Business Cycle literature (see Cooley (1995)). This of course implies that labor's share of total income is constant and equal to .68.

To match Keane and Wolpin's (1997) observation that the late 1970's wage inequality among blue collar workers was about 2, we chose parameter values such that $\phi_{ss} = 2$. In addition to that, we want to produce an annual average wage gain from tenure to 3 per cent, as calculated by Violante (2002). With $\phi = 2$, this implies $\pi_l = 22/23$. Consistent with data on the the separations rate in the U.S., we set $\pi_h = 5/6$ (Blanchard and Diamond, 1990). Finally, using the value $\rho = 1 - .709$, as calculated by Katz and

Autor (1999), implies $\psi = .43549$. Appendix C shows that the results change only somewhat if ρ is lower.

To get turbulence in the model we alter the skill transition probabilities to be consistent with the observations on increased earnings volatility. Juhn, Murphy and Pierce (1993) show that the rise in residual inequality equals about 20 per cent. We thus set $\phi_{ss}^{tur} = 2.4$. Violante (2002) further calculated that the returns from tenure increased to an annual 5 per cent in the 1980's. The increased return to tenure pins down $\pi_l^{tur} = 17/18$ and we set $\pi_h^{tur} = 4.35/6$ to obtain $\phi_{ss}^{tur} = 2.4$. With this calibration we get another result from Violante (2002) by implication, namely that the skill loss is larger and more frequent in turbulent times.

A final key parameter is R, the income if unemployed. We set R to equal 80% of a steady state low skilled wage in an economy without a union, and keep this number fixed when moving to turbulence. Martin (1996) calculated the average European unemployment insurance benefits to 70%. On top of that, the value of leisure should be added. As we have not modelled leisure explicitly, we set the replacement rate for a low skilled to 80 % to take the value of leisure into account.

4.3. Unemployment in Tranquil vs. Turbulent Times

This section is devoted to a numerical experiment using the model described in the preceding section. The experiment is carried out in the following way. Using the baseline calibration of tranquil times, we solve the union's problem as given by (4.6) and simulate it until a steady state is reached. In that steady state, the union is unexpectedly confronted with the new parameter set reflecting turbulence. We then solve the unions problem with the parameter set reflecting turbulence and simulate that economy until it reaches the new steady state. In this way we are able to compare steady states in the union economy to that in the laissez-faire as well as to the data. Moreover, we can also confront the model with the data along the transition to the new steady state. We assume that the economy is in the tranquil time steady state and is hit by turbulence in the beginning of 1980.

TABLE 4.1. STEADY STATES - TRANQUIL VS. TURBULENT TIMES

	Model				Data[a]	
	Union		LF			
	Tranquil	Turbulent	Tranquil	Turbulent	1979	2005
Unemployment (%)	0	6.61	-	-	0	4.31
Fraction of High Skilled	.2069	.1570	.2069	.1681	n.a	n.a
High Skilled Wage	.6749	.7878	.6749	.7707	n.a	n.a
Low Skilled Wage	.3374	.3286	.3374	.3214	n.a	n.a
Inequality	2.00	2.40	2.00	2.40	n.a	n.a

[a]Data refers to the difference in unemployment rates between the United States and a weighted average of France, Germany, Italy and Spain. Source: OECD. n.a = Not available.

4.3.1. Steady states. Table 4.1 depicts the main results in the steady states. As this model contains no frictional unemployment, it is natural to compare the model's predictions to the observed difference between the United States and Europe. In tranquil times, the union chooses to put no workers into unemployment. In turbulent times the quantitative effect of the monopoly union becomes large, with an implied steady state unemployment rate of 6.61 percent. Quantitatively, the model unemployment is a bit too high compared to the data.

We have not calibrated the model to fit the data in terms of unemployment. On the contrary, the model's performance should be judged on how well it fits the unemployment data. In the steady state, the model produces a significant increase in unemployment when turbulence hits he economy. The model reaches its steady state in 2008, but in 2005 it has already reached an unemployment rate of 6.60 percent, a negligible difference compared to the steady state value.

4.3.2. Transitional dynamics. Let us now study what happens along the transition to the new steady state. We are interested if convergence is consistent with the data not only in terms of magnitude, but also speed and monotonicity. The most important variable upon which we evaluate the success of our model is unemployment.

In figure 4.1 it is clear that the transition follows the same pattern as the steady state results. Along the whole transition, unemployment is higher in the model than in

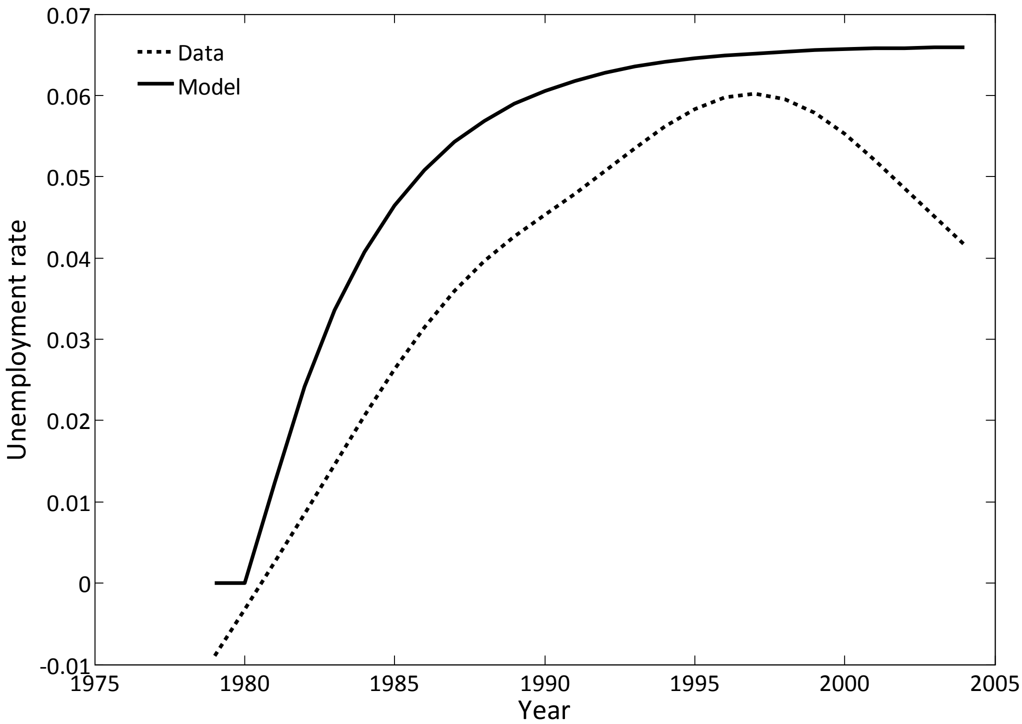

FIGURE 4.1. TRANSITIONAL DYNAMICS OF UNEMPLOYMENT. Data refers to the HP filtered (smoothing parameter = 100) difference in unemployment rates between the United States and a weighted average of France, Germany, Italy and Spain. Source: OECD.

the data. From 1980 until 1997, the model performs reasonably well. In the end of the sample period, the difference in the data drops, whereas the model slowly converges upwards to its steady state value.

4.4. Insurance versus Rent Seeking

The union creates unemployment either for rent seeking or insurance reasons. Withdrawing workers from employment raises the wages of employed workers. The incidence of the reduced production falls partly on the firms. The union effectively consumes a part of the firms' profits, why we call it the rent seeking effect, as in McDonald and Solow (1981).

Table 4.2. Steady States - Risk Neutrality

	Tranquil	Turbulent
Unemployment (%)	1.05	6.58
Fraction of High Skilled	.2047	.1570
Inequality	2.00	2.40

Agell and Lommerud (1991) showed that unemployment also can arise as an insurance device in a world with uneven income and imperfect capital markets. Risk averse workers achieve low variance of wages over time and across states by unemployment, which implies wage compression. When turbulence increases, both the rent seeking and the insurance effect contributes to the increased unemployment.

To single out the rent seeking and the insurance effect, we solve the model under risk neutrality, i.e. we set the coefficient of relative risk aversion to zero, keep the remaining parameters from the previous section, and solve the model in tranquil and turbulent times.[1] With risk neutrality, the insurance motive is moot. Hence, the results from this exercise singles out the rent seeking effect.

Table 4.2 reveals that unemployment is only slightly lower with risk neutrality than with risk aversion. This suggests that the rent seeking effect is important. Indeed, as we shall see, without rent seeking the model is incapable of producing unemployment in the steady state. Although unemployment is higher without the insurance effect, it is not affected by turbulence in the same magnitude as with the insurance motive. With risk aversion and rent seeking, unemployment increases by more than 6.5 percentage points when introducing turbulence. With only the rent seeking effect, unemployment rises by about 5.5 percentage points. The 'difference in difference,' i.e. about 1 percentage point, must then be due to increased insurance motives.

To get further understanding for what happens to the union's choice, Figure 4.2 depicts the decision rules under both risk neutrality and risk aversion in turbulent and tranquil times. We immediately see that the decision rules under risk aversion is much

[1] An alternative way of removing the insurance motive is to assume complete financial markets, maintain risk aversion in preferences, and let the union behvave as the head of a representative family in the spirit of Hansen (1985) and Rogerson (1988). This alternative specification yields exactly the same outcome, as we show in Appendix A.

