LUND UNIVERSITY

Essays on Sickness Insurance, Absence Certification and Social Norms

Dackehag, Margareta

2007

Link to publication

Citation for published version (APA):
Dackehag, M. (2007). *Essays on Sickness Insurance, Absence Certification and Social Norms*. [Doctoral Thesis (compilation), Department of Economics]. Department of Economics, Lund University.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Essays on Sickness Insurance, Absence Certification and Social Norms

Margareta Ekbladh

Lund Economic Studies Number 143

Distributed by the Department of Economics
Lund University
P.O. Box 7082
SE-220 07 Lund
Sweden
Telephone +46 (0)46 222 00 00
Fax +46 (0)46 222 46 13

Copyright © Margareta Ekbladh, 2007
ISSN 0460-0029

Printed in Sweden by
Holmbergs i Malmö AB, 2007

To Fredrik

Acknowledgements

When I started my Ph.D. studies, this moment seemed so far away, but now, suddenly, it is here. There are many individuals I would like to thank for contributing to making this thesis possible and my time at the department enjoyable. First of all, I am greatly indebted to my supervisor Fredrik Andersson. Thank you for your constant support and unwavering patience when reading and discussing my manuscripts. Your initiated comments on contents and language have been invaluable, to say the least.

I am grateful for the helpful suggestions and inspiring ideas from Sten Nyberg, opponent in my licentiate seminar, and Henrik Jordahl, opponent in my final seminar. I would also like to thank participants at the Meeting of the Public Choice Society, New Orleans 2005, the World Congress on Health Economics, Copenhagen 2007 and the IIPF Congress, Warwick 2007 for comments and suggestions. In addition, I wish to express my gratitude for the opportunity of presenting one of the thesis chapters at the Department of Economics, Örebro.

The comments and suggestions that I have received at the Department of Economics in Lund have contributed much to the thesis. A thank you goes to Carl Hampus Lyttkens and Tommy Andersson who took the time to help me improve various aspects of the thesis. Robin Rander has sorted out many mathematical problems for me - cheers! My gratitude to Andreas Bergh for inspiring and encouraging me to apply for the Ph.D. program and giving his support throughout the "ordeal". I would also like to express my appreciation to Lars Söderström for the opportunity of taking part in the project for Socialförsäkringsutredningen. Agneta Kruse has always been helpful not only with insightful comments on my manuscripts but also with information about what's what and who's who regarding social insurances and the National Social Insurance Agency.

I wish to thank Jeanie, Mariana, Peter, Birgitta, Agnetta and Ann-Charlotte for helping me out with different administrative matters and Jaya Reddy for correcting the English in this thesis.

Financial support from the National Social Insurance Agency, the Jan Wallander and Tom Hedelius Foundation and the Royal Swedish Academy of Sciences is gratefully acknowledged.

I have enjoyed many interesting discussions on other matters than economics during the coffee and lunch breaks at EC1. A big thank you to the people who have been sitting around the "Ph.D. candidate table" over the years: Eric, Anna, Martin, Pernilla, Therese, Maria, Fredrik W, Fredrik G, Robin, Erik, Nils, Pär, Frederik, Andreas, Mårten, Victor, Alfredo, Åsa E, Susanna, Fredrik CA and Åsa H. I would like to thank Mårten for coming

round to room 259 for an espresso, a nice chat or a highly interesting drawing exercise on the whiteboard. I especially wish to thank my room-mate Eric for being such a good friend and keeping me from taking things too seriously. Without our laughs and discussions about practically everything, working on the thesis would have been dull.

I am very fortunate to have Louise and Erika as friends. Louise, talking to you always lights up my day. Erika, thank you for always being there for me.

Without the support of my dear mother and father, Inger and Sven-Åke, I could not have written this thesis. Thank you for your love, encouragement and patience. A big hug to Mats, for being the best brother there is. I would also like to thank my "extended family", Eva, Maths, Margit, Anna and Fredrik for their support and care.

To Fredrik, my one and only, thank you for believing in me. I love you more than words can say.

Malmö, September 2007

Margareta Ekbladh

Contents

1 Introduction **1**
1.1 Background . . . 1
1.2 Summary of the thesis . . . 3
1.2.1 The first essay . . . 3
1.2.2 The second essay . . . 4
1.2.3 The third essay . . . 6
References . . . 9

2 Group-Specific Social Norms and Economic Incentives in a General Welfare System **11**
2.1 Introduction . . . 11
2.1.1 Social interactions and social norms . . . 12
2.1.2 Norms and incentives . . . 14
2.1.3 Outline . . . 15
2.2 The model . . . 16
2.2.1 Separate welfare systems . . . 17
2.2.2 Balanced budget . . . 18
2.3 Outline of specifications . . . 18
2.4 Equilibrium states in separate systems . . . 19
2.5 A shared welfare system . . . 24
2.5.1 Geographical differences . . . 25
2.5.2 Equilibrium states in a shared welfare system . . . 26
2.5.3 Subsidisation in a shared welfare system . . . 29
2.5.4 Geographically segregated labour market . . . 30
2.6 Optimal welfare policy . . . 32
2.6.1 First best in separate systems . . . 32
2.6.2 First best in a shared system . . . 36
2.6.3 Second best in separate systems . . . 37
2.7 Concluding discussion . . . 41
References . . . 42
Appendix A . . . 46

3 Sickness Absence Certification **49**
3.1 Introduction 49
3.1.1 Related studies 51
3.1.2 Outline 54
3.2 The model 55
3.2.1 The workers 55
3.2.2 The physicians 57
3.2.3 The insurer 59
3.3 Sickness absence certification 60
3.3.1 Separating and pooling equilibria 60
3.3.2 Absenteeism and presenteeism 63
3.4 Policy instruments 64
3.4.1 Optimal insurance 64
3.4.2 Changes in remuneration 66
3.4.3 Patient allocation 68
3.4.4 Second opinion 69
3.5 Concluding discussion 71
References 73
Appendix A 80
Appendix B 81
Appendix C 82

4 Occupational Physicians and General Practitioners - Implications for Sickness Absence Certification **85**
4.1 Introduction 85
4.1.1 Background 86
4.1.2 Outline 89
4.2 The model 89
4.2.1 The worker's preferences 89
4.2.2 Sickness absence and the firm 91
4.2.3 The physician's certification practice 92
4.3 OP or GP - implications for absence certification 96
4.3.1 Separating and pooling equilibria 97
4.3.2 Absence certification performed by OPs 97
4.3.3 Absence certification performed by GPs 99
4.3.4 GP and OP available for absence certification 102
4.4 Concluding discussion 103
References 106

Chapter 1

Introduction

1.1 Background

Observations on the Swedish public sickness insurance system serve as a background to the thesis. Over a long period of time the sickness absence rate has been high compared to other insurance systems. Furthermore, the absence rate varies across Swedish regions according to a persistent pattern. In the discussion about the underlying causes to these phenomena, the generosity of the insurance and the gate-keeping functions, including the sick-listing practices of the physicians, are considered (Mikaelsson et al., 2003; Nyman et al., 2002). It is also suggested that attitudinal changes concerning the utilization of the sickness insurance play a part in the explanation (Swedenborg, 2003; Modig & Boberg, 2002).

Studies find that the take-up of social insurance programs is influenced by social interactions (Bertrand et al., 2000; Hedström et al., 2003; Ljunge, 2005). The findings indicate that individual behaviour is not only influenced by economic incentives generated by the programs but also by how people around you act and what they consider to be normal. Hence, changes in the social approval of being absent from work could be a factor behind the development of the Swedish absence rate. Similarly, variations in social work norms may contribute to explaining the geographical absence pattern. However, in theoretical models of social insurances, social norms are generally modelled as universally acknowledged codes of behaviour (Lindbeck et al., 1999; Besley & Coate, 1992; Dufwenberg & Lundholm, 2001). In the first essay, we shed light on the effects on a welfare system of social work norms that regulate the behavior of two different groups, using the model of Lindbeck et al. (1999).

The utilization of the sickness insurance not only depends on the claimant's

behaviour. It is also a contingent on the administrators' interpretation of and compliance with the insurance regulations. For example, in the regions characterized by high absence rates, Dutrieux & Sjöholm (2003) observe a tendency among the physicians to sick-list patients for longer periods than elsewhere. The observation emphasizes the physician's essential role in the insurance, in that the physician's "approval" is required for sickness absences extending over the self-certification period. There is some sociomedical research about physician's sick-listing practices (Wahlström & Alexanderson, 2004). However, in economic literature, the certifying physician has received very little attention. One exception is Hesselius et al. (2005) who empirically study the effect on workers' absence behaviour of the requirement that a physician certifies an ability reduction due to sickness before benefits are granted. In the second essay we make a contribution by analysing the physician's behaviour with regard to sickness absence certification, taking the Swedish situation as a starting point. In the essay we recognize that the physician's acting as a gate-keeper of the insurance may conflict with the traditional role as the patient's advocate. On the basis of observations made about patients and physicians, we allow for individual certification practices among the physicians and different motives for sick leave among the workers. In this context, we also consider the fact that the relationship between health state and working ability may be difficult to ascertain not only for the patient but for the physician as well (Alexanderson et al., 2005).

In the third essay, we continue to investigate the effects of absence certification on the utilization of the sickness insurance. The certification practices of the occupational physician, the OP, and the general practitioner, the GP, are compared to each other. In Sweden, where GPs issue most sickness certificates, there is a discussion about increasing the responsibility of the OPs with respect to sickness absence certification (Public Commissions of the State, 2006:86). The OP's better knowledge about the work-place, improving the possibility of qualified assessments of working ability, is one important argument that is brought up in the discussion. In addition, in some insurance systems, e.g. in Finland and the Netherlands, OPs are the principal physician category that certifies sickness absence (Bergendorff & Larheden, 2003). Moreover, the OP's certification practice is of interest considering the fact that the OP's patient is also employed by the OP's client. It is quite likely that the worker and the firm have different attitudes towards sick leave, forcing the absence certifying OP to deal with a "double loyalty", something which we do not expect the GP to experience.

1.2 Summary of the thesis

1.2.1 The first essay

In the first essay, we study the interaction of economic incentives and social influences in a general welfare state context, based on the model in Lindbeck et al. (1999). The individual's decision whether or not to work depends on the welfare policy, in terms of tax rate and transfer level, and the degree of social disapproval against not earning one's income by working. The strength of the work-related norm is contingent on the population share of norm-followers. The more transfer recipients, the weaker the norm.

We modify the model of Lindbeck et al. (1999) by dividing the population into two groups, called the low-wage group and the high-wage group, governed by group-specific social norms. The two groups are studied in separate welfare systems as well as in a single, shared system. The equilibrium states of the economy, in terms of the welfare policy and the share of transfer recipients, are analysed graphically. We also perform a welfare analysis, taking the social norm into account.

Starting with the graphical analysis, we observe with respect to the group-specific social norms that they counteract the incentive effects generated by the welfare policy. However, for sufficiently generous policies, few individuals find work attractive, which undermines the norm considerably. We also find that group affiliation matters for the equilibrium states. For instance, in the case of separate systems, a larger share of low-wage earners than high-wage earners choose not to work for a given tax rate. This leads to a subsidization from the high-wage group to the low-wage group in the shared welfare system. Thus, the behavior of one group affects the other group through the welfare policy. Continuing with the shared welfare system, we take a look at the case where the groups are separated geographically rather than on the basis of wage, in an attempt to reflect the absence pattern in Sweden. We find that individuals identical with respect to the wage, but belonging to different groups make different choices with regard to work.

The welfare analysis begins with the first-best scenario, where we consider a type of command economy. The government is in control of the policy and the share of transfer recipients. In optimum, the share of transfer recipients is restricted in exchange for a generous policy. As to the social norm, we observe two effects that influence the optimal transfer level in different directions. The direct norm-effect implies an increase, while the indirect norm-effect, pertaining to the endogeneity of the norm, implies a decrease in the optimal transfer level. In the second-best scenario, where individual behaviour is regulated only by means of the welfare policy, we find that the social norm

unambiguosly leads to a more generous welfare policy.

1.2.2 The second essay

In the second essay, we study the physician's certification practice and its consequences for the utilization of the sickness insurance. The physician is supposed to verify that the claimant is eligible for sickness benefits by certifying that his working ability is impaired due to sickness.

We assume that in the event of sickness, workers face the risk of an ability impairment. Hence, ill workers are still able to work, but may experience a reduction in future working ability. In the worker population, there are "motivated" type m workers, who only want to be sick-listed if bad health will affect ability, and there are "less motivated" type l workers who want to be sick-listed irrespective of ability. We assume that the worker's type is unobservable. The share of type l workers depends on an underlying distribution of preferences for labour together with the insurance system, consisting of a premium and a sickness benefit.

To get information about working ability and to get certified sickness absence, the workers must consult a physician. However, the physician may certify absence for a worker without verifying that his ability is impaired due to sickness, since writing "sickness certificates" requires no effort but examining working ability does. We consider physicians with strict as well as lax sick-listing routines. The "strict" physician certifies absence after verifying that ability will be negatively influenced by bad health. The "relaxed" physician certifies absence without examining ability, i.e. absence is certified as a matter of routine. Although the worker does not know the physician's type, he can infer from the effort the physician expends during the consultation whether or not working ability is evaluated.

Workers and physicians form expectations about which type of physician and which type of worker they will meet, respectively, in a sick-listing consultation. We assume that physicians are not anonymous, implying that workers have the possibility to attach expectations to individual physicians. Sickness absence certification is evaluated in terms of the "treatment", i.e. routine or justified absence certification, that the workers receive by the physicians. In the essay we consider different combinations of physician types, but we exemplify with the case where both strict and relaxed physicians are available for sickness absence certification.

Depending on the workers' possibility to form physician-specific expectations about treatment, the type l worker chooses to consult relaxed physicians and the type m worker decides to visit strict physicians. Then, the type l worker receives absence certification without ability examination, whereas

the type m worker is sick-listed if a negative effect on ability is observed by the physician. The expectation of separate treatments defines a separating equilibrium. Such interaction between physicians and workers could be one underlying factor to the regional variations in Swedish sickness absence (Dutrieux & Sjöholm, 2003).

If the workers are not able to hold physician-specific beliefs, their expectations about treatments are based on the composition of the physician population. With some probability, the workers will encounter a strict physician and get justified absence certification. With some probability, the workers will encounter a relaxed physician and get routine absence certification. Thus, a pooling equilibrium describes a situation, where the worker types expect identical treatments. Whereas the type l worker always accepts a physician's offer of a sickness certificate, the type m worker only accepts a certificate if an ability reduction is confirmed. In contrast to the separating equilibrium, where eligible as well as ineligible individuals get certified absence, the pooling equilibrium may also imply that eligible individuals are not sick-listed but continue to work.

Hence, sickness absence certification as a form of eligibility screening may be imperfect. Since the insurer cannot observe the physician's effort, it is not possible to distinguish between eligible and ineligible sickness absentees. We consider some policy instruments that the insurer may employ and discuss their welfare implications.

First, however, we analyse the optimal welfare policy. We find that the optimal benefit level is pushed downwards, owing to the fact that the share of type l individuals in the population responds to changes in the insurance policy. Thereafter, we consider the application of a benefit decrease in a situation where the insurer perceives that regulatory measures are necessary. A benefit decrease makes the alternative not to work less attractive. It also reduces the incentive to claim sickness benefit without sufficient reason, since the share of type l workers decreases. We also consider the instrument of randomly allocating workers to certifying physicians. This may be useful in situations where the personal relationship between workers and physicians increase the probability of sick-listing (see Wahlström & Alexanderson, 2004). The scenario of separate treatments can be described as such a situation. Thirdly, the insurer may introduce a system of second opinion. Physicians' sick-listing practices are thereby monitored and negligent physicians are punished.

When discussing the welfare implications, we take into account the effects on the match between workers and physicians with regard to routine and justified absence certification. Considering that ability unaffected by bad health is assumed to have a non-negative effect on willingness to work, the match between the non-working/working status and the ability of the worker

is also included in the discussion. In view of the fact that there are different types of workers, it is not unexpected that we find positive as well as negative effects associated with all three instruments.

1.2.3 The third essay

The third essay deals with the sick-listing practices of the occupational physician, the OP, in comparison with the ones of the general practitioner, the GP. We recognize that, irrespective of the employment form, physicians have conflicting roles in the context of absence certification. The physician's objective to act in the best interest of the patient may be hard to combine with the gate-keeping function in the sickness insurance. To that is added the OP's "double loyalty", which is caused by the OP's relationship with the client and the client's employees.

We again model sickness absence certification by assuming that workers, who fall ill, may experience a future reduction in working ability and that the effect on ability can only be verified by a physician. There are type l and type m workers who differ in their motivation to work and hence in their interest in information about ability before getting sick leave. Sickness absence certification is analysed as a process determined by the interaction between certifying OPs or GPs and workers. All physicians are non-anonymous by assumption, allowing workers to form expectations about their treatments. Equilibrium states are defined by the workers receiving separate or identical treatments in expectation.

We assume that all physicians are able to observe the worker's health state and can form an opinion about the worker's ability without expending effort. However, the physician's ability assessment does not necessarily determine whether or not the worker receives a sickness certificate. The physician's objective to advocate the patient's interest is reflected in a willingness to take the patient's preferences into account when deciding whether or not to certify absence without regard to ability. Such "altruistic" physicians come in the shape of both GPs and OPs. However, the certification practices of the OP may be influenced by the incentives provided by the firm. We illustrate the "double loyalty" of the OP by constructing two types with different certification practices. The OP, who is loyal to the firm, is restrictive with sick-listing in all cases. The altruistic OP, just as the GP, adjusts his certification practice to the worker type.

Whether altruistic or not, the OP observes working ability perfectly, because of his better knowledge of the claimant's place of work. The GP's assessment is not as accurate. We assume that the inaccuracy implies a tendency to sick-list workers whose ability will not be negatively affected by bad

health. Thus, the GP certifies absence "to be on the safe side".

To the workers, altruistic behaviour implies that they can safely state their preferences and receive the desired treatment. If being type l, the worker wants to get sick-listed irrespective of ability. If being type m, the worker wants to get sick-listed provided that an ability reduction is confirmed. If revealing their types to firm-loyal OPs, however, the information may reach the firm, which in its turn may have consequences for their employment. The problem is that workers do not know whether the OP is altruistic or loyal to the firm prior to the consultation.

When analysing the process of sickness absence certification, we first consider the case where workers only have the possibility of consulting OPs. Different scenarios involving different combinations of altruistic and firm-loyal OPs are discussed. Taking the case where both OP "types" are available as an illustration, we identify a scenario where workers are willing to reveal their type to the OPs and receive separate treatments in equilibrium. If workers can form physician-specific expectations about treatment, they can visit the OPs, whom they believe to be altruistic and get absence certification based on, or irrespective of, ability. There are also circumstances which are not conducive to type revelation and therefore imply identical treatments of the workers in equilibrium. If workers are not able attach expectations to the individual OPs, they may encounter a firm-loyal OP when going for a sick-listing consultation. If meeting such an OP, disclosing information does not alter the outcome of the consultation, both types are sent back to work. The lack of information forces the altruistic OP to decide upon a certification practice without being able to consider the preferences of the workers. We find that in equilibrium the altruistic OP certifies absence, taking the worker's ability into account.

In the case where only GPs are available for sick-listing, the workers have no doubts about revealing their preferences, as GPs are altruistic. However, the inaccuracy of the GP's ability assessment poses a problem to the type m worker. If the GP does not recommend sick leave, type m goes back to work knowing that an ability impairment is not to be expected. If the GP advises him to accept the certificate, there is a probability that he is not really eligible for the sickness benefit. However, despite the risk of misidentification, a consultation still implies an update about type m's ability. After analysing the alternatives of the type m worker, we find that it is not certain that he decides to visit the GP and accept absence certification, if recommended. This means that there are either eligible, but working type m workers, or ineligible, but involuntarily sick-listed type m workers. The latter case can be described as "unnecessary" sick leave, which Eklund & Ossowicki (2005) find evidence for when studying Swedish sickness absentees.

We also consider the case where the workers may choose to consult either OPs or GPs. The certification practices of both have elements that are compatible with the preferences of the worker types. Whether or not certified absentees are eligible and working individuals ineligible, depends on the type of OPs available, and type m's response to the GP's certification practice. Lastly, we discuss the effects of controlling the access to OPs and GPs.

References

Alexanderson K, Brommels M, Ekenvall E, Karlsryd E, Löfgren A, Sundberg L & Österberg M, 2005, "Problem inom hälso- och sjukvården kring handläggning av patienters sjukskrivning", rapport, Institutionen för klinisk neurovetenskap, Karolinska Institutet, Stockholm.

Bergendorff S & Larheden H, 2003, "Att förhindra och förkorta sjukfrånvaro - erfarenheter från fyra länder", RFV Analyserar 2003:16, Riksförsäkringsverket.

Bertrand M, Luttmer E F P & Mullainathan S, 2000, "Network Effects and Welfare Cultures", *The Quarterly Journal of Economics*, 115, 1019-1055.

Dufwenberg M & Lundholm M, 2001, "Social Norms and Moral Hazard", *The Economic Journal*, 111, 506-525.

Dutrieux J & Sjöholm K-R, 2003, "Regionala skillnader i sjukskrivning - Hur ser de ut och vad beror de på?", RFV Analyserar 2003:17, Riksförsäkringsverket.

Eklund M & Ossowicki M, 2005, "Sjukskriven i onödan?", Analyserar 2005:5, Försäkringskassan.

Hedström P, Kolm A-S & Åberg Y, 2003, "Social Interactions and Unemployment", working paper 2003:15, Institute For Labour Market Policy Evaluation (IFAU), Uppsala.

Lindbeck A, Nyberg S & Weibull J, 1999, "Social Norms and Economic Incentives in the Welfare State", *The Quarterly Journal of Economics*, 114, 1-35.

Ljunge M, 2005, "Half the Job Is Showing Up: Returns to Work, Taxes, and Sick Leave Choices", working paper, Department of Economics, University of Chicago.

Mikaelsson B, Ekholm J, Kärrholm J, Murray R, Sandberg T, Söderberg J & Nyman K, 2003, "Den svenska sjukan II - regelverk och försäkringsmedicinska bedömningar i åtta länder", Departementsserie 2003:63.

Modig A & Boberg K, 2002, "Är det OK att sjukskriva sig fast man inte är sjuk?", Memorandum T22785, TEMO.

Nyman K, Bergendorff S & Palmer E, 2002, "Den svenska sjukan - sjukfrånvaron i åtta länder", Departementsserie 2002:49.

Public Commissions of the State (Statens offentliga utredningar) 2006:86, "Mera försäkring och mera arbete", Socialdepartementet.

Swedenborg B (ed.), 2003, *Varför är svenskarna så sjuka?*, SNS Förlag.

Wahlström R & Alexanderson K, 2004, "Chapter 11. Physicians' Sick-listing Practices", *Scandinavian Journal of Public Health*, 32, Suppl (63), 222-255.

Chapter 2

Group-Specific Social Norms and Economic Incentives in a General Welfare System

2.1 Introduction

Welfare state institutions such as social insurances and transfer programs generate incentive effects. However, social influences on individual behaviour may affect the impact of moral hazard. For example, non-participation in transfer programs could be explained by stigmatisation (Moffitt, 1983). Empirical findings suggest that social networks and the code of behaviour established in those networks matter for individual use of social insurances and transfer programs (e.g. Bertrand et al., 2000; Rege et al., 2007). Such influences could be an underlying factor behind the observed variations in utilization across regions (Dutrieux & Sjöholm, 2003; Blank & Card, 1991; McCoy et al., 1994).

Social interaction effects have important policy implications, since the direct effect on individual behaviour of a policy change is reinforced by the social multiplier and may generate multiple equilibria.[1]

The theoretical contributions with regard to social interactions and social insurances analyse the effects of a universally acknowledged code of behaviour (Lindbeck et al., 1999; Besley & Coate, 1992; Lindbeck & Nyberg, 2006; Lindbeck & Persson, 2006; Moffitt, 1983; Kolm, 2005; Dufwenberg & Lundholm, 2001; Bird, 1999). Although recognized as important, the effects of selective social networks are not considered (Akerlof, 1980; Lindbeck et al.,

[1]The social multiplier is discussed in more detail by e.g. Brock & Durlauf (2001) and Glaeser et al. (2003).

1999). That is, what is considered normal or deviant behaviour may differ between sub-groups.

In the context of a general welfare system, we study the effects of economic incentives and social influences on individual behaviour. For our purposes we find the model presented in "Social Norms and Economic Incentives in the Welfare State" by Lindbeck, Nyberg & Weibull (1999)(from hereon called LNW) useful. We modify the model by dividing the population into two groups, separated by group-specific social norms that stipulate that individuals should earn their incomes by working. The behaviour of the groups are studied in separate systems as well as in a single, shared welfare system. The policy and the share of transfer recipients will also be considered in a concluding welfare analysis.

2.1.1 Social interactions and social norms

People tend to compare themselves with similar others, those sharing certain characteristics or experiences (Festinger, 1954). Thus, social interaction or network effects hinge on interpersonal dependence. They arise when individual behaviour is influenced by the actions of the reference group, e.g. when the probability of an individual applying for benefits increases as more of his peers apply. Thus, the social influence depends on numbers, in terms of the share of the reference group deviating from what is considered normal, and has a spatial element in that the relevant reference group tends to be selective rather than universal (see Hedström et al. (2003) and Rege et al. (2007)). With regard to the origins of sub-groups with different codes of behaviour Stutzer & Lalive (2004) and Dutrieux & Sjöholm (2003) point to historically weak labour markets when discussing regional variations in unemployment and sickness absence, respectively. Similarly, Frykman & Hansen (2005) argue that there are local sickness absence cultures in Sweden springing from different living conditions that over time have left their mark on the communities.

Empirical studies give evidence of social interaction effects. For example, a US-based study finds that living in areas with many individuals from one's own language group is positively correlated with the mean welfare participation of the language group (Bertrand et al., 2000). When studying the use of social assistance among refugees in Sweden, Fredriksson & Åslund (2003) find that the share of recipients in the group increases individual propensity to draw benefits. However, the size of the ethnic network does not. Hedström et al. (2003) and Lalive (2003) find evidence of social interaction effects in the take-up of unemployment insurance and in the unemployment duration, respectively. Another example is the study by Rege et al. (2007), where

the participation rate in the Norwegian disability pension program of previously employed neighbours is found to increase workers' entry rate to the program. Lastly, with respect to sick leave, Ljunge (2005) observes that the probability of claiming sickness benefits increases with the share of claimants on the municipal level. Lindbeck et al. (2004) find that a person's living area and work place have explanatory power when studying factors behind the sick-listing pattern in Sweden, which could be indicative of effects from social interactions.[2]

Bertrand et al. (2000) identify two important mechanisms by which the behaviour of others affects individual behaviour, information exchange and social norms.[3] The empirical studies are not always able to distinguish between the mechanisms, which, of course, may interact. However, speaking in favour of social pressure is the observation that work tends to be regulated by strong social norms (Elster, 1989). Measuring the share in Swiss communities that voted for a reduction in the unemployment benefit in a referendum as the strength of a social work norm, Stutzer & Lalive (2004) find that a stronger norm is assocatied with shorter unemployment duration and lower subjective well-being among unemployed individuals. Furthermore, findings in Clark (2003) support the hypothesis that there are social and psychological costs connected to being unemployed, and that the size of these costs is related to what is considered normal in the reference group of the individual. Results in Hedström et al. (2003) point in the same direction, in that transition rates out of unemployment are influenced by the unemployment level in the "neighborhood-based reference groups".[4]

A social norm may emerge in situations characterized by a strong positive or negative externality (Coleman, 1990), and may take the form of a shared understanding of a group about what actions are forbidden, permit-

[2]With respect to the work-place, the findings in Lindbeck et al. (2004) could be explained by the effect of the work environment on individual behaviour. In contrast, in an international study on work absence Drago & Wooden (1992) draw the conclusion that norm formations on worker level and firm level are likely explanations for individual absence behaviour and firm absence rates, respectively. See also Arai & Skogman Thoursie (2004).

[3]Social norms are associated with what Hedström et al. (2003) call desire-based interactions. However, such interactions could also be unrelated to social norms. Hedström et al. (ibid) point out that being unemployed is probably more rewarding if you can spend time with other unemployed individuals, a phenomenon Rege et al. (2007) call leisure complementarities. However, we argue that social norms and leisure complementarities tend to reinforce each other, since an increasing number of unemployed probably weakens social stigma and improves the social life, in terms of the number of contacts, of the unemployed individual.

[4]Psychological studies corroborate the detrimental effect of unemployment on life satisfaction (Paul & Moser, 2006; Lucas et al., 2004; Winkelmann & Winkelmann, 1998).

ted or obligatory. In order to uphold the social norm, deviating behaviour is punished (Ostrom, 2000; Elster, 1989). The sanction can be categorized as being either external or internal (Coleman, 1990). In the first case, other individuals punish the deviator by social exclusion, loss of reputation, invoking feelings of shame etc.[5] In the second case, the individual breaking the norm punishes herself by feeling anxiety, embarrassment or guilt. Through the process of internalization, the social norm transforms from a constraint to a preference (Lindbeck, 1995).[6] The classification of the sanction is not necessarily exclusive. Norm-deviating behaviour could trigger both external and internal sanctions (Elster, 1989).[7]

The emergence and the survival of a social norm is also a question of numbers. In an early model of social customs, Akerlof (1980) assumes that the strength of the norm depends on the community share that comply with the norm in terms of behaviour as well as the share that believe in the norm. In the LNW model, the focus lies on the manifestation of beliefs in terms of behaviour. The strength of the work norm depends on the share of norm-followers, i.e. the share of working individuals. In contrast, Besley & Coate (1992) assume that there is a fixed share of potential benefit recipients, needy and non-needy poor. In their taxpayer resentment model, the social stigma increases with the number of non-needy recipients as well as the level of benefits. In a similar fashion, Moffitt (1983) allows stigma to arise from program participation per se as well as from the generosity of the program. The differences between the first two examples and the last to examples with respect to the social norm may originate from which type of welfare state program that is considered, universal or targeted programs (see the discussion in Lindbeck (1995)).

