COMPENSATION IF ILL: IN CASH AND IN KIND

ESSAYS ON HEALTH INSURANCE

ANNE WENCHE EMBLEM

SEPTEMBER, 2003.

A dissertation submitted to the Department of Economics at the Norwegian School of Economics and Business Administration in partial fulfilment of the requirements for the degree of doctor oeconomiae.

ISBN 82-405-0103-6

Preface

A thesis can be read in different ways. My three-year old son Eirik 'read' a friend's thesis with great interest and a big smile on his face, reading aloud about forests, trolls and gigantic motorcars. If you are able to read my thesis with only a fraction of his enthusiasm and positive attitude, I would be very pleased. Indeed, if you do not feel the need to use your imagination to the same extent as Eirik did, I would be thrilled!

I would like to thank the members of my advisory committee, Professor Agnar Sandmo, Professor Vidar Christiansen and Professor Sören Blomquist. I could not have wished for a more qualified committee. My supervisor Agnar Sandmo has provided highly valuable comments and has been encouraging and supportive throughout the writing of this thesis. At times I entered his office feeling quite miserable about my work, but I always left feeling inspired and somewhat emboldened. I am also grateful to Vidar Christiansen for his comments and suggestions.

I would also like to express my gratitude towards Professor Geir B. Asheim and Professor Tore Nilssen. I think that without Geir Asheim this thesis would not have been completed. He has during his visits at Agder University College been a stimulating and encouraging discussion partner. Working with him and Tore Nilssen has been instructive, demanding, and enjoyable. Two of the essays included in this thesis are written together with them.

Financial support from the SiS-programme (Satsing i Sør), sponsored by Kristiansand county and Agder University College, is gratefully acknowledged. I am also indebted to my colleagues at the School of Management at Agder University College for their moral support and encouragement. A special thank to Olav Nygaard for technical assistance.

Last, but not the least, I thank my family for their support throughout the writing of this thesis.

Kristiansand, September 2003.

Anne Wenche Emblem

Contents

Introduction **vii**

Motivation . . . vii

Purpose of Study . . . viii

Outline of Essays . . . xi

References . . . xv

1 Theoretical Foundations: An Overview **1**

1.1 Introduction . . . 1

1.2 Demand for Health . . . 1

1.3 Choice under Uncertainty . . . 4

1.4 Demand for Insurance . . . 7

1.4.1 Markets for Insurance . . . 8

1.5 Information . . . 8

1.5.1 Asymmetric Information . . . 10

1.6 The Role of Government . . . 15

1.6.1 Efficiency Arguments . . . 16

1.6.2 Redistribution . . . 18

1.7 Synopsis . . . 23

References . . . 25

2 Health Insurance: Treatment vs. Compensation **31**

2.1 Introduction . . . 31

2.2 The Model . . . 35

2.3 Preliminary Analysis . . . 38

2.4 Main Results . . . 41

2.5 A Special Case . . . 44

2.6 *Ex-Post* Moral Hazard . . . 45

2.7 Discussion . . . 47

References 51

3 **Deductibles in Health Insurance: Pay or Pain?** **53**
3.1 Introduction 53
3.2 The Model 57
3.3 Separating Equilibrium 59
3.4 Pay or Pain? 63
3.5 Discussion 67
References 69
Appendix 71

4 **Redistribution at the Hospital** **73**
4.1 Introduction 73
4.2 Preliminary Analysis: Allocation of Income Across States of the World. 76
4.3 Main Analysis: Health vs. Subsidy 80
4.4 Discussion 87
References 89

5 **Able or Disabled: Optimal Income Taxation and Public Provision of Health Care** **91**
5.1 Introduction 91
5.2 The Model 95
5.3 Optimal Non-Linear Income Taxation and Public Supply of Medical Treatment 104
5.3.1 Self-Selection is not a Problem 107
5.3.2 Self-Selection is a Problem 113
5.4 Medical Treatment as a Screening Device 120
5.4.1 The Utility Possibilities Frontier 123
5.5 Discussion 124
References 127

List of Figures

1.1 First- and second-best utility possibilities frontier. 21

2.1 The single-crossing property. 40
2.2 The optimal level of treatment. 42

4.1 The single-crossing property. 79
4.2 Efficient bundles of treatment and payment when there is no redistribution. 82
4.3 Efficient bundles of treatment and payment when self-selection is a problem. 86

5.1 Earnings and consumption. 100
5.2 Medical treatment and earnings. 101
5.3 Efficient bundles of treatment and earnings. 111
5.4 Efficient bundles of earnings and consumption. 112
5.5 Information-constrained efficient bundles of earnings and consumption. 114
5.6 Information-constrained efficient bundles of treatment and earnings. 116
5.7 Consumption across states. 117
5.8 The utility possibilities frontier. 124

Introduction

This thesis is comprised of five self-contained essays. In the following, I motivate the choice of research questions and underline the main contributions of my work. Last, I describe the main features of the five essays.

Motivation

The health sector constitutes an important and considerable part of the economic activity in a number of countries. Expenditures on health amounted to an average of 8,4 percent of gross domestic product (GDP) in the OECD countries in 2001, with the United States (US) at the upper end with expenditures amounting to 13.9 percent of GDP.[1] In Norway, health expenditures constitutes about 8,3 percent of GDP. The public sector plays, moreover, an important part in the financing of health care services. This may take the form of specific arrangements for specific groups,*e.g.*, the poor and elderly people, or more general schemes that provide for the whole population. The public share of total health expenditures in the OECD countries is indeed substantial: on average 72 percent of total health spending was publicly financed (ranging from 44 percent in the US to more than 80 percent in the Nordic countries).[1] Moreover, about 90 percent or more of the population in most OECD countries is covered through public programmes (though only about 25 percent in the US).[1]

In terms of economic significance, there is thus little doubt that issues related to the financing of and the demand for health care are highly relevant. These are important issues also at the individual level. In particular, individuals' may suffer a loss in their level of health, and, consequently, encounter substantial expenditures on health care. Risk averse individu-

[1]OECD Health Data 2003. Available at: http://www.oecd.org

als wish to dispose of this risk, hence, the availability and the design of insurance schemes are important to their well-being.

Purpose of Study

My ambition is to contribute to the theoretical literature on health economics, in particular, on issues related to the financing of health care services. Indeed, the topics of the essays included in this thesis all revolve around health insurance. I study both actuarially fair insurance and redistributive tax-financed insurance. The importance of insurance in the case of health care is supported by the empirical fact that there is little 'out of pocket' payment for health care.

Initially, my research questions came about after reading contributions to the literature on public provision of private goods as a means for redistribution in a world of asymmetric information. There, in-kind provision (*i.e.*, financing and supply) of private goods is typically considered to better facilitate the needs of low-income individuals (or, the target group in question), than those of high-income individuals (those outside the target group). In particular, the good in question is assumed to be a substitute for leisure. Health care is often mentioned as an example of such a private good. However, is it really so that public provision of health care better facilitates the needs of low-income individuals, or could it be the other way around? Maybe high-income individuals derive more utility from public provision than do low-income individuals?

This basic question gave rise to the following queries: how do individuals wish to be compensated if ill, and how is their choice of compensation influenced by their level of earnings ability? Moreover, do individuals always choose to restore health when possible? Empirically, we observe that an increasing number of individuals in the working age population are living from disability payments. Typically, these are characterized as low-income earners prior to entering the disability scheme. I do not pursue to give the answer to why this is so. Rather, I wish to study how individuals' level of inherent ability may influence their decision on whether or not to restore health if sick: will they *trade off health for consumption?* In order to shed light on these challenging questions, I expand the concept of health insurance so as to include not only compensation of medical expenditures, but also compensation of income loss due to (permanent) reduced health, and, thus, earnings ability. I argue that these are two types of financial loss arising from the same fundamental risk, namely the risk of falling ill. Thus, I integrate what is traditionally considered two types of insurance, namely medical insurance and disability insurance. To my knowledge, this constitutes a new approach to the problems of health insurance. As will be readily apparent, this approach allows for a number of interesting studies. My co-authors and I construct an innovating and rich model that provides the foundation of the essays included in this thesis. In this model, an individual's health if ill is endogenous. Moreover, non-monetary losses in utility if health is not fully restored, are acknowledged. In what may be considered the traditional approach in health economics, it is assumed that health can be fully restored by means of medical treatment. In the literature on disability insurance, health is assumed not to be fully restorable, or, alternatively, not to be replaceable. Our approach thus provides a link between the two.

Generally, in the case of medical insurance, compensation is usually dependent on (verifiable) consumption of certain goods and services, *e.g.*, visits to a physician, hospital stays, *etc.* Thus, unconditional cash payments are rare. This is mainly so because of moral hazard problems. Moreover, in the literature on insurance, the use of deductibles (co-payments) to mitigate problems of adverse selection is well established. Typically, deductibles are defined in monetary terms. Taking into account that compensation if ill is dependent upon the consumption of health care, I find it natural to allow for deductibles defined also in physical terms. Hence, individuals may be constrained also in the quantity (and quality) of treatment made available if ill.

Based on the significant role of the public sector in the financing and supply of health care, it seems natural to raise the question: What will be the implications of introducing this expanded concept of health insurance to the analysis of public provision of health care? As is often the case in the literature on public provision of private goods, I do not aim to *justify* public provision (although I elaborate somewhat on this in Essay 1). Rather, I take public provision as given and study how the government can use this as an instrument for redistribution of income when there is asymmetric information. As is often the case in studies of information-constrained redistribution, I assume that information about individual earnings ability is private and unobservable by the government. Returning to my initial query regarding who is benefitting the most from a public supply, I show that more redistribution can be carried out if the level of treatment available to low-income individuals is constrained, *i.e.*, health care is 'under-provided'. Moreover, I show that public provision of health care improves redistribution also when preferences are separable in consumption and leisure. This is so because health care improves individual's productivity in the labour market.

Basic problem As will be readily apparent from Essay 1, this thesis encompasses themes from different fields of economics. However, the fundamental problem in Essays 3 - 5 is that of asymmetric information, in particular: adverse selection. Typically, I study situations in which information about individuals' characteristics (*e.g.*, their inherent ability and/or risk of illness) is private and, thus, not observable by others. These are characteristics that influence the outcome of the contractual relationship between informed and uninformed agents (*e.g.*, insurers' profits or a government's redistributional achievements). The design of efficient contracts are thus at issue. Self-selection constraints place restrictions on the range of instruments available, as well as on the design of contracts. (The uninformed agent may in the case of designing redistributional schemes be the government or, in the case of designing competitive insurance schemes; private insurers.) I derive information constrained Pareto-efficient contracts facilitating efficient allocation of risks (in Essay 3) and efficient redistribution of income (in Essay 4 and 5). The uninformed agents offer a menu of contracts that provides the informed agents with incentives to truthfully reveal their characteristics, and, at the same time, promoting the uninformed agents' objectives. As is standard in the literature on insurance, I do not consider the problem of adverse selection and moral hazard simultaneously.

I consider two kinds of uncertainty in these essays. First, individuals face exogenous uncertainty with respect to their health, that is, whether they will be in good or in poor health (and thus in need of health care). This is the issue of Essay 2. Second, in a world of asymmetric information, uninformed agents (*e.g.*, insurers and the government) face uncertainty in that they are not able to observe important characteristics of individuals. The latter uncertainty may thus be considered 'endogenous' in the sense that the outcome of the schemes designed by the uninformed agents are influenced by the behaviour of the informed agents: Individuals may indeed choose not to reveal their private characteristics if it is not in their interest to do so, *i.e.*, if it is not incentive-compatible. This kind of uncertainty is at issue in Essays 3-5.

Outline of Essays

The five essays included in this thesis are briefly outlined in the following.

Essay 1. In this essay, I provide an overview of the theoretical literature from which I have profited in writing this thesis. This overview provides a setting against which my contributions can be viewed.

Essay 2. This essay provides a theoretical analysis of individuals' demand for health insurance, taking explicitly into account that the insuree wishes to hedge against both medical expenditures and loss in income due to reduced health. Hence, disability and medical insurance are integrated. Assuming symmetric information, *i.e.*, a 'first-best' situation, my co-authors and I characterize the optimal choice of insurance compensation and level of compensation when insurance is offered at an actuarially fair premium. We show that individuals with a sufficiently low level of inherent ability choose to restore health only partly, and to hold a contract entitling them to a cash transfer (*i.e.*, disability payment) if ill. Thus, they trade off health for consumption. Individuals with a sufficiently high level of ability, on the other hand, choose to restore their health fully. The findings in this essay provides a benchmark against which the findings of the other essay may be compared.

Essay 3. In this essay, individuals are assumed to differ along two dimensions: inherent ability and risk of illness. My co-authors and I study a competitive insurance market where information about the probability of illness is asymmetrically distributed. As is standard when there is problems of adverse selection, we show that those facing a low probability of falling ill are constrained in their level of insurance coverage, (*i.e.*, they face a strictly positive deductible. Since we allow individuals to differ also with respect to income (ability), the intriguing question is then whether the deductible will be in the form of reduced consumption or reduced medical treatment, or both. We show that individuals with a high ability and a low risk of falling ill will have a deductible in the form of reduced consumption only (*i.e.*, in pay, whereas individuals with a low ability and a low risk of falling ill will have a deductible partly in the form of reduced consumption, *i.e.*, pay, and partly in the form of reduced treatment, *i.e.*, pain.

Essay 4. This essay studies the role of redistributive in-kind transfers. The government is assumed to be unable to observe any individual characteristics, hence, individualized lump-sum transfers are not feasible. I derive a scheme in which the government offers a menu of different combinations of health care and contributions. The benefits from treatment are increasing in ability, thus, the government can offer different combinations of medical treatment combined with different payments in order to separate high-ability individuals from low-ability individuals. Redistribution is hence carried out by means of different levels of medical treatment associated with different lump-sum redistributive payments. I show that if self-selection is a problem, then low-ability individuals are offered partial treatment against a low (or negative) payment, whereas high-ability individuals are offered complete treatment against a high payment. My analysis takes place *ex ante* hence the derived scheme provides individuals with a (partial or complete) hedge against the potential consequences of illness.