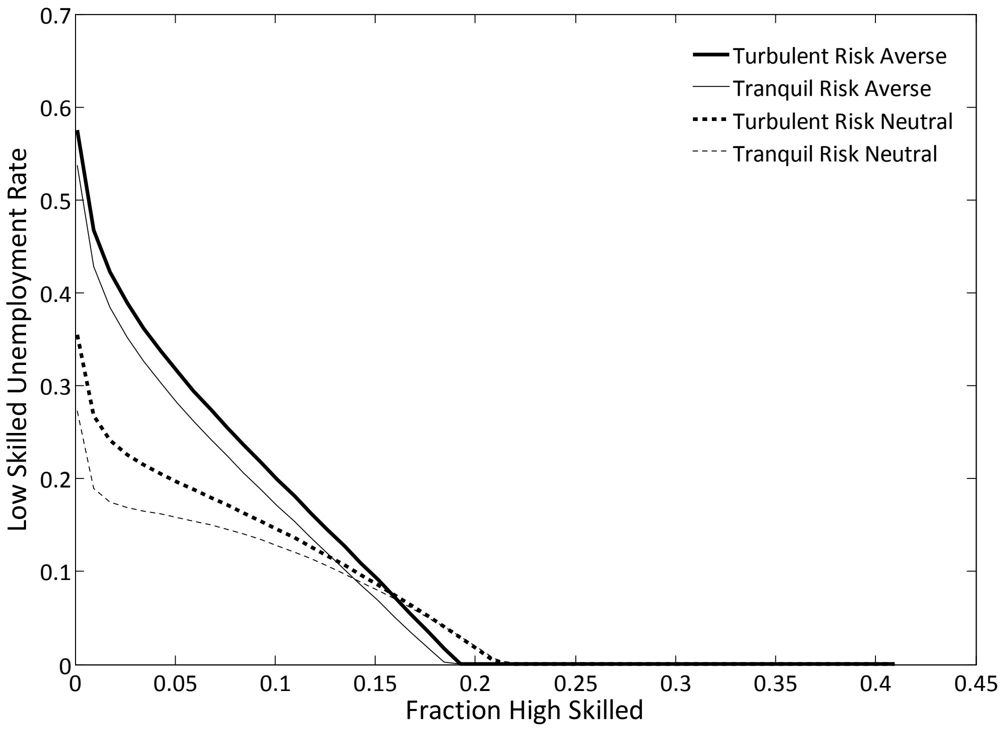

FIGURE 4.2. DECISION RULES - RISK AVERSE VS. RISK NEUTRAL.

steeper than those under risk neutrality. The reason is that for low levels of the measure of high skilled in the economy, the risk averse union has strong motives to compress wages. The reason is that, for low values of χ, inequality is high, as pointed out in section 4.2. Unemployment reduces inequality by raising the marginal productivity of the low skilled. In states with a large fraction of high skilled workers, the chosen unemployment is lower in the risk averse case than with a risk neutral union. Here, the underlying inequality is so low that the union is reluctant to unemployment as unemployment destroys skills and the more high skilled, the lower the inequality.

4.4.1. A pure insurance model. To show that rent seeking is a necessary condition to generate plausible equilibria with unemployment, we modify the model of section 4.2 to obtain a model where rent seeking is not present. To accomplish that, we return to the original specification of the production function, let capital be adjustable, and assume that the economy is a small open one with a constant real interest rate. Formally, the representative firm solves

$$\max_{K,H,L} \left\{F\left(K,H,L\right) - rK - w_h H - w_l L\right\},$$

taking the interest rate and the wages as given. The union solves the same problem as before, with the important difference that they now take different firm first order conditions as given.

Rent seeking is not present in this model as profits are zero. To appreciate this, note that the first order conditions are

$$r = F_K; w_h = F_H; w_l = F_L,$$

where subscripts are partial derivatives. Since $F(\cdot)$ has constant returns to scale, it follows from Euler's theorem that

$$F\left(K,H,L\right) = F_K K + F_H H + F_L L.$$

Combining the last expression with the first order conditions reveals that profits are always zero.

Figure 4.3 depicts the decision rules under tranquil and turbulent times, where we have set the real interest rate to 4 percent and adjusted R to equal 80 percent of the tranquil times laissez-faire low skilled wage where χ and K have reached their steady state values. These decision rules are very steep and hit zero for any χ above .14, a measure associated with a laissez-faire wage ratio of 2.8. Even if turbulence shifts the decision rule to the northeast, there is still no unemployment in the steady state of this model. This can be seen in figure 4.3 as unemployment is zero for all levels of χ between the tranquil times laissez-faire level of .2069 to the turbulent times laissez-faire fraction of high skilled of .1681.

The fundamental reason for the absence of unemployment in this small open economy is the adjustable capital. If the union marginally shifts unemployment upwards, firms respond to this by reducing its employed capital, thereby dampening the increase in wages that would come about if capital was fixed. This effect naturally induces the

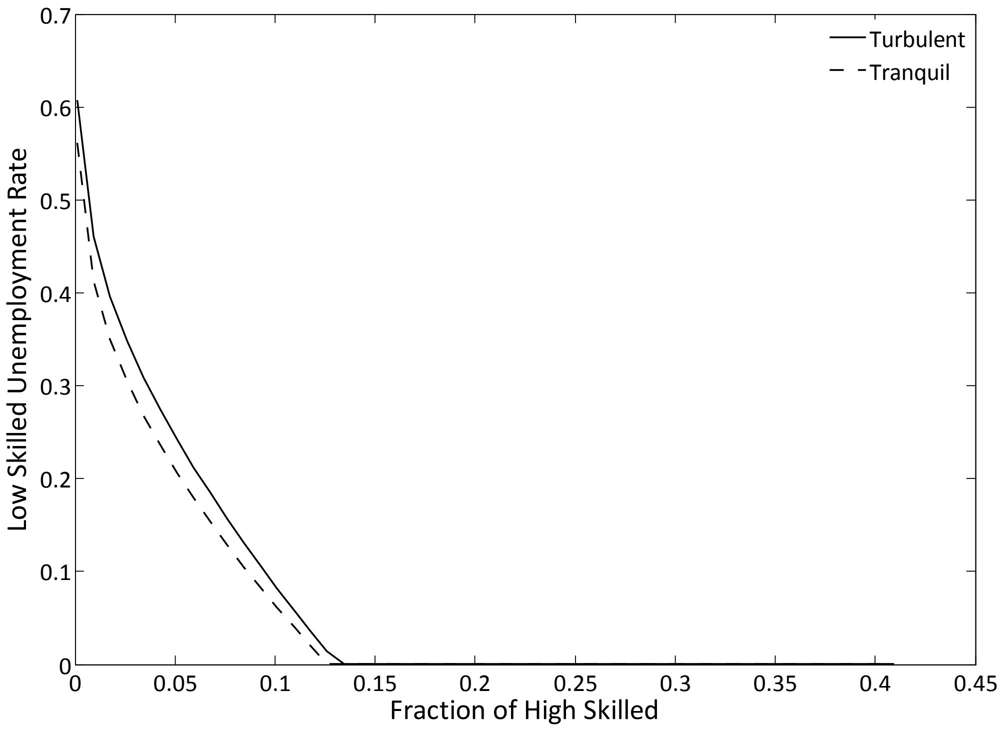

Figure 4.3. Decision Rules in Pure Insurance Model.

union to demand lower wages, and hence lower unemployment rates, compared to the case where capital is fixed. In fact, the union is unable to affect the long run evolution of wages, so its monopoly power vanishes in the long run.

To get further understanding of this, suppose $\rho = 0$ so that the firms' maximization problem becomes[2]

$$\max_{K,H,L} \left\{ K^{\alpha} \left[H^{\psi} L^{1-\psi} \right]^{1-\alpha} - rK - w_h H - w_l L \right\}. \tag{4.11}$$

Using the first order conditions from the program (4.11), we can derive the firms' choice of capital as

$$K = \left(\frac{\alpha}{r} \right)^{\frac{1}{1-\alpha}} H^{\psi} L^{1-\psi}, \tag{4.12}$$

[2] The reasoning generalizes to any ρ, but the algebra gets a bit more tedious. I show the general case in appendix B.

which makes it clear that for higher u, which implies a lower equilibrium L, the equilibrium K is lower. The wages are increasing in K, as demonstrated by

$$w_{h,flex} = (1-\alpha)\,\psi K^{\alpha}\left[H^{\psi}L^{1-\psi}\right]^{1-\alpha}H^{-1}, \tag{4.13}$$

and

$$w_{l,flex} = (1-\alpha)\,(1-\psi)\,K^{\alpha}\left[H^{\psi}L^{1-\psi}\right]^{1-\alpha}L^{-1}. \tag{4.14}$$

This shows that if the union reduces labor supply by increasing unemployment, the firms respond to this by employing less capital, which tends to reduce wages. This effect is not present in the model with a fixed factor of production. Substituting (4.12) in (4.13) and (4.14) yields

$$w_{h,flex} = (1-\alpha)\,\psi\left(\frac{H}{L}\right)^{\psi-1}, \tag{4.15}$$

and

$$w_{l,flex} = (1-\alpha)\,(1-\psi)\left(\frac{H}{L}\right)^{-\psi}. \tag{4.16}$$

Wages are completely determined by the fraction of high to low skilled employed, thanks to the constant returns to scale that is now present in production. As we saw in the steady state characterization of section 4.2, the steady state ratio of employed high skilled to employed low skilled is unaffected by the union's policy. This means that, when capital is flexible, not only the wage ratio, but the absolute level of steady state wages are independent of the union's policy, and given by

$$w^{ss}_{h,flex} = (1-\alpha)\,\psi\left(\frac{1-\pi_l}{1-\pi_h}\right)^{\psi-1},$$

and

$$w^{ss}_{l,flex} = (1-\alpha)\,(1-\psi)\left(\frac{1-\pi_l}{1-\pi_h}\right)^{-\psi}.$$

So, in the long run, the union cannot affect wages and its monopoly power effectively vanishes. As seen in Figure 4.3, the union still creates unemployment in some states but this is only worthwhile in states with extreme inequality where the distribution of wages is more important than the size of the wage bill. From those states, the measure of high skilled converges to the laissez-faire level at a lower pace than in laissez-faire, as unemployment destroys skills.