2.1.2 Norms and incentives

LNW analyse the interaction between economic incentives and a social norm in the context of a general welfare system, which covers the entire population, as do the prescriptions of the social norm. The individual has two choices to make, a political and an economic one. We choose to focus on the latter. The

[5]Due to costly effort to punish norm deviators, Axelrod (1986) finds that a "metanorm", according to which individuals who fail to show disapproval of deviant behaviour are punished too, is required to uphold the norm. See also Bird (1999) for an application of a metanorm in the context of a welfare system.

[6]Socialization is considered an important means of internalizing norms, see Lindbeck & Nyberg (2006) for an application in a model of social insurance.

[7]However, the grounds on which individuals adhere to social norms may be of importance policy-wise. Kreps (1997, p. 359) argues that "opacifying welfare administration" increases the number of recipients in cases where norms work through external pressure.

individual maximizes utility under the restrictions introduced by the welfare state policy, i.e. the income tax rate and the transfer, and by the strength of the norm. When making his economic decision, the individual faces a binary choice, to work full-time and receive his after-tax wage or not to work and live on a public transfer. Opting for the second alternative, however, means breaking the social norm, which favours work. The embarrassment of accepting the public transfer reduces the utility of not working. The social norm is assumed by LNW to be dependent on the number of people adhering to the norm, i.e. the number of people working. The more people who choose not to work, the weaker the social norm gets, which in its turn affects the incentives to work. In other words, the social norm is endogenous to the model. Consequently, multiple equilibria may arise.

2.1.3 Outline

Contrary to LNW we consider a population that is separated into two groups: low-wage earners and high-wage earners. The groups are under the influence of separate social norms. The behaviour of the individuals in one group only affects the strength of the social norm regulating that group. In this manner, we make a rough distinction of the "relevant others", i.e. the people whose opinions and actions may influence the behaviour of the individual.

We study the segregated population in different settings. Equilibrium states, in terms of policy and share of transfer recipients, are analysed graphically. As a first step we consider the case of total separation, where low-wage and high-wage earners share neither social norm nor welfare system. We study the effects of the group characteristics on equilibrium states. This exercise could also be considered as an analysis of the effects of a progressive tax system. In a shared welfare system, the two groups make their decisions subject to the same policy. We consider a separation between groups not only based on wage, but also on another characteristic. We have regional variations in mind, but it could also pertain to other intra-regional differences. Moreover, we extend the model of LNW further, by performing a welfare analysis for the two system settings. We study first-best as well as second-best scenarios, differing with respect to the control of the government over individual behaviour. The effect of the social norm on the optimal welfare policy is analysed in both scenarios.

In the graphical analysis we find, as expected, that the social norm influences individual behaviour. The more severe the sanction, the less attractive it is for the individual to accept the transfer. More importantly, we observe differences between the groups. In separate welfare systems, for instance, more low-wage earners than high-wage earners choose not to work at a given

tax rate. Furthermore, the transfer recipients in the low-wage group accept a comparatively small transfer. In the case of a shared welfare system we observe a substantial subsidisation of the low-wage group. This implies a more frugal treatment of the high-wage group compared to the context of separate systems. As to the welfare analysis we find that in a first-best context it may be optimal to restrict the number of recipients and thereby increase the transfer level. In the first-best as well as in the second-best scenario, we observe that the social norm may imply a more generous welfare policy.

2.2 The model

There is a continuum of individuals, who belong to two groups, distinguished by different wage distributions: the low-wage group L and the high-wage group H. All individuals earn wages distributed according to two continuously differentiable cumulative probability distributions, Φ_L and Φ_H, defined on the separate wage intervals of the groups. The probability densities are ϕ_L and ϕ_H. In addition, no individual earns a zero wage.

Irrespective of group affiliation, the individual has to decide whether or not to work. If she chooses to work, she works full-time and can consume her after-tax wage earnings, and enjoy some spare time, which we normalize to zero. Deciding not to work, the individual receives a government public transfer, which is exempt from taxation and granted to everyone who has no other income. In that case, however, she has to take into consideration the effects of a social norm which favours work and proscribes living off other people. When deviating from the norm, the individual may experience feelings of embarrassment or guilt. However, the disutility is assumed to decrease as more people choose not to work. The smaller the population share of norm-adhering, i.e. working, individuals, the weaker the social norm becomes. Hence, the social norm is endogenous in the model.

The low-wage group and the high-wage group have developed separate norms. Therefore, the decisions made by members of one group are not affected by the strength of the social norm regulating the behaviour of the other group. We denote the population share of transfer recipients in the separate groups x_L and x_H. The disutility of accepting the transfer is a function of the group share of transfer recipients, $v_L(x_L)$ and $v_H(x_H)$. When making her utility-maximizing choice, the individual considers her after-tax wage, the transfer level and the share of transfer recipients in his group, i.e. the strength of the social norm.

The welfare system consists of a policy pair, the tax rate and the transfer level. Our intention is to study the groups in different welfare system set-ups.

Firstly, the behaviour of low-wage and high-wage earners is analysed when they are subject to separate welfare systems. Secondly, we let the two groups be governed by the same welfare system.

2.2.1 Separate welfare systems

Following the LNW model, we assume additively separable utility. Moreover, all individuals have identical preferences over consumption and leisure. An individual with wage w works if and only if

$$u\left((1-t_i)w\right) > u(T_i) + \mu - v_i(x_i), \quad i = L, H, \tag{2.1}$$

where $\mu \in \Re$ is the utility difference between enjoying leisure time and the intrinsic utility an individual may experience by working. The utility function is strictly concave in consumption. Furthermore, we assume that $u(c) \to -\infty$ as $c \to 0$ and $u(c) \to +\infty$ as $c \to +\infty$. With respect to the punitive norm sanction, v_i is non-increasing in the share of transfer recipients x_i.

For every tax rate t_i, transfer level T_i and expected share of transfer recipients x_i, there exists a unique critical wage rate, w_i^*, such that all individuals with lower wages decide to live on public transfer. The critical wage rate is found by solving the equation (2.1) with equality. The unique solution w_i^* is increasing in tax rate and transfer and non-decreasing in the group share of transfer recipients,

$$w_i^* = \omega_i(t_i, T_i, x_i) = \frac{1}{(1-t_i)} u^{-1}[u(T_i) + \mu - v_i(x_i)]. \tag{2.2}$$

Like LNW, we call a point $s_i = (t_i, T_i, x_i)$ a state of the economy for group i and a pair $p_i = (t_i, T_i)$ a policy in the welfare system of group i.

If the expectations about the fraction of transfer recipients in the separate populations are correct, then x_i equals the share of the group whose wages are lower than the critical wage rate w_i^*,

$$x_i = \Phi_i\left(\frac{u^{-1}[u(T_i) + \mu - v_i(x_i)]}{(1-t_i)}\right). \tag{2.3}$$

As LNW point out, the right-hand side is a continuous function of x_i, mapping the interval $[0, 1)$ into itself. Thus, equation (2.3) is a fixed-point equation in x_i, with exogenous parameters t_i and T_i. Consequently, there exists at least one population share x_i in each group, which satisfies equation (2.3) for every policy p_i. Whether there exist multiple solutions or not depends on the characteristics of the functions u and v_i, as well as on the wage distribution Φ_i. A share of transfer recipients x_i^*, which solves equation

(2.3) is called an equilibrium population share of transfer recipients and the associated tax and transfer an equilibrium policy. Hence, $s_i = (t_i, T_i, x_i^*)$ is an equilibrium state of the economy.

2.2.2 Balanced budget

We also require that public spending equals public revenues. We aggregate income by using the truncated expected value function $\Psi_i(w) = \int_w^\infty a\phi_i(a)da$. By means of Ψ_i individual wages above w are summed up and normalized to per capita units. Thus, the tax base in state s_i is Ψ_i evaluated at w_i^*. In equilibrium, the critical wage rate equals $\Phi_i^{-1}(x_i)$. Consequently, public revenues equal $t\Psi[\Phi_i^{-1}(x_i)]$. Public spending depends on the benefit level and the number of transfer recipients, $T_i x_i$. Thus, in an equilibrium state s_i the government budget is balanced if and only if

$$T_i x_i = t_i \Psi_i \left[\Phi_i^{-1}(x_i)\right]. \tag{2.4}$$

By examining equations (2.3) and (2.4), LNW show that for every share of transfer recipients there exists a unique policy pair $p_i = (t_i, T_i)$, such that state s_i is a balanced equilibrium state. They also show that there exists at most one balanced equilibrium share of transfer recipients for every policy. On the basis of these observations there exist two continuous functions, $f_i : [0,1) \to [0,1)$ and $F_i : [0,1) \to \Re_+$, which map $x_i \in [0,1)$ to the unique tax rate, $t_i = f_i(x_i)$, and the unique transfer, $T_i = F_i(x_i)$. The tax rate, t_i, and the transfer, T_i, constitute the elements in a balanced equilibrium state $s_i = (t_i, T_i, x_i)$. When $x_i = 0$, $f_i(x_i) = 0$ and $F_i(x_i) = w_i^{\min}$, where $w_i^{\min}$ is the lower boundary of the wage interval on which the wage distribution is defined (and, consequently, $w_i^{\max}$ is the higher boundary). When $x_i \to 1$, $f_i \to 1$ and $F_i \to 0$. Because of the endogenous social norm, f_i may be non-monotonic and F_i non-concave.

2.3 Outline of specifications

When studying the equilibrium states of the economy, we alternate between illustrating the equilibrium policy $p_i = (t_i, T_i)$ for a given x_i and the equilibrium pair of transfer and share of transfer recipients, (T_i, x_i), for a given tax rate t_i.

For the simulations, we use the parametric specifications of LNW for $f_i(x_i)$ and $F_i(x_i)$ (see Appendix A). Like LNW, we assume that utility from consumption is logarithmic, $u(c) = \ln(c)$ for all $c > 0$. In order to observe the influence of the social norm, we study two contrasting cases, which differ

with respect to the severity of the punishment associated with deviating behaviour. We work with $v_i(x_i) = 0$ and $v_i(x_i) = -\ln(x_i)/2$, where $v_i \geq 0$ and defined for $x_i \in (0, 1)$. For the first function the disutility is equal to zero. The second function inflicts discomfort on the norm-breaking individual, but as more people accept the transfer, the disutility decreases.

We assume that wages are distributed uniformly on $[0, 1]$. The graphical presentation is primarily based on the case when the group-wise distributions, Φ_L and Φ_H, are uniform on $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$, respectively.[8] Furthermore, we contrast the case of equal group sizes with the case where the low wage group is defined on $[0, \frac{3}{4}]$ and the high wage group on $[\frac{3}{4}, 1]$, and also with the case where the group-wise distributions are uniform on $[0, \frac{1}{4}]$ and $[\frac{1}{4}, 1]$. When the relative sizes of the groups are important to the results, they will be commented on.

For the graphical illustrations in LNW, wages are distributed according to the Weibull distribution. Although the underlying assumptions are essentially the same, our graphical results cannot be directly compared to theirs. However, we are able to make some general observation as to effects of the distribution in equilibrium states. In addition, LNW also use the Weibull distribution for the punitive sanction v. Different specifications notwithstanding, the social norm functions similarly in both cases. The punishment is strong for the first transfer recipients but weakens considerably as the majority of the population stops working. We aim to apply LNW's specifications in the context of group-specific norms in further research.

2.4 Equilibrium states in separate systems

Our main focus is on the comparison of group-wise behaviour in terms of the population shares of transfer recipients in response to a certain welfare policy. Thus, analysing the low-wage group and the high-wage group separately, may also be characterized, to some extent, as studying the effects of a progressive tax system. Furthermore, when comparing the low-wage and high-wage groups in separate welfare systems we consider the effects of varying the wage distribution. First, however, the comparative statics effects of the social norm are studied.

In Figures 1 and 2, we illustrate the consequences for the tax rate and the transfer level from a social norm with endogenous intensity compared to the "non-social" case separately for the two groups. In both figures, the

[8] Hence, for the simulations we diverge somewhat from the assumptions of the model. We allow for the wage intervals to include non-negative values and the worker earning $w = \frac{1}{2}$ belongs to both groups, as the wage intervals of the two are not disjunct.

dark solid graphs, called "$T_L, v(x) = 0$" and "$T_H, v(x) = 0$", show the transfer levels in the case of $v_i(x_i) = 0$, while the light solid graphs, called "$T_L, v(x) = -ln(x)/2$" and "'$T_H, v(x) = -ln(x)/2$"', show the transfer levels in the case of $v_i(x_i) = -ln(x_i)/2$ for a given x_i. Correspondingly, the dark, dotted graphs, "$t_L, v(x) = 0$" and "$t_H, v(x) = 0$", describe the tax rates for a given x_i in the non-social case, whereas the light, dotted graphs, "$t_L, v(x) = -ln(x)/2$" and "$t_H, v(x) = -ln(x)/2$", show the tax rates for a given x_i in the "social" case.

It pertains to both groups that, for a given tax rate t_i, the proportion of transfer recipients is lower if norm-deviating behaviour is punished. Hence, the benefit level can be raised, without compromising the budget requirement (2.4). For example, in Figure 1 for $t_L = 0.2$, the share of recipients is around 0.4 in the social case compared to 0.5, approximately, in the non-social case. Consequently, the transfer level is lower in the latter case than in the former case for the tax rate in question. In other words, the constraining effects of the social norm enables the government to set a more generous welfare policy. To finance a larger transfer to the share of recipients a higher tax rate is needed. Again, we exemplify with Figure 1. For $x_L = 0.4$, the transfer level in the social case is higher than in the non-social case, but so is the tax rate. As may be discerned in both figures, for large x_i, the social norm eventually collapses and equilibrium policies in the social and the non-social cases converge.

The strength of the social norm is greatest when norm-adhering individuals are many. Therefore, we expect the transfer to reach its maximum level for the first x_i and then to decrease for larger shares. However, that is only observed for the high-wage group, for which the transfer curve in the social case is monotonically decreasing (see Figure 2). In contrast, the corresponding graph for the low-wage group (see Figure 1), keeps its concave shape, familiar from the non-social case. The (non)monotonicity of the transfer functions is a result of the differing wage distributions. For the first low-wage earner to decide not to work, a smaller transfer is required than for the first transfer recipient in the high-wage group. The equilibrium transfer levels in the social case are also a reflection of the group-wise differences with respect to the average wage.

Figure 1 Equilibrium policy $p_L = (t_L, T_L)$. Effects of the social norm.

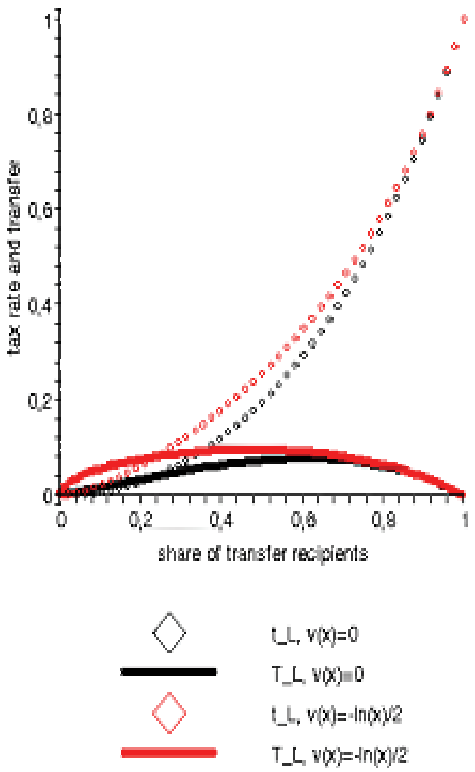
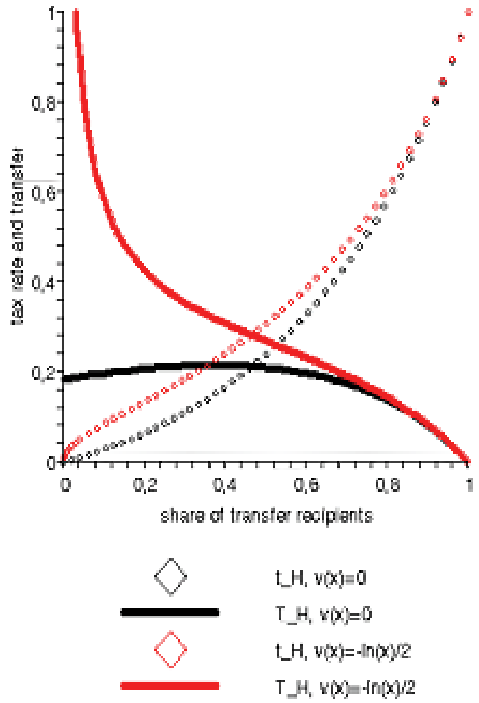


Figure 2 Equilibrium policy $p_H = (t_H, T_H)$. Effects of the social norm.



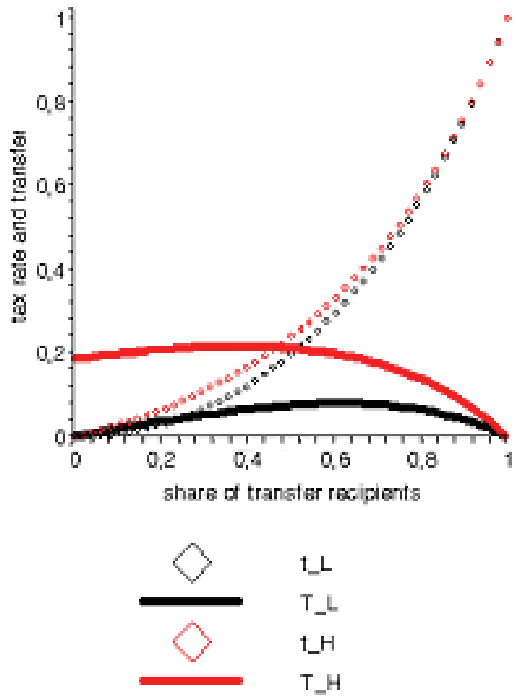
Apart from the effects of the social norm, the low-wage earners and the high-wage earners generally differ in terms of recipient share and the associated welfare policy in equilibrium states. In Figure 3 we compare the equilibrium policies for the two groups when deviating behaviour is not punished. The dark solid graph, called "T_L", shows the transfer level in the low-wage group, whereas the dotted, dark graph "t_L" shows the tax rate for a given share of transfer recipients x_L. Correspondingly, the light, solid graph "T_H" and the dotted, light graph "t_H" illustrate the transfer level and the tax rate in the high-wage group for a given share of transfer recipients x_H. We observe in the figure that for any given recipient share the transfer is smaller in the low-wage group than in the high-wage group. The tax rate is also lower in the former group until the group-wise populations mainly consist of transfer recipients. Then the tax rates of the two systems converge. This means that the share of transfer recipients is larger in the low wage group than in the high wage group for all but high tax rates.[9] The observations are explained by the differences in the wage distribution. In contrast to the low-wage group, a low tax rate on working individuals in the high-wage group generates a considerable contribution to public revenues. Furthermore, more low-wage earners than high-wage earners decide not to work for a given critical wage. This implies that a given tax rate is associated with a lower transfer level but with a larger recipient share in the low-wage group. Thus, Figure 3 shows that the outcome of a certain welfare system depends on the characteristics of the population.

Next, we consider the comparative-statics effects of varying the relative sizes of the groups. This means making alterations to the wage distributions of the two groups. We find that, for a given social norm and a given share of transfer recipients, the groups encounter in general the same tax rate regardless of their relative size. However, more members in the low wage group leads to a higher average wage, which causes the transfer level to increase for every x_L. For the high wage group the opposite effect occurs. The more members, the lower the average wage and the lower the transfer level for every x_H.[10]

[9] When comparing equilibrium states in the case where the low-wage group is regulated by the norm associated with $v_L(x_L) = -\ln(x_L)/2$ and the high wage group by the one associated with $v_H(x_H) = 0$, the pattern alters somewhat. Although the transfer level in the low-wage group is lower, the social norm influences behaviour in such a way that group-wise tax rates and shares of transfer recipients are more or less identical.

[10] Similarly, LNW find that changing the wage distribution by multiplying all individuals' wages with a positive factor λ does not affect the tax function but leads to a proportional change in the transfer function.

Figure 3 Equilibrium policies $p_L = (t_L, T_L)$ and $p_H = (t_H, T_H)$, when $v_L(x_L) = 0$ and $v_H(x_H) = 0$.



2.5 A shared welfare system

In the single, shared welfare system, there are still group-specific norms but all individuals now maximize utility subject to tax rate t and transfer T. The common policies $p = (t, T)$ change the expressions of the critical wages and, hence, the group-wise shares of transfer recipients,

$$w_i^* = \omega_i(t, T, x_i) = \frac{1}{(1-t)} u^{-1}\left[u(T) + \mu - v_i(x_i)\right], \quad i = L, H, \tag{2.5}$$

$$x_L = \Phi_L\left(\frac{u^{-1}[u(T) + \mu - v_L(x_L)]}{(1-t)}\right), \tag{2.6}$$

and

$$x_H = \Phi_H\left(\frac{u^{-1}[u(T) + \mu - v_H(x_H)]}{(1-t)}\right). \tag{2.7}$$

The equilibrium state now comprises tax and benefit, common to all individuals, and group-specific fractions of transfer recipients, $s = (t, T, x_L, x_H)$.

When balancing the budget, there are transfer recipients and tax payers from both groups to consider. We exemplify with the case where the groups are of equal size,

$$T(x_L + x_H) \leq t[\Psi_L(\Phi_L^{-1}(x_L)) + \Psi_H(\Phi_H^{-1}(x_H))].^{11} \tag{2.8}$$

In section 2.5.2 we encounter a certain class of recipient shares, $(x_L, x_H) \in [0, 1) \times 0$ and $(x_L, x_H) \in 1 \times (0, 1]$. For this class the population share of transfer recipients increases with the generosity of the welfare system. For relatively modest welfare policies only low-wage earners decide not to work, but as the tax rate increases we will reach a point where the first member of the high-wage group stops working. For higher tax rates, the share of transfer recipients consists of the entire low-wage group as well as an increasing amount of high-wage earners. Thus, the social norm governing the behaviour of the low-wage earners is not as intense as to generate a situation where low-wage earners choose to work while high-wage earners accept the transfer.

On the basis of examples, it is likely that there exists at least one combination (T, x_L, x_H) for a given t. In addition, we find that there exists a unique policy (t, T) for every pair (x_L, x_H), in the same way as in the case of separate systems (see section 2.2.2).

2.5.1 Geographical differences

So far we have defined the two social networks on the basis of income. However, with regard to the extension of the network, the groups may differ geographically. That is, individuals earning the same wage could belong to different groups depending on where they live.[12] For a given welfare policy, an individual belonging to the low-wage group may choose to accept the transfer, due to a weak social norm. For the same policy, another individual, earning the same wage but belonging to the high-wage group, decides to work, since receiving the transfer is strongly stigmatised. Such a norm structure could play a part in explaining the regional variations that have long been observed for the sickness absence rate in Sweden (Dutrieux & Sjöholm,

[11] The group-specific variables are multiplied by a factor reflecting the relative size of the group.

[12] However, the non-wage based group affiliations need not be only geographical but could also pertain to other "cultural" differences on an intra-regional level, e.g. group formations among blue- and white-collar workers, groups based on ethnicity etc., see e.g. the results in Lindbeck et al. (2004) that the place of residence as well as the work-place influence the sickness absence rate.

2003; Olsson, 2006).[13]

We characterize geographical differences with partly overlapping wage intervals. Thus, under the assumption that the social norm does not stop members of the high-wage group from accepting the transfer until all low-wage earners are transfer recipients, we must also deal with the class $(x_L, x_H) \in (0,1) \times (0,1)$ for the overlapping parts of the wage intervals. Since x_L is relatively large and x_H is relatively small for the relevant wage interval it is likely that $v_L(x_L) < v_H(x_H)$ for a given (t, T), provided that both groups are influenced by punitive sanctions for norm-deviant behavior. This implies that the critical wages in the two groups differ.

2.5.2 Equilibrium states in a shared welfare system

We study the welfare system expressing x_L and x_H as functions of the policy pair (t, T). For tractability reasons, we only allow for two variations concerning the group-specific social norms in the shared system. In the first case, deviating behaviour is not sanctioned in neither the low-wage group or the high-wage group. In the other case, the low-wage group is regulated by the social norm associated with disutility $v_L(x_L) = \frac{-\ln(x_L)}{2}$, while the disutility function $v_H(x_H) = 0$ remains for the high-wage group.[14]

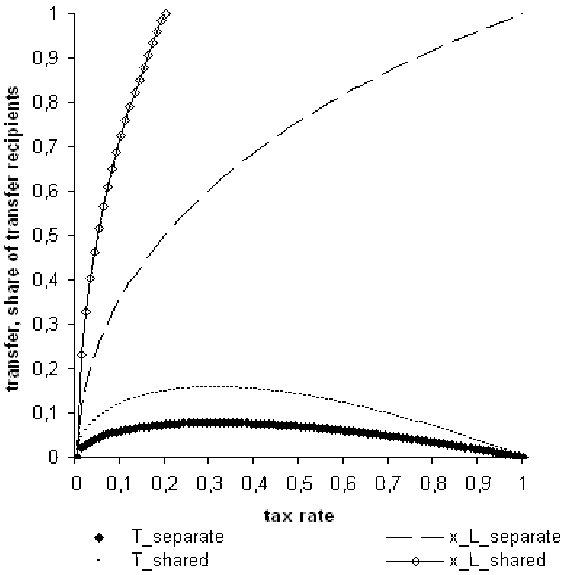
Taking the non-social case as an example, we observe in Figure 4 that the shared welfare system is more generous to the low-wage group than the separate system. The transfer level in the shared system, shown by the graph "T_{shared}", is larger than the level in the separate system, shown by the graph "$T_{separate}$", for practically all tax rates. The increased generosity leads to a larger share of transfer recipients in the shared system for every tax rate, which may be observed by comparing the graph "$x_{L,shared}$" with the graph "$x_{L,separate}$" for a given tax rate t. This result in its turn implies that all members exit the labour market at a lower tax rate in the shared system.

Continuing with the non-social case, in Figure 5 graphs "$x_{H,separate}$" and "$x_{H,shared}$" show that the first transfer recipient in the high-wage group appears at a considerably higher tax rate in the shared system than in the separate system. However, except for large values of t, the transfer level is lower in the system common to both groups. This may be observed by relating the graph "T_{shared}" to the graph "$T_{separate}$". Thus, sharing the system with

[13]Using a spatial diffusion model, Holm & Öberg (2004) find local and regional effects on the duration of sickness absence.

[14]With regard to the relative strengths of the group-wise social norms, there is, to our knowledge, no empirical evidence of low-income earners being less, or for that matter more, inclined to accept a voluntary labour market exit in their group than high-income earners for a given share of transfer recipients.

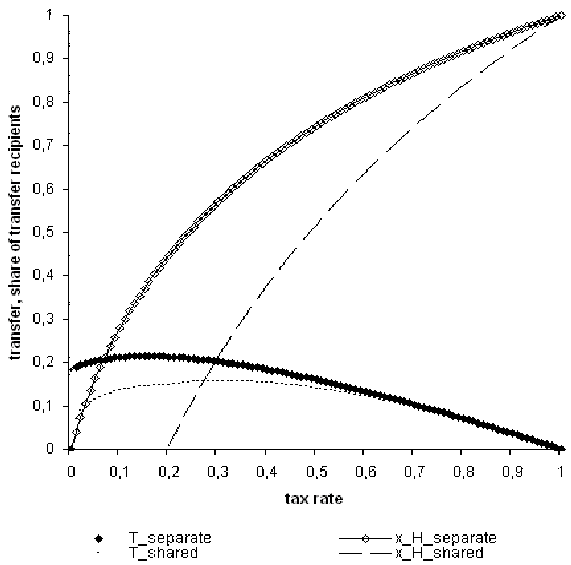
Figure 4 Comparison of separate and shared systems for group L.



the low-wage group means higher tax rates but smaller transfers for nearly all x_H. Hence, in the shared welfare system, members of the low-wage group seem to accept the transfer because it is more attractive, while the individuals from the high-wage group exit the labour market because of the income tax rate. These observations are valid for separate wage intervals, irrespective of wage distribution, as the addition of the high-wage group implies more resources to the transfer recipients in the low-wage group.