Essay 5. In this essay, I characterize a public tax/provision scheme in which nonlinear income taxation does not only provide an instrument for redistribution, but also provides individuals with insurance against potential loss in health, and, thus, earnings. Individuals' level of inherent ability (*i.e.*, productivity) and labour supply are assumed non-observable by the government. I derive contracts in five dimensions: pre- and post-tax income if in good health, and pre- and post-tax income, as well as medical treatment, if in poor health. The information constrained Pareto-efficient scheme is shown to imply a downward distortion in low-ability individuals' labour supply in both health states, and a downward distortion in their level of treatment if sick.

Informational assumptions. As follows from the above, all but one essay deals with problems of asymmetric information and contract design. The informational assumptions made in the different essays vary. When analyzing the outcome of a competitive insurance market in Essay 2 and 3, I assume that information on risk of illness is asymmetric. This is a standard assumption in the literature on insurance. When analyzing public provision of health care in Essay 4 and 5, I assume that information on inherent ability is asymmetric. This is standard in the literature on income taxation. Moreover, in order to simplify the analysis, I assume in these essays that information on risk is symmetric. Naturally, a more realistic situation would be one in which information on both ability and risk of illness are asymmetrically distributed, and possibly also correlated. Extension of the analysis to include also these questions, will be subject to future research.

Note that the contents of Essay 3 and 4 are identical to their published versions.

References

Asheim, G.B., A.W. Emblem and T. Nilssen. (2003). "Health Insurance: Treatment vs. Compensation."

Asheim, G.B., A.W. Emblem and T. Nilssen. (2003). "Deductibles in Health Insurance: Pay or Pain?" *Forthcoming in International Journal of Health Care Finance and Economics.*

Emblem, A.W. (2002). "Redistribution at the Hospital." *International Tax and Public Finance* 9(4), 367-378.

Emblem, A.W. (2003). "Able or Disabled: Optimal Income Taxation and Public Provision of Health Care." Working Paper 2003:3. School of Management, Agder University College.

Essay 1

Theoretical Foundations: An Overview

1.1 Introduction

We provide an overview of the existing literature in the area of this thesis. Focus is placed on issues central to the essays comprised in the thesis, hence we by no means do justice to the substantial literature available within the larger area of the economics of health and health insurance. The paper is organized as follows. In Section 1.2 we discuss issues related to individuals' demand for health and health care. In Section 1.3 we elaborate somewhat on problems of uncertainty and the demand for insurance. Problems of asymmetric information is the subject of Section 1.4, and the role of government is discussed in Section 1.5. Last, we synthesize the main issues in Section 1.6.

1.2 Demand for Health

> Health is not everything in life, but without health, life is nothing.[1]

To be in good health is a major aspiration for most individuals.[2] While some may argue that being in good health is *the* most important objective

[1]Proverb, as cited in Zweifel and Breyer (1992).

[2]Health is defined in the Encyclopedia Britannica as: "the extent of an individual's continuing physical, emotional, mental, and social ability to cope with his environment."

(implying that individuals' ordering of preferences is lexicographic), it is indeed an empirical fact that individuals do trade-off the health objective for other objectives, *e.g.*, when speeding, smoking, *etc.* Thus, being in good health is but one objective in life. Health is, moreover, important not only in its own right: it enables individuals to participate in, and take pleasure from, a number of activities such as consumption, production, leisure activities, *etc.*[3] Consequently, health enters individuals' utility functions not only directly, but also indirectly through its effect on utility from other goods and services. Health is indeed an important determinant of individuals' productivity, both in market and household production. In particular, it constitutes an important part of their innate *human capital*[4], often referred to as 'ability' (Fallon and Verry, 1988).

Individuals are endowed with an *initial level* of health. This inherent level of health constitutes an important determinant of their future health.[5] Changes in health is to some (others would argue large) extent *endogenous* to individuals; their choice of consumption and leisure activities are important determinants of health. Indeed, health may be considered a capital stock in which individuals can invest. Thus, health may constitute both a consumption and investment good.[6] Socioeconomic factors such as sanitary living and working condition are, of course, also important determinants of health. In addition, the distribution of income in the society has an impact on individual health. The influence of absolute and relative income inequalities on health and health inequalities, provides an important and interesting field of study. For instance, Dardanoni and Wagstaff (1987) show that inequality in health may stem more from inequalities in wealth than from inequalities in access to medical care.

---

[3]For instance, as mentioned by Zweifel and Breyer (1997), individuals suffering from a mental illness like depression, takes little pleasure from most types of activities.

[4]Traditionally, innate human capital such as inherent ability and health, is distinguished from acquired human capital such as education and skills.

[5]According to Currie and Madrian (1999): "there is growing evidence that poor health in childhood can have profound effects on future outcomes, both because of effects on adult health, and because of effects on the accumulation of other forms of human capital such as education." (p. 3351).

[6]In particular, one can describe a health production function as well as a health demand function. For more on this in an inter-temporal framework, see Grossman (1972, 2000). For a text-book exposition, see Zweifel and Breyer (1997).

Interpersonal exchange of health is difficult, thus health *per se* does not have a value in exchange. *Health care*, *i.e.*, goods and services that are inputs in the production of health, are however, exchangeable. Health care is supplied by health personnel and may take the form of information (diagnosing), treatment, prescription of drugs or remedies, care, *etc.* Health care is mainly a *private good* and is normally requested when individuals are suffering from an illness. It is, moreover, consumed under the assumption that it has a positive effect on the more fundamental good: health. The demand for health care may thus be considered to be *derived* from the demand for health.

In the literature on health economics, a distinction is drawn between individuals' *need* for care and their *demand* for care. Whereas 'need' depends on individuals' ability to benefit from care, 'demand' depends on their preferences and ability to pay for care (Hurley, 2000). We focus on the latter. Individuals' *demand for health care* may be analyzed according to standard microeconomic theory; when allocating income, individuals will have to choose between spending income on health improving products and other consumption goods. Hence, we fall into the group of economists who believe that health care is not so different from other commodities.[7]

Cullis and West (1979) classify *benefits from health care* into four groups: (i) temporary relief from pain and suffering during treatment and care (or, alternatively, disutility from treatment), (ii) future relief from pain and suffering, (iii) improved productivity, and more labour hours due to improved state of health, and (iv) improved productivity in, and time available for, non-market production. Whereas (i) and (ii) typically are non-monetary benefits from treatment, (iii) is a monetary benefit, and (iv) is a non-market pecuniary gain. Benefits from health care are to some extent uncertain:

> "Even physicians and epidemiologists have a hard time specifying the production function for health. The commodity is produced probabilistically, substantially through the lifestyle choices of individuals." Fuchs and Zeckhauser, 1987 (p. 265).

Health care may in some cases be considered a 'bad' in that it causes pain and suffering to those undergoing it (*e.g.*, treatment like chemotherapy,

[7]According to Hurley (2000), we thus form part of the 'narrows' (as opposed to 'broads' who consider health as distinctly different from other commodities), a typology used on health economists .

surgery, *etc.*). In the theoretical literature it is, however, usually assumed that the ultimate effect on health is positive, and that health care thus is a 'good'.[8] Non-monetary losses such as pain and suffering associated with medical treatment, may be taken into account by indexing the utility function before and after health care is undertaken. In this case, even if health should be fully restored after treatment, individuals place a different value on the restored health than the (identical) pre-illness health. In other words, health may be considered an irreplaceable good, see Cook and Graham (1977) and Schlesinger (1984) for more on this.

Health may be subject to considerable *negative shocks*; individuals may fall ill from communicable diseases, such as the recent Severe Acute Respiratory Syndrome (SARS), suffer severe injuries due to accidents, or fall ill from illnesses that strongly affect their health (*e.g.*, AIDS, Parkinson's disease, *etc.*) Indeed, changes in health and, consequently, expenditures on health care and loss in earnings, are unpredictable and stochastic. The economics of uncertainty thus constitutes an important ingredient in the study of health care demand.

1.3 Choice under Uncertainty

> "...nothing is more obvious than the universality of risks in the economic system." Arrow, 1971 (p. 46).

Uncertainty is an inevitable fact of life: individuals face uncertainty with respect to what are the choices available, and what are the consequences of their decisions. As regards health, individuals face uncertainty with respect to their need for health care, expenditures on treatment, effectiveness of treatment, loss in productivity (temporarily and permanently), as well as loss in earnings. These are all important determinants of individual well-being. Consequently, individuals have to make choices under uncertainty.

In the literature on choice under uncertainty, a distinction is made between (i) the 'expected-utility' approach and (ii) the 'state-preference' approach. In the following, we elaborate somewhat on these alternative ap-

[8] The effect of health care may, obviously, be open to discussion. Whereas for certain illnesses efficient treatment is available, there are others for which no appropriate treatment exists. Indeed, one may even argue that treatment in some cases may reduce health: "Some remedies are worse than the disease," Publilius Syrus.

proaches to individuals' choice under uncertainty.

(i) The expected utility approach

Individuals' preferences over uncertain outcomes may be described by a von Neumann-Morgenstern (vNM) expected utility function:

$$E(U) = \sum_{i=1}^{n} \pi_i u(x_i)$$

where $0 < \pi_i \leq 1$ gives the probability of state i and $u(x_i)$ gives the vNM utility function of outcome x in state $i = 1, ...n$. The (additive) probabilities are well-defined and sum to unity: $\sum_i \pi_i = 1$. Outcomes are measured in cardinal terms. The vNM utility function u is derived through what Hirshleifer and Riley (1992) call 'the assignment of cardinal utilities'. The utility function $u(x_i)$ is, thus, a cardinal function and can be subject to a positive affine transformation only, *i.e.*, $v(u) = au + b$, where $a > 0$.

According to the expected utility model, individuals evaluate an action based on the expected *utility* derived from it, *i.e.*, $E(U) = \sum \pi_i u(x_i)$, and *not* on the basis of the expected *outcome* of the action, *i.e.*, $E(x) = \sum \pi_i x_i$ (Machina, 1987). Individuals hence maximize expected utility, not expected outcome. This finding is accredited Daniel Bernoulli, the utility function $u(x_i)$ is, however, traditionally referred to as a von Neumann-Morgenstern (vNM) expected utility function.

The vNM linear expected utility model is dominating in the literature on individuals' decision under uncertainty. Indeed, it is fundamental in the standard theory of demand for health insurance. Underlying the vNM utility function is axioms of: (i) independence between outcomes, (ii) weak order, and (iii) continuity (Dionne and Harrington, 1990). Individuals are assumed to have preferences over uncertain contingencies and to know the relevant choices, as well as the consequences of these choices. Moreover, they can assign numerical probabilities reflecting their beliefs as to the likelihood that the different outcomes will take place, and, finally, they are able to assess the utility from the different consequences (McGuire *et al.*, 1988).[9] Probabilities may be objective or subjective. [10]

[9]The behavioural assumptions underlying the expected utility model, are being criticized for being to restrictive, and are being challenged, see, *e.g.*, Machina (1987).

[10]Whereas objective probabilities are observable and verifiable, subjective probabilities are based on individuals' available information, prior beliefs and experiences. Thus, objec-

Expected utility functions may be categorized into three types according to the sign of its second-order derivative: negative, zero or positive. A standard assumption in the expected utility model is that individuals' marginal utility is decreasing, *i.e.*, their utility function is strictly concave: $u''(x) < 0$, where $u''(x)$ denotes the second-order derivative of the utility function. A concave utility function implies that individuals are *risk averse*. Hence, individuals prefer a certain outcome to any risky outcome whose (mathematical) expectation equals that of the certain outcome, *i.e.*, $u(\sum \pi_i x_i) > \sum \pi_i u(x_i)$. If, however, they prefer the risky outcome to the certain one, then they are said to be *risk-seeking*, and if they are indifferent between the two, they are *risk-neutral* (Hirshleifer and Riley, 1992).

Indeed, the curvature of the expected utility function provides a measure of individuals' risk attitudes. Individuals' (absolute) risk aversion is given by the Arrow-Pratt measure: $r(x) = -u''(x)/u'(x)$. Hence, the more concave the utility function, the more risk averse is the individual.

(ii) The state-preference approach

This is a more general approach to individuals' choice under uncertainty and is accredited Arrow (1953) and Debreu (1959). Uncertainty is represented by a set of exhaustive, mutually exclusive, and exogenous states of the world. Commodities (or claims) are defined in terms of their physical characteristics, as well as the location, date and state of nature in which they are made available. Receipts and deliveries are at each date dependent on the state of nature and it is assumed that a market exist for delivery of each commodity, at each date, conditional on each state of nature. In particular, all contingent consumption claims are separately tradable at a market price.

Individuals' utility functions reflect their preferences for the contingent consumption claims, their (subjective or objective) appraisal of the likelihood that the different states will occur, and their preferences for risk (Debreu, 1959). Probability beliefs thus form an integral part of individuals' preferences, implying that no specification of probabilities is required when studying individuals' choice under uncertainty.

---

tive probabilities are uniform across individuals (*e.g.*, different persons assess the likelihood that tossing a fair coin will come up head as 50 percent), while subjective probabilities are unique to each individual (*e.g.*, different persons' apprehension about their future health varies substantially).

1.4 Demand for Insurance

Risk averse individuals wish to hedge against the uncertainty facing them. In a market for insurance, risk are transferred from risk-averse individuals to those willing to take on risk for a given payment. There exists a number of excellent papers that provide a thorough review of the development of insurance economics, see, *e.g.*, Dionne and Harrington (1990) and Loubergé (2000). Here, we briefly describe risk-averse individuals' demand for insurance in a basic two-state model.