Unemployment being unwanted by the worker collective in the long run with constant returns to scale is reminiscent the result on hours worked restrictions studied by Marimon and Zilibotti (2000). They showed that with a fixed factor of production, which implies decreasing returns to scale (or rent seeking in the language of this paper), an hours worked restriction can be welfare improving whereas, if factors are flexible and the production function exhibits constant returns to scale, it cannot. The reason that an hours worked restriction can be welfare improving in Marimon and Zilibotti's (2000) matching model is that, just as in this model, fewer hours worked (or, in this paper, more unemployed) increases the marginal product of the employed. In a steady state with constant returns to scale, this effect is not present.

4.5. The Role of the Welfare State

A key parameter in the calibration is the income received when unemployed. We calibrated this parameter to equal 80% of the low skilled laissez-faire steady state wage. This is a reasonable number for the European welfare states, especially considering that it implies a low replacement rate for a laid off high skilled and that the value of leisure should be included in R. Still, it is interesting to inquire the model's response to different values of this parameter to find out if the presence of a welfare state is a necessary institutional difference to generate the different labor market outcomes between Europe and the United States, as Krugman (1994) conjectured.

To figure out the effect of different R's, we solve the model in turbulent times for R equal to 70 percent and 90 percent of the tranquil times laissez-faire steady state wage for the low skilled.

In table 4.3, we see that the steady state unemployment rate is indeed sensitive to the choice of R. With the low level of R, the steady state unemployment disappears while for the high level it increases to a massive 35.37 percent.

The reason why R is such an important parameter is that it determines the loss in income for those getting unemployed. If R is high, those going into unemployment lose little consumption by being put into unemployment by the union.

Table 4.3. Steady States for Different values of R

	70%	90%
Unemployment (%)	0	35.37
Fraction of High Skilled	.1681	.1086

4.6. Welfare

There are a number of reasons to study the welfare consequences of union wage setting in this economy. Given that the union is utilitarian, it is obvious that a hypothetical average member prefers the union equilibrium over the laissez-faire equilibrium. However, the magnitude of these effects and the consequences for firm owners are unclear.

Agell (1999) argues that even though union wage setting may imply unemployment and lower production, it can be welfare improving if unemployment and its implications counteracts some other market rigidities or imperfections. Again, it is the insurance argument. In the economy of this paper, the union creates unemployment partly because workers are unable to self insure by borrowing and saving, for example, a riskless bond.[3] The market imperfection that the lack of access to capital markets constitutes can partly be overcome by another imperfection, namely unemployment created by the union's monopoly power. Lindquist (2005) offers a distinctly different perspective. He uses an overlapping generations model with educational investment to calculate massive welfare losses from union wage setting. A somewhat peculiar property of Linquist's model is, however, that, within the environment of his paper, no one benefits from the presence of a union.

This section is devoted to three things. We first study the welfare consequences for the worker collective. Then, we ask to what extent firms' profits are reduced by the prevalence of union wage setting. Finally, we quantify the total welfare effects of union wage setting in this economy to quantitatively evaluate Agell's (1999) conjecture, and reassess Linquist's (2005) findings in the environment of this paper.

[3] Of course, even better than a riskless bond is complete contingent claims markets. In such an environment Agell's argument goes away completely. Agell acknowledges this fact and argues within an economy with some inherent market frictions.

4.6.1. Workers' welfare. To evaluate the cost of business cycles, Lucas (1987) asked how much extra consumption in every period a worker in an environment with business cycles would demand to be indifferent between living in an economy with business cycles and in an economy without fluctuations. We shall use the same methodology to quantify the welfare effects of union wage setting.

Figure 4.4 depicts the compensation a worker demands for living in the laissez-faire rather than in the union equilibrium.[4] The compensation is calculated as the increase in consumption needed in all states to make the average worker indifferent between the union equilibrium and the laissez-faire. Formally, for every χ we solve for Δ in

$$\sum_{t=0}^{\infty} \beta^t \left(\chi_t \frac{\left(w_{ht}^{LF}(\chi_t) + \Delta\right)^{1-\sigma}}{1-\sigma} + (1-\chi_t) \frac{\left(w_{lt}^{LF}(\chi_t) + \Delta\right)^{1-\sigma}}{1-\sigma} \right) = J(\chi), \qquad (4.17)$$

where $\chi_0 = \chi$, w_{ht}^{LF} and w_{lt}^{LF} are the laissez-faire wages, and

$$\chi_{t+1} = \pi_h \chi_t + (1-\pi_l)(1-\chi_t).$$

From figure 4.4 it is also obvious that the welfare effects of a union are substantial in turbulent times and more modest in tranquil times. In turbulent times, around the steady state measure of high skilled, the average worker demand around 1.5 to 2.0 percent of the laissez-faire steady state wage in every period to be willing to give up union wage setting. In tranquil times however, the welfare effects of union wage setting is miniscule in a neighborhood of the steady state. This is unsurprising; given that there is no steady state unemployment in our calibration of tranquil times, the values are equal across the union and laissez-faire equilibria are equal for all values of χ above the steady state value.

4.6.2. Firm values. Workers gain from the prevalence of union wage setting, but what about firm owners? The possibility of rent seeking suggests that there may be nonnegligible effects on firms' profits. In this subsection, we calculate firms' profits and the value of the firm in all states to figure out these effects. Define by $Y(\chi, u(\chi)) \equiv Y(\chi)$ the equilibrium production in state χ. Thanks to the Cobb-Douglas nature of

[4] Setting $u(\chi) = 0$ for all χ and calculating the wages corresponding to that produces the laissez-faire values.

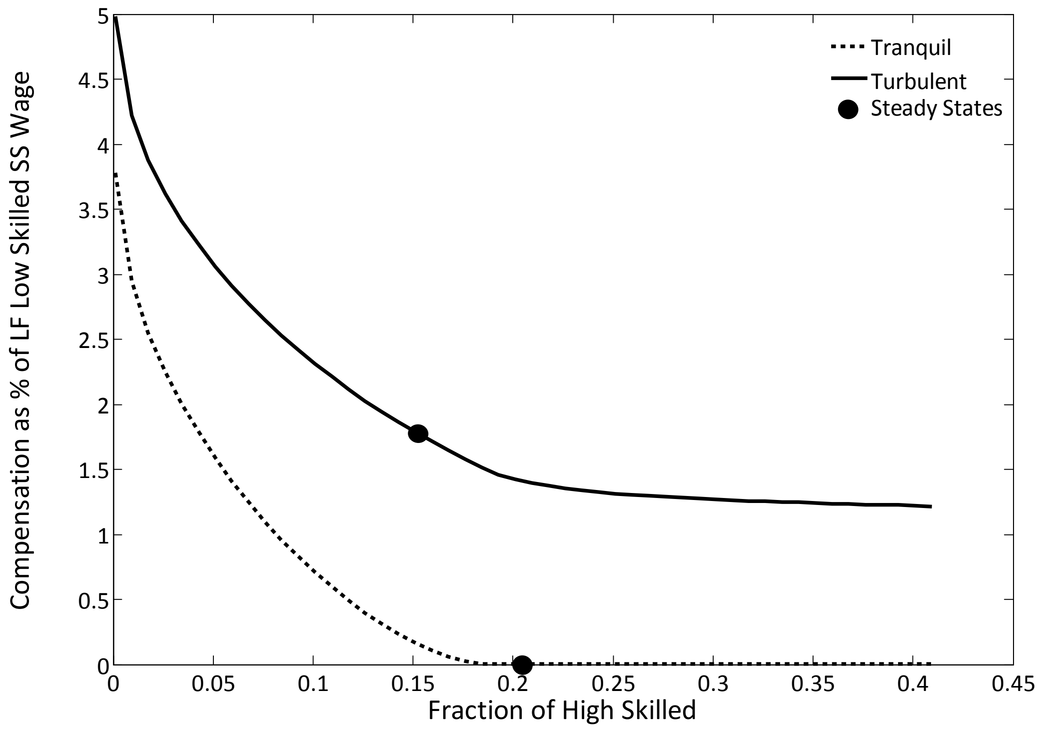

FIGURE 4.4. WORKER GAINS FROM UNION WAGE SETTING.

the production function, the per period profits in state χ is then $\alpha Y(\chi)$. In order to be able to compare total welfare later, we shall assume that unemployment benefits are financed by profit taxation, which is nondistortionary with a fixed capital stock.[5] The value of the firm is given by

$$\Pi(\chi) = \alpha Y(\chi) - (1-\chi)\, u(\chi)\, R + \beta \Pi(\chi'), \tag{4.18}$$

where

$$\chi' = \pi_h \chi + (1 - u(\chi))(1-\pi_l)(1-\chi). \tag{4.19}$$

To make comparisons with the workers' situation under union wage setting, we calculate the firms' willingness to pay to get rid of union wage setting. If the loss in firm value due to unionization is A, the firms could, if they have access to a capital

[5] If we did not make this adjustment we would essentially compare two economies with different potential resources.

market with an interest rate equal to $1/\beta - 1$, pay each worker Δ^{firm} in each period, where $\Delta^{firm} = (1 - \beta) A$.