If norm-deviant behaviour of the low-wage group is punished, a more generous welfare system is achieved. The larger benefit accrues only to transfer recipients in the low-wage group, however. The share of transfer recipients in the low-wage group in the case of a punitive sanction, shown in Figure 6 by the graph "x_L, case 2", is smaller compared to the share in the non-social case, shown by the graph "x_L, case 1" for low tax rates. However, we observe that the punitive sanction has no effect on the policy in which all low-wage earners accept the transfer. That is, the social norm reduces the proportion of transfer recipients in the low-wage group for every t, while the rate at which individuals accept the transfer is higher because every new transfer recipient reduces the disutility of deviating from the norm. In contrast, the share of transfer recipients in the high-wage group remains unchanged, see the graph "x_H". For all policies where there are still working individuals

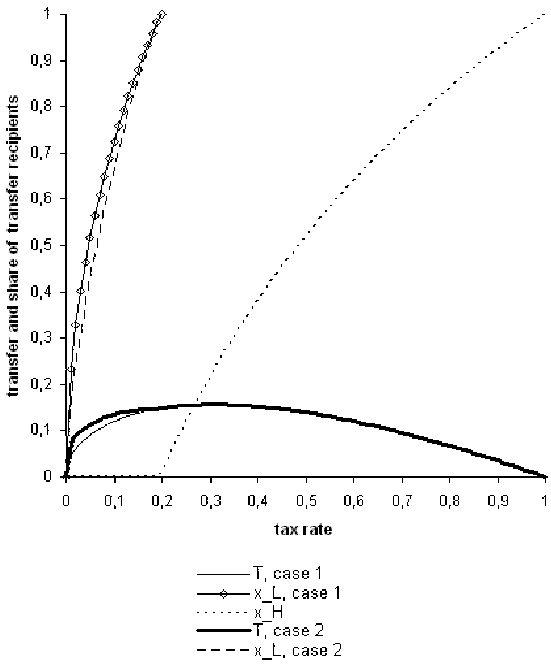
Figure 5 Comparison of separate and shared systems for group H.



in the low-wage group, no high-wage group member accepts the transfer. Thus, we are dealing with the class of recipients $(x_L, x_H) \in [0, 1) \times 0$ and $(x_L, x_H) \in 1 \times (0, 1]$. The first member of the high-wage group accepts the transfer at the tax rate at which the last low-wage earner stops working.

Regarding the relative size of the groups, we observe that when increasing (decreasing) the low-wage group, the tax rate at which the last low-wage earner and the first high-wage earner accept the transfer is raised (reduced). In contrast to the situation in the separate system case, the transfer level for a given t is not influenced by the relative size of the groups, since the group-wise proportion is accounted for.

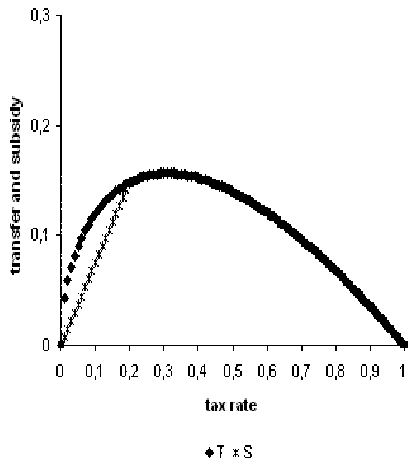
Figure 6 Shared system with equilibrium policy $p = (t, T)$.
Case 1: $v_L(x_L) = 0$ and $v_H(x_H) = 0$.
Case 2: $v_L(x_L) = -ln(x_L)/2$ and $v_H(x_H) = 0$.



2.5.3 Subsidisation in a shared welfare system

The equilibrium states generated in the shared welfare system imply a substantial subsidy being transferred from the high-wage group to the low-wage group. Continuing with the case of equal group size, we obtain the subsidy S as function of the policy and the group-wise fractions of transfer recipients and truncated expected value functions, $S = Tx_L - t\Psi_L = t\Psi_H - Tx_H$. Figure 7 shows the subsidy (graph S) and the transfer level (graph T) for a given tax rate. Because of our specifications, the subsidy increases linearly for policies up to the point at which all low-wage earners accept the transfer. For that interval, where S increases linearly, the high-wage earners and the low-wage earners, who are still working, share the task of providing for the transfer recipients belonging to the low-wage group. For the policy, at which $x_L = 1$ and $\Psi_L = 0$, S converges with the transfer level T, since only high wage earners contribute to public revenues.

Figure 7 Transfer T and subsidy S, case 1: $v_L(x_L) = 0$ and $v_H(x_H) = 0$.



2.5.4 Geographically segregated labour market

In the case when the groups are defined on (partly) overlapping wage intervals, we think of the labour market as geographically segregated. Then, the subsidisation may not be the one-way transfer, from high-wage earners to low-wage earners, that we observe for the vertically segregated labour market.

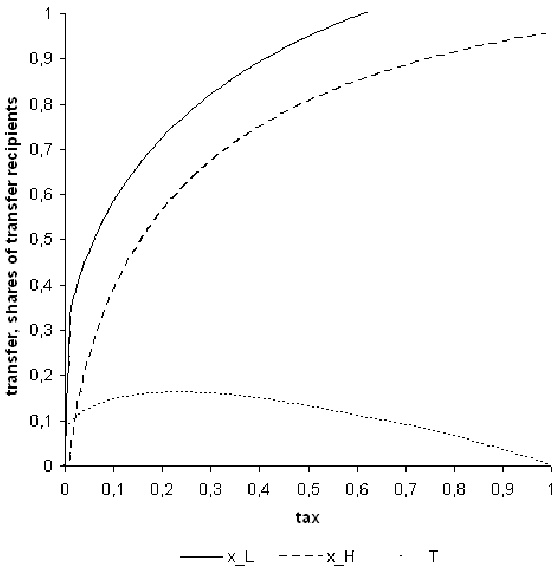
As an example we take a situation where the group-wise distributions, Φ_L and Φ_H, are uniform on $[0, \frac{3}{4}]$ and $[\frac{1}{4}, 1]$, respectively. The norm of the low-wage group is $v_L(x_L) = -\ln x_L/2$, while we assume that the norm of the high-wage group is non-punitive, i.e. $v_H(x_H) = 0$. Figure 8 shows the shares of transfer recipients in the high-wage group and in the low-wage group, graphs "x_H" and "x_L", for a given tax rate. The graph "T" represents the transfer level for tax rate t.

The figure shows that the first individual from the high-wage group accepts the transfer at a higher tax rate than the first recipient in the low-wage group. However, as the member earning the lowest wage in the high-wage group earns a low wage also in the perspective of the low-wage group, he decides not to work already at a low tax rate. Furthermore, high-wage individuals accepting the transfer is facilitated by the lack of punishment for norm-deviant behaviour, while the social norm"delays" the switch from work to non-work for low-wage individuals. Nevertheless, for a given tax rate and

transfer level, the share of transfer recipients in the low-wage group is larger than the recipient share in the high-wage group. Thus, in Figure 8 we observe that individuals with the same wage choose differently depending on which group they belong to.

For a large tax-rate interval, low-wage and high-wage transfer recipients are financed by the taxes paid by working individuals in both groups. Compared to the case of the vertically segregated labour market, it implies a "double" burden for high-wage members in that they subsidise the low-wage group and finance transfer recipients from their own group already at low tax rates. For low-wage members, it means that the additional resources of the high-wage group are not channelled directly to the low-wage group. Moreover, working individuals belonging to the low-wage group contribute to public spending on other individuals than group members. How the groups fare in the shared system depends on how similar they are in terms of wage and on the wage distributions.

Figure 8 Regional variations, case 2: $v_L(x_L) = -\ln x_L/2$ and $v_H(x_H) = 0$.



In a shared welfare system, there is interaction between the two groups. Changes in the degree of social stigma for one group affect the other group through the welfare policy.[15] Whether a centralised system or not, administrative controls can be used to regulate the utilization of the system. This will be discussed in the next section. However, the administrative interpretation of and compliance with the regulations may also be influenced by the sub-national social norms (Dutrieux & Sjöholm, 2003; Frykman & Hansen, 2005).

2.6 Optimal welfare policy

We have now studied the consequences of group-wise social norms for separate and shared welfare systems. We have analyzed how the share of transfer recipients in the high-wage and the low-wage groups relates to tax rate and transfer in balanced equilibrium states. In this section we will consider how the recipient share and the welfare policy relate to one another optimally. The effect of a social norm is also studied in this context.

We assume that the government gives equal weight to all individuals, favouring neither tax payers nor transfer recipients. Thus, it maximizes a utilitarian social welfare function. First, we focus on the first-best solution, where individual behaviour is controlled by the government by means of the policy as well as the critical wage. In a second-best context, the government cannot directly control the population share of recipients by other means but using the welfare policy to increase or decrease the generosity of the system.

2.6.1 First best in separate systems

We let the government control the tax rate t_i, the transfer T_i and the critical wage w_i^*. In this case the critical wage rate is not considered a function of the policy but is set directly by the government to achieve the optimal proportion of transfer recipients in the population. However, it may be the case that the optimal welfare policy would generate a different critical wage than the optimal one, if it were not controlled by the government. Thus, for the optimal critical wage, it is not certain that individuals earning higher wages would like to work or, for that matter, that individuals earning lower wages are keen on accepting the transfer.

[15] Besley & Coate (1992) take the financing of the welfare system as a starting point when modelling welfare stigma as resulting from taxpayer resentment. On account of the assumption of separate norms, however, we do not consider stigmatisation across groups.

Since the individual decision whether or not to work is in the control of the government, social welfare is maximized under the sole restriction of a balanced budget. The population share of transfer recipients is expressed as a function of the critical wage, $x_i = \Phi_i(w_i^*)$.

First, we study the case where norm-deviant behaviour is not punished. The social welfare function that the government maximizes with respect to the tax rate, the transfer level as well as the critical wage is

$$W = \int_{w_i^*}^{w_i^{\max}} u((1-t_i)w_i)\phi_i(w_i)dw_i + \int_{w_i^{\min}}^{w_i^*} (u(T_i)+\mu)\phi_i(w_i)dw_i, \tag{2.9}$$

where $w_i^{\max}$ is the highest value and $w_i^{\min}$ the lowest value of the wage interval on which the group-wise distributions are defined. Welfare is maximized subject to the constraint that the public revenues must at least cover public spending,

$$t_i\Psi_i(w_i^*) - T_i\Phi_i(w_i^*) \geq 0.$$

The Lagrangean function is

$$L = \int_{w_i^*}^{w_i^{\max}} u((1-t_i)w_i)\phi_i(w_i)dw_i + \int_{w_i^{\min}}^{w_i^*} (u(T_i)+\mu)\phi_i(w_i)dw_i + \lambda\big(t_i\Psi_i(w_i^*) - T_i\Phi_i(w_i^*)\big), \tag{2.10}$$

where λ is the Lagrangean multiplier. The first-order conditions are

$$\frac{\partial L}{\partial t_i} = -\int_{w_i^*}^{w_i^{\max}} u'((1-t_i)w_i)w_i\phi_i(w_i)dw_i + \lambda\Psi_i(w_i^*) = 0, \tag{2.11}$$

$$\frac{\partial L}{\partial T_i} = \int_{w_i^{\min}}^{w_i^*} u'(T_i)\phi_i(w_i)dw_i - \lambda\Phi_i(w_i^*) = 0, \tag{2.12}$$

$$\frac{\partial L}{\partial w_i^*} = -u((1-t_i)w_i^*)\phi_i(w_i^*) + (u(T_i)+\mu)\phi_i(w_i^*) + \lambda(t_i\Psi_i'(w_i^*) - T_i\Phi_i'(w_i^*)) = 0, \tag{2.13}$$

and

$$\frac{\partial L}{\partial \lambda} = t_i\Psi_i(w_i^*) - T_i\Phi_i(w_i^*) = 0, \tag{2.14}$$

given $\lambda > 0$. Because of the assumptions we make about the utility function (see section 2.2.1), it is clear that the tax rate, and thereby the transfer

level, cannot be set to zero. For the same reasons, a tax rate of one is not possible. The requirements that $t_i \in (0,1)$ and $T_i > 0$ also set boundaries for the critical wage. Based on the condition (2.14), neither $w_i^* = w_i^{\max}$ nor $w_i^* = w_i^{\min}$ is allowed. In the first case, $\Psi_i(w_i^{\max}) = 0$ and $\Phi_i(w_i^{\max}) = 1$, leading to $T = 0$ in order to achieve balance. In the latter case, it is required that $t = 0$ since $\Psi_i(w_i^{\min}) > 0$ and $\Phi_i(w_i^{\min}) = 0$. Thus, neither a situation where all individuals work, nor a situation where no one works is socially desirable. The individuals earning the lowest wages in the group contribute too little to add to the social welfare by working. The individuals earning the highest wages, on the other hand, contribute too much to the social welfare not to work.

Solving for λ we get

$$\frac{\int_{w_i^*}^{w_i^{\max}} u'((1-t_i)w_i)w_i\phi_i(w_i)dw_i}{\Psi_i(w_i^*)} = \frac{\int_{w_i^{\min}}^{w_i^*} u'(T_i)\phi_i(w_i)dw_i}{\Phi_i(w_i^*)}, \tag{2.15}$$

which says that in optimum, the relationship between the weighted marginal utility of labour income and the tax base equals that between the weighted marginal utility of transfer income and the share of transfer recipients. The condition indicates that if the share of transfer recipients, $\Phi_i(w_i^*)$, were to increase exogenously, the weighted marginal utility of transfer income must also increase, implying a smaller transfer in optimum. In the case of an exogenous increase in the tax base, the optimal tax rate must increase to compensate for the increase in the expected average wage of the working individuals.

Since the welfare policy is separated from the critical wage and thereby from the share of transfer recipients in the population, it may be the case that critical individual would prefer not to work. Condition (2.13) concerns the incentives of the critical individual. We subtract the last term on the left-hand side, leaving us with

$$(u(T_i) + \mu) - u((1-t_i)w_i^*) = \lambda(w_i^* t_i + T_i), \tag{2.16}$$

since $\Psi_i'(w_i^*) = -w_i^*\phi_i(w_i^*)$ and $\Phi_i'(w_i^*) = \phi_i(w_i^*)$. (2.16) shows that for the critical individual, the utility difference between not working and working is positive. If policy and critical wage were not separated, the optimal welfare policy t_i and T_i in a first-best context would lead to a higher critical wage, and hence a larger share of transfer recipients, than the critical wage chosen by the insurer. In other words, the policy that generates w_i^* is less generous than the first-best optimal policy. On the basis of (2.16), we find that the utility of not working can be high, if budgetary considerations allow it. In a

first-best context, the government reduces the share of transfer recipients in exchange for a relatively more generous welfare policy.

Next, we consider the case of optimizing the social welfare function when a social norm with a punitive sanction, $v_i(x_i) = v_i(\Phi_i(w_i^*)) > 0$, influences the behaviour of the population. The Lagrangean function with multiplier λ is

$$L = \int_{w_i^*}^{w_i^{max}} u((1-t_i)w_i)\phi_i(w_i)dw_i + \int_{w_i^{min}}^{w_i^*} (u(T) + \mu - v_i(\Phi_i(w_i^*)))\phi_i(w_i)dw_i + \lambda(t_i\Psi_i(w_i^*) - T_i\Phi_i(w_i^*)). \quad (2.17)$$

The first-order condition with respect to the critical wage is

$$\frac{\partial L}{\partial w_i^*} = -u((1-t_i)w_i^*)\phi_i(w_i^*) - (v_i'(\Phi_i(w_i^*))\Phi_i'(w_i^*)) \int_{w_i^{min}}^{w_i^*} \phi_i(w_i)dw_i + (u(T_i) + \mu - v_i(\Phi_i(w_i^*)))\phi_i(w_i^*) + \lambda(t_i\Psi_i'(w_i^*) - T_i\Phi_i'(w_i^*)) = 0. \quad (2.18)$$

In optimum,

$$(u(T_i) + \mu - v_i(\Phi_i(w_i^*))) - u((1-t_i)w_i^*) = = \lambda(w_i^*t_i + T_i) + (v_i'(\Phi_i(w_i^*)) \int_{w_i^{min}}^{w_i^*} \phi_i(w_i)dw_i. \quad (2.19)$$

Condition (2.19) shows that the social norm influences the optimal welfare policy in two ways. On the left-hand side, we find a direct effect of the norm, the disutility from breaking the norm reduces the utility of transfer income. On the right hand side of the condition, we find added an indirect effect, pertaining to the marginal effect of an increase in the critical wage on the strength of the social norm. Whereas the direct effect implies an increase, the indirect effect implies a decrease in the optimal transfer level. This corresponds well with the intuition of the model and our results from the graphical analysis. Hence, by altering the transfer level, the government influences the strength of the social norm. Based on (2.19) we cannot determine whether the system under the social norm is more or less generous than the "non-social" system.

Furthermore, it is not possible to determine from (2.19) whether or not the critical individual is willing to work at the optimal critical wage. The government can bring about a generous welfare policy by restricting the population share of recipients but, paradoxically, in doing so, the transfer recipients experience higher disutility due to their norm-deviant behaviour. The

question is to what extent the government is prepared to subject the group of non-working individuals to a more intense social sanction.

In a first-best situation, we find that the government may restrict the utilization of the welfare system and thereby benefit the non-working individuals. Our interpretation of the findings is that the use of administrative control mechanisms, which separate justified from unjustified benefit claims, may be advantageous, such as the requirement of a physician's assessment of the claimant's ability to work in the context of the sickness insurance, or the requirement that a sufficiently high search effort is expended by the unemployed. Conclusions of such kind are also drawn by Boone et al. (2002) and Fredriksson & Holmlund (2005) in their analyses of optimal unemployment insurance. They find that by means of monitoring, in conjunction with sanctions, the remuneration level can be higher compared to other program designs.

2.6.2 First best in a shared system

In the context of a shared welfare system, the government maximizes the social welfare function with respect to policy (t, T) and the critical wages w_L^* and w_H^*. We only present the "non-social" case, since the results regarding the social norm are essentially equivalent to the results we find for separate systems. Consequently, the relevant Lagrangean function is

$$\begin{aligned} L = &\int_{w_L^*}^{w_L^{\max}} u((1-t)w_L)\phi_L(w_L)dw_L + \int_{w_L^{\min}}^{w_L^*} (u(T)+\mu)\phi_L(w_L)dw_L + \\ &+ \int_{w_H^*}^{w_H^{\max}} u((1-t)w_H)\phi_H(w_H)dw_H + \int_{w_H^{\min}}^{w_H^*} (u(T)+\mu)\phi_H(w_H)dw_H + \\ &+ \lambda\left(t(\Psi_L(w_L^*)+\Psi_H(w_H^*)) - T(\Phi_L(w_L^*)+\Phi_H(w_H^*))\right), \end{aligned} \tag{2.20}$$

where λ is the multiplier.

Given the first-order conditions, the equivalent of (2.15) is,

$$\frac{\int_{w_L^{\min}}^{w_L^*} u'(T)\phi_L(w_L)dw_L + \int_{w_H^{\min}}^{w_H^*} u'(T)\phi_H(w_H)dw_H}{\Psi_L(w_L^*)+\Psi_H(w_H^*)} = \frac{\int_{w_L^*}^{w_L^{\max}} u'((1-t)w_L)w_L\phi_L(w_L)dw_L + \int_{w_H^*}^{w_H^{\max}} u'((1-t)w_H)w_H\phi_H(w_H)dw_H}{\Phi_L(w_L^*)+\Phi_H(w_H^*)}. \tag{2.21}$$

With respect to the critical individuals, we have that

$$(u(T)+\mu)\phi_L(w_L^*) - u((1-t)w_L^*)\phi_L(w_L^*) = \lambda\phi_L(w_L^*)(w_L^* t + T) \tag{2.22}$$

and

$$(u(T)+\mu)\phi_H(w_H^*) - u((1-t)w_H^*)\phi_H(w_H^*) = \lambda\phi_H(w_H^*)(w_H^* t + T). \qquad (2.23)$$

Similar to the case of separate systems, we find that the group-wise shares of transfer recipients that is associated with the optimal critical wages are smaller than the shares associated with the optimal policy, provided that the government had no control over individual behavior. The difference in utility between not working and working is positive in both (2.22) and (2.23). Hence, the critical individual in both groups would rather accept the transfer than work. From this information we can infer that firstly, the group of high-wage earners is not assigned the role of sole provider for the transfer recipients. Optimally, there are tax payers in both groups. Secondly, considering the observation that neither the critical individual in the low-wage group nor the critical individual in the high-wage group prefers to work, it must be the case that the optimal welfare policy is set at a relatively high level, while the share of transfer recipients in the entire population is relatively small. However, it is not possible to determine the share of transfer recipients in each group or the share of the non-working individuals in one group compared to the other. It seems likely, though, that the optimal critical wage for the low wage group is set higher than for the high wage group, since it is "cheaper" to finance a large share of recipients belonging to the former group.

2.6.3 Second best in separate systems

In a second-best context, the government has no possibility of directly controlling the individual choice. We exemplify with separate welfare systems. Thus, the critical wage is a function of the welfare policy p_i, $w_i^*(t_i, T_i)$. For the case where norm-deviant behaviour is not punished, the government maximizes the following social welfare function with respect to the tax rate t_i and T_i,

$$W = \int_{w_i^*(t_i,T_i)}^{w_i^{\max}} (u(1-t_i)w_i)\phi_i(w_i)dw_i + \int_{w_i^{min}}^{w_i^*(t_i,T_i)} (u(T_i)+\mu)\phi_i(w_i)dw_i \quad (2.24)$$

s.t.

$$B(t_i, T_i) = t_i\Psi_i(w_i^*(t_i, T_i)) - T_i\Phi_i(w_i^*(t_i, T_i)) \geq 0,$$

where the constraint states that the balance of the government budget, $B(t_i, T_i)$, must be positive. The condition that the critical individual must

be indifferent between working and accepting is fulfilled by assumption, and determines the critical wage w_i^*, i.e.

$$u((1-t_i)w_i^*(t_i,T_i)) - (u(T_i)+\mu) = 0.$$

The Lagrangean function is

$$L = \int_{w_i^*(t_i,T_i)}^{w_i^{\max}} (u(1-t_i)w_i)\phi_i(w_i)dw_i + \int_{w_i^{min}}^{w_i^*(t_i,T_i)} (u(T_i)+\mu)\phi_i(w_i)dw_i + \lambda B(t_i,T_i), \quad (2.25)$$

where λ is the multiplier.

Owing to the condition of the critical individual, the effects of w_i^* through the integrals disappear in the first-order conditions. Hence, the first-order condition with respect to the tax rate is

$$\frac{\partial L}{\partial t_i} = -\int_{w_i^*(t_i,T_i)}^{w_i^{\max}} u'((1-t_i)w_i)w_i\phi_i(w_i)dw_i + \lambda\frac{\partial B(t_i,T_i)}{\partial t_i} = 0, \quad (2.26)$$

where

$$\frac{\partial B(t_i,T_i)}{\partial t_i} = \Psi_i(w_i^*) - (t_i w_i^* + T_i)\phi_i(w_i^*)\frac{\partial w_i^*}{\partial t_i} \lesseqgtr 0$$

and

$$\frac{\partial w_i^*}{\partial t_i} = -\frac{-u'((1-t_i)w_i^*)w_i^*}{u'((1-t_i)w_i^*)(1-t_i)} > 0.$$

The first-order condition with respect to the transfer level is,

$$\frac{\partial L}{\partial T_i} = \int_{w_i^{min}}^{w_i^*(t_i,T_i)} u'(T_i)\phi_i(w_i)dw_i + \lambda\frac{\partial B(t_i,T_i)}{T_i} = 0, \quad (2.27)$$

where

$$\frac{\partial B(t_i,T_i)}{\partial T_i} = -\left(\Phi_i(w_i^*) + (t_i w_i^* + T_i)\phi_i(w_i^*)\frac{\partial w_i^*}{\partial T_i}\right) < 0$$

and

$$\frac{\partial w_i^*}{\partial T_i} = -\frac{-u'(T_i)}{u'((1-t_i)w_i^*)(1-t_i)} > 0.$$

We solve for λ, leaving us with

$$-\frac{\int_{w_i^*(t_i,T_i)}^{w_i^{max}} u'\left((1-t_i)w_i\right)w_i\phi_i(w_i)dw_i}{\frac{\partial B(t_i,T_i)}{\partial t_i}} = \frac{\int_{w_i^{min}}^{w_i^*(t_i,T_i)} u'(T_i)\phi_i(w_i)dw_i}{\frac{\partial B(t_i,T_i)}{\partial T_i}}. \quad (2.28)$$

In optimum, the weighted marginal utility of transfer income divided by the marginal budgetary effect of an increase in the transfer level equals the weighted marginal utility of labour income divided by the marginal budgetary effect of an increase in the tax rate. In order to fulfill condition (2.28), a marginal raise in the tax rate has to contribute positively to public revenues, i.e. $\frac{\partial B(t_i,T_i)}{\partial t_i} > 0$, since spending another dollar on financing the transfer recipients implies a budgetary cost, $\frac{\partial B(t_i,T_i)}{\partial T_i} < 0$.

Making a Laffer interpretation of condition (2.28), there is an optimal tax rate at which public revenues are maximized. On either side of the maximum, it is assumed that a given level of revenues can be generated by means of a low tax rate, associated with a large tax base, or by means of a high tax rate, associated with a small tax base. From the conclusion that $\frac{\partial B(t_i,T_i)}{\partial t_i} > 0$ in optimum, we can infer that the optimal tax rate in the second-best context generates revenues that positions the policy to the left of the maximum of a hypothetical Laffer curve.

We contrast with the case of a punitive sanction for norm-deviant behaviour, $v_i(x_i) > 0$ with $x_i = \Phi_i(w_i^*)$. The Lagrangean function with multiplier λ is

$$L = \int_{w_i^*(t_i,T_i)}^{w_i^{\max}} u((1-t_i)w_i)\phi_i(w_i)dw_i + \int_{w_i^{min}}^{w_i^*(t_i,T_i)} (u(T_i) + \mu - v_i(\Phi_i(w_i^*)))\phi_i(w_i)dw_i + \lambda B(t_i,T_i)). \quad (2.29)$$

Again, the critical wage w_i^* is defined by the condition that critical individual is indifferent to working,

$$u((1-t_i)w_i^*(t_i,T_i)) - (u(T_i) + \mu - v_i(\Phi_i(w_i^*))) = 0.$$

The first-order condition with regard to the tax rate is

$$\frac{\partial L}{\partial t_i} = -\int_{w_i^*(t_i,T_i)}^{w_i^{max}} u'\left((1-t_i)w_i\right) w_i\phi_i(w_i)dw_i - \frac{\partial v_i(\Phi_i(w_i^*))}{\partial t_i}\int_{w_i^{min}}^{w_i^*(t_i,T_i)} \phi_i(w_i)dw_i + \lambda\frac{\partial B(t_i,Ti)}{\partial t_i} = 0, \quad (2.30)$$

where

$$\frac{\partial v_i(\Phi_i(w_i^*))}{\partial t_i} = v_i'(\Phi_i(w_i^*))\phi_i(w_i^*)\frac{\partial w_i^*}{\partial t_i}$$

and

$$\frac{\partial w_i^*}{\partial t_i} = -\frac{-u'((1-t_i)w_i^*)w_i^*}{u'((1-t_i)w_i^*)(1-t_i) + v_i'(\Phi_i(w_i^*))\phi_i(w_i^*)}.$$

We assume that the marginal effect of the tax rate on the critical wage is positive, since the derivative of the condition of the critical individual with respect to the wage, found in the denominator, is positive in the relevant equilibrium states. Consequently, the marginal effect of the tax rate on the norm-strength, $\frac{\partial v_i(\Phi_i(w_i^*))}{\partial t_i}$, is negative.

The first-order condition for the transfer level is

$$\frac{\partial L}{\partial T_i} = \left(u'(T_i) - \frac{\partial v_i(\Phi_i(w_i^*))}{\partial T_i}\right) \int_{w_i^{min}}^{w_i^*(t_i,T_i)} \phi_i(w_i)dw_i + \lambda \frac{\partial B(t_i, Ti)}{\partial T_i} = 0. \quad (2.31)$$

The argument applied to the marginal effect of the tax rate on the critical wage and norm-intensity also applies to the marginal effect of the transfer level. Hence

$$\frac{\partial v_i(\Phi_i(w_i^*))}{\partial T_i} = v_i'(\Phi_i(w_i^*))\phi_i(w_i^*)\frac{\partial w_i^*}{\partial T_i} < 0,$$

where

$$\frac{\partial w_i^*}{\partial T_i} = -\frac{-u'(T_i)}{u'((1-t_i)w_i^*)(1-t_i) + v_i'(\Phi_i(w_i^*))\phi_i(w_i^*)} > 0.$$

Solving for λ produces

$$-\left(\frac{\int_{w_i^*(t_i,T_i)}^{w_i^{max}} u'((1-t_i)w_i)w_i\phi_i(w_i)dw_i + v_i'(\Phi_i(w_i^*)) \int_{w_i^{min}}^{w_i^*(t_i,T_i)} \phi_i(w_i)dw_i}{\frac{\partial B(t_i,Ti)}{\partial t_i}}\right) = \frac{(u'(T_i) - v_i'(\Phi_i(w_i^*))) \int_{w_i^{min}}^{w_i^*(t_i,T_i)} \phi_i(w_i)dw_i}{\frac{\partial B(t_i,Ti)}{\partial T_i}}. \quad (2.32)$$

Compared to its non-social counterpart (2.28), we observe in (2.32) that the marginal effect of the tax rate on the norm intensity is added to the marginal utility of labour income, while the marginal effect of the transfer on the norm intensity is subtracted from the marginal utility of transfer income. Consequently, for the equality to hold when a social norm with a punitive sanction is considered, the marginal utility of labour income has to decrease whereas the marginal utility of transfer income has to increase. Thus, it implies a more generous welfare policy, higher t_i and larger T_i, in relation to the non-social case.

2.7 Concluding discussion

Drawing on the model of Lindbeck et al. (1999), we assume that an individual's decision on whether or not to work is influenced by the disapproval from people she relates to. We consider two groups with separate norms and find that group affiliation matters for the welfare system, in terms of the generosity of the welfare policy and the share of transfer recipients.

Formal as well as informal rules may influence individual behaviour. Here, the welfare policy interacts with an endogenous social norm. In the graphical analysis as well as the welfare analysis, we observe that the social value of work may enable the government to establish a more generous welfare policy. We also find that there may be scope for higher benefits when the share of benefit recipients in the population is regulated. Hence, although we assume that individual choice whether or not to utilize the system is voluntary, we find that exogenous controls may play a part in the optimal welfare system. However, just as group-specific norm matters for individual behaviour, variations between groups and regions may influence how administrators apply and comply with the rules set up for the system. This is something which we discuss further in chapter 3 with respect to the physician who certifies sickness absence.