An individual has a level of income Y if in good health. If ill, she suffers a loss L in income, hence, income if ill is given by $Y - L$. The probability of falling ill is given by: $0 \leq \pi \leq 1$. The individual can hedge against the loss in income by buying insurance. Insurers offer coverage βL, where $0 \leq \beta \leq 1$, at a premium $p = \gamma\pi\beta L$, where $\gamma \geq 1$ denotes loading factor (*e. g.*, to cover administrative costs). The individual's expected income if not insured is given by: $\overline{Y} = Y - \pi L$, while her income if insured is given by: $Y^i = Y - p$. (Thus, for $\beta = 1$ and $\gamma = 0$, then $\overline{Y} = Y^i$). Her preferences are assumed to be represented by a vNM expected utility function u, satisfying $u' > 0$, $u'' < 0$. Thus, her expected utility when not insured is given by: $u^{ni} = (1 - \pi)u(Y) + \pi u(Y - L)$, whereas her utility if insured is given by: $u^i = (1 - \pi)u(Y - p) + \pi u(Y - L(1 - \beta(1 - \gamma\pi))$. Now, if $\gamma = 1$, then premium is *actuarially fair*, and the individual would choose *complete insurance coverage*, *i.e.*, $\beta = 1$. It then follows that $u^{ni} < u^i$, since $u'' < 0$. The difference $u^i - u^{ni}$ gives the risk premium. Hence, her willingness to pay for insurance exceeds the actuarially fair premium.[11] Moreover, defining $\widehat{\gamma} \equiv u^{ni} - u^i$, then for $1 < \gamma < \widehat{\gamma}$, she would choose *partial insurance* coverage, *i.e.*, $0 < \beta < 1$. Indeed, β measures rate of coinsurance.

From this it follows that if insurance markets are complete and perfectly competitive, then risk-averse individuals will choose to insure fully at an actuarially fair premium. If, however, individuals have state dependent preferences, then their insurance coverage may fall short of the loss in income, that is: individuals choose to insure partly.

[11]The risk premium is depending on individuals' risk aversion (as measured by: $r = -u''/u'$) and the (statistical) variance of the outcomes.

1.4.1 Markets for Insurance

The general equilibrium model of competitive markets formalized by Arrow and Debreu (1954) constitutes an important part of economic theory. This model, often referred to as the 'Arrow-Debreu' model, is extended so as to include uncertainty about availability of resources, and about consumption and production possibilities. In particular, in his seminal paper, Arrow (1953) analyses optimal allocation of resources under subjective uncertainty and shows that Pareto-optimal allocation of risk is achieved in a complete, competitive financial market.[12] His analysis was later extended and generalized by Debreu (1959). Their approach is that of state-preferences as described in the above.

In the Arrow-Debreu model, assuming symmetric information, perfect competition and a complete set of markets, unregulated competitive markets are shown to ensure *efficiency* in the allocation and use of resources. Moreover, in equilibrium, demand equals supply at every date and in every state. In addition, there is one market clearing price only, referred to as the *Law of One Price*, and this market price conveys all relevant information to market participants.

The optimality of a competitive equilibrium under uncertainty is, however, limited to the special case of an economy possessing a *complete set of future markets* in all commodities. For most commodities, future markets do not exists. This may be due to differences in information, transaction costs, lack of product uniformity among producers, indivisibility in production and increasing returns to scale. In the following, we discuss the implications of abolishing the assumption of symmetric information.

1.5 Information

> "...the fact of differential information as between contracting parties will prevent some efficient contracts from being made."
> Arrow, 1973.

In the economic literature it is now well established that imperfect and asymmetric information has quite dramatic consequences for the functioning and outcome of an unregulated market. Indeed, many of the traditional

[12] The assumption that markets are complete, makes the Arrow-Debreu model under uncertainty equivalent to a model with certainty, Radner (1982).

'dictums' of economics no longer hold. If the assumption of symmetric information is abolished in the general equilibrium model, the existence and optimality of a competitive equilibrium is shown to be non-robust (Stiglitz, 1985). Indeed, asymmetric information can be shown to lead to an equilibrium in which there is *no market clearing*, and *no single market price* (Stiglitz, 1985). Market prices will, moreover, not convey all relevant information about scarcity, characteristics of goods and services, or on behavioral implications. Decentralization through the price system will consequently *not* ensure a Pareto optimum (Stiglitz, 1985, 2000).[13] What is more, an *equilibrium may not exist.*[14] Hence, the assumption of complete contingent markets in the Arrow-Debreu model does not hold.[15]

Following Stiglitz (1985), this may be explained by the following example. A price may reflect non-observable aspects of quality. Indeed, an increase in insurance premium may aggravate the risk-composition (quality) of the pool of insurees, since low-risk individuals as a consequence of higher premium will self-select out of the market. An insurer may thus be reluctant to offering insurance to an individual willing to pay a high premium since the insurer would think that she is a high-risk individual. Thus, the demand for insurance may not equal the supply of insurance. Moreover, if there exists an equilibrium in which markets do clear, then this would not be robust to competition: Another insurance firm may enter the market and offer lower insurance coverage at a lower premium and, thereby, attract low-risk individuals from the other firm. As a consequence, the initial firm suffers a loss since its risk-composition is worsen.

[13]Stiglitz (1985) refers to this as the fundamental non-decentralizability theorem (p. 27).

[14]Non-existence of complete contingent claims markets may not only be due to differences in information, but also due to transaction costs, lack of product uniformity among producers, *etc.*

[15]The Arrow-Debreu model has over the years been extended by others so as to incorporate also informational aspects when evaluating the efficiency of the market under uncertainty. Account has been taken of asymmetric information, the 'production' of information, and the incompleteness of markets. Radner (1968) extends the Arrow-Debreu model to allow for asymmetric information. He concludes that if economic decision makers have unlimited computational capacity for choice among strategies, then even if there is uncertainty about the environment, and different agents have different information and different beliefs about the environment, one can apply the standard theorems on the existence and optimality of competitive equilibrium.

1.5.1 Asymmetric Information

> "The exchange process is intertwined with the process of selection over hidden characteristics and the process of providing incentives for hidden behaviors." Stiglitz, 2000 (p. 1447).

Information, *i.e.*, knowledge, forms an integral part of individuals' decision-making process.[16] In economic theory, it is fundamental that agents have adequate information about the state of the world, characteristics of goods and services, consequences of behaviour, and choices available. Traditionally, it is assumed that individuals do indeed possess (and is able to comprehend) this information. If they do not, then they are equally badly informed. Evidently, information is not perfect and symmetrically distributed. The issue of health care easily demonstrates this: sick individuals may not know what is the appropriate treatment to undertake, what are the treatment options available, what is the efficiency of treatment, and sometimes even: whether indeed they are sick (*e.g.*, hypochondriacs). Since the 1970s, issues of imperfect information and in particular asymmetric information, have become central to most economic studies. Indeed, the 2001 Nobel Prize in Economics was rewarded three of the pioneers in the study of asymmetric information; George Akerlof, Michael Spence and Joseph Stiglitz. In what is often referred to as the economics of information, focus is placed on how to improve or acquire new information, information costs, *etc.* [17]

Here, the point at issue is that of *contracting* between two types of agents, one of which has more information (knowledge) than the other. We will refer to the more and less informed agents as informed and uninformed agents, respectively. The interesting question is how to *design* contracts in order to mitigate informational constraints regarding individuals' (i) behaviour, and (ii) characteristics. Problem (i) is usually referred to as a problem of hidden behaviour, whereas problem (ii) is referred to as a problem of hidden information. We will elaborate somewhat on the two types of asymmetric information in the next subsections.

[16] Hirschleifer and Riley (1992) distinguish between *knowledge* which refers to objective facts, and *beliefs* which refers to subjective knowledge. Individuals' decisions are based upon their beliefs, which, naturally, are updated when knowledge improves.

[17] For a textbook exposition on the economics of information, see Hirschleifer and Riley (1992).

Hidden Behaviour

Unobservable behaviour gives rise to problems of *moral hazard.* According to Dionne and Harrington (1990), the concept of moral hazard was introduced by, among others, Arrow (1963) and Pauly (1968).

When information about individuals' behaviour is private and not verifiable, then behaviour cannot be contractually enforced. Generally, it is distinguished between behaviour that affects the *likelihood* that an insured against event will occur (*e.g.*, the probability of injury), and behaviour that, after the insured-against event is realized, affects the *consequences* of the event (*e.g.*, costs associated with recovery from illness). The former is referred to as *ex ante moral hazard*, while the latter is referred to as *ex post moral hazard.*

Ex ante moral hazard exists because insurance *per se* reduces individuals' incentives to take precautionary actions so as to reduce the likelihood that uncertain outcomes will occur. The fundamental problem is that the uninformed agents (*e.g.*, insurers) cannot separate endogenous risk from exogenous risk. For instance, it is more likely that a person will take less care of her belongings if she knows that her insurance will cover the potential loss. If indeed she is deprived of her belongings, then the insurer will not be able to tell whether this occurred just by accident (*i.e.*, just 'bad luck'), or if it was because the insuree did not behave in a precautionary manner.

Ex post moral hazard is particularly prevalent in the case of health insurance and refers to the situation where risk is resolved. The fundamental problem is that the size of the loss itself may be endogenous to the insured. In the case of health insurance, the existence and severity of illness may not be easily verifiable, especially *prior* to treatment (Pauly, 1968). Insurance changes the relative prices facing the insured, hence, they will change their behaviour relative to a situation without insurance. For instance, a sick individual holding insurance against medical expenditures faces a low (or zero) price on treatment. If her demand for treatment is price elastic, then she will rationally respond to the low(er) price by increasing her demand for treatment. Suppliers of medical treatment acting on behalf of patients may indeed recommend this high level of treatment as long as the benefits from treatment are positive (since they may not face the true costs of the services provided).[18] Consequently, the more extensive insurance coverage, the less

[18]This gives rise to what is called 'third-party payment' problems.

incentives to economize both in the supply of, and demand for, health care.

It follows that the *ex post* moral hazard problem is not that individuals' behaviour is immoral, but that individuals increase their demand to a level at which the marginal production cost of treatment exceeds marginal utility from treatment, Pauly (1968). The more price elastic the demand for health care, the higher is the efficiency loss associated with an more generous insurance coverage.

Efficiency vs. Incentives The problems of moral hazard may be reduced if the appropriate incentives are incorporated in the contracts offered. There is, though, a trade-off between the goal of risk spreading, *i.e.*, efficiency, and the goal of appropriate incentives; on the one hand, the risk of suffering a loss should be allocated efficiently and on the other hand, incentives to behave optimally should be ensured. The conflict between risk prevention and moral hazard is by Fuchs (1996) referred to as the Fundamental Problem of Health Economics.

In the case of insurance, the moral hazard problem gives rise to partial insurance coverage only, since this will provide insurees with incentives to undertake preventive measures *ex ante*, and to behave in a cost-effective manner *ex post*. Usually, this takes the form of positive cost sharing, *e.g.*, in the form of coinsurance where the insured pays a certain percentage of the costs (or in the form of deductibles, *i.e.*, a fixed co-payment), Pauly (1968). Monitoring and regulation also form integral parts of an incentive scheme. For instance, individuals may be rationed in the level of treatment available if ill, Arrow (1968).[19]

Hidden Information

Asymmetric information on particulars of commodities (*e.g.*, medical treatment) and individuals (*e.g.*, proneness to illness), gives rise to problems of *adverse selection*. Information asymmetries exists in a number of markets and situations. For instance, a government may face adverse selection problems when allocating goods and services towards specific groups of individuals that are not easily identified. In the market for insurance, insurer and insuree typically have unequal information about the probability that

[19]Empirically, we observe for example that insurance companies inspects private homes and offers subsidized life vests, children-seats for cars, etc. The purpose of these efforts is, of course, to prevent accidents from occurring and to reduce their consequences.

the insured-against event will occur.[20]

Adverse selection have important implications for the functioning of markets. Kenneth Arrow notes in his seminal article "Uncertainty and the Welfare Economics of Medical Care" (1963) that the equilibrium allocation of risk bearing is *inefficient* relative to a situation where information is equally available to the insurer and the insured. Moreover, George Akerlof explains in his path-breaking article on "The market for lemons" (1970) why markets may fail to *exist*. He shows that in a situation where potential buyers cannot observe the quality of the traded goods, then profitable trade may not take place due to problems of adverse selection. Applied to the market for insurance, adverse selection hence reduces the welfare of those who would like to transfer risk and of those who would be willing to take on the risk. Furthermore, it reduces consumers' desire to consume services that have uncertain outcomes (Arrow, 1963).

> "... adverse selection tends to come about when (a) the insurance pool contains a relatively wide range of risks (all of whom must be charged the same premium per unit of coverage, owing to the insurer's inability to distinguish among them), and (b) risk-aversion is relatively mild. Conversely, a narrow range of risks and a high degree of risk-aversion tend to retain the better risks in the insurance pool and therefore to prevent adverse selection." Hirshleifer and Riley, 1992 (p. 312).

Selection problem There are, however, ways of mitigating the problems of adverse selection. In the literature, a distinction is made between studies in which the informed agents have incentives to *signal* their private information, and studies in which the uninformed agents design systems that *screen* (informed) individuals. In the first case, the informed agents 'move' first, while in the latter, the uninformed agents 'move' first. The article by Michael Spence in 1973 on "Job market signaling" is fundamental to the understanding of *signaling* when there are problems of adverse selection: Informed agents signal their unobservable characteristics by undertaking

[20]It is usually assumed that insurees know their true probability of encountering an accident, whereas insurers do not. The reverse may, however, also be true. Insurers may for instance have access to information which reveals (or improves the estimates of) the 'objective' probability of, *e.g.*, a certain illness, while insurees do not. Hence, insurees' 'subjective' probability (or probability beliefs) may be incorrect.

costly and observable behaviour, *e.g.*, education, in order to convince the uninformed agents about their quality (Nilssen, 2001).[21]

The uninformed agents may mitigate the effects of hidden information by designing a set of contracts that induce the informed agents to reveal their private information, and to behave in a preferred way. In particular, they offer a 'menu' of contracts from which the informed agents can choose. The menu of contracts is designed so that individuals have incentives to choose the contract intended for them, *i.e.*, to *self-select.*

The intriguing question is of course how to *design* such a set of contracts. The influential paper by James Mirrlees in 1971 on "An exploration in the theory of optimal income taxation" is considered to be the first contribution in this respect. He studies the design of an optimal tax scheme when information about ability (productivity) is private and, consequently, individualized lump sum taxes not feasible.

Shortly after, Michael Rothschild and Joseph Stiglitz published their highly influential article on "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information" in 1976. They show how uninformed agents can induce the better informed agents to reveal important information (*e.g.*, risk) by offering insurance contracts defining different *price-quantity* combinations (*i.e.*, bundles of premiums and deductibles). Hence, screening takes place by means of deductibles, *i.e.*, copayments. In equilibrium, if existent, low-risk individuals buy *partial coverage* at a low insurance premium while high-risk individuals buy full coverage at a high premium. Hence, low-risk individuals suffer a loss in welfare relatively to a situation with symmetric information. They summarize their analysis of competitive insurance markets with adverse selection by stating that:

> "...the structure of the equilibrium as well as its existence depended on a number of assumptions that, with perfect information, were inconsequential; and finally, and in some ways most disturbing, under quite plausible conditions equilibrium did not exist." Rothschild and Stiglitz, 1976 (p. 648).