Figure 4.5 plots Δ^{firm}, again related to the laissez-faire low skilled steady state wage. The result is naturally the opposite of the consequences for the workers. Unionization has large negative implications for firm values in states where it has large positive consequences for workers. In tranquil times, the firm has virtually no costs of unionization and is not willing to pay anything in order to abolish union wage setting. During turbulent times, however, the loss in firm value is large, and the firms' per period willingness to pay is around 8 percent of the laissez-faire steady state low skilled wage in a neighborhood of the steady state. The reason for the profit reduction with union wage setting is of course that the union's monopoly power enables it to extract rents from the firm, which deteriorates the firms' values. The effects are larger in turbulent times than in tranquil times as, in tranquil times, the two equilibria share a common steady state. Indeed, as in the case of the workers, there is no difference between the union and the laissez-faire profits for high χ in tranquil times, and the willingness to pay in those states is naturally zero.

4.6.3. Is union wage setting welfare decreasing? Workers gain and firm owners lose from union wage setting, but is there a way to calculate if the union equilibrium is preferable to the laissez-faire or vice versa? In this subsection we ask if the firm owners could 'bribe' the workers not to exercise their monopoly power and still have higher profits than in the union equilibrium. If the workers are as well of as in the laissez-faire equilibrium with the bribe as in the union equilibrium and the firms' profits, net of bribes, are higher in the laissez-faire than in the union equilibria, we have performed a Pareto improvement upon the union equilibrium and union wage setting is deemed as welfare decreasing.

The answer to this question is found by combining the results in figure 4.4 with those of figure 4.5. Figure 4.6 plots the resources left after the firm has compensated the workers not to exercise their monopoly power. It is evident that, except in states where the union wage setting has no influence (in tranquil times for χ larger than its

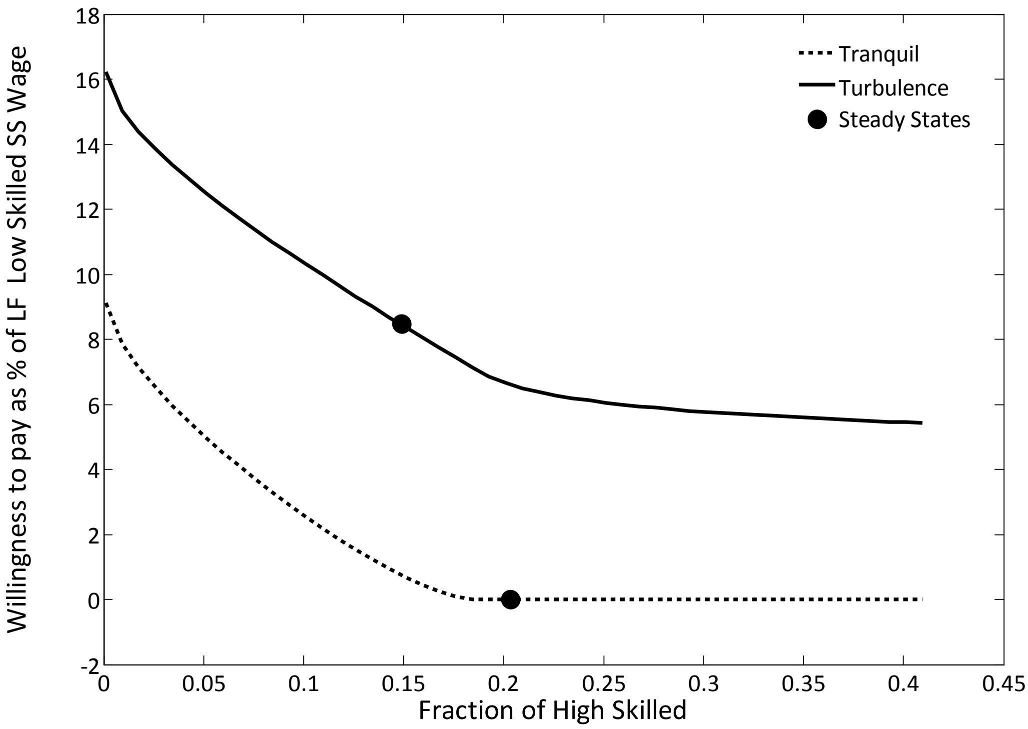

Figure 4.5. Firm Willingness to pay for Laissez-Faire.

steady state value), it is possible to construct an equilibrium where production and wage setting is competitive, with side payments making workers indifferent between the union equilibrium and the laissez-faire. The resources that are left over are then a measure of the welfare cost of union wage setting. In turbulent times, the welfare costs of union wage setting is around 6 percent of the low skilled laissez-faire steady state wage per period. In tranquil times there are, of course, no welfare effects of union wage setting in a neighborhood of the steady state.

To conclude the welfare analysis we have seen that the benefits of union wage setting is outweighed by its cost in terms of hurting firm owners. The gains from wage compression and the insurance that brings along are not big enough to compensate the reduction in production associated with unemployment.

Agell (1999) probably did not have an environment like this in mind when he argued that union wage setting may be welfare enhancing. Rather, he was thinking along the

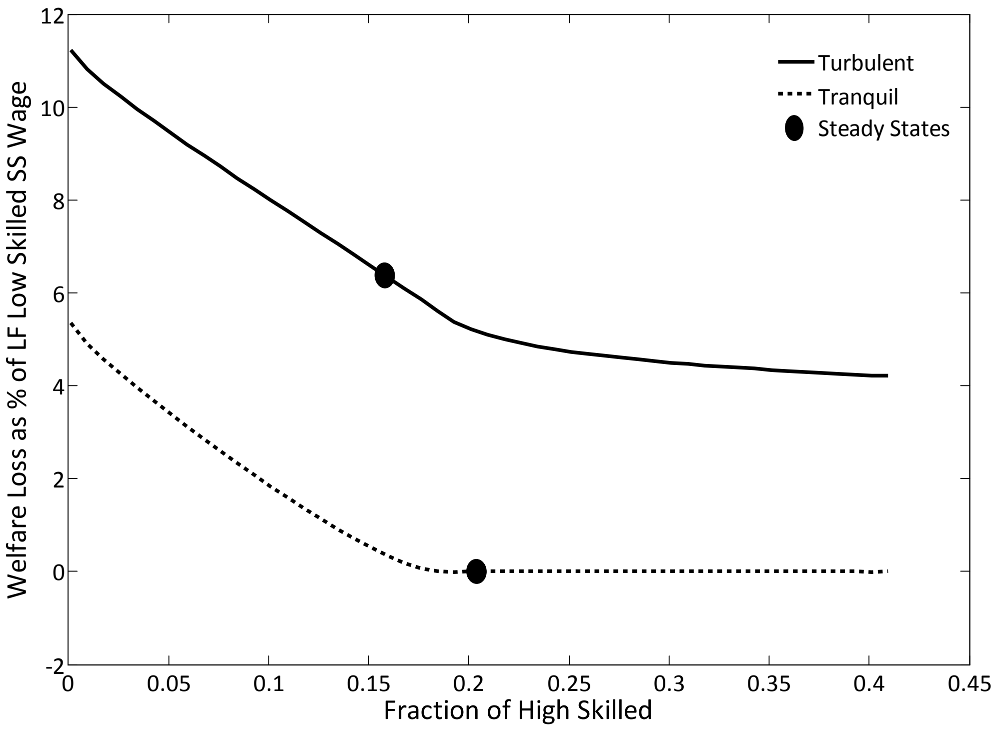

Figure 4.6. The Welfare Costs of Union Wage Setting.

lines of his own work with Kjell Erik Lommerud, Agell and Lommerud (1991). In their model without skill dynamics, it is possible to obtain unemployment without rent seeking effects, and unemployment is then welfare increasing. In this environment, we have seen that this is not possible; rent seeking is imperative for the model's possibility to generate steady state unemployment. The skill dynamics in this paper brings the ratio of high to low skilled employed to a given value independently of the unions strategy. Hence, the effects of unemployment on inequality is at best temporary. In Agell and Lommerud (1991), the skill distribution is unaffected by unemployment, as their model is static. Moreover, in this paper, rent seeking deteriorates production and profits to the extent that it outweighs the benefits of wage compression for the workers that lack of access to conventional consumption smoothing vehicles causes.

We thus qualitatively confirm Lindquist's (2005) assessment that union wage setting reduces welfare. His model incudes educational attainment and, with union wage setting, unemployment. Union wage setting is detrimental to welfare because the education decision is distorted and because unemployment reduces production in the economy. Lindquists welfare losses ranges from 2 to 20 percent for different worker types, and are mostly due to the prevalence of unemployment. Although the calculated welfare losses are reasonably similar to those in this paper, they have different interpretations. In this paper, workers benefit from union wage setting and firm owner lose. In Lindquist (2005), everyone lose from unionization.

4.7. Conclusions

The combination of union wage setting, turbulence and a generous welfare state can generate the increased unemployment in Europe relative to the U.S. that has occurred during the last decades. In order for this story to hold water, a utilitarian union that monopolizes labor has to be able to extract rents from the employers.

If one accepts the rent seeking assumption, the model has a large impact on the welfare of different groups. The average worker gains from the prevalence of union wage setting, but firm owners lose. In the aggregate, union wage setting is welfare reducing.

References

Acemuglu, Daron., Aghion Philippe, and Giovanni L. Violante. (2001). "Deunionization, Technical Change and Inequality," *Carnegie Rochester Conference Series on Public Policy* **55**, pp. 229-264.

Agell, Jonas and Kjell-Erik Lommerud. (1991). "Union Egalitarianism as Income Insurance," *Economica* **59**, pp. 295-310.

Agell, Jonas. (1999). "On the Benefits from Rigid Labour Markets: Norms, Market Failures, and Social Insurance," *Economic Journal* **109**, pp. F143-F164.

Blanchard, Olivier and Peter Diamond. (1990) "The Cyclical Behavior of the Gross Flows of U.S. Workers," *Brookings Papers on Economic Activity* **2**, pp. 85-143.

Blanchard, Olivier and Lawrence Summers (1986). "Hysteresis and European Unemployment," In Stanley Fischer (ed.), *NBER Macroeconomics Annual.* Cambridge MA: The MIT Press, pp. 15-77.