We observe that the interaction between the welfare policy and the endogenous social norm may imply a raise in the insurance coverage. However, a generous welfare policy makes it more attractive not to work, which affects the strength of the social norm. Hence, in this respect, the economic incentives of the welfare system conflict with, or crowd out, social work norms.[16] Thus, from a long-term perspective, it is conceivable that a generous transfer influences the intensity of the social norm and thereby changes the conditions for the optimal policy.[17] Lindbeck (1995) argues that it may be necessary to strengthen the incentives to work and introduce stricter regulatory controls in response to eroding social norms. The restrictions applied to the group of non-working individuals in a first-best context, can be interpreted in such a light. As is also pointed out by Lindbeck et al. (1999), the interaction between the welfare policy and the social norm renders a dynamic application of the model an interesting idea for further research.

[16]For a review on theoretical and empirical findings about cooperation and social norms, see Ostrom (2000).

[17]Henrekson & Persson (2006) find that the long-run effect on sick leave of changes in the insurance policy is larger than the short-run effect, which could be explained by inertia due to social norms. See also Ljunge (2005) and Lindbeck & Nyberg (2006).

References

Akerlof G A, 1980, "A Theory of Social Custom, of Which Unemployment May Be One Consequence", *The Quarterly Journal of Economics*, 94, 749-775.

Arai M & Skogman Thoursie P, 2004, "Sickness Absence, Worker and Establishment Effects", *Swedish Economic Policy Review*, 11, 9-28.

Axelrod R, 1986, "An Evolutionary Approach to Norms", *The American Political Science Review*, 80, 1095-1111.

Bertrand M, Luttmer E F P and Mullainathan S, 2000, "Network Effects and Welfare Cultures", *The Quarterly Journal of Economics*, 115, 1019-1055.

Besley T & Coate S, 1992, "Understanding Welfare Stigma: Taxpayer Resentment and Statistical Discrimination", *Journal of Public Economics*, 48, 165-183.

Bird E J, 1999, "Can Welfare Policy Make Use of Social Norms?" *Rationality and Society*, 11, 343-365.

Blank R & Card D, 1991, "Recent Trends in Insured and Uninsured Unemployment: Is There an Explanation?" *Quarterly Journal of Economics*, 106, 1157-1190.

Boone J, Fredriksson P, Holmlund B & van Ours J, 2002, "Optimal Unemployment Insurance with Monitoring and Sanctions", working paper 21, Institute For Labour Market Policy Evaluation (IFAU), Uppsala.

Brock W A & Durlauf S N, 2001, "Discrete Choice with Social Interactions", *Review of Economic Studies*, 68, 235-260.

Clark A E, 2003, "Unemployment as a Social Norm: Psychological Evidence from Panel Data", *Journal of Labor Economics*, 21, 289-322.

Coleman J S, 1990, *Foundations of Social Theory*, Harvard University Press.

Drago R & Wooden M, 1992, "The Determinants of Labor Absence: Economic Factors and Workgroup Norms across Countries", *Industrial & Labor Relations Review*, 45, 764-778.

Dufwenberg M & Lundholm M, 2001, "Social Norms and Moral Hazard", *The Economic Journal*, 111, 506-525.

Dutrieux J & Sjöholm K-R, 2003, *Regionala skillnader i sjukskrivningar: hur ser de ut och vad beror det på?*, RFV Analyserar 2003:12, Riksförsäkringsverket.

Elster J, 1989, "Social Norms and Economic Theory", *The Journal of Economic Perspectives*, 3, 99-117.

Festinger L, 1954, "A Theory of Social Comparison Processes", *Human Relations*, 7, 117-140.

Fredriksson P & Holmlund B, 2005, "Optimal Unemployment Insurance Design: Time Limits, Monitoring, or Workfare", working paper 13, Institute For Labour Market Policy Evaluation (IFAU), Uppsala.

Fredriksson P & Åslund O, 2005, "Påverkas socialbidragsberoende av omgivningen?", rapport 3, Institute For Labour Market Policy Evaluation (IFAU), Uppsala.

Frykman J & Hansen K, 2005, "Att leva på kassan. Allmän försäkring och lokal kultur", Analyserar 2005:4, Försäkringskassan.

Glaeser E L, Sacerdote B I & Scheinkman J A, 2003, "The Social Multiplier", *Journal of the European Economic Association*, 1, 345-353.

Hedström P, Kolm A-S & Åberg Y, 2003, "Social Interactions and Unemployment", working paper 15, Institute For Labour Market Policy Evaluation (IFAU), Uppsala.

Henrekson M & Persson M, 2004, "The Effect on Sick Leave of Changes in the Sickness Insurance System", Journal of Labor Economics, 22, 87-113.

Holm E & Öberg S, 2004, "Contagious Social Practice?" Geografiska Annaler, Series B, 86, 297-314.

Kolm A-S, 2005, "Work Norms and Unemployment", *Economic Letters*, 88, 426-431.

Kreps D M, 1997, "Intrinsic Motivation and Extrinsic Incentives", *The American Economic Review*, 87, 359-364.

Lalive R, 2003, "Social Interactions in Unemployment", discussion paper 803, Institute for the Study of Labour (IZA), Bonn.

Lindbeck A, 1995 "Welfare State Disincentives with Endogenous Habits and Norms", *The Scandinavian Journal of Economics*, 97, 447-494.

Lindbeck A & Nyberg S, 2006, "Raising Children to Work Hard: Altruism, Work Norms and Social Insurance", *The Quarterly Journal of Economics*, 121, 1473-1503.

Lindbeck A, Nyberg S & Weibull J, 1999, "Social Norms and Economic Incentives in the Welfare State", *The Quarterly Journal of Economics*, 114, 1-35.

Lindbeck A, Palme M & Persson M, 2004, "Sjukskrivning som socialt fenomen", *Ekonomisk Debatt*, 4, 50-62.

Lindbeck A & Persson M, 2006, "A Model of Income Insurance and Social Norms", seminar paper 742, Institute for International Economic Studies (IIES), Stockholm University.

Ljunge M, 2005, "Half the Job Is Showing Up: Returns to Work, Taxes, and Sick Leave Choices", working paper, Department of Economics, University of Chicago.

Lucas R E, Clark A E, Georgellis Y & Diener E, 2004, "Unemployment Alters the Set Point for Life Satisfaction", *Psychological Science*, 15, 8-13.

McCoy J L, Davies M & Hudson R E, 1994, "Geographic Patterns of Disability in the United States", *Social Security Bulletin*, 57, 25-36.

Moffitt R, 1983, "An Economic Model of Welfare Stigma", *The American Economic Review*, 73, 1023-1035.

Olsson S, 2006,"Attityder och regionala skillnader i sjukförsäkringen" in Palmer E (ed.), *Sjukförsäkring, kultur och attityder. Fyra aktörers perspektiv*, Analyserar 2006:16, Försäkringskassan, 586-620.

Ostrom E, 2000, "Collective Action and the Evolution of Social Norms", *The Journal of Economic Perspectives*, 14, 137-158.

Paul K I & Moser K, 2006, "Incongruence as an Explanation for the Negative Mental Health Effects of Unemployment: Meta-Analytic Evidence", *Journal of Occupational and Organizational Psychology*, 79, 595-621.

Rege M, Telle K & Votruba M, 2007, "Social Interaction Effects in Disability Pension Participation", discussion paper 496, Research Department, Statistics Norway.

Stutzer A & Lalive R, 2004, "The Role of Social Work Norms in Job Searching and Subjective Well-Being", *Journal of the European Economic Association*, 2, 696-719.

Winkelmann L & Winkelmann R, 1998, "Why Are the Unemployed So Unhappy? Evidence from Panel Data", *Economica*, 65, 1-15.

Appendix A: Parametric specifications

We assume that utility from consumption is logarithmic, $u(c) = ln(c)$ for all $c > 0$. Using that assumption with (2.4), we get expressions for f_i and F_i:

$$f_i(x_i) = \frac{x_i \Phi_i^{-1}(x_i)}{x_i \Phi_i^{-1}(x_i) + \Psi_i[\Phi_i^{-1}(x_i)] exp[\mu - v_i(x_i)]} \tag{2.33}$$

$$F_i(x_i) = \frac{\Phi_i^{-1}(x_i)\Psi_i[\Phi_i^{-1}(x_i)]}{x_i \Phi_i^{-1}(x_i) + \Psi_i[\Phi_i^{-1}(x_i)] exp[\mu - v_i(x_i)]}, \tag{2.34}$$

where $\mu = 1$.

Wages are distributed uniformly on $[0, 1]$ and hence, the group-wise distributions, Φ_L and Φ_H, are uniform. Taking the case when the group-wise distributions are uniform on $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ as an example, the probability density functions of the conditional distributions are

$$\phi_L(w_i) = 2 \tag{2.35}$$

for $w_L \in [0, \frac{1}{2}]$ and 0 otherwise, and similarly

$$\phi_H(w_H) = 2 \tag{2.36}$$

for $w_H \in [\frac{1}{2}, 1]$ and 0 otherwise.

The cumulative distribution functions are

$$\Phi_L(w_L) = 2w_L \tag{2.37}$$

for $w_L \in [0, \frac{1}{2}]$ and 0 for $w_L < 0$ and 1 otherwise and

$$\Phi_H(w_H) = 2w_H - 1 \tag{2.38}$$

for $w_H \in [\frac{1}{2}, 1]$ and 0 for $w_H < \frac{1}{2}$ and 1 otherwise.

Consequently, we can construct the equations for the critical wages as functions of group-wise proportions of transfer recipients:

$$\omega_L = \Phi_L^{-1}(x_L) = \frac{x_L}{2} \tag{2.39}$$

for $x_L \in [0, 1]$ and

$$\omega_H = \Phi_H^{-1}(x_H) = \frac{x_H + 1}{2} \tag{2.40}$$

for $x_H \in [0, 1]$.

We calculate the respective truncated expectations as functions of the critical wages:

$$\Psi_L(\omega_L) = \int_{\omega_L}^{\frac{1}{2}} 2\omega d\omega = \omega^2 \mid_{\omega_L}^{\frac{1}{2}} = \frac{1}{4} - \omega_L^2 \tag{2.41}$$

for $\omega_L \in [0, \frac{1}{2}]$, and

$$\Psi_H(\omega_H) = \int_{\omega_H}^{1} 2\omega d\omega = \omega^2 \mid_{w_H}^{1} = 1 - \omega_H^2 \tag{2.42}$$

for $\omega_H \in [\frac{1}{2}, 1]$.

By means of the equations (2.39) and (2.40), we express the group-wise incomes as functions of x_L and x_H:

$$\Psi_L(x_L) = \frac{1 - x_L^2}{4} \tag{2.43}$$

and

$$\Psi_H(x_H) = 1 - (\frac{x_H + 1}{2})^2. \tag{2.44}$$

When exploring the perspective of the policy $p_i = (t_i, T_i)$, we express the group-wise fractions of transfer recipients as functions of t_i and T_i. We exemplify by using $v_i(x_i) = 0$ and the group-wise distributions uniform on $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$. Starting with the low wage group, we apply our specifications to equation (2.3)

$$x_L = \Phi_L(\frac{u^{-1}[u(T_L) + \mu - v_L(x_L)]}{1 - t_L}) = \Phi_L(\frac{\exp[\ln(T_L) + \mu]}{1 - t_L}) = \Phi_L(\frac{T_L \exp(\mu)}{1 - t_L}). \tag{2.45}$$

Taking the inverse of the cumulative distribution function Φ_L, we get the equivalent of equation (2.39), which expresses the critical wage for the low wage group at given policy (t_L, T_L) as a function of x_L. Equating the two expressions and solving for x_L we get:

$$\omega_L = \Phi_L^{-1}(x_L) = \frac{x_L}{2} = \frac{T_L \exp(\mu)}{1 - t_L} \Rightarrow x_L = \frac{2T_L \exp(\mu)}{1 - t_L}. \tag{2.46}$$

The expression for x_H,

$$\omega_H = \Phi_H^{-1}(x_H) = \frac{x_H + 1}{2} = \frac{T_H \exp(\mu)}{1 - t_H} \Rightarrow x_H = \frac{2T_H \exp(\mu)}{1 - t_H} - 1. \tag{2.47}$$

Chapter 3

Sickness Absence Certification

3.1 Introduction

Sickness insurance compensates an individual for the income loss incurred when he is unable to work due to sickness. The physician's assessment of the applicant's ability impairment is an essential component in the process of granting sickness benefits, possibly after a period of self-certification. Hence, the physician has a decisive influence on the utilization of the insurance. However, the role as gatekeeper of the insurance may conflict with the physician's traditional role as the patient's advocate.[1]

Despite the physician's pivotal role in the insurance system, it is quite unexplored by economists. We model sickness absence certification in a single-period model, considering the physician's relationship with the worker as well as the insurer. The physician may encounter two types of individuals whose motives for a sick-listing consultation differ. Before the physician certifies absence, he is supposed to verify that his patient's ability to work is reduced. However, ability verification requires effort, which the physician dislikes. We study the physician's incentives in the process of sickness absence certification and in what way his behaviour influences the utilization of sickness insurance. The certification process is evaluated in terms of the treatments the two types of workers receive from the physicians. We also consider a couple of relevant policy instruments with respect to their effects on the workers' utilization of the insurance as well as to their welfare implications.

[1]The intention to act in the best interest of the patient is a manifestation of ethical considerations. The emphasis put on medical ethics is intended to discourage the expert from taking advantage of the layman, see McGuire (2000). The element of trust is a feature that distinguishes the relationship between physician and patient from other agency relationships characterized by asymmetric information (Evans, 1984; Arrow, 1963).

The Swedish experience

Experiences from Swedish sickness insurance serve as a background to the model. In comparison with other insurance systems in Europe, the Swedish system is distinguished by a high sickness absence rate. Several factors are discussed as contributory to the development, among them the generosity of the insurance and the gate-keeping functions in the system (Mikaelsson et al., 2003; Nyman et al., 2002).

Studies and reports on Swedish sickness insurance show that physicians have considerable scope to create individual certification practices. The lack of general guidelines and monitoring practices are considered to be contributory factors (Bergendorff & Larheden, 2003). The physicians forming individual certification practices also allow some liberty of action for the patient to influence the outcome of the consultation. There are reports of patients intent on obtaining a so-called sickness certificate, a medical statement forming the basis for the decision of granting sickness benefits. However, there are also patients who are reluctant to become sick-listed and feel that certificates are offered to them off-hand (Alexanderson et al., 2005). This may be an indication of patients being under the influence of a social work norm (Arrelöv, 2006). Besides showing that patients have different motives for consulting physicians, the observations could also indicate that the state of a patient's working ability is unclear not only to the patient but also to the physician.

Sickness and inability

The relationship between health state and working ability is straightforward in some cases; in others it is not. The kinds of sicknesses we consider here do not automatically render a person incapable of working. Moreover, the connection between sickness and lack of ability is not identified readily. To the worker, this means that falling ill involves a risk of subsequent reduction in his ability to work.[2] Hence, he does not know whether or not it is advisable to work. For the physician, it implies that he cannot observe the worker's ability effortlessly, but, with some effort, he can make a qualified decision regarding the worker's ability to work.

It is not always the case that absence can be medically certified, i.e. be based on observable medical findings. In many instances symptoms described by the patients form the basis for the diagnosis, which, we argue, rather increases the importance of the physician's assessment of (remaining) working

[2]His falling ill and staying at work does not involve a risk of infecting co-workers. We focus on ailments that health- and ability-wise affect only the individual worker. Contagion effects are considered by Skåtun (2003) and Barmby & Larguem (2004).

ability.[3] In either case, the physician makes an assessment whether or not it is reasonable to certify absence in relation to individual assignments and demands at work.

On a more general level, it is not established whether or not sick-listing is beneficial as a kind of medical treatment. Englund (1997) argues that the health effects of sickness absence may be negative. Moreover, the physician's decision to "prescribe" sick leave could be understood in a context where the choice of medical treatment is not straightforward, and the influence of economic incentives on the physician's behaviour therefore may increase (see McGuire, 2000). As a matter of fact, Jansson & Johansson (2003) point out that sick leave is a cheap form of treatment for the Swedish public health care organisation, since the costs do not burden its budget.

3.1.1 Related studies

Quite a number of empirical studies find that sickness absence is influenced by the costs of being absent. Swedish studies report that increased replacement rates are associated with increased absence rates (Johansson & Palme, 1996; Henrekson & Persson, 2004). Remuneration is also positively correlated with the incidence and the duration of sick leave (Johansson & Palme, 2002, 2005; Andrén, 2005). Andrén (2001, 2007) and Lidwall et al. (2005), studying long-term absentees specifically, also find support for a positive relationship between sickness benefits and absence rates. Similarly, using international data, Osterkamp & Röhn (2005) and Frick & Maló (2005) observe that institutional generosity, including factors such as replacement rate, waiting period, possibility of self-certification etc., is an important determinant of workers' absence behaviour.[4]

In contrast, there are but a few studies that analyse sickness insurance theoretically, e.g. Lindbeck & Persson (2006) and Rikner (2002). Both studies assume that the individual's decision to attend work depends on how much

[3]For psychiatric and musculoskeletal diagnoses, the two most prevalent sick-listing diagnoses in the Swedish context, Alexanderson et al. (2005) maintain that the possibility for making medical diagnoses is restricted (see also Englund (2001) and Samtal om socialförsäkring Nr 7 (2006)). The Swedish National Board of Health and Welfare (2006a) reports that for three quarters of the examined journals in primary care, diagnoses were medically rather than symptom-based. In a Norwegian study, the corresponding proportion of absence certification based on clinical findings was considerably smaller (Reiso et al., 2000).

[4]See also Barmby et al. (1991, 1995) for analyses of employment contracts and absence behaviour among employees and Brown (1994), who analyses an experience-rated sick-pay scheme. Brown & Sessions (1996) survey economic literature on work absence. For an interdisciplinary review on employee absence behaviour, see Kaiser (1998).

effort working involves. An individual's momentary motivation to work is not only influenced by his health but also by other things, such as leisure-time activities. Our model focuses on how working ability, influenced by individual health, affects willingness to work. Furthermore, the physician's decision on whether or not to sick-list a patient is an issue well-suited for a single-period model. Lindbeck & Persson (2006) also use a single-period model to analyse the optimal insurance contract, while Rikner (2002) extends his analysis in a later dynamic model.

Lindbeck & Persson (2006) allow for a heterogeneous population in that the mean of the stochastic preference variable for work may vary among individuals. We assume that the workers differ in their preferences for labour too, but non-stochastically. However, in both cases the number of absentees is the result of the underlying distribution of preferences and the insurance policy. Heterogenous individuals are also analysed in Whinston (1983), where workers differ in the probability of becoming disabled. In an insurance model with health-related disability Lantto (1991) considers, among other things, the case where workers differ in the disutility of malingering. However, in the latter cases, the group sizes are not influenced by the insurance policy.

Insurance remuneration is generally contingent on health-related ability impairment. Hence, the physician is required to establish the relationship between the individual's health state and working ability before granting sick leave. However, the health-ability connection is little discussed in the relevant economic literature. Lindbeck & Persson (2006), Rikner (2002) and Lantto (1991) appear to equate sickness with inability to work. Furthermore, in Diamond & Mirrlees (1978), although originally a model of variable retirement, it is simply assumed that a worker risks a complete loss of his earning ability without any mention of an underlying factor (see also Whinston (1983)). Another example is Diamond & Sheshinski (1995), where individuals are, by assumption, disabled but to varying degrees. We recognize that sickness does not always and not automatically lead to a loss of earning ability. The starting point of our model is the assumption that bad health may affect a person's ability to work but does not make him incapable of work instantly. Instead, the potential impact of sickness is delayed. To receive information about working ability and whether it is recommendable to attend work, a physician's examination is required.

Rikner (2002) and Lindbeck & Persson (2006) illustrate that the decision to be absent from work is the result of a complex process involving multiple factors. However, they do not take into account that in most sickness insurance systems the choice is not entirely up to the individual. In our model, self-certification is disregarded. Instead, we stress the role of the

physician as gatekeeper in the sickness insurance.[5] Control mechanisms influence the behaviour of insured individuals. Specifically, it has been observed that the requirement of a sickness certificate has a deterrent effect (Hesselius et al., 2005; Henrekson et al., 1992; Johansson & Palme, 2005). Through employing screening devices, by which an applicant's eligibility is "actively" identified, the insurance program may be made more generous compared to self-screening. However, an individual's "true" (in)ability status is difficult to observe, leading to imperfections in the screening mechanism. Parsons (1996) and Diamond & Sheshinski (1995) recognize that imperfect screening devices give rise to a situation where some disabled individuals are denied benefits while some able-bodied individuals are granted benefits.[6] Parsons (1991) points out that physical and/or mental impairments that are observable for the administrators may be used as the basis for an active screening mechanism. Lantto (1991) also addresses the issue of monitoring, in one case assuming that the screening costs vary with different diseases. Contrarily, we study the physician dealing with a group of ailments, for which the information costs are identical. We assume that the physician can make a correct diagnosis and relate the patient's health state to his working ability perfectly. However, the effort required from the physician to establish that relationship leads to imperfect screening, since the sickness certificate, a prerequisite for receiving benefits, can be issued by the physician without verifying the worker's eligibility. The workers respond by either accepting or declining the certificate offer. Thus, the physician's screening may lead to situations where eligible workers continue to work but also to situations where ineligible workers get certified sickness absence. In a sense, the screening mechanism in our model produces type 1 as well as type 2 errors.

We consider two types of physicians differing in willingness to expend effort to examine their patients' working abilities before certifying absence. The physicians are employed by the public sector and paid salaries reflecting the general situation in Sweden. Physician behaviour is primarily studied in the context of health care consumption (see e.g. Zweifel & Breyer, 1997; Blomqvist, 1991), but there are some connections between our model and e.g. models of supplier-induced demand (Evans, 1974; McGuire, 2000). The effort-minimizing, salaried physician is also found in Giuffrida & Gravelle (2001) and Grytten et al. (1995), where physicians may induce as well as ration health care consumption. As far as physicians' reimbursement forms are concerned, there have been few attempts to study the effects on sick-

[5]Garcia Mariñoso & Jelovac (2003) study the physician's screening of patients concerning referrals to specialists in a gate-keeping and a non-gate-keeping health-care system.

[6]The two-sided classification error for the US Social Security Disability Insurance is empirically estimated by Smith & Lilienfeld (1971) and Benitez-Silva et al. (2004).

listing behaviour. Physicians involved in a project of financial co-operation in Sweden, giving the primary health care organization incentives to keep the costs of sick-listing down, certified shorter absence periods (Arrelöv, 2003). Furthermore, a report from the Norwegian National Insurance Organization (2006) regarding the sick-listing practices of physicians whose reimbursement schedule consists of a capitation and a fee-for-service component, finds that physicians with large patient lists or working in physician-dense areas issue more certificates. The findings indicate that economic incentives matter for physician behaviour, but the implications for sick-listing are not clear.

With regard to certification practices, most research is done in the field of social medicine. For research reviews, see Söderberg & Alexanderson (2003) and Wahlström & Alexanderson (2004). In search of explanations for the observed variations in physicians' sick-listing practices, the potential influence of several factors, such as gender, age and specialist category, is analysed (Arrelöv, 2003; Englund, 2000; Tellnes et al., 1990). Several studies offer evidence of the patient's influence on the physician's decision to certify sickness absence. For example, in cases where the patient initiates the discussion of sick leave, the physician tends to certify absence to a greater extent, even if the wishes of the patient do not coincide with the opinion of the physician (Englund, 2001; Gulbrandsen et al., 2002; see also Larsen et al., 1994). However, there are both patients who are very interested in getting a sickness certificate and patients who want to avoid sick-listing, if possible (Alexanderson et al., 2005). In our model, we consider motivated as well as unmotivated workers who take different attitudes towards sickness absence. The interaction between workers and physicians determines the outcome of the process of absence certification, i.e. whether or not the worker's sickness absence is based on the physician's assessment of working ability.

3.1.2 Outline

In the next section we discuss the preferences of the workers and the physicians' sick-listing behaviour. The insurer's objective is also discussed. The manner in which the workers are treated by the physicians is the main issue in section 3.3, where equilibria in the certification process are studied. In section 3.4 we discuss the policy measures available for the insurer to control the number of benefit recipients. Section 3.5 concludes the paper.

3.2 The model

We identify three key agents in the process of sickness absence certification, the insured workers, the absence-certifying physicians and the insurer. There are two types of workers, differing with respect to their motives for sick-listing but both facing uncertainty about the health-related ability impairment. Physicians certify absence, but to verify that sick leave is justified, effort must be expended. Variations in the effort-cost lead to two types of physicians, who perform absence certification either based on the observation of a low ability to work or without examining the worker's ability. The interaction between workers and physicians determines the outcome of the certification process. Concerned with the utilization of sickness insurance, the insurer maximizes welfare under the constraint that expected production covers expected consumption.

3.2.1 The workers

The worker encounters the risk P of falling ill. We assume that he can observe whether his health state, σ, is good, $\sigma = 0$, or bad, $\sigma = 1$. His health state is related to his ability to work, α. Working ability is either low, $\alpha = L$, or high, $\alpha = H$. If the worker is healthy he knows that his ability is high. However, if he falls ill he only knows that it increases the probability of a reduced working ability. That is, we assume that the worker is aware of his health state, but is uncertain about the extent to which his future working ability is affected.[7] Π represents the probability of having a low ability to work and depends on the health status such that $\Pi(\sigma = 0) = 0$ and $\Pi(\sigma = 0) < \Pi(\sigma = 1) < 1$. To verify working ability in case the worker falls ill, an examination performed by a physician is necessary.

Falling ill does not imply an instant ability impairment, nor does it lead to an immediate reduction in productivity. Hence, the worker cannot infer his working ability by observing his level of productivity. All working individuals, whether ill or healthy, are equally productive. However, future productivity, just as future ability, may be affected in the event of health deterioration.

The insured individual may either work full-time or be on sick leave. We assume that the worker consumes the net labour income y_1 in the working state after paying the insurance premium. If the individual is on sick leave,

[7] Rikner (2002) analyses a sickness-insurance model, where the effort the individual expend at work in the first period affects his distribution of health in the second period. Absence from work can be regarded as an investment in individual health (see Gilleskie, 1998; Bolin & Lindgren, 2001). However, the general applicability of the statement that sickness absence generates positive health effects is questioned by Englund (1997).

he enjoys full-time leisure and receives compensation y_2. We assume that $y_1 \geq y_2 \geq 0$. For the lion's share of the analysis, the insurance policy is fixed.

The utility of the individual depends on his income as well as his working ability. Although suppressed in the notation, utility is also contingent on whether or not he works. Hence, the utility function is $u(y, \alpha)$, where $y \in \{y_1, y_2\}$. Labour is associated with disutility whereas utility is strictly increasing and concave in income, $u_y > 0$ and $u_{yy} < 0$. Furthermore, an increase in working ability has a positive influence on individual willingness to work, implying that $u(y_1, L) - u(y_2, L) \leq u(y_1, H) - u(y_2, H)$.

There are two types of workers, the "less motivated" type l and "motivated" type m. Information about individual type is private. Behind the distinction of type l and type m lies a distribution of preferences for labour, and possibly a social norm supporting motivation to work. Every individual weighs the utility of working against the utility of not working. For every insurance system there is a cut-off point, below which individuals prefer not to work. For the income levels y_1 and y_2, we assume that the proportion of type l individuals in the worker population is $0 < \lambda < 1$.[8]

The "less motivated" worker l prefers leisure to work, irrespective of his working ability

A 1. $u_l(y_2, \alpha) \geq u_l(y_1, \alpha)$, $\alpha \in \{L, H\}$.

In contrast, the "motivated" worker m, prefers to work if he is able to do so:

A 2. $u_m(y_2, L) \geq u_m(y_1, L)$, $u_m(y_1, H) \geq u_m(y_2, H)$.

In the case of illness, the worker's expected utility function U is

$$U_\theta(y, \Pi(1)) = (1 - \Pi(1))u_\theta(y, H) + \Pi(1)u_\theta(y, L), \quad \theta = l, m, \tag{3.1}$$

where $y \in \{y_1, y_2\}$. In the expression above, it is shown that whether he works or is sick-listed, earning y_1 or y_2, respectively, the worker is uncertain about his ability to work. The expression below describes the situation where the worker knows his future working ability after consulting a physician,

$$U_\theta(y_1, y_2, \Pi(1)) = (1 - \Pi(1))u_\theta(y_1, H) + \Pi(1)u_\theta(y_2, L), \quad \theta = l, m. \tag{3.2}$$

The primary interest of a worker of type l is to attain sick leave and avoid working (see Assumption 1). Information about his ability is of no particular

[8]The population share of type l individuals λ is discussed in Appendix A.

concern to him. However, irrespective of type, the worker recognizes whether or not the effects on his ability have been evaluated by the physician, e.g. based on the type and the number of questions asked about health and work conditions, the physician's effort to contact the employer etc. For type l it is the case that

$$U_l(y_2, \Pi(1)) \geq U_l(y_1, y_2, \Pi(1) \geq U_l(y_1, \Pi(1)). \tag{3.3}$$

Contrarily, the motivated worker m prefers to be certain about his working ability rather than not to work, which is in line with Assumption 2. The expected utility associated with knowing his ability is at least as great as the expected utility without such information,

$$U_m(y_1, y_2, \Pi(1)) \geq U_m(y, \Pi(1)), \quad y \in \{y_1, y_2\}. \tag{3.4}$$

Springing from Assumption 2, there is a value of $\Pi(1)$ at which the type m worker is indifferent between working and not working. We assume that $\Pi(1)$ is smaller than this value. Consequently, if he is not able to acquire information about his ability, he will not accept a sickness certificate,

$$U_m(y_1, \Pi(1)) \geq U_m(y_2, \Pi(1)). \tag{3.5}$$

3.2.2 The physicians

The insurance rules stipulate that only workers with low working ability are entitled to sickness benefits. A sickness certificate is a prerequisite for remunerated sick leave and as such is supposed to confirm the reduced earning ability. Hence, we disregard any initial period of self-certification. However, whether or not the certificate conveys the intended information is contingent on the behaviour of the physician.