Thus, asymmetric information may hamper efficient allocation of risks and, indeed, make a competitive insurance market *non-existent.* If an equilib-

[21] Riley (2001) provides an extensive overview of the literature on screening and signaling.

rium exists, then a Pareto improvement can be achieved through revelation and symmetric distribution of relevant information (Rothschild and Stiglitz, 1976). Their model is later extended by, among others, Wilson (1977) and Spence (1978) who show that a separating equilibrium may indeed exists if insurers anticipate competitors' reaction to their behaviour. For more on this, see Dionne and Doherty (1992).

1.6 The Role of Government

Governments are in a number of countries involved in the financing and supply of health care. We will in the following discuss efficiency and equity arguments in favour of public interventions in the market for health insurance.[22]

'First-Best'

Economic theory emphasizes the role of private markets in the efficient supply of goods. In a 'first-best' situation with perfect information, perfect competition and absence of market failures, an unregulated competitive market will ensure *efficiency* in the allocation and use of (scarce) resources. Thus, resources cannot be reallocated so as to improve the situation of one individual without making the situation for another individual worse. This is referred to as *The First Fundamental Welfare Theorem*, and implies that public interventions aiming at improving efficiency are futile. If the income distribution is considered undesirable by the society, then a government may alter the *distribution* by means of individualized lump-sum transfers and taxes. Indeed, any Pareto-efficient allocation is attainable in a perfectly competitive market when the appropriate redistribution of initial wealth is implemented. This is *The Second Fundamental Welfare Theorem.* Individualized lump-sum taxes and transfers thus allow the government to pursue its distributional ambitions without causing any price distortions and efficiency losses.

However, as we have shown in the above, the conditions on which these argument are based are quite strong, and are often not met. What will be the implications of asymmetric information for the role of the government? This is discussed below.

[22]Historically, though, the main motivation for public involvement in the financing and supply of health care in the Nordic countries is paternalism, Lundholm (1991).

'Second-Best'

Studies of public interventions when the informational assumption of the Arrow-Debreu general equilibrium model is abolished, are often referred to as studies of 'second best'. There is some dispute whether indeed it is appropriate to refer to the outcome of a model in which information is assumed not to be costly, and to be symmetrically distributed, as a 'first-best', see, *e.g.*, Stiglitz (1985). Still, such a theoretical first-best provides a useful reference point when discussing the implications of asymmetric information among market participants. In the following, we will thus use the terms first-best and second-best when suitable.

According to Boadway (1994), the concept of the theory of *second-best* was formalized by Lipsey and Lancaster in the mid 1950s. The theory initially revolved about problems of market failures causing the First Fundamental Theorem of Welfare Economics not to hold (Boadway, 1994). This implies that the economy cannot operate on its first-best utility possibilities frontier (*upf*). Public interventions may, thus, be *Pareto-improving*. However, the government may face the same informational restrictions as does the private market. The resulting allocation is thus referred to as *constrained Pareto efficient*. Geometrically, this amounts to moving from a point inside a hypothetical first-best *upf* to a point closer to it.

More recent contributions to the literature on second-best focus on *redistributive* policies when the Second Fundamental Welfare Theorem is violated. Private information about characteristics on which to condition the redistributional policy (*e.g.*, risk and ability) precludes non-distortionary distribution. Redistribution can thus only take place by use of instruments affecting the efficiency of the markets. Consequently, efficiency and equity considerations cannot be separated, in particular: there will be a trade-off between efficiency and equity.

In the following, we will briefly provide efficiency and equity arguments in favour of a public supply of health insurance in a world of asymmetric information.

1.6.1 Efficiency Arguments

A private market for health insurance may be incomplete or non-existent. Moral hazard and adverse selection cause individuals to be restrained in their level of insurance coverage, *e.g.*, because of a strictly positive co-payment or co-insurance. Risk is thus only partially covered. Moreover, certain types of individuals may be excluded from the market, *e.g.*, because of congenital (high-cost) medical conditions such as HIV and heart failure, or because of revealed health risks (*e.g.*, smoking, BASE-jumping, *etc.*).[23]

Inefficient allocation of risks in the private insurance market provides an *efficiency argument* for government interventions (Barr, 1989). Of course, governments will face many of the same informational constraints as do the private agents, *e.g.*, about individuals' private health risk. Still, the government is able to impose and enforce *compulsory* insurance, thereby reducing the problem of adverse selection. A compulsory *public* insurance may indeed be more successful in reducing adverse selection than is a compulsory and competitive private insurance market. This is so since the latter would (still) seek to improve its risk pool by 'cream-skimming' potential insurees.[24] Also, it allows for provision of long-term contracts, thus avoiding problems associated with contract re-writing. Consequently, public interventions in the market of insurance may improve the allocation of risk and ensure all individuals access to (some) health care if sick.

While the problem of adverse selection may provide an argument for public supply of health *insurance*, it is not obvious whether the supply of *health care* should be private or public. However, it may be argued that:

> "... the imperfect information of consumers justifies regulation of quality and that of insurance companies regulation of quantity. Both forms of policing might be more effective if production itself were public." Barr, 1989 (p. 75).

Regulation of the supply of health care relates to problems of *consumer sovereignty*. The assumption of consumer sovereignty is fundamental to economic theory; individuals are assumed to be capable of making rational choices between different goods and services when maximizing their utility. However, in order to make rational choices, individuals must both have access to, and the ability to comprehend the necessary information. In the case of health, and health care demand, the assumptions of rationality and consumer sovereignty may not be fulfilled for several reasons. For instance,

[23]Alternatively, the insurance premium is set so high that individuals are not able to enter the market.

[24]For a textbook exposition of Pareto improving compulsory insurance, see Zweifel and Breyer (1997), chapter 5.

sick individuals may indeed not be able to fully apprehend (i) whether they are in need of care, (ii) which services are appropriate, and to what extent, (iii) what is the efficiency of treatment, and (iv) what is the quality of the services provided. This is information that usually suppliers of health care (at least to some extent) possess. The lack of information consequently makes the individuals (patients) 'dependent' on the suppliers of care in that they must assist them in defining their medical needs and making decisions. Hence, there are arguments in favour of public regulations of quality and quantity of care; indirectly in the form of licensing of health personnel and prosecution of malpractice, or directly in the form of a public supply.

Also, there are efficiency arguments in favour of integrating the financing and supply of health care. This is because of the so called ' *third-party payment problem*': by merging the supply side and the financing side, then suppliers of care will face the full marginal costs of their decisions. While in the Scandinavian countries health care is publicly provided, private insurers and suppliers in the USA are increasingly forming parts of Health Maintenance Organizations (HMOs) in order to contain costs.

Moreover, *ex post* moral hazard may be reduced if insurees are reimbursed in the form of health care services, *i.e.*, in kind, rather than cash. Indeed, indemnity insurance, that is; insurance that provides cash compensation if the insured against event occurs, is seldom observed in health insurance. More often, insurance benefits are stated in physical terms, *e.g.*, payments for consultations with physicians, treatment, *etc.* Intuitively, compensation in kind will strengthen individuals' incentives to take precautionary actions. Moreover, it seems reasonable to assume that individuals take less pleasure from unnecessary treatment than from cash payments, hence, their incentive to exaggerate loss is lower.

1.6.2 Redistribution

Promotion of equality may be considered one of the major tasks of a government. Redistributional considerations are particularly prevalent in the case of health care. In economic theory, it is often distinguished between *general* and *specific* egalitarianism, the first referring to equalization of incomes, the latter to the equalization of consumption or access to specific resources (*e.g.*, health care), Sandmo (1991). Here, focus is placed on a general distributional ambition, namely that of income equality (*i.e.*, ability to pay for care).

Distortionary Taxation

> "..distortionary taxation need not be imposed as an exogenous constraint on the problem; it arises naturally as an optimal form of policy in a world of imperfect information." Boadway and Keen, 2000 (p. 738).

In the literature on optimal income taxation it is well established that asymmetric information about characteristics on which a tax is to be based, restricts the *range* of policies available to the government and the *design* of these.[25] Moreover, it restricts the *extent* of redistribution possible. This is so because taxes and transfers must be conditioned on observable characteristics, characteristics that may indeed be endogenous to individuals. Thus, individuals' may alter their behaviour as a response to the policy: they may choose to reduce their labour supply in the formal labour market when income taxes increases, and to not comply with the tax regulations (for instance by misreporting earnings). The government thus face problems of adverse selection and moral hazard. Consequently, redistribution implies an efficiency loss.

Mirrlees (1971) is cited to have formalized the problem of optimal income redistribution when information is private (Boadway, 1994). He derives a scheme which provides individuals with incentives to behave in accordance with the governments's intentions. The government must balance its ambition to achieve equity against the loss in efficiency imposed by the policy. Taking into account the effect of the scheme on individuals' *incentives*, he characterizes a non-linear optimal tax scheme that takes this equity-efficiency trade-off into account. A prerequisite is that individuals' preferences are such that marginal rate of substitution between post- and pre-tax income is decreasing in ability (productivity). This is referred to as an *agent-monotonicity property* and implies that pre-tax income is increasing in ability (Myles, 1995).[26] Geometrically, this means that the slope of an indifference curve through a given post- and pre-tax income point is less steep the higher the level of ability. Thus, indifference curves entail a single-crossing property.

[25] See Sandmo (1999) for an excellent presentation of James Mirrlees' and William Vickrey's contribution to the literature on public policy under asymmetric information.

[26] In the literature on screening, this condition is referred to as the 'Spence-Mirrlees condition' (Macho-Stadler and Pérez-Castrillo, 1997.)

Stiglitz (1982, 1987) analyses Pareto efficient non-linear income taxation, viewing the optimal tax problem as an adverse selection problem. Analogously to an insurer's adverse selection problem, the set of contracts from which individuals can choose, provides a screening device that induces individuals to reveal private information through their choice of contract. The self-selection constraints facing the government amount to restricting the tax schemes to those guaranteeing that individuals with certain characteristics (*i.e.*, skills) do not have an incentive to 'mimic', that is, to choose contracts intended for individuals with different characteristics. The self-selection constraints thus constitute an 'upper limit' to redistribution. Stiglitz (1987) shows that the government may relax the self-selection constraints by introducing distortionary taxation. In particular, assuming that the government pursues income redistribution towards low-ability individuals, and that individuals' labour supply are perfect substitutes, then the Pareto-efficient non-linear tax scheme entails a zero marginal tax rate on high ability individuals, and a positive marginal tax rate (ranging from 1 to 100 percent) on low ability individuals. Thus, the efficiency conditions are distorted for only one of the ability types; the one that others may wish to mimic.[27] Stiglitz (1987) refers to this as a property of Pareto-efficient tax structures in a second-best world (often referred to as 'non distortion at the top'). Moreover, if the government raises taxes for distributional purposes only, then, in this setting, the Pareto-efficient tax scheme is shown to be progressive.

Now, assuming that there exists only two types of individuals: those with a high innate ability (H) and those with a low such ability (L), and that the number of high-ability individuals equals the number of low-ability individuals. The single-crossing property ensures that at most only one self-selection constraint is binding. Following Boadway and Keen (2000), we can then depict the self-selection constraint by means of a utility possibility frontier (*upf*). In Figure 1.1, the curve FF illustrates a first-best *upf* (assuming Pareto-efficiency). The initial situation without redistribution is given by the *laissez faire* point L on the curve FF. If individualized lump-sum taxation would be possible, then a government with a utilitarian social welfare function would choose point U on the curve (where the slope of the

[27]Our finding resembles the results found in the literature on asymmetric information in insurance markets where high- and low-risk individuals are offered different contracts and where low-risk individuals are constrained in the extent of coverage offered, see for example Rothschild and Stiglitz (1976) who address the private market mechanism.

upf is -1), while a government pursuing a maxi-min welfare function would choose point M (where the *upf* intersects the 45° line).

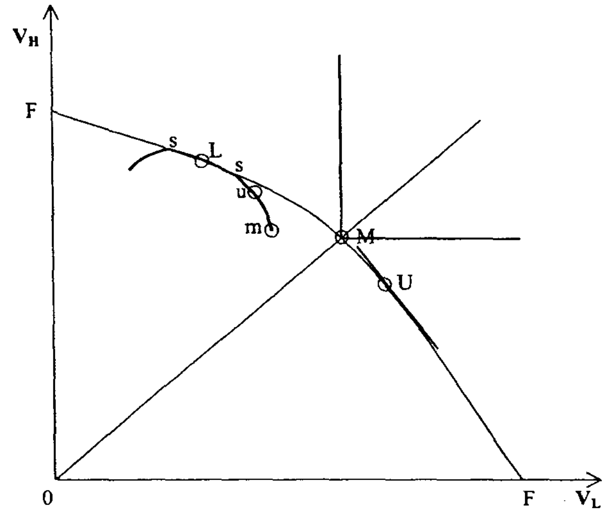

Figure 1.1: First- and second-best utility possibilities frontier.

Now, the government is unable to issue such lump-sum taxes. Redistribution can be carried out to some extent without self-selection being a problem. This is illustrated by the segment ss on the FF curve. Further redistribution cannot be accomplished without the self-selection constraint binding. Thus, beyond ss, the utility possibilities are restricted to those illustrated by the heavily drawn curve. This is the second-best *upf* and includes only points such that high-ability individuals enjoy higher levels of utility than do low-ability individuals. This is so because high-ability individuals will spend less hours in the labour market in order to generate a certain level of earnings than do low-ability individuals. Assuming a quasi-concave social welfare function, the relevant part of the second-best *upf* is the segment sm. Indeed, the second-best utilitarian optimum is now given by point u, whereas the second-best maxi-min point is given by m. As shown, when information is imperfect, the utilitarian optimum (u) entails less redistribution than does the information constrained maxi-min (m), (while the opposite is the case when information is perfect).