Cooley, Tom (ed.) (1995) *Frontiers of Business Cycle Research*, Princeton, NJ: Princeton University Press.

Diamond, Peter A. (1982). "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies* **55**, pp. 337-360.

Gottfries, Nils and Henrik Horn (1987). "Wage Formation and the Persistence of Unemployment," *Economic Journal* **97**, pp. 877-884.

Gottschalk, Peter and Robert Moffit. (1994). "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity* **2**, pp. 217-272.

Hansen, Gary D. (1985). "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics* **16**, pp. 309-32.

Juhn, Chinhui., Kevin M. Murphy and Brooks Pierce. (1993). "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* **101**, pp. 410-442.

Katz, Lawrence and David Autor. (1999). "Changes in the Wage Structure and Earnings Inequality," ch. 26 of *Handbook of Labor Economics* vol. **3**, ed. Orley Ashenfelter and David Card. Elsevier Science.

Keane and Wolpin (1997) "The Career Decisions of Young Men," *Journal of Political Economy* **105**, pp. 473-522.

Lindbeck, Assar and Dennis Snower. (1988). *The Insider-Outsider Theory of Employment and Unemployment*. Cambridge, MA: MIT Press.

Lindquist, Matthew J. (2005). "The Welfare Costs of Union Wage Compression," *European Economic Review* **49**, pp. 639-658.

Ljungqvist, Lars and Thomas J. Sargent. (1998). "The European Unemployment Dilemma," *Journal of Political Economy* **106,** pp. 514-550.

Ljungqvist, Lars and Thomas J. Sargent. (2004). "European Unemployment and Turbulence Revisited in a Matching Model," *Journal of the European Economic Association* **2**, pp. 456-468.

Ljungqvist Lars and Thomas J. Sargent. (2007a). "Understanding European Unemployment with Matching and Search-Island Models," *Journal of Monetary Economics* **54**, pp. 2139-2179.

Ljungqvist Lars and Thomas J. Sargent. (2007b). "Understanding European Unemployment with a Representative Family Model," *Journal of Monetary Economics* **54**, pp. 2180-2204.

Ljungqvist, Lars and Thomas J. Sargent. (2008). "Two Questions about European Unemployment," *Econometrica* **76**, pp. 1-29.

Lucas, Robert E., Jr. (1987). *Models of Business Cycles*, Oxford: Basil Blackwell.

Lucas, Robert E., Jr and Edward C. Prescott. (1974). "Equilibrium Search and Unemployment," *Journal of Economic Theory* **7**, pp. 188-209.

Marimon, Ramon and Fabrizio Zilibotti (2000). "Employment and Distributional Effects of Restricting Working Time," *European Economic Review* **44**, pp. 1291-1326.

Martin (1996). "Measures of Replacement Rates for the Purpose of International Comparisons: a Note", *OECD Economic Studies* **26**, pp. 99-115.

McCall, John J. (1970). "Economics of Information and Job Search," *Quarterly Journal of Economics* **84**, pp. 113-126.

McDonald, Ian M. and Solow, Robert M. (1981). "Wage Bargaining and Employment," *American Economic Review* **71**, pp. 896-908.

Mortensen, Dale T. (1982). "The Matching Process as a Noncooperative Bargaining Game." In John J McCall (ed.), *The Economics of Information and Uncertainty.* Chicago, IL: University of Chicago Press and National Bureau of Economic Research, pp. 233-258.

Pissarides, Christopher A. (1990). *Equilibrium Unemployment Theory.* Cambridge, MA: Basil Blackwell.

Rogerson, Richard. (1988). " Indivisible Labor, Lotteries, and Equilibrium," *Journal of Monetary Economics* **21**, pp. 3-16.

Violante, Giovanni L. (2002). "Technological Acceleration, Skill Transferability, and the Rise in Residual Inequality," *Quarterly Journal of Economics* **117**, pp. 297-338.

Appendix A - A Representative Family Interpretation of Risk Neutrality

I section 4.4 we compared the baseline case with risk averse workers to the case where workers are risk neutral. Comparisons across different preference specifications may seem unappealing for singling out different effects within one of the specifications. This appendix shows that if workers are risk averse and the union is able to costlessly transfer resources across individuals and over time and save or borrow at a gross interest rate equal to the reciprocal of the workers' discount factor, such a union will behave *as if* its members were risk neutral. In other words, the union just described will implement exactly the same unemployment rate in every state as the union with risk neutral members. Of course, different individual consumption allocations and values will be attained, but that does not affect the analysis of section 4.4.

Implicit in the formulation of (4.6) is that the union cannot transfer consumption across individuals, save or borrow. If the union had the possibility to transfer resources across individuals and over time and save or borrow at a gross interest rate equal to the reciprocal of the workers' discount factor, it solves

$$\max_{\{u_t, c_{ht}, c_{lt}, c_{ut}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left\{ \chi_t \frac{c_{ht}^{1-\sigma}}{1-\sigma} + (1-\chi_t) \left[(1-u_t) \frac{c_{lt}^{1-\sigma}}{1-\sigma} + u_t \frac{c_{ut}^{1-\sigma}}{1-\sigma} \right] \right\}, \tag{4.A.1}$$

where c_j is time t consumption for a $j-$type, the maximization is subject to (4.2), (4.3), (4.5), and a present value budget constraint given by

$$\sum_{t=0}^{\infty} \beta^t \left[\chi_t c_{ht} + (1-\chi_t)(1-u_t) c_{lt} + (1-\chi_t) u_t c_{ut} \right] \leq \sum_{t=0}^{\infty} \beta^t y_t, \tag{4.A.2}$$

where $y_t \equiv [\chi_t w_{ht} + (1-\chi_t)(1-u_t) w_{lt} + (1-\chi_t) u_t R]$ is the period t total income of the union members. With risk aversion and perfect transfer possibilities, the solution to this problem involves $c_{jt} = \overline{c}$ for all j, t. This means that the solution to the above program solves

$$\max_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \overline{c}, \tag{4.A.3}$$

subject to (4.2), (4.3), (4.5), and

$$\sum_{t=0}^{\infty} \beta^t \overline{c} \leq \sum_{t=0}^{\infty} \beta^t y_t. \tag{4.A.4}$$

This program has the same solution, i.e. the same $\{u_t\}_{t=0}^{\infty}$ as

$$\max_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t y_t \qquad (4.A.5)$$

subject to (4.2), (4.3), (4.5), which is what we solved in section 4.4. Hence, we get the same unemployment sequence when using the risk neutral preference specification as when using a representative family interpretation.

Appendix B - Steady State Wages with Capital Adjustment

This appendix shows that the steady state wages are independent of the unions choice in the model with adjustable capital, i.e. in the pure insurance model of section 3.1.

The firms' maximization problem becomes

$$\max_{K,H,L}\left\{K^{\alpha}\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1-\alpha}{\rho}}-rK-w_hH-w_lL\right\}.$$

The choice of capital obeys the first order condition

$$\alpha K^{\alpha-1}\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1-\alpha}{\rho}}=r.$$

Hence

$$K=\left(\frac{\alpha}{r}\right)^{\frac{1}{1-\alpha}}*\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1}{\rho}}.$$

The high skilled wage is given by

$$\begin{aligned} w_h &= (1-\alpha)\,\psi K^{a}\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1-\alpha-\rho}{\rho}}H^{\rho-1} \\ &= (1-\alpha)\,\psi\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}}\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1-\rho}{\rho}}H^{\rho-1}, \end{aligned}$$

where the last equality follows from insertion of the expression for optimal capital. We can rewrite this expression in the following way

$$w_h=\frac{(1-\alpha)\,\psi\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}}\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1}{\rho}}H^{-1}}{\left[\psi+(1-\psi)\left(\frac{L}{H}\right)^{\rho}\right]}, \tag{4.B.1}$$

now as the function $g(H,L)\equiv\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1}{\rho}}$ is linearly homogenous, we know that

$$g(H,L)\,H^{-1}=\left[\psi H^{\rho}+(1-\psi)\,L^{\rho}\right]^{\frac{1}{\rho}}H^{-1}=g\left(1,\frac{L}{H}\right).$$

Hence, the numerator of (4.B.1) depends only on L/H, as does the denominator. In other words, the wage is given solely by the fraction of high to low skilled employed, which, in the steady state, is independent of the unions's policy. The same operations can be performed for the low skilled wage.

Appendix C - Robustness

In this section we perform robustness checks for our results. Specifically, we solve the model under a different elasticity of substitution between high skilled and low skilled workers. The reason for doing this is that the elasticity used in the baseline calibration is the measured elasticity of substitution between white collar and blue collar workers estimated by Katz and Autor (1999). The elasticity of substitution between workers with the same education but with different job tenure, which is the interpretation of high and low skilled workers in this economy, could be different.

Table 4.C.1 depicts the steady state unemployment rates when the elasticity of substitution is 2 instead of 1.4. We changed the rest of the calibration in the following way. We reset the elasticity of substitution and then recalibrated ψ so as to maintain the steady state wage inequality. Evidently, the model is not that sensitive to a slightly higher elasticity of substitution than in the main calibration. In fact, when $\rho = 1/2$, the model fits the data slightly better than with $\rho = 1 - .709$.

TABLE 4.C.1. ROBUSTNESS. STEADY STATES

	$\rho = 1/2$	
	Tranquil	Turbulent
Unemployment (%)	0.00	5.28

Discrimination

CHAPTER 5

Is there Discrimination in Swedish High Schools?

Erik Höglin and Björn Tyrefors♠

Abstract. Recent research has found evidence of males being discriminated in school. We reassess this finding in a field experiment where a random sample of tests in the Swedish language is subject to blind and non-blind grading. We find no statistically significant evidence of discrimination.