The physician is a utility maximizer. He enjoys his salary s, which satisfies his participation constraint by assumption. He dislikes effort from which he experiences a non-negative disutility d. When performing regular health examinations, the physician observes his patient's health state with a low level of effort, $e = 0$. To observe his patient's working ability, however, more effort is required, $e = 1$.[9]

Whereas the examination to observe the patient's working ability is strenuous, the physician can issue a sickness certificate without effort. That is,

[9]Considering the earlier discussion about vague diagnoses, another model, where the relationship between health and working ability is clear, but where the physician misidentifies an ill, high-ability worker as a low-ability one with some probability, would have similar implications, see chapter 4.

he can certify sickness absence without verifying that the worker really is eligible for the benefit. The potential certificate receiver is aware of the physician's level of ambition and therefore knows how to respond to his offer. The insurer, on the other hand, can distinguish a working individual from a non-working individual but cannot observe the physician's effort. Consequently, the insurer does not know if absence is certified according to the insurance rules. Hence, we assume that the patient and the physician can misrepresent information to the insurer (cf. Ma & McGuire, 1997).

If the worker is healthy, the physician refrains from absence certification and thereby avoids the risk of getting caught certifying absence for an individual who is undoubtedly able to work. Thus, physicians offer patients sickness certificates contingent on $\sigma = 1$ and possibly on the condition of a confirmed ability reduction. If the physician observes his patient's ability he acts in accordance with the insurance rules and certifies only eligible patients, i.e. patients with $\alpha = L$.

We consider two types of physicians. The types differ in willingness to examine their patients' working abilities. Again, information about type is private. "Relaxed" physicians, type R, experience a positive disutility $d_R > 0$ and therefore do not exert effort (voluntarily), but issue sickness certificates as a matter of routine. "Strict" physicians, type S, experience the effort cost $d_S = 0$. We assume that strict physicians only certify absence if $\alpha = L$ is observed. Hence, they behave in concordance with the role of gatekeeper of the insurance system. In general terms, the physician maximizes his utility function Γ with respect to effort,

$$\max_e \Gamma_\rho = s - d_\rho e, \quad \rho = R, S, \quad e \in \{0, 1\}. \tag{3.6}$$

We assume that the physicians' objective is to provide adequate treatment concerning both health and working ability. However, Alexanderson et al. (2005), studying absence certification at public health clinics in Sweden, maintain that the physician's high work-load affects his sick-listing behaviour negatively. They also find examples of an implicit policy to give priority to pure health-care cases over sick-leave cases. Hence, the separation of physicians into two types could be explained by differences in demand for health care. Physicians with many patients on their waiting lists are forced to spend less resources on sick-leave claimants, while physicians with less "patient pressure" can "afford" to examine the relationship between health and ability properly.[10]

[10]With respect to sick-listing as a kind of medical treatment, the physicians can be distinguished by their decision on whether or not to evaluate the health effects of sickness absence (cf. Englund, 1997).

3.2.3 The insurer

We assume that the insurer's aim is to maximize social welfare, which refers to the workers' situation.

With probability $1-P$, the worker is healthy, which is associated with a high working capacity, i.e. $\Pi(0)=0$. Since physicians do not offer certificates in that case, the individual stays at work, experiencing utility $u_\theta(y_1, H)$. With probability P, the worker falls ill, which is associated with a risk of a reduced working ability, $0 < \Pi(1) < 1$. Depending on the physician's sick-listing practice, the individual may work or not work, possibly without regard to his true capacity.

In the welfare function we let the utility functions u_l and u_m describe the average utility for each group. Thus, the welfare function expresses the average expected utilities of the worker types weighted by the population shares. For the combination of two available physician types, the social welfare function is

$$\begin{aligned} W = \lambda\big((1-P)u_l(y_1,H) + P((1-\Pi(1))u_l(y,H) + \Pi(1)u_l(y,L))\big) + \\ (1-\lambda)\big((1-P)u_m(y_1,H) + \\ P((1-\Pi(1))u_m(y,H) + \Pi(1)u_m(y,L))\big), \quad y \in \{y_1, y_2\}. \end{aligned} \tag{3.7}$$

Note that the income levels of type l and type m may differ, depending on the certification practices of the physicians.[11]

With regard to the budget restriction, it is required that the expected production from working individuals exceeds or at least meets the expected consumption, in the form of net labour income and sickness benefits. Working individuals produce one unit, while non-working individuals produce none. Healthy workers in both groups contribute to production. Whether sick workers do the same or receive sickness benefits, is contingent on their responses to the physicians' certification practices. Consequently, returns and expenses vary with the behaviour of the physician, and as we will see in section 3.4, with the policy instruments used by the insurer to influence the outcome of the absence certification process. We exemplify with the budget expression for the situation where all high-ability individuals work. The budget constraint for this case is

$$P\Pi(1)y_2 + (1 - P\Pi(1))y_1 \leq (1-P) + P(1-\Pi(1)), \tag{3.8}$$

which can be reduced to

$$y_1 + \frac{P\Pi(1)}{(1-P\Pi(1))}y_2 \leq 1. \tag{3.9}$$

[11] We present an additive specification of the welfare function in Appendix B.

3.3 Sickness absence certification

We consider different scenarios, where one as well as both of the physician types are available for sick-listing consultations. We assume that the worker has a free choice concerning which physician to consult. The essential question concerns the treatment that the patient receives from the physician, with regard to absence certification and thereby to ability verification.

Due to incomplete information, the equilibrium state is characterized by strategies as well as beliefs on behalf of all players. Workers decide on a strategy based on their expectations about which type of physician they will meet. If the type l worker visits a relaxed physician, he agrees to be put on the sick-list. If he meets a strict physician, he accepts the certificate, if it is offered to him (c.f. (4.5)). As far as the type m worker is concerned, sickness absence is only an interesting alternative to working if the physician confirms that his health state will affect his ability (see (4.6) and (4.7)).

The physicians issue sickness certificates either with or without an examination of working ability, depending on their being of the strict or the relaxed type. Thus, the physician's behaviour is not linked to the type of worker, which implies that the workers cannot influence the outcome during the consultation but by choosing an absence-certifying physician according to his expectations.

The physicians are non-anonymous, a quite natural assumption considering how health care is sought and supplied at clinics. The fact that there is some information about the physicians makes it possible for the workers to distinguish between them.

3.3.1 Separating and pooling equilibria

A separating equilibrium describes a situation where the two worker types are treated differently in a predictable way. Consequently, one worker type is offered a certificate without an examination while the other is put on the sick list if a reduced working ability is confirmed. A prerequisite for a separating equilibrium is that the workers consistently expect that some physicians sick-list as a matter of routine, while the other physicians certify absence on the basis of future ability impairment. Thus, the workers are able to consult physicians that "match" their preferences.

A pooling equilibrium describes a situation where the workers cannot attach expectations about which, if any, among the available physicians are relaxed and which are strict. Instead, the workers entertain expectations about treatment solely based on the physician population shares. The share of relaxed physicians corresponds to the probability that the worker will be

offered a sickness certificate without examination, whereas the share of strict physicians is associated with absence certification conditioned on an examination. Thus, the worker types receive identical treatments in expectation. The certification process is illustrated further in Appendix C.

The distinguishing factor between a separating and a pooling equilibrium is the amount of information, which renders the dissemination of information and the time aspect an interesting issue.

Pooling equilibria

The proportion of type l in the worker population is $0 < \lambda < 1$. Let β represent the proportion of type R physicians in the physician population.

A pooling equilibrium is characterized by the individual worker expecting to encounter a type R physician with probability β and a type S physician with probability $1-\beta$. With the same probability weights he will get routine and justified absence certification, respectively.

For pooling equilibria regarding the scenario where both physician types are available, i.e. $0 < \beta < 1$, it is the case that the worker types are randomly allocated between two different certification treatments. This means that a type l worker may end up consulting a diligent physician who offers a sickness certificate with probability $\Pi(1)$. Correspondingly, a type m worker may end up at the office of a relaxed physician, who certifies absence indiscriminately. In such a case, and contrary to a type l worker, a type m worker will decline the offer and resume work.

In the scenario of a single physician type, where either $\beta = 0$ or $\beta = 1$, the workers expect all physicians to behave identically. If $\beta = 0$, the optimal strategy for both worker types is to accept the certificate if the physician's examination proves they are eligible for insurance remuneration. In the opposite situation, the workers themselves practically determine whether or not absence is to be certified. Type l gets a sickness certificate automatically by visiting a physician, while type m has no hope of a consultation resulting in more information about his working ability. Consequently, he declines the offer, which in effect means that he does not visit a physician but continues to work. Thus, whether or not the individual is put on the sick list differs for the worker types, even though the actual certification treatment is identical.

Separating equilibria

The existence of separating equilibria is necessarily associated with multiple types, i.e. $0 < \beta < 1$. If it is possible to attach expectations about type to the physicians, the workers can get an idea about what treatment awaits

them behind each office door. They believe that some specific physicians have lax sick-listing routines and some verify working ability. By consulting the former it is certain that the worker is sick-listed without further ado. By visiting the latter, the worker is sure to receive information about his working ability before a sickness certificate is issued. It is thereby possible to achieve a match between patients and physicians.

Type l workers visit relaxed physicians and accept the certificates offered. Type m workers also decide on consultations at, what they expect to be, strict physicians' offices and are sick-listed if is recommended. The separate treatments result in type l virtually engaging in self-certification, while the working/non-working status of type m is determined exogenously.

We hypothesize that there are variations in the social acceptance of claiming sickness benefits, which in its turn implies that underlying social norms distinguish one worker type from the other. The interaction between different types of physicians and workers could play a part in explaining the geographical variations in sickness absence observed for Sweden. Dutrieux & Sjöholm (2003) find that regions characterized by weak labour markets tend to have high levels of sickness absence. Similarly, Johansson & Palme (2002) find that the local unemployment rate is positively correlated with work absence, which goes counter to most empirical studies that use the unemployment rate at country level (Henreksson & Persson, 2004; Arai & Skogman Thoursie, 2001; Askildsen et al. 2002). According to Hesselius' (2006) interpretation it is likely that there is a connection between the local unemployment rate and local social norms. Observations that support Hesselius' line of reasoning are made by ethnologists Frykman & Hansen (2005) who study two Swedish municipalities, one characterized by a high sickness absence rate and the other by a low rate, and find different local cultures with regard to sick-listing. In addition, Arrelöv et al. (2005) find that physicians' certification practices vary geographically, which could be explained by local absence cultures. A similar pattern is observed for physicians in Norway (National Insurance Organization, 2006).

Capacity restrictions The scenario of separate treatments is built upon the assumption that there is a complete match between workers and physicians. However, there may be capacity restrictions in terms of the number of patients per physician type. Take for example a situation where there is one physician of each type and the two physicians are able to receive five patients each, while there are six workers of type l who are allocated to the physicians on a "first come, first serve"-basis. Then, we are talking of a semi-separating equilibrium.

3.3.2 Absenteeism and presenteeism

In our model, the occurrence of incorrect absence certification is the result of the interaction between workers and physicians. The individuals "tagged" (see Akerlof, 1978) as eligible for sickness benefits may have working ability unaffected by illness but be disinclined to work. That kind of identification problem is present in all scenarios, where physicians are "accommodating" towards unmotivated workers. Assuming that physicians are able to diagnose ailments, unjustified absence implies excess utilization rather than explicit malingering.

Imperfect screening may also involve possibly eligible workers refraining from consultations, which occurs when motivated workers expect routine absence certification. Thus, for pooling equilibria, it could result in a situation where eligible individuals are not granted certified sick leave, whereas ineligible individuals are.

Presenteeism, as opposed to absenteeism, is a term that describes the behaviour of individuals attending work despite having ailments that justify sickness absence.[12] Results from a survey among Swedish employees show that the financial consequences of not attending work when feeling ill are essential for the individual's decision on whether or not to report ill (Eek & Rikner, 2005). Using data on U.S. employees, Goetzel et al. (2004) find that the hidden costs of presenteeism, in terms of productivity losses, are considerable. In addition, across 10 health conditions included in their study, presenteeism was the largest component of the annual costs per employee and year. Many firms in the UK offer sick pay exceeding the statutory minimum level, which Chatterji & Tilley (2002) argue may be explained by the firms' efforts to avoid the costs associated with sickness presenteeism (see also Skåtun, 2003).

The effects of presenteeism on individual long-term health are ambiguous (Vingård et al., 2004; Aronsson et al., 2000; Kivimäki et al., 2005). Likewise, empirical studies on the Swedish sickness insurance come to different results with regard to the health effects from changes in the remuneration level (Johansson & Palme, 2005; Bäckman, 1998). In contrast, Englund (1997) maintains that sickness absence, as a medical treatment, could have negative health effects. For insurances fraught with high costs, such as the Swedish system, presenteeism is probably not so much of a problem as absenteeism. Nevertheless, the effects on productivity and individual health of presenteeism should not be considered as insignificant a priori.

[12] We disregard the possibility that workers contract with physicians to examine their working abilities, creating a private market for absence certification.

3.4 Policy instruments

Imperfect screening is attributed to the physician's behaviour. However, to influence the utilization of sickness insurance the insurer may employ various policy instruments. There is a balance to strike between granting sickness benefit in too many cases on the one hand, and risking an exclusion of eligible individuals on the other.

A number of policy instruments may be used in the context of sickness insurance, but not all of them can be analysed in our model, e.g. the period of self-certification or the number of qualifying days. Another alternative, to shift the insurance to the employer, is discussed in chapter 4.

We choose to discuss three relevant policy instruments. Firstly, the insurer can make changes to the insurance policy, in terms of the remuneration level. Besides affecting the worker's economic incentives to apply for sick leave, there is also the possibility of influencing the interaction between patient and physician. We identify a potential trade-off between, on the one hand, the physician acquiring greater knowledge about his patient and, on the other, his developing sentiments of loyalty towards his patient (Mikaelsson et al., 2003). There is evidence showing that a personal relationship between patient and physician increases the probability of the patient receiving a sickness certificate (Wahlström & Alexanderson, 2004). Taking that into consideration, we analyse the effects of the insurer appointing the patient to a physician, instead of letting the patient choose freely.

Insurance systems with low and stable sickness absence rates, such as the systems in Finland and Germany, tend to emphasize control mechanisms (Nyman et al., 2002; Bergendorff & Larheden, 2003). Together with the allegation that the monitoring practices of the Swedish insurance administration are weak, the observation motivates our third choice of instrument (Bergendorff & Larheden, 2003). By means of a second opinion by a randomly appointed physician, the prevalence of incorrect absence certification is attenuated.[13]

3.4.1 Optimal insurance

Before discussing the effects of a remuneration change on the extent of insurance utilization, we analyse the optimal insurance policy, taking into consideration that the population share of type l individuals depends directly on the underlying distribution of preferences for labour and indirectly on the

[13]Söderberg & Alexanderson (2004) remark that the monitoring practices of the social insurance officers are little explored.

prospect (y_1, y_2) generated by the insurance policy.[14]

We exemplify with the scenario of separating equilibria. The insurer maximizes the welfare function

$$\begin{aligned}\max_{y_1,y_2} \quad W(y_1,y_2) = \lambda(y_1,y_2)\big((1-P)u_l(y_1,H)+ \\ P((1-\Pi(1))u_l(y_2,H)+\Pi(1)u_l(y_2,L))\big)+ \\ (1-\lambda(y_1,y_2))\big((1-P)u_m(y_1,H)+ \\ P((1-\Pi(1))u_m(y_1,H)+\Pi(1)u_m(y_2,L))\big)\end{aligned} \tag{3.10}$$

subject to the budget constraint which can be expressed as

$$y_1 + \frac{P(\lambda(1-\Pi(1))+\Pi(1))}{1-P(\lambda(1-\Pi(1))+\Pi(1)}y_2 = 1.^{15} \tag{3.11}$$

The expression above defines net labour income as an implicit function of the population share of type l individuals and the benefit, $y_1 = \eta(\lambda(y_1,y_2),y_2)$. Differentiating y_1 with respect to y_2 produces

$$\begin{aligned}\frac{dy_1}{dy_2}\left(1+\frac{P(1-\Pi(1))}{[1-P(\lambda(1-\Pi(1))+\Pi(1))]^2}\frac{\partial\lambda}{\partial y_1}y_2\right)+ \\ \frac{P(1-\Pi(1))}{[1-P(\lambda(1-\Pi(1))+\Pi(1))]^2}\frac{\partial\lambda}{\partial y_2}y_2+\frac{P(\lambda(1-\Pi(1))+\Pi(1))}{1-P(\lambda(1-\Pi(1))+\Pi(1)}=0.\end{aligned} \tag{3.12}$$

The factor multiplied with dy_1/dy_2

$$\Delta = 1+\frac{P(1-\Pi(1))}{[1-P\left(\lambda\left(1-\Pi(1)\right)+\Pi(1)\right)]^2}\frac{\partial\lambda}{\partial y_1}y_2 \tag{3.13}$$

is less than one since we assume that λ depends negatively on y_1 but is positive, unless a raise in y_1 is self-financing, which we assume is not the case. Hence,

$$\begin{aligned}\frac{dy_1}{dy_2} = -\frac{1}{\Delta}\Bigg(\frac{P(1-\Pi(1))}{[1-P(\lambda(1-\Pi(1))+\Pi(1))]^2}\frac{\partial\lambda}{\partial y_2}y_2+ \\ \frac{P(\lambda(1-\Pi(1))+\Pi(1))}{1-P(\lambda(1-\Pi(1))+\Pi(1)}\Bigg)<0.\end{aligned} \tag{3.14}$$

[14]See Appendix A.

[15]In Appendix B we sketch the welfare function according to an additive specification, where the direct welfare effects of y_1 and y_2 operate only through the average utilities.

Comparing (3.14) with the case where λ is fixed,

$$\frac{dy_1}{dy_2} = -\frac{P\left(\lambda\left(1-\Pi(1)\right)+\Pi(1)\right)}{1-P\left(\lambda\left(1-\Pi(1)\right)+\Pi(1)\right)} < 0, \tag{3.15}$$

we draw the conclusion that the endogeneity of λ increases the magnitude of dy_1/dy_2, through the first term and the factor $1/\Delta$.

The first-order condition is

$$\begin{aligned}\frac{\partial W(y_1,y_2)}{\partial y_2} = \left(\frac{\partial\lambda}{\partial y_2}+\frac{\partial\lambda}{\partial y_1}\frac{dy_1}{dy_2}\right)\big((1-P)u_l(y_1,H)+PU_l(y_2,\Pi(1))-\\ ((1-P)u_m(y_1,H)+PU_m(y_1,y_2,\Pi(1)))\big)+\\ \lambda\big((1-P)u_l'(y_1,H)\frac{dy_1}{dy_2}+P((1-\Pi(1))u_l'(y_2,H)+\Pi(1)u_l'(y_2,L))\big)+\\ (1-\lambda)\big((1-P)u_m'(y_1,H)\frac{dy_1}{dy_2}+\\ P((1-\Pi(1)u_m'(y_1,H)\frac{dy_1}{dy_2}+\Pi(1)u_m'(y_2,L))\big)=0.\end{aligned} \tag{3.16}$$

The first term in (3.16), referring to the movement between types, is zero in equilibrium. The marginal change in λ has no effect, since the marginal worker switching types is indifferent to "type memberships". As can be observed in the following condition, the optimal benefit level is pushed downwards, due to the endogeneity of λ,

$$\begin{aligned}\big(\lambda(1-P)u_l'(y_1,H)+\\ (1-\lambda)((1-P)u_m'(y_1,H)+P(1-\Pi(1))u_m'(y_1,H))\big)(-\frac{dy_1}{dy_2})=\\ \lambda P\big((1-\Pi(1))u_l'(y_2,H)+\Pi(1)u_l'(y_2,L))\big)+\\ (1-\lambda)P\Pi(1)u_m'(y_2,L)).\end{aligned} \tag{3.17}$$

3.4.2 Changes in remuneration

The insurer's need to control the (over-)utilization of the insurance is connected to the proportion of type l workers and to the presence of accommodating physicians. One way to limit the use of the insurance is to make it less attractive not to work. A decrease in the remuneration, from y_2 to $\hat{y_2}$, affects those individuals whose dislike for labour make them prefer leisure to work under the original insurance policy in either of two ways, lowering their income or making them switch from type l to type m. That is, the share of type l workers in the population decreases from λ to $\hat{\lambda}$, as the group size is

endogenous to income levels in the working and the non-working states. This leads to a smaller premium and consequently an increase in the worker's net labour income to $\hat{y_1}$ from y_1.

Decreased over-utilization

Continuing with the example of separate treatments, we again have the budget constraint,

$$P\hat{y_2} + (1-\hat{\lambda})P\Pi(1)\hat{y_2} \leq \big[(1-P) + (1-\hat{\lambda})P(1-\Pi(1))\big](1-\hat{y_1}). \quad (3.18)$$

The remuneration decrease leads to a larger share of type m workers, $1-\hat{\lambda}$, that consults the physicians with strict routines and receives a sickness certificate with probability $P\Pi(1)$. With probability $P(1-\Pi(1))$ type m continues working, earning $\hat{y_1}$. The smaller share of type l workers seek help from relaxed physicians, who certify absence conditioned on their patients being ill.

In terms of behaviour, increasing the co-insurance rate affects those individuals that switch from type l to type m. However, compared to other consequences of the policy change, the direct welfare effect of marginal changes in λ is small. With a larger share of motivated workers, the likelihood of eligible certificate holders increases and the share of non-working individuals decreases, due to the decreasing number of type l individuals, who receive routine absence certification.

Individuals already of type m or remaining type l are not influenced behaviorally. However, non-working and working individuals are affected by the benefit decrease in terms of consumption possibilities but the effect goes in different directions for the two groups. Considering that $\hat{y_1} > y_1$, working pays off to a greater extent, while the absentees have to make do with less than before, as $\hat{y_2} < y_2$.

From the insurer's perspective, the improved possibilities of justified absence certification are a positive effect of increasing the co-insurance rate. Considering that the utility of working is non-decreasing in the ability to work, the transfer of individuals with a high working capacity from the sick list is another welfare improvement. However, it comes at the cost of making the insurance coverage less generous for the population of risk-averse workers.

Increased under-utilization

In scenarios where workers are not separated treatment-wise, the benefit decrease leads to an increased under-utilization. For example, in the case where only physicians with lax routines are available, a consultation does

not lead to ability verification, a fact that prompts the larger share of type m individuals to refrain from a physician's visit. In the budget constraint,

$$\hat{\lambda} P \hat{y_2} \leq \big[(1-P) + (1-\hat{\lambda})P\big](1-\hat{y_1}), \tag{3.19}$$

we observe that more individuals work in spite of a possible future ability reduction, as $(1-\hat{\lambda})P > (1-\lambda)P$. Simultaneously, the share of individuals with well-founded absence certification decreases, as absentees constitute a smaller group of the worker population, $\hat{\lambda}P < \lambda P$.

An increase in the co-insurance rate influences the utilization in different directions across scenarios. Its introduction is based on how the insurer weighs the extent and costs of presenteeism against those of absenteeism.[16]

3.4.3 Patient allocation

The insurer could also consider introducing an allocation mechanism. The mechanism allocates the workers to the available physicians in a completely random fashion. We base our discussion on the assumption that the situation prior to the introduction of the instrument is characterized by type-specific treatment of the workers. Hence, by using the allocation instrument separating equilibria are ruled out.

First we consider the outcome of the unregulated certification process. Workers of type l are sick-listed by relaxed physicians conditional on sickness but regardless of working ability, while absence for type m individuals is certified only if strict physicians observe a negative effect on future ability. The random allocation causes workers to expect a physician of type R with probability $0 < \beta < 1$. Hence, some type l (type m) individuals end up at the office of a strict physician and some at the office of a physician with lax sick-listing routines. The policy instrument brings about a negative change for those type m individuals, who choose to work because they do not get any information about their working capacities from the physicians. The situation also changes for the worse for those type l individuals who are subjected to ability examination and miss out on sick-listing due to having unaffected working abilities (c.f. (4.5) and (4.7)). As can be seen from the

[16] A "non-deterrent" insurance, where the insurer tolerates some degree of moral hazard, is also considered by Lindbeck & Persson (2006) and Lantto (1991).

social welfare function,

$$\begin{aligned} W = \lambda\big[&(1-P)u_l(y_1,H)+ \\ &P\big(\beta U_l(y_2,\Pi(1)+(1-\beta)U_l(y_1,y_2,\Pi(1))\big)\big]+ \\ &(1-\lambda)\big[(1-P)u_m(y_1,H)+ \\ &P\big(\beta U_m(y_1,\Pi(1)+(1-\beta)U_m(y_1,y_2,\Pi(1))\big)\big] \end{aligned} \tag{3.20}$$

type l workers cannot expect routine absence certification conditional on sickness anymore, since it is also dependent on meeting relaxed physicians, which occurs with probability $\beta < 1$. Correspondingly, the probability of ability verification for type m workers is now $(1-\beta) < 1$.

More type l individuals and more type m individuals are denied and decline routine absence certification, respectively. Consequently, for a given insurance policy, the introduction of the instrument is associated with an expected increase in production and a corresponding decrease in benefit expenditures, as the share of absentees decreases. Then, the insurer can fix y_2 and increase the net labour income y_1. Besides leading to a larger share of type m individuals, the premium decrease benefits the increased number of working individuals without deteriorating the financial circumstances of the absentees. As noted earlier, the welfare effect of the movement between the two groups is small, whereas increased consumption possibilities have positive but in magnitude unknown welfare implications. Alternatively, the insurer could make the system more generous.

Another effect of the random allocation of patients to physicians is the "mismatch" between patients and physicians. The fact that some type l and type m workers receive treatments that are better suited for the other type implies a cost in terms of welfare. In addition, the mismatch has implications for the relationship between ability and labour status. By making type l consult physicians with strict certification practices, the insurer is able to separate justified from unjustified sickness absence to some extent. Simultaneously, the allocation of motivated workers to physicians with lax routines implies an exclusion of individuals with possibly low working ability from the sick list.

3.4.4 Second opinion

Another way to control the utilization of sickness insurance is to randomly appoint a physician to have a second look at the worker's case before granting sickness benefits. As physicians do not deny eligible workers absence certification, a second opinion alleviates the problem of type 2 errors. Another

motive for using the control mechanism may also be identified, that of giving the worker information about his ability.[17]

Introducing a system of second-opinion influences scenarios where both physician types are available for absence certification. Taking the case of separating equilibria as an example, another physician's "inspection" may alter the outcome of the first consultation. A worker of type l who is granted sick leave in the first stage by a physician with lax routines, may meet a strict physician in the second stage. Thus, there is a probability $(1-\beta)(1-\Pi)$ that the worker is not granted sick leave. For the worker of type m, neither the second assessment of a strict nor that of a relaxed physician changes the behaviour of the worker, since ability is verified in the first stage. Thus, the requirement of a second opinion leads to a reduced frequency of incorrect absence certification without increasing the prevalence of presenteeism.

As the effort of the physician is unobservable for all but the treated patient, his behaviour is monitored indirectly by means of the second opinion. In addition, the physician who certifies absence in the first stage but is found to be incorrect in the second stage could receive a sanction in order to prevent routine absence certification. If a high working ability is observed by the physician giving the second opinion, the certificate is null and void, the compensation is withdrawn and the physician receives a punitive sanction $\Omega > 0$. If a low ability is observed, no changes are made. In other words, the physician is not punished for expending too little effort if the worker in question is found to be eligible at the second stage. A physician who fails to perform the examination runs a risk $(1-\beta)(1-\Pi(1))$ that he will be sanctioned. In effect, only the behaviour of relaxed physicians is influenced by the policy instrument. The condition for type R physicians to perform ability examinations is $(1-\Pi(1))(1-\beta)\Omega \geq d_R$.

When deliberating whether or not to introduce a second-opinion system combined with sanctions, the insurer takes into account the enforcement cost when evaluating the welfare effects. We assume that the physician's salary is not included in the administrative costs of the sickness insurance. Irrespective of the equilibrium played, type l workers are negatively influenced by the system of second opinion, while type m workers benefit from ability verification. Furthermore, the monitoring policy leads to a better match between ability and labour status, which adds to the positive effects in terms of welfare. Depending on the costs incurred, the decreased extent of unjustified absence certification could be compatible with a more generous transfer for a given premium.

In the literature on unemployment insurance there is an increasing in-

[17]See the discussion in Lantto (1991).

terest for active screening policies. Despite the uncertainty about the costs, theoretical and empirical findings cautiously support the introduction of control mechanisms in order to increase job search effort among the unemployed (Fredriksson & Holmlund, 2003).[18]

3.5 Concluding discussion

Empirical studies find that there is a positive relationship between the replacement rate and the sickness absence rate (e.g. Johansson & Palme, 2005). However, they do not recognize that, for all absence periods extending the self-certification period, a physician's assessment of the claimant's working ability is required. Hence, the findings on individual absence behaviour imply that the physician's certification practices do not function as a perfect control mechanism.