Public Provision of Private Goods

> "...viewing the failure of the Second Theorem as a problem of asymmetric information turns out to have rather dramatic consequences for the extent of redistribution that can be achieved by the tax-transfer system. Not only that, it leads to some rather surprising policy prescriptions." Boadway, 1994 (p. 2).

Generally, taxes and transfers are considered the most appropriate instruments for redistribution. Transfers in cash are, moreover, viewed as superior to transfers in kind. Public provision (*i.e.*, financing and supply) of private goods is traditionally justified on the grounds of market failures, merit wants arguments and redistribution (Hare, 1988). However, in their seminal article, Guesnerie and Roberts (1984) show that quantity constraint are Pareto improving if the economy initially is in a second-best situation. Indeed, transfers in kind may slacken the self-selection constraints facing the government thus improving the redistributive efficiency of a non-linear income tax scheme. In-kind transfers may thus, enhance welfare.[28]

In a world of asymmetric information, public supply of private goods may serve as a means of redistribution, depending upon its effect on the constraints facing the government when pursuing its distributional goals. Indeed, the efficient scheme is contingent upon how the private good in question is valued by individuals, and how it influences individuals' labour supply. In particular: how is their marginal willingness to pay for the good influenced by their earnings ability. The key idea is that different types of individuals value the good in question differently (Besley and Coate, 1991).

It is well established that if the marginal valuation of the private good is increasing in the number of hours spent in the labour market (implying that the low-ability individuals value the increase in treatment more than do high-skilled mimicker), then (more) public supply will slacken the self-selection constraint, thus increasing the extent of redistribution possible. Correspondingly, if the marginal valuation of the private good is decreasing in the number of labour hours, (implying that low-ability individuals value the increase in treatment less than do mimickers), then (more) public supply

[28]Public provision may also be justified on the grounds that the the good in question, *e.g.*, health care and child care, may induce individuals to increase their labour supply, thereby improving the tax base. See, *e.g.*, the study by Bergstrom and Blomquist (1993) on public provision of day care.

tighten the self-selection constraint, Christiansen (1998). Blomquist and Christiansen (1998) show that if the self-selection constraint is binding, then a public provision of private goods may strictly Pareto dominate the optimal income tax optimum. The private good in question should, however, not be re-tradable or supplementable, and it should be a substitute for leisure (Blomquist and Christiansen, 1995). Balestrino (1999) provides an excellent survey of the different arguments raised in favour of in-kind transfers in a second-best world. He distinguishes between papers that makes an argument for public provision without taking the tax system into account, and those that do take the tax system into consideration.

1.7 Synopsis

The preceding overview provides a background against which the essays included in this thesis can be viewed. Although the overview spans a large number of topics, the fundamental problem underlying the questions at issue is the same; namely that of asymmetric information and contract design. Indeed, this is at the core of the essays included in this thesis.

References

Akerlof, G. (1970)."The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(3), 488-500.

Arrow, K.J. (1953). "Le rôle des valeurs boursières pour la répartition la meilleure des risques."In *Econométrie*, CNRS, Paris, 41-47. English version: "The Role of Securities in the Optimal Allocation of Risk-Bearing." (1964). *The Review of Economic Studies* , 31, 91-96.

Arrow, K.J. (1963). "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review*, 53(5), 941-973.

Arrow, K.J. (1968). "The Economics of Moral Hazard: Further Comment." *The American Economic Review*, 58(3), 537-538.

Arrow, K.J. (1971)."Insurance, Risk and Resource Allocation." In *Aspects of The Theory of Risk-Bearing.* Yrjö Jahnsson Lectures, Helsinki, 45-56.

Arrow, K.J. (1973). "Information and Economic Behaviour." Lecture at The Federation of Swedish Industries, Stockholm, May 30.

Arrow, K.J. and G. Debreu. (1954). "Existence of an Equilibrium for a Competitive Economy." *Econometrica*, 22(3), 265-290.

Balestrino, A. (1999). "The Desirability of In-kind Transfers in the Presence of Distortionary Taxes." *Journal of Economic Surveys*, 13(4), 333-354.

Barr, N. (1989). "Social Insurance as an Efficiency Device." *Journal of Public Policy* 9, 59-82.

Besley, T. and S. Coate. (1991). "Public Provision of Private Goods and The Redistribution of Income." *The American Economic Review*, 81(4), 979-984.

Blomquist, S. and V. Christiansen. (1995). "Public Provision of Private Goods as a Redistributive Device in an Optimum Income Tax Model." *Scandinavian Journal of Economics*, 97 (84), 547-567.

Blomquist, S. and V. Christiansen. (1998). "Price Subsidies Versus Public Provision." *International Tax and Public Finance*, 5(3), 283-306.

Blomqvist, Å. and H. Horn. (1984). "Public Health Insurance and Optimal Income Taxation." *Journal of Public Economics*, 24(3), 353-371.

Boadway, R. (1994). "The Role of Second-Best Theory in Public Economics." Working Paper no. 95-06, Economic Policy Research Unit (EPRU), University of Copenhagen, Denmark.

Boadway, R. and M. Keen. (1993). "Public Goods, Self-Selection and Optimal Income Taxation." *International Economic Review*, 34(3), 463-478.

Boadway, R. and M. Keen. (2000). "Redistribution." In: *Handbook of Income Distribution*. Atkinson and Bourguignon (eds). North-Holland. Vol. 1, 677-789.

Boadway, R. and M. Marchand. (1995). "The Use of Public Expenditures for Redistributive Purposes." *Oxford Economic Papers*, 47(1), 45-59.

Breyer, F. and A. Haufler. (2000). "Health Care Reform: Separating Insurance from Income Redistribution." *International Tax and Public Finance*, 7(4-5), 445-461.

Christiansen, V. (1998). "Public Provision and Distortionary Taxes: Two Second Best Regimes." Paper prepared for Symposium in honour of Agnar Sandmo. NHH 9-10 January.

Cook, P.J. and D.A. Graham. (1977). "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities." *The Quarterly Journal of Economics*, 91, 143-156.

Cullis, J.G., and P.A. West. (1979). *The Economics of Health. An Introduction.* Martin Robertson and Company Ltd.

Currie, J. and B.C. Madrian (1999). "Health, Health Insurance and the Labor Market." In: *Handbook of Labor Economics*. Ashenfelter and Card (eds). Elsevier Science, North-Holland. Volume 3C, 3309-3416.

Dardanoni, V. and A. Wagstaff. (1987). "Uncertainty, Inequalities in Health and the Demand for Health." *Journal of Health Economics* 6(4), 283-290.

Debreu, G. (1959). *Theory of Value*, Chapter 7, 98-102. New York: Wiley.

Dionne, G. and N. Doherty. (1992). "Adverse Selection in Insurance Markets: A Selective Survey." In: *Contributions to Insurance Economics.* Dionne (ed.). Kluwer Academic Publishers.

Dionne, G. and S.E. Harrington. (1990). "An Introduction to Insurance Economics." In: *Foundations of Insurance Economics. Readings in Economics and Finance.* Dionne and Harrington (eds). Kluwer Academic Publishers, 1-48.

Doherty, N.A. and H.J. Jung. (1993). "Adverse Selection When Loss Severities Differ: First-Best and Costly Equilibria." *The Geneva Papers on Risk and Insurance Theory*, 18(2), 173-182.

Fallon, P. and D. Verry. (1988). *The Economics of Labour Markets.* Philip Allan.

Fuchs, V.R. (1996). "What Every Philosopher Should Know About Health Economics." *Proceedings of the American Philosophical Society* , 140(2), 186-195.

Fuchs, V.R. and R. Zeckhauser. (1987). "Valuing Health - A "Priceless" Commodity." *The American Economic Review*, 77(2), 263-268.

Guesnerie R. and K. Roberts. (1984). "Effective Policy Tools and Quantity Controls." *Econometrica*, 52(1), 59-86.

Hare, P.G. (1988). "Economics of Publicly Provided Private Goods and Services." In: *Surveys in Public Sector Economics.* Hare (ed). Blackwells, Oxford.

Hirshleifer, J. (1965). "Investment Decision Under Uncertainty: Choice-Theoretic Approaches." *The Quarterly Journal of Economics*, 79(4), 509-536.

Hirshleifer, J. and J.G. Riley. (1992). *"The Analytics of Uncertainty and Information."* Cambridge University Press.

Loubergé, H. (2000). "Developments in Risk and Insurance Economics: the Past 25 Years." In: *Handbook of Insurance.* Dionne (ed.). Kluwer Academic Publishers.

Lundholm, M. (1991). "Compulsory Social Insurance: A Critical Review." Department of Economics, 1991:1. Uppsala University, Sweden.

Machina, M.J. (1987). "Choice Under Uncertainty: Problems Solved and Unsolved." *Economic Perspectives*, 1(1), 121-154.

Macho-Stadler, I. and D. Pérez-Castrillo. (1997). *An Introduction to the Economics of Information. Incentives & contracts.* Oxford University Press.

McGuire, A., J. Henderson and G. Mooney. (1988). *The Economics of Health Care: An Introductory Text.* Routledge and Kegan Paul, London.

Mirrlees, J. (1971). "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies*, 38, 175-208.

Mushkin, S.J. (1962). "Health as an Investment." *Journal of Political Economy*, 70, 129-57.

Myles, G.D. (1995). *Public Economics.* Cambridge University Press.

Nichols, A.L. and R. J. Zeckhauser. (1982). "Targeting Transfers Through Restrictions on Recipients." *The American Economic Review*, 72(2), 372-377.

Nilssen, T. (2001). "*De vet no' som ikke vi vet: Nobel-prisvinnerne i økonomi.*" Økonomisk Forum, 55(9).

OECD (2003). Health data. Available at: http://www.oecd.org/

Pauly, M.V. (1968). "The Economics of Moral Hazard: Comment." *The American Economic Review*, 58(3), 531-537.

Pauly, M.V. (1971). *Medical Care at Public Expense. A Study in Applied Welfare Economics.* Praeger Publishers.

Radner, R. (1968)."Competitive Equilibrium Under Uncertainty." *Econometrica*, 36(1), 31-58.

Radner, R. (1982). "Equilibrium under Uncertainty." In: *Handbook of Mathematical Economics.* Arrow and Intriligator (eds.). North-Holland Publishing Company. Vol. 2, 923-1006.

Riley, J. G. (2001). "Silver Signals: Twenty-Five Years of Screening and Signaling." *Journal of Economic Literature*, 39, 432-478.

Rothschild, M. and J.E. Stiglitz. (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics*, 90(4), 629-49.

Rothschild, M. and J.E. Stiglitz. (1997). "Competition and Insurance Twenty Years Later." *The Geneva Papers on Risk and Insurance Theory*, 22, 73-79.

Sandmo, A. (1976). "Optimal Taxation. An Introduction to the Literature." *Journal of Public Economics,* 6(1-2), 37-54.

Sandmo, A. (1991). "Economists and the Welfare State." *European Economic Review*, 35(2-3), 213-239.

Sandmo, A. (1995). "Introduction: The Welfare Economics of the Welfare State." *Scandinavian Journal of Economics*, 97(4), 469-476.

Sandmo, A. (1999). "Asymmetric Information and Public Economics: The Mirrlees-Vickrey Nobel Prize." *Journal of Economic Perspectives*, 13(1), 165-180.

Schlesinger, H. (1984). "Optimal Insurance for Irreplaceable Commodities." *The Journal of Risk and Insurance,* 51, 131-137.

Spence, M. (1978). "Product Differentiation and Performance in Insurance Markets." *Journal of Public Economics*, 10, 427-447.

Stiglitz, J.E. (1982). "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics*, 17, 213-40,

Stiglitz, J.E. (1985). "Information and Economic Analysis: A Perspective." *The Economic Journal*, 95, 21-41.

Stiglitz, J.E. *et al.* (1989). *The Economic Role of the State.* A. Heertje (ed). Basil Blackwell in association with Bank Insinger de Beaufort NV.

Stiglitz, J.E. (2000). "The Contribution of the Economics of Information to Twentieth Century Economics." *Quarterly Journal of Economics*, 115(4), 1441-1478.

Zweifel, P. and F. Breyer. (1997). *Health Economics.* Oxford University Press.

Varian, H.R. (1996). *Intermediate Microeconomics. A Modern Approach.* Fourth edition. W.W. Norton and Company.

Wilson, C. (1977). "A Model of Insurance Markets with Incomplete Information." *Journal of Economic Theory*, 16, 167-207.

Essay 2

Health Insurance: Treatment vs. Compensation

Abstract

In this paper, we view health insurance as a combined hedge against the two consequences of falling ill: treatment expenditures and loss in income. We discuss how an individual's ability when healthy affects her decision on whether to buy health insurance with treatment to full recovery if ill or with partial treatment combined with cash compensation for the resulting loss in income. We find that a high-ability individual demands full recovery and equalise utility across states, while a low-ability individual demands partial treatment and cash compensation and suffers a loss in utility if ill.

2.1 Introduction

Who are the individuals choosing to be partially disabled and to live from disability payments if ill, and who are the individuals choosing to fully restore ability if ill? This is the focus of our paper. In particular, we study how different types of individuals prefer to be compensated if illness occurs. The novelty of our paper lies in the integration of individuals' demand for insurance against medical expenditures, *i.e.*, medical insurance, and their demand for insurance against (permanent) loss in earnings due to impaired health, *i.e.*, disability insurance. Medical insurance and disability insurance provide coverage against different consequences of the same risk, namely that of falling ill. Integrating the two and solving for the optimal design of insurance, we show that individuals may indeed trade off health for consumption.

Our paper is motivated by the empirical fact that health insurance and disability insurance are integrated in a number of real-life health care systems. This is particularly prevalent in European countries where health insurance with in kind compensation, and disability insurance with cash compensation typically form parts of a public tax-financed (social) insurance system. Also in the United States, public programs include medical insurance and disability payments: low-income (poor) individuals are insured through a public program against both medical expenditures (*e.g.*, Medicare or Medicaid) and against loss in income due to disability (*e.g.*, through public disability programs, or income support programs toward disabled such as Social Security Disability Insurance (SSDI), Supplementary Security Income (SSI), or state-based Workers' Compensation systems). High-income individuals in the US are, on the other hand, typically (privately) insured against medical expenditures but only few have private disability insurance. Moreover, whereas low-income individuals falling into the Medicare or Medicaid program are usually considered to be restrained in the level (and quality) of health care provided, high-income individuals with private health insurance have access to high-quality care. Thus, we observe that different income groups are compensated differently if ill. It is an empirical fact, both in the US and in Europe, that individuals with less schooling are more likely to be disabled than those with more schooling, see for instance Haveman and Wolfe (2000). Naturally, there are many explanations to why this is so: different proneness to illness and thus different educational 'carriers', different types of work and thus different exposure to health risks, etc. The causal effects are also not readily apparent. We do not aspire to provide a definite answer to this challenging question, but hope to shed some light on the question by studying how individuals' inherent earnings capabilities influence their decision regarding how to be compensated if ill.