5.1. Introduction

Gender differences are present both in school and in the labor market. A puzzling empirical regularity is that while females outperform males in school, they generally have lower wages in the labor market. A large body of literature has studied gender differences and discrimination in the labor market.[1] Much less is known about the cause of gender differences among individuals before entering the labor market.

In this paper, we ask if gender differences in school performance can be explained by discrimination. To this end, we compare randomly drawn test scores graded with full information about the student (non-blind grading) to test scores graded by teachers with no knowledge of the student's gender (blind grading).

Applying a difference-in-difference approach allow us to single out the effect of discrimination. Since we compare the same test twice, differences in ability across gender are differenced out. We find that female students perform better in the tests than their male counterparts and that blind grading leads to lower test scores for both genders. However, we find no statistically significant evidence of discrimination.

[0] ♠Dep. of Economics, Århus University, btyrefors@econ.au.dk. We have benefited from comments and suggestions from Magnus Johannesson and David Strömberg. Karolina Wallin provided excellent research assistance. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

[1] See Altonji and Blank (1999) for a survey.

In a related study using a large data set from Israel, Lavy (2004) finds that male high school students are discriminated in a number of subjects. By comparing two similar tests, graded with and without knowledge of student identity, he argues that the gender differences are due to discrimination.[2] Lavy (2004) is the first, and to our knowledge the only, paper investigating discrimination in schools using a natural experiment. Thus, there is need for more studies in other countries in order to evaluate the generality of Lavy's results. In addition, we make a methodological improvement by comparing not similar, but the very same tests.[3]

Gender differences in Swedish schools have previously been studied, but not in a natural experiment setting. Lindahl (2007) uses data on Swedish junior high school students to investigate how the difference between the grade on the school leaving certificate and the grade in a national test is related to gender. She finds that given a specific test score on the national test, female students get higher grades than male students on the school leaving certificate.[4] However, even though there is solid evidence of gender differences in school performance, it does not follow that the differences are caused by discrimination. To credibly make that argument, we need to rely on some type of experimental design.[5]

The next section describes the Swedish school system and the sampling procedure. In section 5.3 we discuss our empirical strategy. The results are presented in section 5.4 and section 5.5 concludes the paper.

---

[2] Similar identification strategies have been applied by Blank (1991) and Goldin and Rouse (2000), investigating discrimination in the labor market.

[3] A drawback is that for cost reasons, we can only draw a sample of 1000 students. Lavy (2004) makes use of all secular Jewish schools eligible for the test in the years of 2000-2002.

[4] Lindahl's results are in line with international evidence. See, for example, Emanuelsson and Fischbein (1986). Moreover, a number of studies have investigated if the effect is related to the gender of the teacher and the gender/ethnic congruence between student and teacher, e.g. Dee (2005). However, Holmlund and Sund (2005) find no such effects when evaluating the school leaving certificates in Sweden.

[5] There is no formal relation between the test score on a national test and the final grade in the subject, which makes a comparison between the two types of grades difficult to use for investigating discrimination

5.2. The Swedish High School and Sampling Procedure

After nine years of compulsory schooling, the vast majority of the Swedish youth enroll in high school education. High school lasts for three years and can be either vocational training or on an academic track. Both the academic track and the vocational programs offer the same set of core subjects, comprising Swedish, English, math, and social studies. Basic courses in the core subjects are compulsory and, upon completion, the student earns basic eligibility for college education.[6] In addition to the core subjects, students on the academic track complete advanced courses in either math/science or humanities/social studies. Students in vocational programs specialize in their field, e.g. cooking, construction and automobile mechanics.

Students' achievements in different subjects are graded on a four-tiered ordinal scale: Fail, Pass, Pass with Distinction and Excellent. To calculate a GPA, the grades are translated into a cardinal scale with 0 for Fail, 10 for Pass, 15 for Pass with Distinction and 20 for Excellent. Grades are absolute and the core subjects have nationally stipulated prerequisites for each grade. The prerequisites are exclusively based on knowledge criteria. Hence, conditional on the level of knowledge, grades should not reflect participation, diligence or ambition. In practice however, teachers enjoy great discretion when setting grades. Grades are not externally evaluated, so teachers could base their grades on anything they observe.

Compulsory national tests are given in the core parts of Swedish, English and math. We focus on the test in Swedish, since we believe that grading a Swedish test allows for more arbitrariness than, for example, Math. Every academic year, two national tests in Swedish are constructed by the National Agency of Education in conjunction with the Department of Scandinavian Languages at Uppsala University. The tests have three parts, one oral and two written. We use data from the second, more extensive, written test for the academic year 2005/2006. In this test, students are asked to write an essay

[6] Some college educations, e.g. medical schools and college programs aiming at a degree in Engineering, have additional requirements, such as completed high school courses in Science and/or advanced Math.

based on one out of nine topics within a common theme.[7] Students choose their topic with full discretion.

The written part of the national test is graded on the same scale as the subjects. Teachers are given written guidelines stating the prerequisites for each grade. Once more, they have great discretion in the actual grading. Moreover, the teachers grade their own students. No means are taken by the national authorities to ensure that the guidelines are followed, nor are any evaluations of the schools conducted.

The Swedish school system thus provides us with one of the components needed, the non-blind grade. To obtain blind test scores, we drew a random sample of 1000 students eligible to take the test.[8] Out of the 1000 students in the sample, we received complete information, that is the actual test, the test score and the student's identity, for 677 students.[9] Absenteeism is the main cause for not taking the test, but tests were also missing due to inferior administrative routines at the schools. As many as 96 out of 100 schools participated, but not all schools had proper filing procedures in line with the guidelines of the National Agency of Education. In the end, 94 school were included.

[7] We use the fall test of 2005 and the spring test of 2006. The common themes in these semesters were, *Leva Livet* (Live Your Life, our translation) and *Hur mår du?* (How are you?, our translation), respectively.

[8] Being eligible means that a student attends a class that is participating in the course Swedish B. To perform the random sample, we obtained a complete list of all 467 Swedish public high schools for 2005/06 and the schools' enrollment data from the National Agency of Education. Based on this data, we used a two-step procedure to ensure that each student is equally likely to end up in our sample. In the first step, we weighted all schools by the number of enrolled students in the final year 2005/06. We then chose 100 schools, where the probability of each school being chosen corresponds to its weight in the population. Since Swedish public high schools are subject to a law requiring that documents produced at the schools should be made available to anyone asking for them, we phoned these 100 schools and asked for the classes that took the test either in the fall of 2005 or the spring of 2006. After receiving the lists of students in each class, we randomly drew 10 students from each school. Using this procedure, we thus ended up with a sample of 1000 students where all students in the population had the same probability of being sampled.

[9] It is worth noticing that the National Agency of Education requires that all tests and test results should be properly filed and also handed out to any citizen according to the Swedish constitution. As compared to the statistics from the yearly collection of test scores, not tests, that Statistics Sweden does for 200 representative High Schools, we have a somewhat higher response rate. For Swedish B, their total response rate for 2006 was about 62 %, as compared to 68 % for our study. According to their definitions, about 10 % of the missing values are due to administrative causes. The rest is due to the fact that eligible students do not take the test, i.e. absenteeism. See http://www.skolverket.se/content/1/c4/20/08/kursprovrapport%20vt06.pdf

We had all tests rewritten on a word processor and the students' identities as well as their teachers' notes deleted. We did this to ensure the re-graders would not be able to identify the students' gender or be influenced by the non-blind grade. Naturally, nothing else was changed.

As a final step, we randomly selected 35-50 tests into groups and hired 15 teachers from a teachers' agency to regrade one group each. The re-grading teachers did not know which student's test they regraded and they had no information regarding the purpose of the study. We required re-graders to have been grading national tests in Swedish before. We also accepted a few newly licensed teachers. Since there is a slight majority of female teachers in Swedish high schools, we required the share of female teachers to be 50-60 %. Moreover, we required that 75 % of the teachers were licensed in order to match the corresponding national share. Table 5.1 presents some of the re-graders' characteristics.[10]

5.3. Empirical Design

In this section, we develop a difference-in-difference approach in order to investigate discrimination. Let a non-blind (NB) test score be determined by student i's ability in a broad sense, the examiner's potential prejudice of gender and an error term. Assume it to be linearly related as

$$Testscore_{iNB} = \alpha_{NB} + ability_i + \beta Male_i + u_{iNB}, \tag{5.1}$$

where $Male_i$ is an indicator taking the value of 1 if student i is male and 0 otherwise. Thus, we can interpret β as a discrimination effect. If β is negative, then male students are discriminated and if β is positive, female students are discriminated. The classical problem with this formulation is that we do not observe ability. If ability is correlated with gender, e.g. if female students of school age are more mature or for some reason study harder, then estimating (5.1) without conditioning on ability would bias β downwards and we could falsely conclude that boys are discriminated, when in fact female students are more able.

[10] For more information, please see the appendix.

Table 5.1. Re-grading Teachers' Characteristics

Share men	40 %
Share licensed high school	60 %
Share licensed	80 %
Share newly certified	7 %
Share Immigrated	7 %

Thanks to the set up of the study, this endogeneity problem can be taken care of. Consider an examiner that has no information about gender (B). Then, we simply have $\beta = 0$, and

$$Testscore_{iB} = \alpha_B + ability_i + u_{iB}. \tag{5.2}$$

Taking the difference between (5.1) and (5.2) yields

$$Testscore_{iB} - Testscore_{iNB} = (\alpha_B - \alpha_{NB}) - \beta Male_i + (u_{iB} - u_{iNB}). \tag{5.3}$$

Thanks to the non-blind/blind design and randomization, the OLS estimator of β is unbiased, since all other factors affecting the test score are differenced out. The strengths of this method with respect to non-respondents are worth mentioning. Our estimate of β could be biased through selection. However, only 4 out of 100 schools did not respond which makes selection very unlikely to be problematic at the school level. For students being absent on the test date to create a problem, we need their potential difference in test scores to be related to gender. It is not a problem for our identification strategy that this group would perform differently from the students taking the test, since this is taken care of by differencing.