The medical expert plays an essential role in sickness insurance, which so far has received little attention in the economic literature. We model the absence certification process, taking into consideration that screening procedures are imperfect. There are two types of physicians who differ in their efforts to verify whether or not claimants are eligible for sickness benefits. The physician's certification practices have implications for the worker's decision for or against a sick-listing consultation. The two types of workers deal with the uncertainty regarding the health-related ability impairment in different ways. The "less motivated" worker prefers leisure to labour regardless of his working ability, whereas the "motivated" worker prefers to work if it is medically confirmed that his ability is high despite illness.

If circumstances allow it, the workers can attach expectations to physicians about their certification practices. Then, if there are physicians who are restrictive with absence certification as well as physicians who certify generously, the certification process is associated with separate treatments of the worker types. Such an interaction between physicians and workers could be one underlying cause of the persistent geographical pattern of sickness absence found in Sweden (Dutrieux & Sjöholm, 2003).

With respect to remuneration, the endogeneity of the group size of type l individuals affects the optimal insurance contract by decreasing the optimal benefit level. Besides the co-insurance rate, we consider another two relevant policy measures which the insurer can take in a situation where the sickness

[18]There are examples of "non-administrative" punitive sanctions with respect to sickness insurance. In Lantto (1991) a healthy individual who reports sick experiences feelings of shame if he is detected. Lindbeck & Persson (2006) study the effect of a social norm as a control mechanism involving sanctions.

insurance is perceived to be over-utilized. Firstly, a decrease in the insurance remuneration causes the number of unjustified absence certifications to drop, but involves a less generous insurance coverage. Secondly, introducing a random allocation of patients to physicians addresses the problem that may occur when personal relationships are formed between patient and physician. Starting from a separating equilibrium, the allocation mechanism leads to a decrease in routine sick-listing of type l. However, justified absence certifications of type m also decrease. Thirdly, the insurer can appoint physicians to give second opinions on benefit applications. The outcome is beneficial for type m workers but is to the detriment of type l workers. It also involves more administrative costs than the other instruments considered. All three measures influence the number of sickness absences but involve positive as well as negative effects for the workers.

Further work is required to analyse physicians' sick-listing practices and to evaluate measures to control sickness absence. For the Swedish context, general guidelines have been discussed as a way to support certifying physicians (Alexanderson et al., 2005).[19] However, it may also be considered a means to better monitor their certification practices. Moreover, little is known about the effects of reimbursement forms on sick-listing behaviour. Gathering data on sick leave cases (see e.g. Hetzler et al., 2005) appears to be a fruitful way to increase our knowledge about physician behaviour in the context of sickness absence certification.

[19]Such guidelines are now being introduced (National Board of Health and Welfare, 2006b).

References

Akerlof G A, 1978, "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning", *The American Economic Review*, 68, 8-19.

Alexanderson K, Brommels M, Ekenvall E, Karlsryd E, Löfgren A, Sundberg L & Österberg M, 2005, "Problem inom hälso- och sjukvården kring handläggning av patienters sjukskrivning", rapport, Institutionen för klinisk neurovetenskap, Karolinska Institutet, Stockholm.

Andrén D, 2001, *Work, Sickness, Earnings, and Early Exit from the Labor Market: An Empirical Analysis Using Swedish Longitudinal Data*, dissertation, Göteborg University.

Andrén D, 2005, "Never on a Sunday: Economic Incentives and Short-Term Sick Leave in Sweden", *Applied Economics*, 37, 327-338.

Andrén D, 2007, "Long-Term Absenteeism Due to Sickness in Sweden. How Long Does It Take and What Happens After?", *The European Journal of Health Economics*, 8, 41-50.

Arai M & Skogman Thoursie P, 2005, "Incentives and Selection in Cyclical Absenteeism", *Labour Economics*, 12, 269-280.

Aronsson G, Gustafsson K & Dallner M, 2000, "Sick But Yet at Work. An Empirical Study of Sickness Presenteeism", *Journal of Epidemiology and Community Health*, 54, 502-509.

Arrelöv B, 2003, *Towards Understanding of Determinants of Physicians' Sick-listing Practice and their Interrelations. A Population-Based Epidemiological Study*, dissertation, Uppsala University.

Arrelöv B, 2006,"Läkarna i sjukskrivningsprocessen" in Palmer E (ed.), *Sjukförsäkring, kultur och attityder. Fyra aktörers perspektiv*, Analyserar 2006:16, Försäkringskassan, 423-477.

Arrelöv B, Borgquist L & Svärdsudd K, 2005, "Influence of Local Structural Factors on Physicians' Sick-Listing Practice: A Population-Based Study", *European Journal of Public Health*, 15, 470-474.

Arrow K J, 1963, "Uncertainty and the Welfare Economics of Medical Care", *The American Economic Review*, 53, 941 - 973.

Askildsen J E, Bratberg E & Nilsen Ø, 2002, "Unemployment, Labour Force Composition and Sickness Absence: A Panel Data Study", discussion paper 466, Institute for the Study of Labor (IZA), Bonn.

Barmby T & Larguem, M, 2004, "Worker Absenteeism: A Study of Contagion Effects", discussion paper 1, Centre for European Labour Market Research, Aberdeen.

Barmby T, Sessions J G & Treble J, 1994, "Absenteeism, Efficiency Wages and Shirking", *Scandinavian Journal of Economics*, 96, 561-567.

Benitez-Silva H, Buchinsky M & Ruser J, 2004, "How Large Are the Classification Errors in the Social Security Disability Award Process", working paper 10219, National Bureau of Economic Research (NBER), Cambridge, Massachusetts.

Bergendorff S & Larheden H, 2003, "Att förhindra och förkorta sjukfrånvaro - erfarenheter från fyra länder", RFV Analyserar 2003:16, Riksförsäkringsverket.

Blomqvist Å, 1991, "The Doctor as Double Agent: Information Asymmetry, Health Insurance, and Medical Care", *Journal of Health Economics*, 10, 411-433.

Bolin K & Lindgren B, 2001, "How Stable are the Empirical Results of the Grossman Model? Testing Different Indicators of Health Capital and Health Investments in Sweden 1980/81, 1988/89, and 1996/97", Studies in Health Economics 36, Lund University Centre for Health Economics (LUCHE), Lund.

Brown S, 1994, "Dynamic Implications of Absence Behaviour", *Applied Economics*, 26, 1163-1176.

Brown S & Sessions J G, 1996, "The Economics of Absence: Theory and Evidence", *Journal of Economic Surveys*, 10, 23-53.

Bäckman O, 1998, *Longitudinal Studies on Sickness Absence in Sweden*, Socialforskningsinstitutet 34, Stockholms universitet.

Chatterji M & Tilley C J, 2002, "Sickness, Absenteeism, Presenteeism, and Sick Pay", *Oxford Economic Papers*, 54, 669-687.

Diamond P A & Mirrlees J A, 1978, "A Model of Social Insurance with Variable Retirement", *Journal of Public Economics*, 10, 295-336.

Diamond P & Sheshinski E, 1995, "Economic Aspects of Optimal Disability Benefits", *Journal of Public Economics*, 57, 1-23.

Dutrieux J & Sjöholm K-R, 2003, "Regionala skillnader i sjukskrivning - Hur ser de ut och vad beror de på?" RFV Analyserar 2003:17, Riksförsäkringsverket.

Eek D & Rikner K, 2005, "What Determines People's decisions Whether or Not to Report Sick?", *Applied Economics*, 37, 533-543.

Englund L, 1997, "Är sjukdom ett bra skäl för att bli sjukskriven?" Försäkringskasseförbundet.

Englund L, 2000, *Sick-Listing - Attitudes and Doctor's Practice. With Special Emphasis on Sick-Listing Practice in Primary Health Care*, dissertation, Uppsala University.

Englund L, 2001, "Förändringar i distriktläkares sjukskrivningspraxis mellan åren 1996 och 2001 i ett svenskt landsting", rapport, Centrum för Klinisk Forskning, Dalarna.

Evans R G, 1974, "Supplier Induced Demand: Some Evidence and Implications" in Perlman M (ed.), *The Economics of Health and Medical Care*, MacMillan.

Evans R G, 1984, *Strained Mercy: The Economics of Canadian Medical Care*, Butterworths.

Fredriksson P & Holmlund B, 2003, "Improving Incentives in Unemployment Insurance: A Review of Recent Research", working paper 922, Center for Economic Studies & Institute of Economic Research (CESIfo), München.

Frick B & Maló M, 2005, "Labour Market Institutions and Individual Absenteeism in the European Union: The Relative Importance of Sickness Benefits Systems and Employment Protection Legislation", working paper, Universität Witten/Herdecke.

Frykman J & Hansen K, 2005, "Att leva på kassan. Allmän försäkring och lokal kultur", Analyserar 2005:4, Försäkringskassan.

Garcia Mariñoso B & Jelovac I, 2003, "GP's Payment Contracts and Their Referral Practice", *Journal of Health Economics*, 22, 617-635.

Gilleskie D B, 1998, "A Dynamic Stochastic Model of Medical Care Use and Work Absence", *Econometrica*, 66, 1-45.

Giuffrida A & Gravelle H, 2001, "Inducing or Restraining Demand: The Market for Night Visits in Primary Care", *Journal of Health Economics*, 20, 755 - 779.

Goetzel R Z, Long S R, Ozminovski R J, Hawkins K, Wang S & Lynch W, 2004, "Health, Absence, Disability, and Presenteeism Cost Estimates of Certain Physical and Mental Health Conditions Affecting U.S. employers", *Journal of Occupational and Environmental Medicine*, 46, 398-412.

Grytten J, Carlsen F & Sørensen R, 1995, "Supplier Inducement in a Public Health Care System", *Journal of Health Economics*, 14, 207 - 229.

Gulbrandsen P, Forde R & Aasland O G, 2002, "What Does It Feel Like for a Physician to Be a Gatekeeper?", *Tidsskrift for den Norske laegeforening*, 122, 1874-1879.

Henrekson M & Persson M, 2004, "The Effects on Sick Leave of Changes in the Sickness Insurance System",*Journal of Labor Economics*, 22, 87-113.

Henrekson M, Llanto K & Persson M, 1992, "Bruk och missbruk av sjukförsäkringen", SNS Förlag.

Hesselius P, 2006, "Work Absence and Social Security in Sweden", working paper, second version 2006-09-06, presented at the conference *Reforming the Welfare State*, Saltsjöbaden, September 9-10 2006, arranged by Center for Business and Policy Studies (SNS) and the National Bureau of Economic Research (NBER).
http://www.sns.se/document/patrik_hesselius_1199.pdf

Hesselius P, Johansson P & Larsson L, 2005, "Monitoring Sickness Insurance Claimants: Evidence from a Social Experiment", working paper 15, Institute for Labour Market Policy Evaluation (IFAU), Uppsala.

Hetzler A, Melén D & Bjerstedt D, 2005, *Sjuk-Sverige, Försäkringskassan, rehabilitering och utslagningen från arbetsmarknaden*, Brutus Östlings Bokförlag Symposion.

Jansson K & Johansson P, 2003, "Sjukvården och de sjukskrivna - kan pengarna användas bättre?" Memorandum, Sveriges läkarförbund.

Johansson P & Palme M, 1996, "Do Economic Incentives Affect Work Absence? Empirical Evidence Using Swedish Micro Data". *Journal of Public Economics*, 59, 195-218.

Johansson P & Palme M, 2002, "Assessing the Effect of Public Policy on Worker Absenteeism". *Journal of Human Resources*, 37, 381-409.

Johansson P & Palme M, 2005, "Moral Hazard and Sickness Insurance". *Journal of Public Economics*, 89, 1879-1890.

Kaiser C, 1998, "What Do We Know about Employee Absence Behaviour? An Interdisciplinary Interpretation", *Journal of Socio-Economics*, vol. 27, 79-96.

Lantto K, 1991, *Optimal Deterrents to Malingering. The Role of Incentives, Attitudes and Information Costs in Social Insurance, especially Sickness Benefit and Welfare*, dissertation, Stockholm University.

Larsen B A, Forde O H & Tellnes G, 1994, "Physician's Role in Certification for Sick Leave", *Tidsskrift for den Norske laegeforening*, 114, 1442-1444.

Larsson L, Kruse A, Palme M & Persson M, 2005, "En hållbar sjukpenningförsäkring", Välfärdsrådets rapport, Studieförbundet Näringsliv och Samhälle (SNS).

Lidwall U, Marklund S & Skogman Thoursie P, 2005, "Sickness Absence in Sweden" in Gustafsson R (ed.), *Work Life and Health in Sweden 2005*, Arbetslivsinstitutet.

Lindbeck A & Persson M, 2006, "A Model of Income Insurance and Social Norms", seminar paper 742, Institute for International Economic Studies (IIES), Stockholm.

Ma C A & McGuire T G, 1997, "Optimal Health Insurance and Provider Payment", *American Economic Review*, 87, 685-704.

McGuire T G, 2000, "Physician Agency" in Culyer, A J and Newhouse, J P (eds.), *Handbook of Health Economics*, vol. 1A, Elsevier, 461-536.

Mikaelsson B, Ekholm J, Kärrholm J, Murray R, Sandberg T, Söderberg J & Nyman K, 2003, "Den svenska sjukan II - regelverk och försäkringsmedicinska bedömningar i åtta länder", Departementsserie 2003:63.

Mooney G & Ryan M, 1993,"Agency in Health Care: Getting beyond first principles", *Journal of Health Economics*, 12, 125-136.

National Board of Health and Welfare (Socialstyrelsen), 2006a, "Socialstyrelsens tillsyn av sjukskrivningsprocessen år 2003-2005 - sammanfattande slutsatser", rapport 2006-03-15.

National Board of Health and Welfare (Socialstyrelsen), 2006b, "Beslutsstöd i form av försäkringsmedicinska riktlinjer - En del av en mer kvalitetssäkrad, enhetlig och rättssäker sjukskrivningsprocess", rapport.

National Insurance Organisation (Rikstrygdeverket), 2006, "Fastlegers sykemeldingspraksis I: Variasjoner", rapport 05, http://www.nav.no/805317044.cms.

Nyman K, Bergendorff S & Palmer E, 2002, "Den svenska sjukan - sjukfrånvaron i åtta länder", Departementsserie 2002:49.

Palmer E, 2004, "Sjukskrivningen i Sverige - en inledande översikt", in Hogstedt C, Bjurvald M, Palmer E & Theorell T (eds.), *Den höga sjukfrånvaron - sanning och konsekvens*, report 2004:15, Statens Folkhälsoinstitut, 27 - 80.

Parsons D O, 1991, "Self-Screening in Targeted Public Transfer Programs", *Journal of Political Economy*, 99, 859-876.

Parsons D O, 1996, "Imperfect 'Tagging' in Social Insurance Programs", *Journal of Public Economics*, 62, 183-207.

Osterkamp R & Röhn O, 2005, "Being on Sick Leave - Possible Explanations for Differences of Sick-leave Days Across Countries", working paper 19, Institute for Economic Research (Ifo), München.

Rikner K, 2002, *Sickness Insurance: Design and Behaviour*, dissertation, Lund University.

Ryan M, 1994, "The Agency Relationship in Health Care: Lessons for Economists from Sociologists", *American Journal of Economic Sociology*, 53, 207-218.

Samtal om socialförsäkring Nr 7, 2006, "Sjuk eller ledsen", Socialförsäkringsutredningen.

Skåtun J D, 2003, "Take Some Time Off, Why Don't You? Endogenous Sick Leave and Pay"*Journal of Health Economics*, 22, 379-402.

Smith R T & Lilienfeld A M, 1971, "The Social Security Disability Program, an Evaluation Study", research report 39, Social Security Administration, Office of research and statistics, Washington D C.

Söderberg E & Alexanderson K, 2003, "Sickness Certification Practices of Physicians: A Review of the Literature", *Scandinavian Journal of Public Health*, 31, 460-474.

Söderberg E & Alexanderson K, 2004, "Gatekeepers in Sickness Insurance: A Systematic Review of the Literature on Practices of Social Insurance Officers", *Health and Social Care in the Community*, 13, 211-223.

Söderberg E, Dahlqvist H & Shahirnejad B, 2002, "Underlag för beslut - en studie av läkarintyg ur ett socialförsäkringsperspektiv", rapport 2002-10-10, Oktaven kassorna.

Tellnes G, Sandvik L & Moum T, 1990, "Inter-Doctor Variation in Sickness Certification", *Scandinavian Journal of Primary Health Care*, 8, 45-52.

Vingård E, Alexanderson K & Norlund A, 2004, "Chapter 10. Sickness Presence", *Scandinavian Journal of Public Health*, 32, Suppl. 63, 216-221.

Wahlström R & Alexanderson K, 2004, "Chapter 11. Physicians' Sick-Listing Practices", *Scandinavian Journal of Public Health*, 32, Suppl. 63, 222-255.

Whinston M D, 1983, "Moral Hazard, Adverse Selection, and the Optimal Provision of Social Insurance", *Journal of Public Economics*, 22, 49-71.

Zweifel P & Breyer F, 1997, *Health Economics*, Oxford University Press.

Appendix A: The population share of type *l* individuals

The population of workers is heterogeneous with respect to the disutility of labour, $\mu \leq 0$, which is distributed according to a continuously differentiable cumulative probability distribution $G(\mu)$. For a given insurance policy there is a cut-off value μ^*, under which we have the individuals of type l. Thus, $\lambda = G(\mu^*)$.

Let us assume that the direct utility of working is $v(y_1, \alpha) + \mu$, while the direct utility of not working is $v(y_2, \alpha)$ for $\alpha = L, H$. In contrast to type m, type l prefers sick leave to work when having high working ability. For a given insurance policy, the values of μ for which workers are of type l are defined by

$$v(y_1, H) + \mu \leq v(y_2, H). \tag{*}$$

Hence,

$$\lambda = G(v(y_2, H) - v(y_1, H)).$$

For a given insurance policy λ increases if disutility of labour increases. For a given distribution of μ, λ tends to increase as the benefit level is raised, and to decrease as the premium level is reduced. We have that

$$\frac{\partial \lambda}{\partial y_1} = -g(v(y_2, L) - v(y_1, H))v'(y_1, H)$$

and

$$\frac{\partial \lambda}{\partial y_2} = g(v(y_2, L) - v(y_1, H))v'(y_2, L)).$$

If (*) is violated, i.e. if the worker is willing to work when his ability is high, we assume that he receives a strictly positive benefit z from this, such that (4.7) is satisfied for a given $Pi(1)$. This discontinuity excludes the possibility of a third type of worker who is not willing to work without an update about his ability. The addition of a third type to the model would not provide more insight to the analysis. The background of the benefit z may be that type m is under the influence of a social norm. Type m individuals receive social approval when showing motivation to work, even when they cannot ascertain the effects of bad health on working ability. However, we disregard the norm-related benefit in the welfare analysis.

Appendix B: Welfare function

We exemplify with the case where both types are sick-listed on the basis of a health-related ability impairment, under the assumption that utility is $v(y_1, H) + \mu$ if working, and $v(y_2, L)$ if not working. The welfare function W is

$$W = (1-P)\int_{\mu^{\min}}^{\mu^*}(v(y_1,H)+\mu)g(\mu)d\mu +$$
$$P(1-\Pi(1))\int_{\mu^{\min}}^{\mu^*}(v(y_1,H)+\mu)g(\mu)d\mu + P\Pi(1)\int_{\mu^{\min}}^{\mu^*}(v(y_2,L))g(\mu)d\mu +$$
$$(1-P)\int_{\mu^*}^{\mu^{\max}}(v(y_1,H)+\mu)g(\mu)d\mu +$$
$$P(1-\Pi(1))\int_{\mu^*}^{\mu^{\max}}(v(y_1,H)+\mu)g(\mu)d\mu + P\Pi(1)(v(y_2,L))g(\mu)d\mu \Rightarrow$$
$$W = \lambda(1-P)v(y_1,H) + \lambda P(1-\Pi(1))v(y_1,H) +$$
$$\lambda P\Pi(1)v(y_2,L) + (1-P\Pi)\int_{\mu^{\min}}^{\mu^*}\mu g(\mu)d\mu +$$
$$(1-\lambda)(1-P)v(y_1,H) + (1-\lambda)P(1-\Pi(1))v(y_1,H) +$$
$$(1-\lambda)P\Pi(1)v(y_2,L) + (1-P\Pi)\int_{\mu^*}^{\mu^{\max}}\mu g(\mu)d\mu \Rightarrow$$
$$W = (1-P\Pi(1))v(y_1,H) + P\Pi(1)v(y_2,L) + (1-P\Pi)\int_{\mu^{min}}^{\mu^{\max}}\mu g(\mu)d\mu$$

Appendix C: Absence certification - interaction between agents with incomplete information

The sickness certification process is modelled as a game of incomplete information. The relationship between the players, i.e. the workers and the physicians, is characterized by asymmetric information with respect to the other players' utility functions and, hence, type. They have beliefs about what type of worker and what type of physician to encounter in order to adopt the optimal strategy. Furthermore, the information set arrived at by playing the equilibrium strategies is reached with positive probability. That is, the equilibrium beliefs are determined by the equilibrium strategies.

Strategies and expectations

In the context of a consultation, the physician can decide to issue the sickness certificate without verifying his patient's working ability or condition the offer on his observing a low ability. The worker in his turn decides whether or not to accept an offer of absence certification. He can visit the physician and accept the certificate, provided that it is offered to him. He may also decline the offer, whereupon he resumes work. The decision to stay at work is also possible in the case when the physician is not prepared to certify absence.

The type l worker sticks to the strategy of visiting a physician and accepting sickness absence certification. Contrarily, the optimal strategy of the type m worker is contingent on the type of physician he expects to meet, since he is only interested in consulting a physician if his work ability is examined.

The behaviour of the two physician types, however, is not influenced by the type of worker. The relaxed type R physicians certify absence without any examination while strict type S physicians certify absence on the basis of ability verification.

The expectations of the workers and the physicians are based on the composition of the physician population and worker population, together with their equilibrium strategies.

Separate and non-separate treatments

To elucidate the role of the players' expectations and strategies in equilibrium we exemplify by describing two cases, one characterized by separate treatments of the workers, the other characterized by non-separate treatments.

We illustrate the process of sickness absence certification in a game with four players. There are two workers, one of type l and one of type m and two

physicians, one of type R, called "1", and one of type S, called "2". Hence, the proportion of type l in the worker population is $\lambda = 1/2$ and the proportion of type R in the physician population is $\beta = 1/2$.

For a separating equilibrium it is required that the workers hold the belief that physician 1 certifies absence as a matter of routine while physician 2 controls his patient's ability first. Then, it is optimal for the type l worker to visit physician 1 and for the type m worker to consult physician 2. Thus, given the worker's equilibrium strategies, the type l (m) worker ends up at the office of physician 1 (2) with certainty and receive separate treatments. Note that the physician types have dominant strategies by assumption. In effect, the belief entertained by a physician about what type of worker/patient he will meet does not affect his behaviour.

In the case of a pooling equilibrium, the workers receive identical treatments in expectation. Their expectations about treatment are based on the composition of the physician population. That means that the type l worker risks correct absence certification performed by the strict physician 2, while the type m worker risks routine absence certification performed by the relaxed physician 1. Thus, the workers expect absence certification with probability $\beta + (1 - \beta)\Pi(1)$, where $\Pi(1)$ is the probability of low working ability given bad health.

Separating Equilibria

Table 1 Absence certification with two physician types: separate treatments.

	Belief	**Strategy**
Physician 1 of type S	Encounter only type m worker	Verify working ability, certify absence if $\alpha = L$
Physician 2 of type R	Encounter only type l worker	Certify absence without verifying working ability
Worker type l	Physician 1 (2) certifies absence conditioned on (without) examination	Visit physician 2 and accept the certificate
Worker type m	Physician 1 (2) certifies absence conditioned on (without) examination	Visit physician 1 and accept the certificate if it is offered

Pooling Equilibria

Table 2 Absence certification with two physician types: identical treatments.

	Belief	**Strategy**
Physician 1 of type S	Encounter type l worker with probability λ	Verify working ability, certify absence if $\alpha = L$
Physician 2 of type R	Encounter type l worker with probability λ	Certify absence without verifying work ability
Worker type l	With probability β physician 1 (2) certifies absence at request	Visit physician 1 (2) and accept the certificate if it is offered
Worker type m	With probability β physician 1 (2) certifies absence at request	Visit physician 1 (2) and accept the certificate if it is offered

Chapter 4

Occupational Physicians and General Practitioners - Implications for Sickness Absence Certification

4.1 Introduction

An occupational physician's (OP) relationship with employers and workers is of special interest in the context of sickness insurance. The OP assesses the functional impairment of ill workers while being an employee himself, which may affect his decision to certify absence. In contrast, the firm's influence on the sick-listing routines of the general practitioner (GP) is very likely weaker. The "double loyalty" of the OP adds to his conflicting roles of acting as gatekeeper and advocating the patient's interests.

Findings from studies on consumer satisfaction concerning occupational health services (OHS) suggest that low satisfaction scores for OHS are related to a conflict of interest between employers and workers (Verbeek et al., 2001). A Dutch study reports that the independence of the OP is called into question in circumstances, where the OP is perceived to play the role of mediator, such as consultations regarding assessment of working ability (Plomp, 1999). In expectation thereof the workers "may take a cautious and calculating approach" to the OP (ibid, p. 187).[1]

[1]Draper (2002) describes the problems of U.S. company physicians in striking a balance between, on one side, loyalty towards workers, which subjects the OP's position and possibly also employment in the firm to risks, and, on the other, loyalty towards management, which causes the workers to mistrust the OP.

Despite the dilemma of conflicting interests, the effect of which on the OP's certification practices is ambiguous, the possibilities of the OP to make qualified ability assessments seem to be better than the GP's, considering the former's closer relation to the place of work.

In spite of the fact that the OP has an influential position in many sickness insurance systems, the role of the OP, and of physicians in general, in sickness insurance has received little attention in economic literature.[2]

In our model of sickness absence certification, we contrast the sick-listing routines of the OP with those of the GP. The certification practices are characterized by how the physicians balance concern for their patients' welfare against the firm's interests and by the assessment precision. With respect to the workers, the random event of illness is associated with the risk of a subsequent reduction in working ability. There are two types of workers, whose motives for sick leave differ. We analyse the interaction between physicians and workers and its consequences for the utilization of sickness insurance.

4.1.1 Background

Economic incentives in sickness insurance

Sickness insurance compensates for income losses due to health-related ability impairments. However, empirically, it has been shown that there is a negative correlation between the costs of being absent, in terms of sickness benefits, and the utilization of sickness insurance, indicating that economic incentives matter for the behaviour of insured individuals (Johansson & Palme, 1996, 2002; Henrekson & Persson, 2004; Andrén, 2005; Edgerton et al., 2004).[3] Speaking of economic losses in a wider sense, studies also show that sickness absence tends to increase with job security and decrease with the risk of unemployment (Riphahn, 2004; Lindbeck et al., 2006; Arai & Skogman Thoursie, 2005; Dyrstad & Ose, 2002; Askildsen et al., 2002). On a theoretical basis, there are still few contributions with sickness-insurance models.

[2]In Germany, OPs are not allowed to certify sickness absence, while their Dutch colleagues took on the task of absence certification, formerly a responsibility of insurance physicians, in 1994 (Weel et al., 1999). According to Whitaker (2001), occupational health professionals (physicians as well as nurses) in the UK are often asked by employers to verify whether or not the absence, usually certified by a GP, is justified. In Sweden, all physicians are authorized to certify absence, but the bulk of certificates is issued by GPs (see e.g. Söderberg et al., 2002). In contrast, OPs are the main certifying body in Finland, where, as in the Netherlands, the provision of OHS is mandatory for employers (Bergendorff & Larheden, 2003).

[3]Whether or not restrictive reforms have consequences for individual health has not been established, see Johansson & Palme (2005) and Bäckman (1998).

Examples are Rikner (2000), Lindbeck & Persson (2006) as well as Llanto (1991), which we discuss further in relation to our model of sickness absence certification in chapter 3.

While individual absence behaviour has been explored to some extent, little work has been done on the certifying physician. The prerequisite of absence certification by a physician has been found to function as a deterring mechanism (Henrekson et al., 1992; Johansson & Palme, 2005; Hesselius et al. 2005). For more information about physicians' certification practices, we turn to findings from, i.a., sociomedical research.

Sick-listing behaviour and patient influence

Recently, the Swedish National Board of Health and Welfare (2006) inspected OHS as well as public and private health clinics, with specific focus on absence certification. The study found that in all or almost all cases OPs considered their knowledge about the patient's working conditions to be sufficient to assess the working ability of the latter. Only about half of the responding GPs working at public health clinics considered this to be the case. The majority of the OPs also maintained that they were the decision-makers in the matter of sick-listing a patient, while one third of the GPs declared that their judgements were the determining factor. The position of the OP appears to imply better access to information that is valuable for the sick-listing consultation. However, sociomedical research has yet to establish whether or not OPs certify sickness absence for shorter periods compared to GPs (Arrelöv et al., 2001; Tellnes et al., 1990). Nor is it clarified whether or not OPs deny sick-listing requests from patients to a larger extent than GPs (Arrelöv, 2006).