Generally, individuals face an inevitable risk of falling ill. Illness entails both monetary and non-monetary consequences. Firstly, ill individuals purchase health care services in order to alleviate the health consequences of the illness, *i.e.*, they incur medical expenditures. Secondly, during the period of illness, individuals suffer a (temporarily) loss in earnings due to reduced productivity in the labour market. Thirdly, if the illness cannot be cured, or if the quantity and quality of care undertaken is insufficient, then individuals suffer a permanent loss in health, *i.e.*, they are disabled. Impaired health reduces productivity in the labour market and, subsequently, reduces earnings. (Naturally, the magnitude of the loss in earnings depends on the severity of the disability, *e.g.*, whether it is total or partial.) Fourthly, illness implies a non-monetary loss to the individuals. To be in 'good' health is of value in itself and, moreover, influences the utility derived from other activities, *e.g.*, from consumption.

Illness hence entails diverse and quite substantial losses, all of which the individuals would like to hedge against *ex ante*. In the literature on health insurance, focus is mainly placed on insurance against medical expenditures, assuming that illness entails only monetary losses (e.g., medical expenditures and temporary loss in earnings). The desire to restore health is, by and large, taken for granted.[1] When non-monetary consequences of illness are taken into account, it is assumed that utility is state dependent and that health is either not restorable or irreplaceable.[2] In this paper, we allow for both monetary and non-monetary consequences of illness without imposing the assumption that health is irreplaceable. Rather, we assume that health if ill is endogenous: health if ill is improved if individuals receive medical treatment. We assume that treatment is divisible, and that individuals can choose to which extent their health is to be restored (with certainty) if ill. Health is thus insurable and, the non-monetary consequences of illness endogenous. Our model may thus provide a bridge between models taking only monetary consequences of illness into account, and models postulating that health is irreplaceable. A recent paper by Flochel and Rey (2002) supplements our analysis in that they study individuals' *ex post* demand for health care and, given the level of health care demanded, derive their optimal level of insurance against medical expenditures. Like us, they assume that utility is a function of both consumption and health. However, they do not allow labour earnings to depend on health state, and so their analysis does

[1] The only exceptions we know of are Byrne and Thompson (2000) and Graboyes (2000), who argue that, when the probability of successful treatment is small, the insured may be better off with cash compensation if ill, rather than going through the treatment.

[2] Analysis based on health being non-restorable include Zeckhauser (1970), Arrow (1974), Viscusi and Evans (1990), Evans and Viscusi (1991), and Frech (1994). Health is irreplaceable if individuals value restored health lower than pre-illness health; see Cook and Graham (1977) and Schlesinger (1984).

not encompass disability insurance.

To our knowledge, the health-insurance literature does not discuss compensation of (permanent) income loss due to illness, *i.e.*, a disability payment. Little is thus known about individuals' *choice between different types of compensation* in the case of illness. Indeed, Pauly (1986) recognizes the absence of studies of the relationship between individuals' demand for medical insurance and their demand for insurance that pays cash if illness occurs.[3] Our model facilitates an integrated analysis of what are traditionally considered a medical insurance and a disability insurance. It allows individuals to choose from different types of compensation; whether to receive (i) medical treatment (and thereby improve health), (ii) cash compensation of permanent income loss (*i.e.*, no health improvement), or (iii) a combination of both. We do not discuss contracts providing compensation against temporarily loss in earnings as our model is atemporal (*i.e.*, we assume instant recovery if treatment is received). We show that individuals with low ability (productivity) will choose type (iii) compensation whereas individuals with high ability choose type (i).

Our main analysis takes place in a world of symmetric information about health risks and health states. However, we show that our findings are robust, even in a situation where health state is *not* verifiable, *i.e.*, when insurers face problems of *ex-post* moral hazard, since the integration of in-kind provision of medical treatment and cash compensation reduces an insured individual's incentive to falsely claim to be ill when in good health. Hence, integrating medical treatment, *i.e.*, medical insurance, and cash compensation, *i.e.*, disability insurance, is not intrinsic to a public health-care system, but may also grow out of a totally unregulated system. Indeed, this possibility may be viewed in light of the increasing importance in the US during the last decades of Health Maintenance Organizations (HMOs) where financing and supply of medical treatment are integrated (mainly) in order to reduce moral-hazard problems.[4] Our findings suggest that also financing

[3]As examples of the latter, Pauly (1986) mentions, among others, salary continuation insurance, disability insurance and life insurance (see his note 4).

[4]More than 70 percent of the insured in US were enrolled in some form of managed care plan in 1993 (Glied, 2000). The term managed care organizations comprises organizations that mediate between the insured and the providers of care, *e.g.*, by regulating the services available, as in HMOs, or that restrict the insured's choice of providers, *e.g.*, as in independent practice associations (IPAs) and preferred provider organizations (PPOs), Glied (2000).

of income loss due to illness could form part of such schemes. In fact, if *ex-post* moral hazard is a problem, then integrating medical insurance (with in kind provision) and disability insurance (with cash compensation) may induce self-selection. Integrating the two may indeed incite a larger private provision of disability insurance.

The outline of this paper is as follows. The model is presented in Section 2.2 and a preliminary analysis is provided in Section 2.3. Our main findings are derived in Section 2.4. In Section 2.5, we consider a special case where the individual has Cobb-Douglas preferences. In Section 2.6, we discuss the case where health state is not verifiable, *i.e.*, *ex-post* moral hazard. Our results are discussed in a concluding Section 2.7.

2.2 The Model

Consider an individual who has preferences over consumption, c, and health, h. The individual faces exogenous uncertainty with respect to her state of health. She may either be healthy, which corresponds to state 1, or she may fall ill, which corresponds to state 2. The two states are mutually exclusive, jointly exhaustive, and verifiable. In state 1, the level of health is normalized to 1: $h_1 = 1$. In state 2, the individual is ill and suffers a complete loss in health: $h_2 = 0$. Health if ill may, however, be partially or fully restored (with certainty) if the individual receives medical treatment: $t \in [0, 1]$, *i.e.*, treatment is assumed to be a continuous variable. Medical treatment leading to full recovery is available at cost C, while treatment at cost tC leads to partial recovery.[5] It follows that if an individual receives treatment at a level leading to complete recovery, *i.e.*, if $t = 1$, then her level of health if ill is equal to 1: $h_2 = 1$. If no treatment is received, then $t = 0$ and health equals zero: $h_2 = 0$. Health in state 2 is henceforth measured by the fraction of C spent on treatment, that is, $h_2 = t$. Consumption in the two states are denoted c_1 and c_2, respectively.

The objective probability of falling ill is known to the individual and given by π, where $0 < \pi < 1$. The individual seeks to maximize the

[5]The cost of curing an illness is assumed to depend on the characteristics of the illness, rather than the characteristics of the individual suffering from it. Since all individuals face the same health risk, the cost of treatment is constant across individuals.

von Neumann-Morgenstern expected utility:

$$(1-\pi)u(c_1, 1) + \pi u(c_2, t),$$

where $u(c, h)$ is a Bernoulli utility function. We assume that $u : \Re^2_+ \rightarrow \Re$ is twice continuously differentiable, strictly concave, and satisfies: $\forall (c, h) \in \Re^2_{++}$, $u_c > 0$ and $u_h > 0$, where partial derivatives are denoted by subscripts. In particular, a strictly concave u implies that the individual is risk averse. Moreover, health and consumption are assumed to be complements in utility: $u_{ch} > 0$. This implies that individuals take more pleasure in consumption when health is good than if poor. We also assume that $u_c(c, h) \rightarrow \infty$ as $c \downarrow 0$ whenever $h > 0$, and $u_h(c, h) \rightarrow \infty$ as $h \downarrow 0$ whenever $c > 0$. Finally, $u_c(c, h) \rightarrow \infty$ or $u_h(c, h) \rightarrow \infty$ as $c \downarrow 0$ and $h \downarrow 0$. Note that our assumptions on u imply normality. As specified in the above expected utility function, utility from treatment is positive and decreasing ($u_h > 0$ and $u_{hh} < 0$). This implies that marginal utility from treatment is higher at a low treatment ratio, than at a high such ratio. (Hence, if we would measure treatment in terms of utility, then cost of treatment would be strictly convex.) Also, it follows from the properties of u that the individual prefers to receive an intermediate level of treatment to an uncertain prospect of receiving either complete or zero treatment with the same expected cost.

There exists a competitive insurance market in which profit maximizing insurers offer insurance at an actuarially fair premium. Information about the individual's probability of falling ill (π), which disease she is suffering from, and consequently, the associated costs of treatment, is symmetrically distributed among the market participants. Moreover, health state is *verifiable*, thus insurance policies can be made contingent on it. We hence rule out problems of *ex-ante* and *ex-post* moral hazard. (A situation where health state is non-verifiable is, however, discussed in Section 2.6). The market for health insurance will, therefore, be efficient.

The individual's inherent capacity to generate earnings is henceforth referred to as 'ability', and is given by A. If well, the individual enjoys her full ability, A. We choose our monetary unit so that total earnings are equal to A when well. Since, by assumption, leisure is not included in the utility function, we implicitly assume that leisure is constant (and thus, labour supply is fixed) across states. We assume that the individual, by spending tC on medical treatment, will generate earnings equal to tA when ill. Hence, labour earnings in state 2 are proportional to the rate of cost spent on treatment (recalling that t gives the rate of total treatment cost (C) spent on recovery). Note that the following analysis does not require insurance companies to know the individual's ability when healthy; hence A may be private information.

In this model, there are two types of consequences of an illness. Firstly, the individual suffers a loss in earnings due to reduced ability (productivity), and entails medical expenditures due to illness, *i.e.*, financial losses. Secondly, the individual suffers a direct loss in utility because of lower health (since $u_h > 0$), *i.e.*, a non-financial loss. Since health if ill is endogenous, the size of the non-financial loss is also endogenous. Indeed, if $t = 1$, then she suffers a financial loss only, *viz.*, the costs of treatment, while if $t < 1$, then she suffers both a financial and a non-financial loss. Consequently, a positive level of treatment reduces the non-financial loss from illness relative to a situation without treatment, in fact, if complete treatment is received, the non-financial loss is eliminated.

The risk-averse individual wishes to insure against the consequences of falling ill. Her insurance decision takes place *prior* to her knowing which state has occurred. Her budget constraints in states 1 and 2 are respectively given by:

$$c_1 + \pi I = A,$$

and

$$c_2 + \pi I + tC = tA + I,$$

where πI denotes the insurance premium to be paid in both states of the world in order to receive compensation equal to I if ill. From this, it follows that $I = tC + c_2 + (c_1 - A) - tA$, that is, the compensation consists of a compensation of medical expenditures equal to: tC, and a cash compensation equal to: $(c_2 + \pi I - tA)$.[6] Insurance is the only way to transfer income across the two states of the world. Combining the two budget constraints, it follows that the individual is constrained by:

$$A - c_1 = \pi[tC + (c_2 - tA + A - c_1)] \tag{2.1}$$

when *ex ante* making her choice of c_1, c_2, and t.

[6]Since the premium $A - c_1$ must be paid in both states, disposable income net of the premium equals $tA - (A - c_1)$ if no cash compensation is received. Hence, the cash compensation equals $c_2 - [tA - (A - c_1)]$.

We make the additional assumption about the utility function u that the marginal rate of substitution of consumption for health, u_h/u_c, is higher at full than at partial recovery from illness, given that the expected cost of treatment is subtracted:

$$\frac{u_h(tA - \pi tC, t)}{u_c(tA - \pi tC, t)} \leq \frac{u_h(A - \pi tC, 1)}{u_c(A - \pi tC, 1)}. \tag{2.2}$$

This assumption means that, for a fixed relative price of health in terms of consumption across states, the individual wants to shift the expected cost of treatment towards the healthy state if treatment does not lead to complete recovery (*i.e.*, if $t < 1$). A homothetic utility function satisfies this for any non-negative expected cost of treatment, but the assumption is also satisfied by other demand systems.

In the following, we characterize the individual's demand for insurance, both with respect to level, and type of, coverage. In particular, we analyze how the individual's level of inherent ability, A, influences her choice of compensation: whether to be compensated in the form of health restoration, *i.e.*, medical treatment, and/or in the form of cash, *i.e.*, compensation of loss in income due to incomplete recovery.

2.3 Preliminary Analysis

As explained earlier, we assume that treatment leading to a health level t is available at a cost tC when ill. For the purpose of our analysis, however, let us be more general and ask what is the maximum utility achievable if the individual has to pay P (≥ 0) for treatment t:

$$U(t, P, A) := \max_{(c_1, c_2)} \{(1 - \pi)u(c_1, 1) + \pi u(c_2, t)\}$$
$$s.t. \ (1 - \pi)c_1 + \pi(c_2 + P) = (1 - \pi)A + \pi tA,$$

where $U : \Re_{++} \times [0, (1/\pi - (1 - t))A) \times \Re_{++} \rightarrow \Re$. The individual is offered a positive level of treatment that may, for the purpose of defining and analyzing the U function, exceed one. As specified, the maximum price she is able to pay for this level of t is given by $(1/\pi - (1 - t))\,A$, hence the price of treatment, P, satisfies: $0 \leq P < (1/\pi - (1 - t))\,A$. Naturally, the higher the level of inherent ability (A), the higher the price she can pay for treatment. Also, the higher the probability of falling ill, the less she is able to pay for treatment.