Let $Testscore_{ij}$ be the test score of student i if graded with information $j \in \{NB, B\}$. Then, (5.3) could be reformulated equivalently as

$$Testscore_{ij} = \alpha + \gamma Male_i + \delta NB_j + \beta Male_i * NB_j + \varepsilon_{ij}, \tag{5.4}$$

where the coefficient β of the interaction term is the coefficient of interest and has the same interpretation as before. If β is negative, then male students are discriminated.

The strength of this equivalent formulation is that it reveals the estimates of the average effect of performance on a test for male relative to female students (γ). Moreover, the indicator NB_j, which takes the value of one if the test is graded in non-blindly and zero otherwise, identifies the effect of differences in grading behavior between blind and non-blind grading that is not related to gender. In the subsequent analysis, we will use equation (5.4) since the estimates of γ and δ can be matched with previous studies and hence earn credibility for the representativeness of our sample.

5.4. Results

5.4.1. Descriptives. Out of the 100 schools that we asked for the required material for the 1000 students, 96 schools participated and 474 female and 486 male students were sampled. 147 female and 136 male students were absent or missing due to substandard administrative routines. Table 5.2 contains summary statistics for our final sample of 327 female and 350 male students. In this table and in the rest of the paper (if not pointed out explicitly), we use the cardinal scale used to calculate GPAs, that is 0, 10, 15, 20 for Fail, Pass, Pass with Distinction and Excellent.

In line with previous studies, we conclude that female students on average get approximately 20 percent higher grades than males when graded non-blindly. This difference is highly significant.[11] As expected, blind grading decreases the average score. Both female and male students get about 18 percent lower grades in the blind grading procedure than in the non-blind procedure. The difference between genders is once more highly significant. Looking at the differences of the difference, we see that the estimate of -0.487 indicates that there might be some discrimination against boys. However, the difference between genders is not statistically significant at any reasonable level.

It is of further interest to inspect the distributions at a more detailed level. Figures 5.1 and 5.2 depict the distributions of blind and non-blind test scores. In Figure 5.1, we clearly see that female students have higher grades than male students in the non-blind test. A comparison of Figures 5.1 and 5.2 shows a clear tendency of an overall down

[11] We perform a standard t-test.

TABLE 5.2. SUMMARY STATISTICS OF FEMALE AND MALE STUDENTS

Sample Statistics	N	Mean	Std. Dev.
Female	474		
Male	486		
Female Absent	147		
Male Absent	136		
Variables of Interest			
Non-Blind Grade, Female	327	13.119	4.689
Non-Blind Grade, Male	350	10.890	5.346
Difference		2.229***	
Blind Grade, Female	327	10.703	5.395
Blind Grade, Male	350	8.957	5.751
Difference		1.746***	
$Grade_{iB} - Grade_{iNB}$ Female	327	-2.415	5.885
$Grade_{iB} - Grade_{iNB}$ Male	350	-1.928	6.002
Difference		-0.487	

*** Significant at the 1% level,**at the 5% level,*at the 10% level

grading for both genders. Moreover, figure 5.3 measures the difference between blind and non-blind test scores and indicates that men have a higher mass around zero, i.e. were subject to less change of grade. It is worth noticing that more than 50 percent of the blind grades differ from the non-blind grades and that four female students received the highest grades in the non-blind procedure, while they received the lowest grade when graded blindly.

5.4.2. Regression results. We now turn to the estimation. Table 5.3 presents the results from the estimation of (5.4).[12] The variable of interest, the interaction between the male and the non-blind indicator, measures the potential discrimination. The interpretation of the point estimate in column 1 is that male students get a .487 lower grade on average because of their gender. This effect is not driven by ability or the nature of the grading procedure. However, the estimate is far from significant.

[12] We cluster the standard errors at the school level. However, the level of clustering could also be chosen to be the level of the re-grading teacher or no clustering. The standard errors are quite similar across the different alternatives.

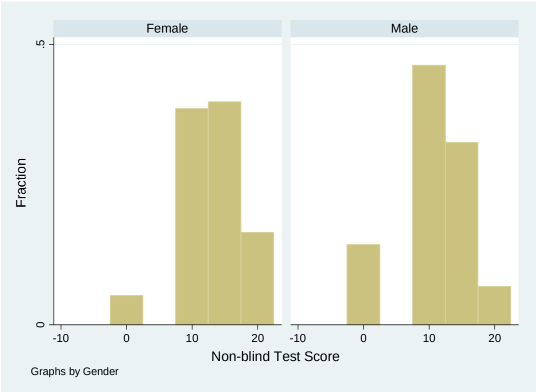

FIGURE 5.1. NON-BLIND GRADE.

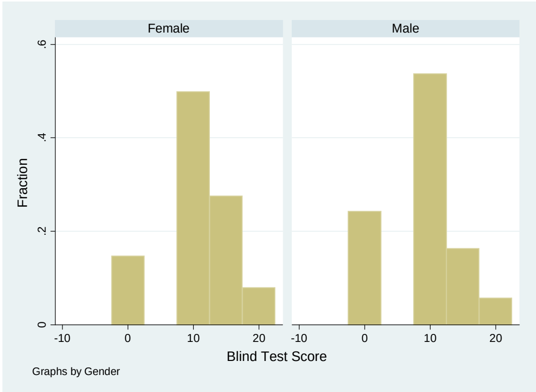

FIGURE 5.2. BLIND GRADE.

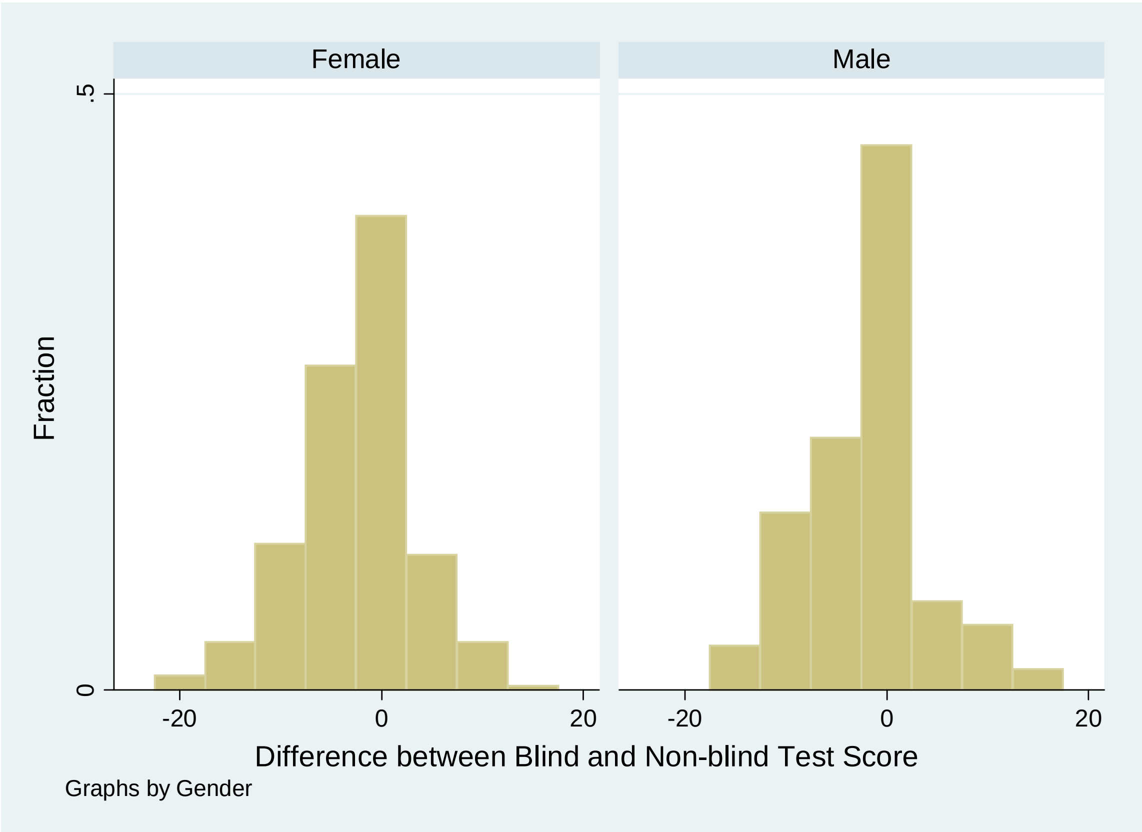

FIGURE 5.3. DIFFERENCES

Moreover, the point estimate of -.487 is rather low, indicating an effect of 4.8 percent of the average test score. Thus, the only conclusion we can draw is that there is no evidence in favor of discrimination of either males or females. The other estimates in column 1 correspond to previously known results. Males do perform worse. The point estimate is highly significant and the size of 1.75 means that males get about a 1.75 lower grade on average, when controlling for the potential discrimination effect and the grading procedure. Finally, the estimate of 2.416 on the variable Non Blind Test means that, on average, the non-blind score is about 2.416 higher than the blind-score. We also add controls for schools and re-grading teacher, which could improve efficiency and also serve as a check for whether the sampling procedure was appropriate. Columns 2 and 3 display the results. The results are qualitatively and quantitatively unchanged, and the point estimates are once more far from significant.