Besides informational restrictions, ethical concerns contribute to the patient's influence on the outcome of the consultation.[4] For example, findings from a study on the Swiss disability insurance, where benefits are related to specific threshold values of disability, support the hypothesis that physicians take the financial consequences of the claimant into account when assessing working capacity (Cassis et al., 1996). The physician's advocating the patient's interests appears to extend beyond purely medical issues. Moreover, it is reported that physicians have apprehensions about the possible damage to the physician-patient relationship when they advise against sick leave (Alexanderson et al., 2005; Hussey et al., 2004; Mayhew & Nordlund, 1988). Studies find that the patient's attitude towards sick-listing as well as taking the initiative to discuss sick leave are two factors with considerable leverage

[4]Trust is considered an important element of the relationship between physician and patient, see Arrow (1963).

over absence certification, even in cases where physicians would not recommend sick leave (Englund et al., 2000; Englund & Svärdsudd, 2000; Larsen et al., 1994). Surveys show that there are individuals who are intent on leaving the physician's office with a sickness certificate. On the other hand, there are also accounts of patients who are reluctant to take sick leave (Alexanderson et al., 2005; Arrelöv, 2006). Thus, to the extent that the physician accommodates the interests of the patient, the latter influences the outcome of the consultation. Adjusting sick-listing behaviour to avoid conflicts of loyalty bears resemblance to findings on the physician's prescribing behaviour, which tends to meet the (perceived) expectations of the patients (Cockburn & Pit, 1997; Stevenson et al., 1999). However, the findings of physicians' "flexible" certification practices are not unchallenged (see Campbell & Ogden, 2006; Claussen, 1998).

Implications for the model

We focus on short-term absence certification, which allows us to make a rough distinction between the workers and the firm on the basis of their attitudes for or against sick leave. We assume that the workers are interested in getting sick leave, although for different reasons. The firm is opposed to sickness absence, as it is costly, administratively and production-wise. However, long-term absence may be considered and handled differently by the firm, which we will discuss later.

Thus, having two clients with conflicting interests may influence the behaviour of the OP. We argue that the independence of the OP is put to the test in cases where the outcome of the sick-listing consultation is a matter of judgement. However, the position of the OP, with access to first-hand information about his patient's work-place, is conducive to a correct assessment of the patient's ability to work. The consequences of these observations on the OP's certification practice in terms of restrictive or generous sick-listing, depends on how the OP relates the patient's interest to the firm's interest.

We think of the GP as a salaried employee at a public health clinic. The GP's income is not directly connected to the outcome in sick-listing consultations, nor indirectly, since the costs of sick-listing per se are not borne by a public health care organization. Thus, the only party considered by the GP when certifying absence is the patient. However, the GP has a disadvantage compared to the OP, in that he knows less about the patient's place of work and work tasks. This affects his capacity to assess working ability perfectly.

Hence, the outcome of a sick-listing consultation may differ, depending on whether OPs or GPs are available. It also depends on the preferences of

the workers, who differ in their interests in sick leave. There are workers who consider sick leave a viable option if it is medically recommended, and there are workers to whom justified absence certification, i.e. absence certification based on health-related ability impairment, is of no importance for their interest in getting sick-listed.

4.1.2 Outline

We start by discussing the worker's preferences, before discussing the employer's attitude towards sickness absence. Next, we describe the situation of the OP, which is followed by an account of the GP. Thereafter, the outcome of absence certification is analyzed for the OP and the GP, separately as well as jointly. A discussion about the consequences of the OP's and the GP's involvement in absence certification concludes the paper.

4.2 The model

We consider short-term sickness absence in a single-period model of absence certification. We assume that workers face uncertainty with regard to health-related ability reduction. Some workers are interested in sick leave regardless of the extent to which bad health will affect working ability, whereas some workers only want to be sick-listed in case working ability will be reduced. To get certified sickness absence, an OP or a GP is supposed to verify that the worker's health state affects his ability negatively. Based on better knowledge of the work-place, the OP observes his patient's ability perfectly, while the GP does not. In general, physicians are concerned with the patient's welfare, which implies patient-adjustable certification practices. However, assuming that the firm is reluctant to allow sickness absence for its workers, the incentives generated by the firm may induce the OP to take up a restrictive attitude towards absence certification.

4.2.1 The worker's preferences

We got acquainted with the workers in chapter 3, to which we refer the reader for full details of the model with respect to the workers.

We assume that a healthy worker may fall ill, which could lead to a reduction in his ability to work. The influence on his working ability cannot be ascertained by the worker. That is, in case the worker contracts an illness or suffers an injury, the effect on his ability to work cannot be observed by himself. Hence, we are interested in the cases where ailments, e.g. backache

or a racking cough, do not render the individual unable to work, at least not in the short run. However, with respect to future working ability, there may be cause for sick leave. Correspondingly, we assume that the event of an ability impairment has future, but not immediate, consequences for individual productivity. This implies that the worker cannot infer from his productivity whether or not he is in fact eligible for sick leave.

On a more technical level, we assume that the worker's health is either good or bad, $\sigma \in \{0, 1\}$. His health status is related to his ability to work, which is low or high, $\alpha \in \{L, H\}$. A healthy worker is aware of the fact that his work ability is high, but if he falls ill he only knows that it increases the probability of an ability reduction, Π. Consequently, Π is conditioned by the health state of the worker. If the worker is well, $\Pi(\sigma = 0) = 0$ and if he is ill, $0 < \Pi(\sigma = 1) < 1$.

Working is a full-time activity and we normalize the remaining time for leisure to zero. In the working state, the individual consumes the labour income y_1, net of the insurance premium. If the individual is not working, he is sick-listed and enjoys full-time leisure. Then he receives compensation y_2. The relationship between the income levels in the working state and the non-working state is $y_1 \geq y_2 \geq 0$. In addition, we assume that the insurance policy is fixed.

The individual's utility depends on his income, working ability and whether or not he works. However, the utility function takes the form $u(y, \alpha)$, where $y \in \{y_1, y_2\}$, since labour activity, generating disutility, is suppressed in the notation. Utility is strictly increasing and concave in income while an increase in working ability, from a low to a high level, has a positive effect on the willingness to work.

The worker population is heterogeneous with respect to preferences for labour. For the prospect (y_1, y_2), there is a share $0 < \lambda < 1$ of workers who would rather stay at home than work. They are the "less motivated" workers, called type l. The rest of the population considers working to be a more attractive option and is the "motivated" type m. How an individual's disutility of labour relates to the insurance policy, and hence which type a worker belongs to, is not observable. Consequently, productivity is not type-specific.

We assume that type l prefers leisure to work, irrespective of his working ability,

$$u_l(y_2, \alpha) \geq u_l(y_1, \alpha), \alpha \in \{L, H\}. \tag{4.1}$$

In contrast, type m prefers working if his ability is high,

$$u_m(y_2, L) \geq u_m(y_1, L), \quad u_m(y_1, H) \geq u_m(y_2, H). \tag{4.2}$$

On account of the uncertainty about ability, given bad health, the expected utility U of the worker is

$$U_\theta(y, \Pi(1)) = (1 - \Pi(1))u_\theta(y, H) + \Pi(1)u_\theta(y, L), \quad \theta = l, m. \tag{4.3}$$

He benefits from income $y \in \{y_1, y_2\}$, depending on whether or not he works. Assuming that the worker expects to find out if his ability will be reduced or not, we have that

$$U_\theta(y_1, y_2, \Pi(1)) = (1 - \Pi(1))u_\theta(y_1, H) + \Pi(1)u_\theta(y_2, L), \quad \theta = l, m. \tag{4.4}$$

Then, he earns net labour income y_1 if his ability is high and receives sickness benefit y_2 if his ability is low.

It follows from (4.1) that type l does not find the uncertainty about his working capacity problematic. First and foremost, type l wants to avoid working whether ability is high or not,

$$U_l(y_2, \Pi(1)) \geq U_l(y_1, y_2, \Pi(1) \geq U_l(y_1, \Pi(1)). \tag{4.5}$$

Type m, on the other hand, prefers to know whether or not the bout of illness causes an ability reduction, see (4.2). Hence, the expected utility of knowing his ability is at least as great as the expected utility without knowing,

$$U_m(y_1, y_2, \Pi(1)) \geq U_m(y, \Pi(1)), \quad y \in \{y_1, y_2\}. \tag{4.6}$$

If type m fails to get information about his ability, we assume that he prefers to work instead of taking sick leave. The assumption implies that $\Pi(1)$ is sufficiently small to make

$$U_m(y_1, \Pi(1)) \geq U_m(y_2, \Pi(1)). \tag{4.7}$$

We assume that the worker's claim for compensation is conditional on a physician issuing a so-called sickness certificate, stating that working ability is reduced due to sickness. The physician is able to observe the patient's ability to work. As mentioned in chapter 3, the worker can infer if the effect of bad health on ability is evaluated or not, based on the physician's behaviour, e.g. the types and number of questions asked.

4.2.2 Sickness absence and the firm

We assume that the firm is opposed to sick leave, at least in the short run. An important reason for the firm's attitude is that the worker taking sick leave may be an indication of his health state not being compatible with his

work assignments. If that is the case, sick leave will probably not improve the match noticeably, meaning that sickness absence now does not affect the occurrence of sickness absence later.

Furthermore, sickness absence involves administrative costs, such as the cost of finding a replacement for the absent worker and, more importantly, a loss in production. Having to pay sick pay to absent workers adds to the costs, of course. As noted earlier, the types of ailments we deal with here do not render the individual incapable of working. Furthermore, an ill worker is still a productive worker. Therefore, there is cause for the firm to keep the employee at work, whether or not sickness absence is justified. The "solution" of the firm is a restrictive approach, keeping workers with possible ability impairments at work.

From a long-term perspective, the firm's attitude towards sickness absence may have other consequences. Despite incurring costs, the firm can use the sickness insurance as a means to transfer unproductive workers from the pay-roll into another insurance system.[5]

4.2.3 The physician's certification practice

What influences the physician's decision to put his patient on the sick list? First of all, the health state of the latter. Being an expert, the physician observes his patient's health state effortlessly. If the patient is well, i.e. if $\sigma = 0$, the physician refuses to issue a certificate. Absence certification is only considered if the patient suffers from some kind of medical ailment, i.e. if $\sigma = 1$. Thus, with regard to the health state, the physician abides by the insurance rules, which stipulate that compensation is paid in case the worker suffers a loss of income due to a health-related reduction in working ability. The physician is also able to form an opinion about the connection between his patient's health state and working capacity without effort costs. However, he may certify absence without taking his assessment into account.

Altruism and incentives

A fundamental guiding principle of the medical profession is to act in the best interest of the patient. In our case the physician's concern for the patient is manifested in his certification practices. The physician cares for the

[5] A longer sick-pay period may induce more effort to reintegrate long-term absentees. However, just as in the case of the consequences for e.g. recruitment policies, employment conditions and the work environment, evidence is scarce and uncertain (de Jong & Lindeboom, 2004; Johansson, 2004; Høgelund & Veerman, 2000; Geurts et al, 2000).

welfare of his patients in a general sense, implying that he takes the preferences of the patient into account when deciding whether to certify absence based on future ability impairment or not.[6] This kind of non-paternalistic, altruistic behaviour, depends on the physician knowing the type of patient he treats. The model implies that the workers are willing to reveal their preferences to altruistic physicians. The internalization of the patient's preferences motivates routine sick-listing, i.e. sick-listing which is not based on ability impairment, of the type l worker and an evaluation of the type m worker's ability before issuing a sickness certificate. Thus, the physician treats his patients differently depending on type, despite observing individual ability. The preferences of the patient carry greater weight than the professional opinion of the physician.

In models of physician-induced demand, supplying health care that diverges from the optimal level is associated with disutility for the physician. Thus, the ethical considerations work as a counterbalance to his self-interest (McGuire, 2000). For the altruistic physician in our model, however, self-interest is not necessarily opposed to his concern for his patient.

We consider two employment forms for the physicians. They may be employed in the public sector, as a GP, or in the private sector, as an OP. Whereas the GP's attitude towards sick-listing remains altruistic, the position of the OP, as a firm employee, may have consequences for his behaviour.[7] Assuming that the OP adopts a restrictive certification practice, in line with the interests of the firm, there is cause for the worker to conceal his type from the OP. A disclosure implies a risk that the firm will be informed about the worker's preferences, which in turn could put his job at risk.

Assessment precision

The probability that the physician will deem the patient unable to work depends on the patient's health state as well as the precision, p, of the assessment, $A(\sigma, p)$. The precision parameter takes the values zero and one, $p \in \{0, 1\}$. The higher the value, the more accurate the assessment.

[6] Our model reflects, to some extent, the interdependency of patients' and physicians' utility functions and also illustrates that the agency relationship between patient and physician may concern objectives other than health maximization (Evans, 1984; Ryan, 1994, Mooney & Ryan, 1993).

[7] It is not inconceivable that the choice of career depends on some underlying differences between physicians, which e.g. Sørensen & Grytten (2003) find for salaried GP and contract physicians, primarily reimbursed on a fee-for-service basis. However, it is well known that physicians respond to economic incentives, which could also be expected to apply in our case.

We assume that the OP observes working ability perfectly. This is a result of the OP's better knowledge of his patient's place of work. In contrast, the GP 's lack of first-hand information affects the accuracy of his judgement. Thus, we assume that OP's assessments are characterized by high precision, and that GP's assessments are of low precision. The working ability of a healthy patient is accurately assessed by the GP and OP alike, but their assessments diverge for a sick patient. The OP correctly identifies cases where illness leads to future ability impairment, $A(1,1) = \Pi(1)$. The GP, on the other hand, is disposed to certify absence in cases where it is not justified, i.e. $A(1,0) > \Pi(1)$.

The OP

The OP is expected to act in the best interest of the patient, but is simultaneously under the influence of the firm through his employment contract. Thus, the OP has two clients with different interests concerning sickness absence. The "double loyalty" of the OP may vary, which we recognize by assuming that the OP adjusts his certification practices to either the patient's interests, i.e. the OP is altruistic, or the interests of the firm. The latter we call "firm-loyal".[8]

We assume that the OPs are paid salaries. Our choice of reimbursement form depends primarily on the fact that an analysis of the optimal payment contract is not our main priority. Instead we consider more "subtle" means of the firm to induce a certain behaviour. Variation among OPs is generated by different degrees of susceptibility to economic incentives created by the firm. One instrument that may be used to affect the behaviour of the OP is the possibility of allocating fewer resources to OHS. Ultimately, the firm may refrain from renewing the contract with the external OHS or, with regard to in-house OHS, may choose to wind up the organization.[9]

An altruistic OP considers the patient's preferences when deciding whether or not to issue a sickness certificate. In contrast, an OP who is loyal to the firm does not take them into account. He refrains from absence certification in all cases, without considering the type of patient or the outcome of the

[8]Indeed, Weel et al. (1999) report from the public debate in the Netherlands, that OPs are both accused of having too permissive certification practices and of forcing employees back to work too soon.

[9]It is possible that the organizational form matters for the services provided. In a study of Dutch OHS, mainly based on interviews with the CEOs, de Valk et al. (2006) find that in-house OHS mostly offer proactive measures, while external OHS are oriented towards reactive measures, primarily sickness absence consultations. Furthermore, in-house OHS tend to view the employees as their customers. In contrast, a majority of the external OHS consider the employers their main clients.

ability assessment. Again, the behaviour of the OPs is explained by the relationship between altruism and economic incentives, but also by the kinds of ailments the workers suffer from. In cases where the worker's health state definitely motivates sick-listing, though not considered here, neither "type" of OP hesitates to certify absence.

In line with our earlier discussion of the employer's attitude towards sickness absence, we expect a more generous certification practice from the OP, with respect to long-term absentees, in order to transfer the costly worker into another insurance system. Indeed, when examining the insurance files of certified absentees in a Swedish province and concentrating on absence periods extending over more than 60 days, Melén et al. (2005) find that OP-patients are much more likely to be granted disability benefits and are less likely to return to work compared to patients sick-listed by other physician categories, despite a greater probability of receiving rehabilitation.[10]

Whether the OP is altruistic or firm-loyal is not observable, but his behaviour is. However, it may take long before there is any information for the firm on which to base his decision to fire or rehire the physician. It depends on the composition and general health status of the workforce.

OP's utility The altruistic OP's certification practice is completely determined by the patient's preferences, whereas the firm-loyal OP ignores them. Hence, given $\Pi(1)$, the utility-maximizing OP decides whether or not to certify absence, and whether or not consider the patient's ability when certifying, taking into account the salary and, if being altruistic, the patient's preferences.

The GP

We assume that the GP is salaried and has public employment. In view of the fact that sickness absence among workers has no direct financial consequence for the public health care organization, we assume that it does not actively influence the GP's altruistic sick-listing behaviour. However, compared to the OP, the GP has less insight into his patient's place of work, which implies an imprecise ability assessment. Facing uncertainty regarding his sick patient's working ability, the GP certifies absence to be on the safe side. That is, he tends to sick-list individuals retaining high working abilities despite illness.

The GP's assessment precision is associated with a probability a of judging sick, but high-ability patients to be eligible for benefits. Hence, the

[10]There may be other underlying factors behind these findings, such as the kind of the diagnoses and the severity of the ailments that OPs primarily encounter, see Hetzler (2005) and Andrén (2007).

probability of the GP identifying a low working ability is

$$A(1,0) = \Pi(1) + (1 - \Pi(1))a, \tag{4.8}$$

where $0 < a < 1$. From the expression it is clear that $\Pi(1) < A(1,0) < 1$. In other words, the GP's screening is characterized by type 2 error, whereas type 1 error is successfully avoided.

GP's utility Like the altruistic OP, the GP bases his certification practice on the patient's preferences. Thus, provided that $A(1,0)$, the utility-maximizing GP considers the patient's ability when certifying absence depending on preferences of the latter.

Information and (mis)representation

The worker receives correct information about diagnosis from both the GP and OP. Moreover, information about ability is withheld by neither the altruistic OP nor the GP, although the assessments of the latter are imprecise. In contrast, the firm-loyal OP declares all ill workers to have intact abilities, without taking his assessment into account. In the first case, as far as type l workers are considered, you could say that physicians and workers misrepresent information to the firm and the insurer. The latter do not know whether or not claimants are truly eligible for sickness benefits.[11] In the second case, physicians misrepresent information to the workers too.[12]

4.3 OP or GP - implications for absence certification

The aim of our paper is to study the effects of physicians' employment forms on absence certification. The treatments that the type l and m workers receive from the physicians vary in terms of routine or justified absence certification, and in terms of assessment precision. This has implications for the utilization of the sickness insurance.

[11] Cf. Ma & McGuire (1997), where incentives for false treatment reporting are identified for both the patient and physician.

[12] Jelovac (2001) analyses the physician's optimal payment contract under the assumptions that neither the effort to assess the diagnosis accurately nor the treatment strategy is verifiable. However, in contrast to our model, the patient can come back for a second visit if it turns out that the physician's choice of treatment was not suitable for the patient's illness.

First of all, the worker can accept or decline an offer of sick-listing. Furthermore, we assume, as a starting point, that the workers are free to consult whatever physician they like. The workers can separate GPs from OPs, but do not know which are altruistic and which are loyal to the firm among the latter. Similarly, neither GPs nor OPs know whether they meet type l or type m workers, if the workers do not reveal their types. Consequently, the treatment the workers receive at the physician's office is dependent on the expectations entertained by both parties about each other.

Considering that in real life physicians are neither nameless nor faceless, we find it quite reasonable to assume that there is at least some information about the physicians, making it possible for the workers to distinguish between them. That is, the workers perceive the physicians as a group of separate individuals, the characteristics of whom the workers form expectations about. Especially, if a worker expects to encounter a GP or an altruistic OP who adjusts his certification practice to the patient, he can safely state his type and receive the preferred treatment.

4.3.1 Separating and pooling equilibria

If the information structure is such that the workers can form expectations about each physicians's character, they can choose to visit physicians with compatible certification practices. Then, the two worker types receive different treatments in expectation, defining a separating equilibrium.

In a pooling equilibrium, the workers are unable to form physician-specific beliefs. They expect a certain treatment with a probability corresponding with the share of e.g. altruistic or firm-loyal OPs in the physician population. Thus, identical treatments of type m and type l are expected.

4.3.2 Absence certification performed by OPs

We start with the case where the workers can only consult OPs. The physician population consists of a share β of firm-loyal OPs and a share $1 - \beta$ of altruistic OPs. We also assume that there are both type l and type m workers in the worker population, i.e. $0 < \lambda < 1$.

Separating equilibria

We identify some scenarios where workers are able to attach expectations about certification practices to individual physicians. In equilibrium, type l and type m are sure to get routine absence certification and certification conditioned on low working ability, respectively.

If $\beta = 0$, all available OPs are altruistic. If the workers recognize the composition of the physician population, they can go for a consultation at any OP's office and reveal their preferences. Thus, type l is sick-listed without regard to his working ability, while type m is sick-listed if truly eligible for compensation.

Separate treatments of the two worker types may also be expected in the scenario where $0 < \beta < 1$, on the condition that the available information enables the workers to form expectations about individual physicians' certification practices. Both firm-loyal and altruistic OPs are available, but all workers prefer to visit altruistic physicians. Hence, provided consistent expectations, the workers receive different treatments. As a consequence of separate treatments, there are both high- and low-ability workers among the certified absentees. As expected, the observed absenteeism, i.e. the occurrence of high-ability workers getting certified absence, is caused by the type l workers, in combination with altruistic OPs' certification practices.

Continuing with the scenario of $0 < \beta < 1$, we consider the case where the OP's services are rationed. In view of the fact that all workers prefer altruistic OPs and that all workers seek such OPs at the same footing, the rationing implies that some type l and type m workers are forced to go elsewhere for sick-listing consultations. This can be called a semi-separation between types.

Pooling equilibria

There are also situations where workers are not willing to reveal their types. We are referring to scenarios where workers' expectations about treatment are not physician-specific. Consequently, it is not expected that the certification process results in separate treatments of the worker types.[13]

If $\beta = 1$, the OP advises patients of type l as well as of type m against sick leave. Hence, the OP's expectations about his patients do not influence his behaviour. If the workers expect that the OPs refrain from sick-listing, they decide against a consultation and continue working. This is a clear case of under-certification, since there are low-ability individuals in the group of people present at work as a result of the OP's certification practice.

In the scenario where $0 < \beta < 1$, workers expect to encounter a firm-

[13]However, scenarios where separation arise despite non-revelation are conceivable. Separation can be based on workers' and physicians' expectations, as in chapter 3. In such scenarios, some altruistic physicians may form expectations about meeting only type m or type l individuals, while the workers expect some altruistic physicians to certify absence without regard to working ability and some with regard to working ability. Provided mutually consistent expectations, the worker types receive separate treatments.

loyal OP with probability β and an altruistic OP with probability $1-\beta$.[14] For altruistic OPs, the lack of information may imply an identical treatment of all workers. The explanation lies in the workers not knowing to which client the physician adjusts his certification practice. The risks involved in unintentionally revealing his type to a firm-loyal OP cause the worker to be cautious.[15] The altruistic OP then resorts to the composition of the worker population in order to decide on a certification practice. Justified absence certification takes place if the expected utility generated by doing so is at least as large as the expected utility resulting from routine absence certification. This is the case if the proportion of patient type l is

$$\lambda \leq \frac{U_m(y_1, y_2, \Pi(1)) - U_m(y_1, \Pi(1))}{U_m(y_1, y_2, \Pi(1)) - U_m(y_1, \Pi(1)) + U_l(y_2, \Pi(1)) - U_l(y_1, y_2, \Pi(1))}, \quad (4.9)$$

which is less than one. If the condition is fulfilled, type m and type l workers receive and accept absence certification with probability $\Pi(1)$, provided that they encounter altruistic OPs. With probability $1-\Pi(1)$ it is found that the workers are not eligible for sickness benefits and are sent back to work. However, as the consultation at the firm-loyal OP also ends with the worker resuming work, the latter will not know his ability. It also means that if a certificate is not offered, neither type of worker can infer from the OP's behaviour which "type" of OP he happens to meet.

If the condition (4.9) is not fulfilled, the pooling equilibrium collapses. The reason for this is the response of type m to altruistic OPs' certification practices. The offer of a sickness certificate, irrespective of working ability, is indicative of an altruistic OP, which means that type m can reveal his preferences and thereby make the OP adjust his behaviour accordingly. Thus, non-separate treatments involve under-certification, in that some low-ability workers of both types continue working.

4.3.3 Absence certification performed by GPs

We now assume that GPs have the exclusive responsibility for sick-listing. The GP is altruistic, which implies that type l workers receive sickness benefits whether or not they are eligible for it, while absence for motivated workers

[14] For small firms it is not unlikely that a physician's services are bought when needed, which could mean that different physicians show up on different occasions. Thus, physicians are associated with multiple firms. In that case, it may be hard not only for workers but also for physicians to get an overview.

[15] Given a sufficiently large probability of meeting an altruistic physician, it is possible that workers will take the risk of revealing their type to the OPs to increase their chances of getting routine or justified sick-listing. However, we disregard the possibility in this context.

is only certified if it is justified, provided that the GP has information about the worker type at hand. However, the GP's assessment does not reflect the worker's future working ability correctly, as $A(1,0) > \Pi(1)$. That considered, the worker has to decide whether or not to see a GP and, if he does so, how to respond to the offer of a sickness certificate.

The patient's decision

Despite his falling ill, the worker may decide not to visit the GP due to the uncertainty involved. In that case, he continues working without receiving information about his ability. If the worker decides on a consultation, he is declared fit with probability $1 - A(1,0)$. Then he may go back to work knowing that the GP's judgement is reliable. With probability $A(1,0)$ the GP finds that the worker's ability will be reduced due to bad health. However, the GP's recommendation of sick leave does not exclude the possibility that the worker's ability will be high, despite illness.

Despite the GP's imprecise assessment, a sick-listing consultation implies an update about the worker's ability. Using Bayes' rule, the probability of having a future low working ability, conditional on being sick-listed, is

$$\frac{\Pi(1)}{A(1,0)} > \Pi(1).^{16} \tag{4.10}$$

Although not perfect, the information makes the worker more "convinced" that he will experience a low ability due to bad health.

As far as the "less motivated" worker is concerned, he undoubtedly decides to visit the GP and accepts the certificate, since he prefers not to work irrespective of his true ability (see (4.5)). However, the imprecision of the GP's assessment poses a problem to the "motivated" worker.

First, type m may choose not to consult a GP and instead continue working,

$$U(y_1, \Pi(1)) = (1 - \Pi(1))u_m(y_1, H) + \Pi(1)u_m(y_1, L). \tag{4.11}$$

If he decides to visit the GP, there is another choice to make. Provided that the physician is willing to certify absence, type m can choose to accept the certificate and thereby decide not to work,

$$\begin{aligned} U(y_1, y_2, \Pi(1), A(1,0)) = (1 - A(1,0))u_m(y_1, H) + \\ (A(1,0) - \Pi(1))u_m(y_2, H) + \Pi(1)u_m(y_2, L). \end{aligned} \tag{4.12}$$

[16] Let E be the event that the worker is unable to work and F the event that the worker is sick-listed. According to Bayes' rule, $P(E \mid F) = \frac{P(F|E)P(E)}{P(F)} = \frac{P(F|E)P(E)}{P(F|E))P(E)+P(F|E^c)P(E^c)}$.

Alternatively, type m can decline the offer and go back to work,

$$U(y_1, \Pi(1), A(1,0)) = (1 - A(1,0))u_m(y_1, H) + \\ (A(1,0) - \Pi(1))u_m(y_1, H) + \Pi(1)u_m(y_1, L). \quad (4.13)$$

Since (4.11) and (4.13) are identical, the worker type m is indifferent between remaining at work and returning to work after disregarding the GP's recommendations. Thus, type m does not consult the GP only for the sake of information but maybe to get sick-listed. The next step is to determine whether or not the GP's assessment is reliable enough to make the worker decide not to work, conditional on the GP offering to put him on the sick list. Considering that $A(1,0) > \Pi(1)$, it is not certain that the assumption in (4.7), which states that the "motivated" worker who is uncertain about his ability prefers working to not working, is applicable. Hence, when relating (4.12) to (4.13) we are not able to produce any conclusive results. The inequality can go in either direction. Our findings imply that the update about his ability that the "motivated" worker receives from the GP may make him decide on a consultation, but it is not certain that he is willing to change his behaviour, i.e. accept the sickness certificate.

Separating equilibria

All GPs are willing to adjust their sick-listing practices to their patient's preferences. Thus, workers can reveal their preferences without any risk to their jobs. However, the type m worker prefers accurate assessments, while precision is of no importance to the type l worker.

At the GP's type l receives a certificate more or less automatically, while the GP's assessment of working ability forms the basis for absence certification of type m. The ability assessment notwithstanding, its imprecision causes ambiguity concerning type m's decision to take sick leave on the GP's advice. It is not clear whether or not type m workers decide to visit the GP and, if so, whether or not they will accept sickness certificates. The first option leads to under-certification, as working individuals in the type m group could rightfully draw sickness benefits instead. The second option implies over-certification. Then, type m workers contribute to the group of ineligible but certified absentees. However, in contrast to type l, the absenteeism of type m is involuntary. A related observation is made by Eklund & Ossowicki (2005) in a recent study of Swedish sickness absentees. One third of the absentees responded that they could return to work, at least on a part-time basis. In our model, the finding would describe a second period, where it is revealed whether sick-listing was justified or not. Eklund & Ossowicki (2005)

argue that "unnecessary sick leave" could be partly remedied by better ability assessments.[17]

4.3.4 GP and OP available for absence certification

Next, we consider the case where GPs as well as OPs are allowed to certify sickness absence and workers can choose freely between physicians. The differences between the OP and the GP could imply that consultation at the latter physician's is preferred by workers. Andrea et al. (2004) study the behaviour of Dutch employees when seeking help at the physician's while active at work. They observe a slight tendency of separation between the decision to visit the OP or the GP depending on the type of work-related problem.[18]

OPs share the ability to make precise assessments but are distinguished from each other by their willingness to issue sickness certificates. Altruistic OPs and GPs both adjust their sick-listing practices to suit the patients, but differ in their skills at assessing their patient's ability. Thus, the treatment of the workers can be categorized according to the possibility of sick-listing or the assessment precision.