To investigate the optimization problem, form the corresponding Lagrangian:

$$\mathcal{L}(c_1, c_2, \lambda; t, P, A) = (1-\pi)u(c_1, 1) + \pi u(c_2, t) + \lambda[(1-(1-t)\pi)A - (1-\pi)c_1 - \pi(c_2 + P)].$$

Given our assumptions on u, the first-order necessary conditions (FOCs) give the consumption demand function in each of the two states of the world:

$$(\mathbf{c}_1(t, P, A), \mathbf{c}_2(t, P, A)) \in \Re^2_{++} ,$$

satisfying

$$u_c(\mathbf{c}_1(t, P, A), 1) = u_c(\mathbf{c}_2(t, P, A), t) = \lambda \tag{2.3}$$

and the constraint (*cf.* equation (2.1)). As shown above, optimal consumption in each of the two states of the world is a function of treatment (*i.e.*, the degree of recovery in state 2), price of treatment (P), and income (A). Equation (2.3) implies that, in optimum, the individual's marginal utility of consumption is equal in the two states.

The utility function U can now be written:

$$U(t, P, A) = (1-\pi)u(\mathbf{c}_1(t, P, A), 1) + \pi u(\mathbf{c}_2(t, P, A), t) .$$

We have that U is strictly increasing in t, strictly decreasing in P, and strictly increasing in A. Hence, we can define an indifference curve in (t, P)–space, call it $\mathcal{P}(t, A; \bar{t}, \bar{P})$, going through $(\bar{t}, \bar{P})$ and showing combinations of t and P yielding a constant level of utility. Hence, $U(t, P, A)$ is equal to $U(\bar{t}, \bar{P}, A)$ if and only if $P = \mathcal{P}(t, A; \bar{t}, \bar{P})$. The slope of the indifference curve is given by:

$$\frac{\partial \mathcal{P}(t, A; \bar{t}, \bar{P})}{\partial t} = -\frac{\frac{\partial U}{\partial t}}{\frac{\partial U}{\partial P}} = -\frac{\frac{\partial \mathcal{L}}{\partial t}}{\frac{\partial L}{\partial P}} = \frac{\pi(u_h(c_2, t) + \lambda A)}{\pi \lambda} = \frac{u_h(c_2, t)}{u_c(c_2, t)} + A , \tag{2.4}$$

where the second equality follows from the envelope theorem, and the fourth equality is implied by equation (2.3). This means that the *marginal willingness to pay for treatment* is equal to the marginal rate of substitution of consumption for health plus the additional earnings capacity generated by treatment. Since, by construction, $\mathcal{P}(t, A; \bar{t}, \bar{P})$ is the indifference curve going through $(\bar{t}, \bar{P})$, it follows that:

$$\frac{\partial \mathcal{P}(\bar{t}, A; \bar{t}, \bar{P})}{\partial A} = 0.$$

I.e., even though the indifference curve through $(\bar{t}, \bar{P})$ will shift when ability increases, it will still go by $(\bar{t}, \bar{P})$. Since $u_{cc} < 0$, $u_{ch} > 0$, and $\partial \mathbf{c}_2/\partial A > 0$, it follows that the shift will be an anti-clockwise rotation, with $(\bar{t}, \bar{P})$ as fixed point, so that the slope at $(\bar{t}, \bar{P})$, $\partial \mathcal{P}(\bar{t}, A; \bar{t}, \bar{P})/\partial t$, increases:

$$\frac{\partial \mathcal{P}(\bar{t}, A; \bar{t}, \bar{P})}{\partial A \partial t} = \frac{\partial}{\partial A}\left[\frac{u_h(\mathbf{c}_2(\bar{t}, \bar{P}, A), \bar{t})}{u_c(\mathbf{c}_2(\bar{t}, \bar{P}, A), \bar{t})} + A\right] > 1.$$

Hence, the slope of an indifference curve through any point $(\bar{t}, \bar{P})$ is increasing in ability A. We will refer to this as the *single-crossing property*. The single-crossing property is illustrated in Figure 2.1 for two different values of ability, $A_l < A_h$, where l and h denote low and high ability, respectively. As illustrated, the slope of the indifference curve going through $(\bar{t}, \bar{P})$ in (t, P)-space is steeper if ability is high than if it is low.

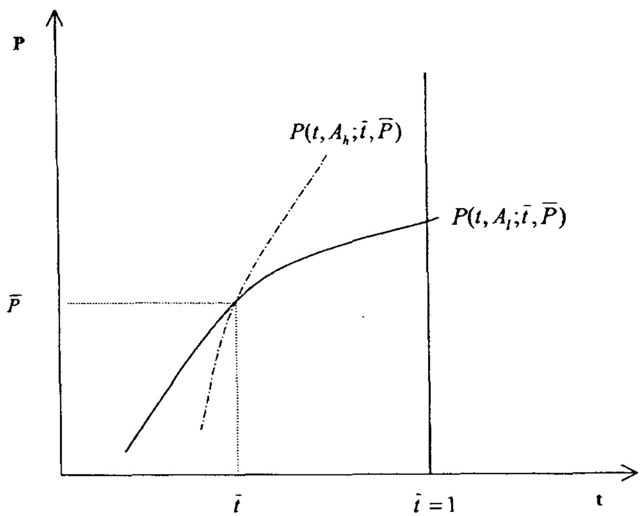

Figure 2.1: The single-crossing property.

It remains to show that $\mathcal{P}(t, A; \bar{t}, \bar{P})$ is a strictly concave function of t, so that an individual being faced with the possibility of purchasing treatment t at cost $P = tC$ constrained by $t \leq 1$, will have a unique level of treatment maximizing $U(t, tC, A)$. This will be shown by demonstrating that, if (t', P') and (t'', P'') are different combinations yielding the same utility level given A, then any interior convex combination

$$(t, P) = (\alpha t' + (1-\alpha)t'', \alpha P' + (1-\alpha)P''), \quad 0 < \alpha < 1,$$

will yield a strictly higher utility level. Accordingly, assume $U(t', P', A) = U(t'', P'', A) = U(\bar{t}, \bar{P}, A)$, and introduce some notation:

$$\begin{array}{ll} c_1' = \mathbf{c}_1(t', P', A) & c_1'' = \mathbf{c}_1(t'', P'', A) \\ c_2' = \mathbf{c}_2(t', P', A) & c_2'' = \mathbf{c}_2(t'', P'', A). \end{array}$$

Also, let $(c_1, c_2) = (\alpha c_1' + (1-\alpha)c_1'', \alpha c_2' + (1-\alpha)c_2'')$. Since (c_1', c_2') satisfies the *ex-ante* budget constraint (2.1) given (t', P', A) and (c_1'', c_2'') satisfies constraint (2.1) given (t'', P'', A), it follows that also (c_1, c_2) satisfies constraint (2.1) given (t, P, A), implying that (c_1, c_2) is feasible. Hence,

$$\begin{aligned} U(t, P, A) &\geq (1-\pi)u(c_1, 1) + \pi u(c_2, t) \\ &> (1-\pi)[\alpha u(c_1', 1) + (1-\alpha)u(c_1'', 1)] \\ &\quad + \pi[\alpha u(c_2', t') + (1-\alpha)u(c_2'', t'')] \\ &= \alpha U(t', P', A) + (1-\alpha)U(t'', P'', A) = U(\bar{t}, \bar{P}, A) \end{aligned}$$

where the first inequality follows since (c_1, c_2) is feasible, and the second inequality follows since u is strictly concave. This means that $\mathcal{P}(t, A; \bar{t}, \bar{P})$ is a strictly concave function of t. We will henceforth refer to this property as *diminishing willingness to pay for treatment.*

2.4 Main Results

Due to the diminishing willingness to pay for treatment, an individual being faced with the possibility of purchasing treatment t at cost $P = tC$, constrained by $t \leq 1$, will have a unique level of treatment $\mathbf{t}(A)$ maximizing $U(t, tC, A)$. Furthermore, due to the single-crossing property, $\mathbf{t}(A)$ will (weakly) increase with A. In fact, whenever $0 < \mathbf{t}(A) < 1$, $\mathbf{t}(A)$ is determined by

$$\frac{\partial \mathcal{P}(\mathbf{t}(A), A; \mathbf{t}(A), \mathbf{t}(A)C)}{\partial t} = C.$$

I.e., marginal willingness to pay for treatment equals marginal cost of treatment (note that marginal cost equals total cost). It follows that $\mathbf{t}(A)$ is a strictly increasing function of A when $0 < \mathbf{t}(A) < 1$.

We have that $t(A) = 1$ for all $A \geq A^*$, where A^* satisfies that the indifference curve through $(1, C)$ has slope C, so that unconstrained maximization of $U(t, tC, A^*)$ leads to $t = 1$. By the single-crossing property, A^* is unique. Hence, we can define A^* by

$$\frac{\partial \mathcal{P}(1, A^*; 1, C)}{\partial t} = C\,.$$

Since $\partial \mathcal{P}(t, A; \bar{t}, \bar{P})/\partial t > A$ for all values of t, $\bar{t}$, and $\bar{P}$, we have that $A^* < C$. Moreover, it follows from equation (2.3) and constraint (2.1) that $c_1 = c_2 = A - \pi C$ when $t = 1$ and $P = C$, implying that $t = 1$ is not feasible when $A < \pi C$. Finally, since $u_c(c, h) \to \infty$ as $c \downarrow 0$ whenever $h > 0$, it follows that $\partial \mathcal{P}(1, A; 1, C)/\partial t \to 0$ as $A \downarrow \pi C$. This means that $A^* > \pi C$. Note that the individual may choose a level of treatment that enables her to fully recover (*i.e.*, $h_2 = h_1$) even if $A < C$, provided that A is greater than, or equal to, the insurance premium.

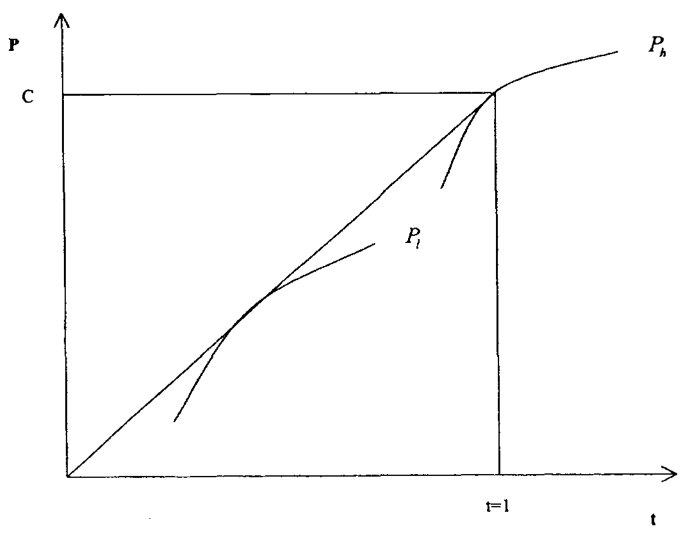

Figure 2.2: The optimal level of treatment.

The individual's optimal level of treatment is illustrated in Figure 2.2 for two different values of ability: $A_l < A^*$ and $A_h = A^*$, where l and h denotes low and high ability, respectively. As illustrated, a high-ability individual's indifference curve in (t,P)-space is tangent to the marginal-cost line for $t = 1$, while that of a low-ability individual is tangent to the marginal-cost line for some $t \in (0, 1)$.

These observations partially prove the proposition below. The proposition states that the individual's utility is constant across states if she chooses full treatment: $u(c_1, h_1) = u(c_2, h_2)$. Her utility if ill is lower than that if well if she chooses less than full treatment (*i.e.*, $t < 1$), even though she receives an insurance indemnity: $u(c_1, h_1) > u(c_2, h_2)$. Moreover, with full treatment, she will not receive any cash payment in addition to what is required to pay for treatment, while in the case of partial treatment, her compensation will exceed the amount spent on medical treatment.

Proposition 1. *There exists a level of inherent ability if healthy, A^*, where $\pi C < A^* < C$, such that:*

1. *If the individual's ability when healthy, A, is high, in particular, if $A \geq A^*$, then her optimal level of treatment is equal to one and does not vary with A: $\mathbf{t}(A) = 1$. Moreover, her level of consumption is identical in the two states: $\mathbf{c}_1(1, C, A) = \mathbf{c}_2(1, C, A) = A - \pi C$, as is her utility: $u(c_1, h_1) = u(c_2, h_2) = u(A - \pi C, 1)$. Her insurance coverage is in the form of medical treatment only.*

2. *If, however, the individual's ability when healthy, A, is low, in particular, if $0 < A < A^*$, then her optimal level of treatment is positive but less than one, $0 < \mathbf{t}(A) < 1$, and increasing in A: $\partial \mathbf{t}(A)/\partial A > 0$. Moreover, her level of consumption if ill is lower than if healthy: $\mathbf{c}_2(\mathbf{t}(A), \mathbf{t}(A)C, A) < \mathbf{c}_1(\mathbf{t}(A), \mathbf{t}(A)C, A)$, and her utility if ill is lower than if healthy: $u(c_2, h_2) < u(c_1, h_1)$. Her insurance coverage is partly in the form of medical treatment and partly in the form of cash.*

Proof. Part (1). Given the observations prior to the Proposition, it remains to show that the individual's utility is constant across states, and that she has insurance coverage in the form of medical treatment only. Constant utility across states follows since $c_1 = c_2 = A - \pi C$ and $h_1 = h_2 = 1$, implying that $u(c_1, h_1) = u(c_2, h_2) = u(A - \pi C, 1)$. Since cash payment equals $c_2 - tA + A - c_1$ (*cf.* footnote 6), it follows that cash payment is zero.