Although the estimated effects are not significantly different from zero, it might still be of interest to investigate the sensitiveness of the point estimates with respect to outliers, since OLS puts high weight on such observations. Column 4 presents the results when removing the four students that were given the highest grade in the non-blind grading and the lowest in the blind grading. Moreover, column 5 shows the result when also removing the group that was either upgraded from Fail to Pass with Distinction or vice versa. It is interesting that removing four outliers decreases the point estimate by approximately one half (column 4). Moreover, when removing 37 students, i.e. about 5 percent of the sample, the point estimate is very close to zero (column 5).

5.4.3. Extensions and robustness. This sub-section first checks for different treatments of the grade scale. Second, we extend the analysis to the students in vocational training and on the academic track. The results are presented in table 5.4.[13]

So far, we have used the national specification of 0, 10, 15, 20 instead of the ordinal grade scale. Thus, we only observe count data and not the underlying continuous test score. Below we run two regressions with different scaling. First, we re-scale the test

[13] Summary statistics is found the appendix.

TABLE 5.3. THE EFFECTS OF DISCRIMINATION ON THE NON-BLIND TEST SCORE

Variables	(1)	(2)	(3)	(4)	(5)
Discrimination	−.487	−.487	−.487	−.270	−.007
	(.487)	(.491)	(.510)	(.481)	(.394)
Male	−1.75***	−1.988***	−2.123***	−1.879***	−2.078***
	(.421)	(.391)	(.428)	(.418)	(.392)
Non Blind Test	2.416***	2.416***	2.416***	2.199***	1.715***
	(.365)	(.366)	(.380)	(.351)	(.315)
Re-grading teacher					
fixed effect	No	Yes	Yes	No	No
School fixed effect	No	No	Yes	No	No
N	1354	1354	1354	1346	1280
R^2	0.071	0.142	0.273	0.069	0.065

Notes: Because of pooled data, the number of observations is twice the number of test takers. Standard errors are clustered at the school level.
*** Significant at the 1% level,**at the 5% level,*at the 10% level

score into 0 for Fail, 1 for Pass, 2 for Pass with Distinction and 3 for Excellent. The results presented in column 1 of table 5.4 indicate a small insignificant effect of −.103, or about 8 percent of the average score.[14] Column 2 presents the result when we define a dependent variable that is 1 if the individual were downgraded and zero otherwise, which we evaluate with OLS using (5.3). Thus, we can interpret the estimate as the probability of being downgraded. The estimated probability is −.049, i.e. the probability of downgrading is .049 lower for male students. Once more, this effect is not significantly different from zero.

It might also be the case that the effects differ across other groups. Columns 3 and 4 present the result when estimating the discrimination effect for two sub-groups; students in vocational training and those on an academic track.[15] Since we split the sample, it is not surprising that the precision becomes worse. However, the point estimates indicate a larger effect in the group of students in vocational training. On the academic track, however, the estimate is very close to zero and of a different sign than in our baseline specification. Once more, we must point out that the effects are

[14] We can also make use of an ordered probit where no cardinal properties are assumed. The results are not qualitatively different from those in column 1.
[15] There are 33 missing observations on the variables vocational training and academic track.

Table 5.4. Alternative Specifications

	(1)	(2)	(3)	(4)
Variables	Re-scale	Prob. of down-grad.	Vocational	Academic
Discrimination	−.103 (.073)	−.049 (.040)	−.860 (.663)	.039 (.700)
Controls	No	No	No	No
N	1354	677	636	684
R^2	0.074	0.003	0.083	0.060

Notes: Because of pooled data, the number of observations is twice the number of test takers, except for (2). Standard errors are clustered at the school level. SE clustered at the school level.
*** Significant at the 1% level,**at the 5% level,*at the 10% level

not significantly different from zero.

Finally, we use (5.4) to investigate whether students on the academic track are discriminated as compared to students in vocational training. We repeat the same steps as in the main analysis and the results are presented in table 5.5.

There is a slightly higher discrimination effect of students on the academic track versus students in vocational training (column 1). The point estimate of –0.849 is about the size of 8 percent of the average score. The point estimate is only significant at the 12 % level, however. Hence, we should interpret the effect with caution. Adding fixed effects for re-grading teacher, column (2), and schools, column (3), neither affects the estimate nor the standard errors in any substantial way. In columns (4) and (5), we have, as previously, removed the outliers. In contrast to gender, the point estimate of discrimination of students on the academic track is not as sensitive. Finally, we add a male dummy control, as presented in column (6), which does not affect the estimates, however.

5.5. Conclusions

In this paper, we ask if gender differences in school performance can be explained by discrimination. The causes of gender differences among individuals before entering the labor market are still not known. Previous studies have suggested that some of the inferior outcomes of male students could be explained by discrimination.

TABLE 5.5. DISCRIMINATION ON THE ACADEMIC TRACK

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Discrimination	−.849 (.543)	−.849 (.545)	−.849 (.563)	−.736 (.531)	−.672 (.466)	−.849 (.543)
Academic	3.604*** (.542)	3.503*** (.502)	2.655*** (.685)	3.562*** (.539)	3.634*** (.535)	3.507*** (.538)
Non Blind Test	2.516*** (.373)	2.516*** (.375)	2.516*** (.387)	2.349*** (.374)	1.987*** (.331)	2.516*** (.373)
Re-grading teacher fixed effect	No	Yes	Yes	No	No	No
School fixed effect	No	No	Yes	No	No	No
Male dummy	No	No	No	No	No	Yes
N	1320	1320	1320	1312	1248	1320
R^2	0.1212	0.177	0.206	0.120	0.123	0.149

Notes: Because of pooled data, the number of observations is twice the number of test takers.
Standard errors are clustered at the school level.
$***$ Significant at the 1%level, **at the 5% level,*at the 10%level

However, when we compare randomly drawn test scores graded with full information about the student (non-blind grading) with test scores graded by teachers with no knowledge of the student's gender (blind grading), we find no evidence of systematic discrimination. In many specifications, the point estimate is close to zero and all our estimates are insignificant. In line with previous studies, we find that females are performing better in the tests than men and that a blind grading procedure leads to more conservative grading. However, these effects are unrelated to discrimination.

References

Altonji, Joseph. G. and Rebecca M. Blank (1999). "Race and Gender in the Labor Market," Chapter 48 of *Handbook of Labor Economics* vol. **3**, eds. Orley Ashenfelter and David Card. Elsevier Science.

Blank, Rebecca M. (1991). "The Effects of Double-blind versus Single-blind Refereeing: Experimental Evidence from the American Economic Review". *American Economic Review* **81**, pp. 1041–1067.

Dee, Thomas S. (2005). "A Teacher Like Me: Does Race, Ethnicity or Gender Matter?," *American Economic Review* **95**, pp. 158–165.

Emanuelsson, Ingemar. and Siv Fischbein. (1986). "Vive la difference? A Study on sex and Schooling,". *Scandinavian Journal of Educational Research* **30**, pp71-84.

Goldin, Claudia. and Cecilia Rouse. (2000). "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians", *American Economic Review* **90**, pp. 715–742.

Holmlund, Helena. and Krister Sund . (2005). "Is the Gender Gap in School Performance Affected by Sex of the Teacher", *Labour Economics* **15**, pp. 37-53.

Lavy, Victor. (2004). "Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment," NBER *Working Paper* (10678).

Lindahl, Erica. (2007). "Comparing teachers' assessments and national test results – evidence from Sweden," IFAU *Working Paper* 2007:24.

Skolverket. (2006). "Könskillnader i måluppfyllelse och utbildningsval," *Technical Report* 286.

Appendix A. Information given to Re-graders.

Hello!

You have now received a package that, besides this document, consists of a number of (35-50) files in Word. Every file represents one student's national test in Swedish. In addition, you should also have received the instruction materials and guidelines provided for the test by the National Agency of Education

The Word documents you have received are computer written versions of the original. The staff that re-wrote the tests on a computer has to the largest possible extent portrayed the language in the original. Naturally, all original errors have been kept. The re-writers have indicated with a [..] when they influenced the text. For example, something that is not readable is denoted [not readable]. In the same way [new page] indicates a new page in the original. Moreover, [-] indicates that the student has hyphenated or tried to hyphenate the word into syllables. For example, if the original says cracker-bread, then the re-writer has written cracker[-]bread.

We want you to prepare yourself by studying the material from the National Agency of Education, in the same manner as you would have done if you had graded the tests while working at a school. Then, we want you to grade the texts on a scale from Fail – Excellent. Please note that a student might have chosen a text that could render no more than a Pass with Distinction (see the instructions from the National Agency of Education).

Moreover, it is of importance that you read the tests on a print-out from the computer and not on a computer screen. This is because we want you to act as you do normally when you grade a national test in Swedish B. However, we are only interested in the grade, so you need not send us your additional comments.

When you have graded a test, you should register the grade in the attached Excel file. Please, note that every test has an Id-number which should be matched with your grade.

Thank you.

Appendix B. Summary Statistics. Vocational Training and Academic Track

Table 5.B.1. Summary Statistics of Female and Male Students

Sample Statistics	N	Mean	Std. Dev.
Vocational	456		
Academic	448		
Vocational Absent	147		
Academic Absent	136		
Variables of Interest			
Non-Blind Grade, Vocational	318	10.535	5.331
Non-Blind Grade, Academic	342	13.289	4.603
Difference		2.754^{***}	
Blind Grade, Vocational	318	8.019	5.417
Blind Grade, Academic	342	11.623	5.243
Difference		3.604^{***}	
$Grade_{iB} - Grade_{iNB}$ Vocational	318	-2.516	6.098
$Grade_{iB} - Grade_{iNB}$ Academic	342	-1.667	5.782
Difference		-0.849^{*}	

*** Significant at the 1% level,**at the 5% level,*at the 10% level