If the OPs are all loyal to the firm, the workers consult the GPs. This leads to separate treatments in equilibrium but it does not necessarily imply that type m workers accept being sick-listed. In what way type m workers influence the utilization of the sickness insurance, in terms of absenteeism or presenteeism, i.e. the occurrence of low-ability individuals staying at work, depends on their response to the certification practices of the GPs. If the workers expect the OPs to be altruistic, it is conceivable that they will receive different treatments from different physicians. "Motivated" workers consult the OPs, since type m prefers perfect information about working ability, while "less motivated" workers are sick-listed by the GPs.

Control of physician access

So far we have assumed that the worker can choose to consult any of the physicians who are available. A continuous relationship between patient and physician is usually considered conducive to correct and justified absence

[17] They also call for more flexibility at the place of work, but as we discussed earlier, it is not certain that the employer considers efforts to reintegrate the absentee financially worthwhile.

[18] Another Dutch study (Anema et al., 2006) finds that, when treating long-term sickness absentees, OPs primarily apply work-related interventions, while GPs recommend medical interventions.

certification. However, there is evidence indicating that a personal relationship with the physician increases the probability of sick-listing (Wahlström & Alexanderson, 2004).

In chapter 3 we discuss the introduction of an allocation mechanism. If the employer has the possibility of randomly assigning workers to a physician in the scenario where both firm-loyal and altruistic OPs are available, it leads to more workers meeting firm-loyal OPs. It also renders it difficult for altruistic OPs to adjust their certification practices to each patient's wishes, as workers are not keen on disclosing information about their type. In the pooling equilibrium discussed in section 4.3.2, the altruistic OPs certify in a manner that is in line with the preferences of type m. However, as was noted in section 4.3.2, the workers may take a chance and reveal their types if the probability of meeting an altruistic OP is high enough. Nevertheless, a random allocation mechanism signifies a reduction in the potential number of absentees compared to workers choosing independently.

The GP's absence certification is associated with either over-utilization or a combination of over- and under-utilization of the insurance, depending on the attitude of type m towards the GP's imprecise ability assessment. Hence, at least in the former case there is cause for the employer to restrict the workers' access to GPs, if possible. Generally, the worker is required to take some time off for an appointment at the GP's office, if not scheduled after working-hours. The disincentives to visit the GP could be amplified by facilitating OP consultations, e.g. by letting workers have free access to the OPs during work time with retained salary. Contradictory to our conclusions, a considerable minority of Swedish employers require their approval before employees are allowed to make use of the occupational health services, according to a survey conducted by the Swedish Agency for Public Management (2001). However, the economy of the restriction can be called into question, considering that it makes a visit to the GP an obvious choice for the workers with regard to both health care and absence certification.

4.4 Concluding discussion

We study the implications of the OP's and the GP's absence certification practices on the utilization of sickness insurance. We focus on cases where bad health involves a risk of a reduction in working ability. Ill workers have to consult physicians to acquire information about their working ability and to receive certified sickness absence. To some workers, type m workers, sick leave is only conceivable if the physician's assessment confirms a low working ability. Some workers, type l workers, are interested in getting sick-listed

regardless of what the physician's assessment says.

The physician's traditional role as the patient's advocate influences his certification practice. It adjusts to the preferences of the patient. However, working as an OP for a firm may also affect the physician's behaviour. In that case, the firm's reluctance towards sickness absence is reflected in his certification practice. Hence, we consider altruistic GPs and OPs, who sick-list with or without consideration of working ability, depending on the type of worker, as well as firm-loyal OPs, who refrains from sick-listing in all cases. Moreover, the position of the OP implies better knowledge of the patient's place of work making his assessment of working ability perfect. The GP, on the other hand, "over-certifies", due to lack of information.

Separate treatments of the worker types describe a situation where type m receives absence certification based on low working ability and type l is sick-listed without consideration to his working ability. Primarily, separate treatments depend on the workers' willingness to reveal their types. If expecting altruistic physicians, the worker can safely state his preferences. If expecting firm-loyal OPs, the worker perceives a risk of indirectly disclosing his type to the firm.

Whereas absence certification by the OP with the restrictive attitude implies presenteeism, absence certification by altruistic OPs and GPs is associated with absenteeism. However, in the case of altruistic OPs, absenteeism is entirely voluntary, in the case of GPs it may be involuntary. Since altruistic OPs identify the relationship between bad health and working ability correctly, only applicants who wish to be sick-listed irrespective of ability get that kind of treatment. As a possible consequence of the GP's imprecise assessment, type m workers, ignorant of their having a high ability, accept sick leave. Evidence of such "unnecessary" sickness absence is found for the Swedish sickness insurance.

We find that by increasing the OP's responsibility for absence certification the occurrence of unnecessary sickness absence may be avoided. Owing to the closer connection to workers' place of work, the OP is in a better position to make qualified assessments than the GP. On the other hand, considering that the OP is contracted by the workers' employer, his position concerning independent decision-making is worse than that of the GP. The OP's mandate, here primarily referring to absence certification, ultimately rests with management policies and financial interests. Hence, in the context of raising the employer's stake in sickness insurance, effects on the sick-listing behaviour of the OP may also be expected. A possible attenuating measure against the problem of dependence is to distribute the financing of OHS over multiple sources, something which has been proposed for the Swedish case (Public Commissions of the State, 2006:86), and already is a part of the

Finnish system (Hämäläinen et al., 2001).

In conclusion, when discussing absence certification and its effect on the utilization of sickness insurance, one should keep in mind that there are advantages as well as disadvantages attached to the OP's as well as to the GP's certification practices.

References

Alexanderson K, Brommels M, Ekenvall E, Karlsryd E, Löfgren A, Sundberg L & Österberg M, 2005, "Problem inom hälso- och sjukvården kring handläggning av patienters sjukskrivning", rapport, Institutionen för klinisk neurovetenskap, Karolinska Institutet, Stockholm.

Andrea H, Metsemakers J F M, Kant IJ, Beurskens A J H M, Swaen G M H & van Schayck C P, 2004, "Seeking Help in Relation to Work - Visiting the Occupational Physician or the General Practitioner", *Occupational Medicine*, 54, 419-421.

Andrén D, 2007, "Long-Term Absenteeism Due to Sickness in Sweden. How Long Does It Take and What Happens After?", *The European Journal of Health Economics*, 8, 41-50.

Anema J R, Jettinghoff K, Houtman L L D, Shoemaker C G, Buijs P C & van den Berg R, 2006, "Medical Care of Employees Long-Term Sick Listed Due to Mental Health Problems: A Cohort Study to Describe and Compare the Care of Occupational Physician and the General Practitioner", *Journal of Occupational Rehabilitation*, 16, 41-52.

Arai M & Skogman Thoursie P, 2005, "Incentives and Selection in Cyclical Absenteeism", *Labour Economics*, 12, 269-280.

Arrelöv B, 2006,"Läkarna i sjukskrivningsprocessen" in Palmer E (ed.), *Sjukförsäkring, kultur och attityder. Fyra aktörers perspektiv*, Analyserar 2006:16, Försäkringskassan, 423-477.

Arrelöv B, Borgquist L, Ljungberg D & Svärdsudd K, 2001, "Do GPs Sick-List Patients to a Lesser Extent than Other Physician Categories? A Population-Based Study", *Family Practice*, 18, 393-398.

Arrow K J, 1963, "Uncertainty and the Welfare Economics of Medical Care", *The American Economic Review*, 53, 941 - 973.

Askildsen J E, Bratberg E & Nilsen Ø, 2002, "Unemployment, Labour Force Composition and Sickness Absence: A Panel Data Study", discussion paper 466, Institute for the Study of Labor (IZA), Bonn.

Bergendorff S & Larheden H, 2003, "Att förhindra och förkorta sjukfrånvaro - erfarenheter från fyra länder", RFV Analyserar 2003:16, Riksförsäkringsverket.

Bäckman O, 1998, *Longitudinal Studies on Sickness Absence in Sweden*, Socialforskningsinstitutet 34, Stockholms universitet.

Campbell A & Ogden J, 2006, "Why Do Doctors Issue Sick Notes? An Experimental Questionnaire Study in Primary Care", *Family Practice*, 23, 123-130.

Cassis I, Dupriez K, Burnand B & Vader J P, 1996, "Quality of Work Incapacity Assessments in the Swiss Disability Insurance System", *International Journal for Quality in Health Care*, 8, 567-575.

Claussen B, 1998, "Physicians as Gatekeepers: Will They Contribute to Restrict Disability Benefits?" *Scandinavian Journal of Primary Health Care*, 16, 199-203.

Cockburn J & Pit S, 1997, "Prescribing Behaviour in Clinical Practice: Patients' Expectations and Doctors' Perceptions of Patients Expectations - a Questionnaire Study", *British Medical Journal*, 315, 520-523.

de Jong P R & Lindeboom M, 2004, "Privatisation of Sickness Insurance, Evidence from the Netherlands", *Swedish Economic Policy Review*, 11, 123-143.

de Valk M M A, Ostrom C & Schrijvers A J P, 2006, "An Assessment of Occupational Health Care in the Netherlands (1996-2005)", *Occupational Medicine*, 56, 475-479.

Draper E, 2002, *The Company Doctor: Risk, Responsibility, and Corporate Professionalism*, Russel Sage Foundation, New York.

Dyrstad J & Ose S, 2002, "Worker Absenteeism: Discipline or Composition Effects?" Working paper 25, Department of Economics, Norwegian University of Science and Technology.

Edgerton D, Kruse A & Wells C, 2004, "An Analysis of Sick Leave in Sweden Using Panel Data 1985-1997", working paper 3, Department of Economics, Lund University.

Eklund M & Ossowicki M, 2005, "Sjukskriven i onödan?" Analyserar 2005:5, Försäkringskassan.

Englund L & Svärdsudd K, 2000, "Sick-Listing Habits among General Practitioners in a Swedish County", *Scandinavian Journal of Primary Health Care*, 18, 81-86.

Englund L, Tibblin G & Svärdsudd K, 2000, "Variations in Sick-Listing Practice among Male and Female Physicians of Different Specialities Based on Case Vignettes", *Scandinavian Journal of Primary Health Care*, 18, 48-52.

Geurts S, Kompier M & Gründemann R, 2000, "Curing the Dutch Disease? Sickness Absence and Work Disability in the Netherlands", *International Social Security Review*, 53, 79-103.

Henrekson M, Llanto K & Persson M, 1992, *Bruk och missbruk av sjukförsäkringen*, SNS Förlag.

Hesselius P, Johansson P & Larsson L, 2005, "Monitoring Sickness Insurance Claimants: Evidence from a Social Experiment", working paper 15, Institute for Labour Market Policy Evaluation (IFAU), Uppsala.

Hetzler A, 2005, "Läkare och deras sjukskrivningsbeteende", in Hetzler A, Melén D & Bjerstedt D, *Sjuk-Sverige, Försäkringskassan, rehabilitering och utslagningen från arbetsmarknaden*, Brutus Östlings Bokförlag Symposion, 185-210.

Hussey S, Hoddinott P, Wilson P, Dowell J & Barbour R, 2004, "Sickness Certification System in the United Kingdom: Qualitative Study of Views of General Practitioners in Scotland", *British Medical Journal*, 328, 88-92.

Hämäläinen R-M, Husman K, Räsänen K, Westerholm P & Rantanen J, 2001, "Survey of the Quality and Effectiveness of Occupational Health Services in the European Union, Norway and Switzerland", People and Work Research Report 45, Finnish Institute of Occupational Health, Helsinki.

Høgelund J & Veerman T J, 2000, "Reintegration: Public or Private Responsibility, the Dutch and the Danish Reintegration Policy towards Work Incapacitated Persons", presented at the conference of International Social Security Associations, held in Helsinki, 25-27 September 2000.

Jelovac I, 2001, "Physicians' Payment Contracts, Treatment Decisions and Diagnosis Accuracy", *Health Economics*, 10, 9-25.

Johansson P, 2004, "Comment on Philip de Jong and Maarten Lindeboom: Privatisation of Sickness Insurance: Evidence from the Netherlands", *Swedish Economic Policy Review*, 11, 145-149.

Johansson P & Palme M, 1996, "Do Economic Incentives Affect Work Absence? Empirical Evidence Using Swedish Micro Data", *Journal of Public Economics*, 59, 195-218.

Johansson P & Palme M, 2002, "Assessing the Effect of Public Policy on Worker Absenteeism". *Journal of Human Resources*, 37, 381-409.

Johansson P & Palme M, 2005, "Moral Hazard and Sickness Insurance", *Journal of Public Economics*, 89, 1879-1890.

Lantto K, 1991, *Optimal Deterrents to Malingering. The Role of Incentives, Attitudes and Information Costs in Social Insurance, especially Sickness Benefit and Welfare*, dissertation, Stockholm University.

Larsen B A, Forde O H & Tellnes G, 1994, "Legens kontrollfunksjon ved sykmelding", *Tidsskrift for den Norske Lægeforening*, 114, 1442-1444.

Lindbeck A & Persson M, 2006, "A Model of Income Insurance and Social Norms", seminar paper 742, Institute for International Economic Studies (IEES), Stockholm.

Lindbeck A, Palme M & Persson M, 2006, "Job Security and Work Absence: Evidence from a Natural Experiment", working paper 660, Research Institute for Industrial Economics (IFN), Stockholm.

Ma C A & McGuire T G, 1997, "Optimal Health Insurance and Provider Payment", *American Economic Review*, 87, 685-704.

Mayhew H E & Nordlund D J, 1988, "Absenteeism Certification: The Physician's Role", *Journal of Family Practice*, 26, 651-655.

Melén D, Bjerstedt D & Hetzler A, 2005, "Rehabiliteringens verklighet" in Hetzler A, Melén D & Bjerstedt D, *Sjuk-Sverige, Försäkringskassan, rehabilitering och utslagningen från arbetsmarknaden*, Brutus Östlings Bokförlag Symposion, 211-236.

National Board of Health and Welfare (Socialstyrelsen), 2006, "Socialstyrelsens tillsyn av sjukskrivningsprocessen 2003-2005 - Sammanfattande slutsatser", rapport 2006-03-15, http://www.socialstyrelsen.se/Publicerat/2006/9103/2006-107-4.htm

Public Commissions of the State (Statens offentliga utredningar) 2006:86, "Mera försäkring och mera arbete", Socialdepartementet.

Rikner K, 2002, *Sickness Insurance: Design and Behaviour*, dissertation, Lund University.

Riphahn R, 2004, "Employment Protection and Effort among German Employees", *Economic Letters*, 85, 353-357.

Plomp H N, 1999, "Evaluation of Doctor-Worker Encounters in Occupational Health: An Explanatory Study", *Occupational Medicine*, 49, 183-188.

Stevenson F A, Greenfield S M, Jones M, Nayak A & Bradley C P, 1999, "GP's Perceptions of Patient Influence on Prescribing", *Family Practice*, 16, 255-261.

Swedish Agency for Public Management (Statskontoret), 2001, "Utnyttja företagshälsovården bättre", rapport 2001:29.

Söderberg E, Dahlqvist H & Shahirnejad B, 2002, "Underlag för beslut - en studie av läkarintyg ur ett socialförsäkringsperspektiv", rapport 2002-10-10, Oktaven kassorna.

Sørensen R J & Grytten J, 2003, "Service Production and Contract Choice in Primary Physician Services", *Health Policy*, 66, 73-93.

Tellnes G, Sandvik L, Moum T, 1990, "Inter-Doctor Variation in Sickness Certification", *Scandinavian Journal of Primary Health Care*, 8, 45-52.

Verbeek J, van Dijk F, Räsänen K, Piirainen H, Kankaanpää E & Hulshof C, 2001, "Consumer Satisfaction with Occupational Health Services: Should It Be Measured?"*Occupational and Environmental Medicine*, 58, 272-278.

Wahlström R & Alexanderson K, 2004, "Chapter 11. Physicians' Sick-Listing Practices", *Scandinavian Journal of Public Health*, 32, 222-255.

Weel A N H, van der Beek A J, Kroon P J, Verbeek J H A M & van Dijk F J H, 1999, "Recent Changes in Occupational Medicine in the Netherlands", *International Archives of Occupational and Environmental Health*, 72, 285-291.

Whitaker S C, 2001, "The Management of Sickness Absence", *Occupational and Environmental Medicine*, 58, 420-424.

Lund Economic Studies

1. Guy Arvidsson	Bidrag till teorin för verkningarna av räntevariationer, 1962
2. Björn Thalberg	A Trade Cycle Analysis. Extensions of the Goodwin Model, 1966
3. Bengt Höglund	Modell och observationer. En studie av empirisk anknytning och aggregation för en linjär produktionsmodell, 1968
4. Alf Carling	Industrins struktur och konkurrensförhållanden, 1968
5. Tony Hagström	Kreditmarknadens struktur och funktionssätt, 1968
6. Göran Skogh	Straffrätt och samhällsekonomi, 1973
7. Ulf Jakobsson och Göran Norman	Inkomstbeskattningen i den ekonomiska politiken. En kvantitativ analys av systemet för personlig inkomstbeskattning 1952-71, 1974
8. Eskil Wadensjö	Immigration och samhällsekonomi. Immigrationens ekonomiska orsaker och effekter, 1973
9. Rögnvaldur Hannesson	Economics of Fisheries. Some Problems of Efficiency, 1974
10. Charles Stuart	Search and the Organization of Marketplaces, 1975
11. S Enone Metuge	An Input-Output Study of the Structure and Resource Use in the Cameroon Economy, 1976
12. Bengt Jönsson	Cost-Benefit Analysis in Public Health and Medical Care, 1976
13. Agneta Kruse och Ann-Charlotte Ståhlberg	Effekter av ATP - en samhällsekonomisk studie, 1977
14. Krister Hjalte	Sjörestaureringens ekonomi, 1977
15. Lars-Gunnar Svensson	Social Justice and Fair Distributions, 1977
16. Curt Wells	Optimal Fiscal and Monetary Policy - Experiments with an Econometric Model of Sweden, 1978
17. Karl Lidgren	Dryckesförpackningar och miljöpolitik - En studie av styrmedel, 1978
18. Mats Lundahl	Peasants and Poverty. A Study of Haiti, London, 1979
19. Inga Persson-Tanimura	Studier kring arbetsmarknad och information, 1980
20. Bengt Turner	Hyressättning på bostadsmarknaden - Från hyresreglering till bruksvärdesprövning, Stockholm 1979
21. Ingemar Hansson	Market Adjustment and Investment Determination. A Theoretical Analysis of the Firm and the Industry, Stockholm 1981
22. Daniel Boda Ndlela	Dualism in the Rhodesian Colonial Economy, 1981

23. Tom Alberts	Agrarian Reform and Rural Poverty: A Case Study of Peru, 1981
24. Björn Lindgren	Costs of Illness in Sweden 1964-75, 1981
25. Göte Hansson	Social Clauses and International Trade. An Economic Analysis of Labour Standards in Trade Policy, 1981
26. Noman Kanafani	Oil and Development. A Case Study of Iraq, 1982
27. Jan Ekberg	Inkomsteffekter av invandring, 1983
28. Stefan Hedlund	Crisis in Soviet Agriculture?, 1983
29. Ann-Marie Pålsson	Hushållen och kreditpolitiken. En studie av kreditrestriktioners effekt på hushållens konsumtion, sparande och konsumtionsmönster, 1983
30. Lennart Petersson	Svensk utrikeshandel, 1871-1980. En studie i den intraindustriella handelns framväxt, 1984
31. Bengt Assarsson	Inflation and Relative Prices in an Open Economy, 1984
32. Claudio Vedovato	Politics, Foreign Trade and Economic Development in the Dominican Republic, 1985
33. Knut Ödegaard	Cash Crop versus Food Crop Production in Tanzania: An Assessment of the Major Post-Colonial Trends, 1985
34. Vassilios Vlachos	Temporära lönesubventioner. En studie av ett arbetsmarknadspolitiskt medel, 1985
35. Stig Tegle	Part-Time Employment. An Economic Analysis of Weekly Working Hours in Sweden 1963-1982, 1985
36. Peter Stenkula	Tre studier över resursanvändningen i högskolan, 1985
37. Carl Hampus Lyttkens	Swedish Work Environment Policy. An Economic Analysis, 1985
38. Per-Olof Bjuggren	A Transaction Cost Approach to Vertical Integration: The Case of Swedish Pulp and Paper Industry, 1985
39. Jan Petersson	Erik Lindahl och Stockholmsskolans dynamiska metod, 1987
40. Yves Bourdet	International Integration, Market Structure and Prices. A Case Study of the West-European Passenger Car Industry, 1987
41. Krister Andersson and Erik Norrman	Capital Taxation and Neutrality. A study of tax wedges with special reference to Sweden, 1987
42. Tohmas Karlsson	A Macroeconomic Disequilibrium Model. An Econometric Study of the Swedish Business Sector 1970-84, 1987
43. Rosemary Vargas-Lundius	Peasants in Distress. Poverty and Unemployment in the Dominican Republic, 1989
44. Lena Ekelund Axelson	Structural Changes in the Swedish Marketing of Vegetables, 1991

45. Elias Kazarian	Finance and Economic Development: Islamic Banking in Egypt, 1991
46. Anders Danielson	Public Sector Expansion and Economic Development. The Sources and Consequences of Development Finance in Jamaica 1962-84, 1991
47. Johan Torstensson	Factor Endowments, Product Differentiation, and International Trade, 1992
48. Tarmo Haavisto	Money and Economic Activity in Finland, 1866-1985, 1992
49. Ulf Grönkvist	Economic Methodology. Patterns of Reasoning and the Structure of Theories, 1992
50. Evelyne Hangali Maje	Monetization, Financial Development and the Demand for Money, 1992
51. Michael Bergman	Essays on Economic Fluctuations, 1992
52. Flora Mndeme Musonda	Development Strategy and Manufactured Exports in Tanzania, 1992
53. Håkan J. Holm	Complexity in Economic Theory. An Automata Theoretical Approach, 1993
54. Klas Fregert	Wage Contracts, Policy Regimes and Business Cycles. A Contractual History of Sweden 1908-90, 1994
55. Per Frennberg	Essays on Stock Price Behaviour in Sweden, 1994
56. Lisbeth Hellvin	Trade and Specialization in Asia, 1994
57. Sören Höjgård	Long-term Unemployment in a Full Employment Economy, 1994
58. Karolina Ekholm	Multinational Production and Trade in Technological Knowledge, 1995
59. Fredrik Andersson	Essays in the Economics of Asymmetric Information, 1995
60. Rikard Althin	Essays on the Measurement of Producer Performance, 1995
61. Lars Nordén	Empirical Studies of the Market Microstructure on the Swedish Stock Exchange, 1996
62. Kristian Bolin	An Economic Analysis of Marriage and Divorce, 1996
63. Fredrik Sjöholm	R&D, International Spillovers and Productivity Growth, 1997
64. Hossein Asgharian	Essays on Capital Structure, 1997
65. Hans Falck	Aid and Economic Performance - The Case of Tanzania, 1997
66. Bengt Liljas	The Demand for Health and the Contingent Valuation Method, 1997
67. Lars Pålsson Syll	Utility Theory and Structural Analysis, 1997
68. Richard Henricsson	Time Varying Parameters in Exchange Rate Models, 1997

69. Peter Hördahl	Financial Volatility and Time-Varying Risk Premia, 1997
70. Lars Nilsson	Essays on North-South Trade, 1997
71. Fredrik Berggren	Essays on the Demand for Alcohol in Sweden - Review and Applied Demand Studies, 1998
72. Henrik Braconier	Essays on R&D, Technology and Growth, 1998
73. Jerker Lundbäck	Essays on Trade, Growth and Exchange Rates, 1998
74. Dan Anderberg	Essays on Pensions and Information, 1998
75. P. Göran T. Hägg	An Institutional Analysis of Insurance Regulation – The Case of Sweden, 1998
76. Hans-Peter Bermin	Essays on Lookback and Barrier Options - A Malliavin Calculus Approach, 1998
77. Kristian Nilsson	Essays on Exchange Rates, Exports and Growth in Developing Countries, 1998
78. Peter Jochumzen	Essays on Econometric Theory, 1998
79. Lars Behrenz	Essays on the Employment Service and Employers' Recruitment Behaviour, 1998
80. Paul Nystedt	Economic Aspects of Ageing, 1998
81. Rasha M. Torstensson	Empirical Studies in Trade, Integration and Growth, 1999
82. Mattias Ganslandt	Games and Markets – Essays on Communication, Coordination and Multi-Market Competition, 1999
83. Carl-Johan Belfrage	Essays on Interest Groups and Trade Policy, 1999
84. Dan-Olof Rooth	Refugee Immigrants in Sweden - Educational Investments and Labour Market Integration, 1999
85. Karin Olofsdotter	Market Structure and Integration: Essays on Trade, Specialisation and Foreign Direct Investment, 1999
86. Katarina Steen Carlsson	Equality of Access in Health Care, 1999
87. Peter Martinsson	Stated preference methods and empirical analyses of equity in health, 2000
88. Klas Bergenheim	Essays on Pharmaceutical R&D, 2000
89. Hanna Norberg	Empirical Essays on Regional Specialization and Trade in Sweden, 2000
90. Åsa Hansson	Limits of Tax Policy, 2000
91. Hans Byström	Essays on Financial Markets, 2000
92. Henrik Amilon	Essays on Financial Models, 2000
93. Mattias Lundbäck	Asymmetric Information and The Production of Health, 2000

94. Jesper Hansson	Macroeconometric Studies of Private Consumption, Government Debt and Real Exchange Rates, 2001
95. Jonas Månsson	Essays on: Application of Cross Sectional Efficiency Analysis, 2001
96. Mattias Persson	Portfolio Selection and the Analysis of Risk and Time Diversification, 2001
97. Pontus Hansson	Economic Growth and Fiscal Policy, 2002
98. Joakim Gullstrand	Splitting and Measuring Intra-Industry Trade, 2002
99. Birger Nilsson	International Asset Pricing, Diversification and Links between National Stock Markets, 2002
100. Andreas Graflund	Financial Applications of Markov Chain Monte Carlo Methods, 2002
101. Therése Hindman Persson	Economic Analyses of Drinking Water and Sanitation in Developing Countries, 2002
102. Göran Hjelm	Macroeconomic Studies on Fiscal Policy and Real Exchange Rates, 2002
103. Klas Rikner	Sickness Insurance: Design and Behavior, 2002
104. Thomas Ericson	Essays on the Acquisition of Skills in Teams, 2002
105. Thomas Elger	Empirical Studies on the Demand for Monetary Services in the UK, 2002
106. Helena Johansson	International Competition, Productivity and Regional Spillovers, 2003
107. Fredrik Gallo	Explorations in the New Economic Geography, 2003
108. Susanna Thede	Essays on Endogenous Trade Policies, 2003
109. Fredrik CA Andersson	Interest Groups and Government Policy, A Political Economy Analysis, 2003
110. Petter Lundborg	Risky Health Behaviour among Adolescents, 2003
111. Martin W Johansson	Essays on Empirical Macroeconomics, 2003
112. Joakim Ekstrand	Currency Markets - Equilibrium and Expectations, 2003
113. Ingemar Bengtsson	Central bank power: a matter of coordination rather than money supply, 2003
114. Lars Pira	Staples, Institutions and Growth: Competitiveness of Guatemalan Exports 1524-1945, 2003
115. Andreas Bergh	Distributive Justice and the Welfare State, 2003
116. Staffan Waldo	Efficiency in Education - A Multilevel Analysis, 2003
117. Mikael Stenkula	Essays on Network Effects and Money, 2004

118. Catharina Hjortsberg	Health care utilisation in a developing country -The case of Zambia, 2004
119. Henrik Degrér	Empirical Essays on Financial Economics, 2004
120. Mårten Wallette	Temporary Jobs in Sweden: Incidence, Exit, and On-the-Job Training, 2004
121. Tommy Andersson	Essays on Nonlinear Pricing and Welfare, 2004
122. Kristian Sundström	Moral Hazard and Insurance: Optimality, Risk and Preferences, 2004
123. Pär Torstensson	Essays on Bargaining and Social Choice, 2004
124. Frederik Lundtofte	Essays on Incomplete Information in Financial Markets, 2005
125. Kristian Jönsson	Essays on Fiscal Policy, Private Consumption and Non-Stationary Panel Data, 2005
126. Henrik Andersson	Willingness to Pay for a Reduction in Road Mortality Risk: Evidence from Sweden, 2005
127. Björn Ekman	Essays on International Health Economics: The Role of Health Insurance in Health Care Financing in Low- and Middle-Income Countries, 2005
128. Ulf G Erlandsson	Markov Regime Switching in Economic Time Series, 2005
129. Joakim Westerlund	Essays on Panel Cointegration, 2005
130. Lena Hiselius	External costs of transports imposed on neighbours and fellow road users, 2005
131. Ludvig Söderling	Essays on African Growth, Productivity, and Trade, 2005
132. Åsa Eriksson	Testing and Applying Cointegration Analysis in Macroeconomics, 2005
133. Fredrik Hansen	Explorations in Behavioral Economics: Realism, Ontology and Experiments, 2006
134. Fadi Zaher	Evaluating Asset-Pricing Models in International Financial Markets, 2006
135. Christoffer Bengtsson	Applications of Bayesian Econometrics to Financial Economics, 2006
136. Alfredo Schclarek Curutchet	Essays on Fiscal Policy, Public Debt and Financial Development, 2006
137. Fredrik Wilhelmsson	Trade, Competition and Productivity, 2006
138. Ola Jönsson	Option Pricing and Bayesian Learning, 2007
139. Ola Larsson	Essays on Risk in International Financial Markets, 2007
140. Anna Meyer	Studies on the Swedish Parental Insurance, 2007
141. Martin Nordin	Studies in Human Capital, Ability and Migration, 2007

142. Bolor Naranhuu	Studies on Poverty in Mongolia, 2007
143. Margareta Ekbladh	Essays on Sickness Insurance, Absence Certification and Social Norms, 2007