Part (2). By the definition of A^*, $0 \leq \mathbf{t}(A) < 1$ whenever $0 < A < A^*$. Moreover, since $u_h(c, h) \to \infty$ as $h \downarrow 0$ whenever $c > 0$, and $u_c(c, h) \to \infty$ or $u_h(c, h) \to \infty$ as $c \downarrow 0$ and $h \downarrow 0$, it follows from $A > 0$ and equation (2.4) that $\partial \mathcal{P}(t, A; t, tC)/\partial t > C$ if t is sufficiently small; hence, $\mathtt{t}(A) > 0$. Now, the single-crossing property implies that $d\mathtt{t}(A)/dA > 0$. From equation (2.3) and the properties of u, it follows that $c_1 > c_2$, since $h_1 = 1$, and $h_2 = \mathtt{t}(A) < 1$. This in turn implies that $u(c_1, h_1) > u(c_2, h_2)$. To show that cash payment is positive, *i.e.*, that $c_2 - tA + A - c_1 > 0$, we start out with the observation that $\mathtt{t}(A)$ is determined by $\partial \mathcal{P}(\mathtt{t}(A), A; \mathtt{t}(A), \mathtt{t}(A)C)/\partial t = C$ whenever $0 < \mathtt{t}(A) < 1$. Thus, marginal willingness to pay for treatment equals cost of treatment: $u_h(c_2, t)/u_c(c_2, t) + A = C$. In the hypothetical case where treatment were available also if healthy, or inversely, where health could be sold at price $C - A$, the access to actuarially fair insurance would imply the same level of health in both states. Since this is not the case, it is a binding constraint that health if healthy cannot be sold at price $C - A$, implying that marginal willingness to pay for health if healthy is less than $C - A$: $u_h(c_1, 1)/u_c(c_1, 1) < C - A = u_h(c_2, t)/u_c(c_2, t)$. Hence, effectively, the relative price of health in terms of consumption is lower if healthy than if ill. Combining this finding with constraint (2.1) and condition (2.2), and recalling that $u_{cc} < 0$ and $u_{ch} > 0$, we have that $c_1 < A - \pi tC$, and $c_2 > tA - \pi tC$. This in turn means that $c_1 - A < c_2 - tA$, or $c_2 - tA + A - c_1 > 0$. ■

2.5 A Special Case

The following Cobb-Douglas function is a Bernoulli utility function that satisfies all assumptions listed in Section 2.2:

$$u(c, h) = c^r h^s, \text{ with } r > 0,\ s > 0 \text{ and } r + s < 1.$$

With this function, it is possible explicitly to calculate A^*. We have that

$$\begin{aligned} \frac{\partial \mathcal{P}(1, A; 1, C)}{\partial t} &= \frac{u_h(c_2, 1)}{u_c(c_2, 1)} + A \\ &= \frac{u_h(A - \pi C, 1)}{u_c(A - \pi C, 1)} + A = \frac{s}{r}(A - \pi C) + A, \end{aligned}$$

where the second equality follows since $c_2 = A - \pi C$ when $t = 1$ and $P = C$, and the third equality follows since

$$\frac{u_h(c, h)}{u_c(c, h)} = \frac{s}{r} \cdot \frac{c}{h}$$

when u is given by the Cobb-Douglas function above. Since A^* is defined by $\partial \mathcal{P}(1, A^*; 1, C)/\partial t = C$, we can find A^* by solving

$$\frac{s}{r}(A^* - \pi C) + A^* = C,$$

which implies that

$$A^* = \frac{r + \pi s}{r + s} C.$$

Thus, A^* is increasing in the probability of falling ill, π, and in the cost of treatment, C.

2.6 *Ex-Post* Moral Hazard

We have, in Proposition 1, shown that an ill individual with ability A lower than the critical level A^* receives partial treatment, $\mathbf{t}(A) < 1$, and, in addition, a positive cash compensation:

$$\mathbf{c}_2(A) - \mathbf{t}(A)A + A - \mathbf{c}_1(A),$$

where we from now on write $\mathbf{c}_1(A) = \mathbf{c}_1(1, C, A)$ and $\mathbf{c}_2(A) = \mathbf{c}_2(\mathbf{t}(A), C, A)$. If, contrary to what we have assumed in the formal analysis, the two states (healthy/ill) are *not observable*, the availability of such a disability insurance will tempt the individual to claim that she has fallen ill, although she is in fact in good health. In this section, we show that our analysis goes through even if we allow for such *ex-post* moral hazard,[7] provided that:

- the cash compensation is paid only in combination with treatment, and
- the disutility of receiving treatment while healthy is sufficiently great.

Consequently, in order to prevent the individual from falsely claiming to be ill, she should suffer a loss in expected utility from undergoing redundant medical treatment. Moreover, the disutility should at least balance the gains in expected utility from masquerading as ill.

Denote by $v(c, t)$ the (direct) disutility of receiving treatment t while healthy and consuming c. Assume that v satisfies, $\forall (c, t) \in \Re^2_{++}$, $v(c, t) > 0$

[7] *Ex-post* moral hazard refers to the effect of insurance on insured individuals' incentives to reveal their true health state (*i.e.*, the insured individual knows the state of the world, while the insurer does not, or verification of health state is too costly for the insurer). The analysis of *ex-post* moral hazard was pioneered by Spence and Zeckhauser (1971). For a textbook exposition, see Zweifel and Breyer (1997, Chapter 6.5).

and $v_t \geq 0$.[8] To ensure that a healthy individual with ability A will not falsely claim to be ill, the following inequality must hold:

$$\begin{aligned}&(1-\pi)u(\mathbf{c}_1(A),1)+\pi u(\mathbf{c}_2(A),\mathbf{t}(A)) \geq \\ &(1-\pi)\big(u(c_2',1)-v(c_2',\mathbf{t}(A'))\big)+\pi u(\mathbf{c}_2(A')+\mathbf{t}(A')(A-A'),\mathbf{t}(A'))\,,\end{aligned} \tag{2.5}$$

where $c_2' = \mathbf{c}_2(A') + (A - \mathbf{t}(A')A')$ is the consumption that a healthy individual with ability A will receive having purchased the optimal insurance contract of an individual with ability A' and masquerading as ill, while $\mathbf{c}_2(A') + \mathbf{t}(A')(A - A')$ is the consumption that an ill individual with ability A will receive having purchased the optimal insurance contract of ability A' and truly claiming to be ill. An individual with ability A higher (lower) than A' generates higher (lower) earnings and can, therefore, sustain a higher (lower) level of consumption than can an individual with ability A'.

In condition (2.5), we must consider that the individual with ability A may *not only misrepresent her health state but also her ability*, in order to receive the higher cash compensation designed to be paid to an individual with a different ability A'. However, we know that an individual with ability A will not misrepresent her ability unless she intends to misrepresent her health state, since $(\mathbf{c}_1(A), \mathbf{c}_2(A), \mathbf{t}(A))$ maximizes expected utility over all triples (c_1, c_2, t) satisfying the constraint in equation (2.1); *i.e.*,

$$\begin{aligned}&(1-\pi)u(\mathbf{c}_1(A),1)+\pi u(\mathbf{c}_2(A),\mathbf{t}(A)) \geq \\ &(1-\pi)u(c_1',1)+\pi u(\mathbf{c}_2(A')+\mathbf{t}(A')(A-A'),\mathbf{t}(A'))\,,\end{aligned} \tag{2.6}$$

where $c_1' = \mathbf{c}_1(A') + (A - A')$ is the consumption that a healthy individual with ability A will receive having purchased the optimal insurance contract of ability A' and *not* masquerading as ill.

Notice that the increase in consumption that an individual with ability A, having purchased the optimal insurance contract of ability A', attains by masquerading as ill, $c_2' - c_1'$, is equal to the cash compensation designed to be paid to an individual with ability A':

$$c_2' - c_1' = \mathbf{c}_2(A') - \mathbf{t}(A')A' - \mathbf{c}_1(A') + A'.$$

[8]The disutility $v(c, t)$ is to be subtracted from the utility derived from consumption c when healthy and *not* receiving unnecessary treatment. Since this utility is a function of both consumption and treatment, so is also the disutility.

Hence, since, by equation (2.6), it is a sufficient condition for equation (2.5) to be satisfied that

$$u(c_2', 1) - u(c_1', 1) \leq v(c_2', \mathrm{t}(A')), \tag{2.7}$$

the *ex-post* moral hazard of an healthy individual masquerading as ill in order to obtain cash compensation does *not* constitute an incentive problem if the additional utility obtained from the cash compensation does not exceed the disutility from undergoing redundant treatment when healthy.

The *ex-post* moral hazard problem associated with cash compensation is solved through the *integration of treatment for illness and payments for disability*. Hence, our analysis presents an argument in favor of such integration. Furthermore, the lack of such integration can help explain why private markets for disability insurance are of little empirical significance.

In reality, the true health state is simple to verify for some medical conditions, while difficult for others; moreover, the disutility from receiving unnecessary treatment is significant for some types of conditions, while insignificant for others. Hence, the severity of the *ex-post* moral hazard problem discussed in this section varies according to the medical conditions.

2.7 Discussion

Our focus of attention has been on how an individual's inherent ability at full functionality (*i.e.*, when healthy) influences her *ex ante* choice of insurance contract and her optimal level of coverage. Insurance allows the individual to allocate income between the two states of the world *prior* to knowing which state has occurred. Moreover, it enables her to achieve her optimal distribution of income on consumption and health when ill. Since the individual is assumed to have perfect foresight, her optimal allocation *ex ante* will be optimal also *ex post*.

The novelty of this paper is the integration of what is usually thought to be different types of insurance, namely insurance against the risk of losing income due to (permanently) reduced health (ability) and insurance against the risk of incurring medical expenditures. We argue that a health insurance should offer a hedge against both potential loss in income due to reduced health and potential expenditures on medical treatment. Contrary to what is assumed in most of the health insurance literature, we allow the individual to choose whether or not to restore health if ill. We show that the individual's marginal willingness to pay for treatment is increasing in ability, and we derive a critical level of ability at which an individual prefers to fully restore health if ill. If the individual's inherent level of ability is sufficiently low, then she chooses to restore health only partly, thus suffering a loss in ability. In order to obtain the preferred level of consumption if ill, she will hold a contract that in addition entitles her to a cash transfer in the event of illness. She will, however, suffer a loss in utility if ill. If, on the other hand, the individual has a sufficiently high level of inherent ability, then she prefers a contract that provides for complete medical treatment and thus full restoration of health.

It is of no importance, in a world of symmetric information, whether the coverage for medical costs is paid in cash intended to cover medical bills, or directly in the form of medical treatment. The individual's *ex-ante* decision concerning what level of treatment to choose is unaffected by the way she is compensated; the fundamental decision concerns to what extent health is to be restored.[9] However, as discussed in Section 2.6, if health state is not easily verifiable, then, due to the *ex-post* moral hazard problem, it becomes essential whether medical expenditures are compensated in cash or in kind. When information about health state is asymmetric, integration of a cash compensation of income loss and an in-kind compensation of medical expenditures reduces the individual's incentive to falsely claim to be ill.

Our findings are driven by the fact that the potential loss in income is larger, the higher the ability at full functionality. This implies that the prices of the two types of contracts differ depending on the individual's ability. The higher the potential income loss due to reduced ability, the cheaper is the contract offering indemnity in kind (*i.e.*, treatment), compared to a contract offering cash compensation of income loss. Thus, the cost-benefit ratio on medical treatment is lower the higher the level of ability at full functionality.

The preceding analysis is based on a highly stylized model. We largely disregard any informational constraints causing the familiar problems of

[9] Arrow (1963) mentions three different ways in which costs of medical care can be covered in an insurance contract: payment directly in medical services, a fixed cash payment, and a cash payment that covers the actual costs involved in providing the necessary medical treatment. In a perfect market, individuals wishing to receive medical treatment would be indifferent between a payment directly in the form of medical treatment and its cash equivalent.

adverse selection.[10] Furthermore, the individual is assumed to be fully informed *ex ante* about health consequences of illness as well as about treatment options. The insurers need not, however, know the individual's ability at full functionality, since, even without such knowledge, first-best, zero-profit insurance contracts lead in an undistorted way to self-selection. Transaction costs associated with gathering of information about relevant treatment options and treatment costs for all types of diseases are ignored. Moreover, we make a somewhat strong assumption regarding the treatment technology: the individual recovers instantly and proportionally to the level of treatment received, and treatment is effective with respect to health. However, in spite of these limitations, our model still provides interesting findings that may be subject to further studies.

Acknowledgments

We would like to thank Agnar Sandmo, Sören Blomquist, and participants at the 1999 EEA Conference and the 1999 Oslo workshop on health economics, for helpful comments on earlier versions. Part of Asheim's and Nilssen's research is financed by the Research Council of Norway through HERO - Health Economics Research in Oslo.

[10] In Asheim, *et al.* (2003), we expand our analysis to a situation of asymmetric information about the probability of illness.

References

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

Arrow, K.J. (1974). "Optimal Insurance and Generalized Deductibles." Scandinavian Actuarial Journal, 1-42; reprinted in: *Collected Papers of Kenneth J. Arrow. Vol. 3: Individual Choice under Certainty and Uncertainty.* Blackwell (1984), 212-260.

Asheim, G.B., A.W. Emblem and T. Nilssen. (2003). "Deductibles in Health Insurance: Pay or Pain?" Available at: http://home.hia.no/~annewe/Pay-Pain.pdf

Byrne, M.M. and P. Thompson. (2000). "Death and Dignity: Terminal Illness and the Market for Non-Treatment." *Journal of Public Economics* 76(2), 263-294.

Cook, P.J. and D.A. Graham. (1977). "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities." *The Quarterly Journal of Economics,* 91(1), 143-156.

Evans, W.N. and W.K. Viscusi. (1991). "Utility-Based Measures of Health." *American Journal of Agricultural Economics,* 73(5), 1422-1427.

Flochel, L. and B. Rey. (2002). "Health Care Demand and Health Insurance." Unpublished manuscript. Groupe d'Analyse et de Théorie Economique (GATE), Université Lumière Lyon II.

Frech, H.E. III. (1994). "State-Dependent Utility and the Tort System as Insurance: Strict Liability versus Negligence." *International Review of Law and Economics,* 14(3), 261-271.

Glied, S. (2000). "Managed Care." In A.J. Culyer and J.P. Newhouse (eds.), *Handbook of Health Economics*, Vol. 1A. Elsevier Science, 707-753.

Graboyes, R.F. (2000). "Our Money or Your Life: Indemnities vs. Deductibles in Health Insurance." Working Paper 00-4, Federal Reserve Bank of Richmond; available at http://www.rich.frb.org/pubs/-wpapers/ pdfs/wp00-4.pdf.

Haveman, R. and B. Wolfe. (2000). "The Economics of Disability and Disability Policy." In A.J. Culyer and J.P. Newhouse (eds.), *Handbook of Health Economics*, Vol. 1B. Elsevier Science, 995-1051.

Pauly, M.V. (1971). *Medical Care at Public Expense: A Study in Applied Welfare Economics.* Praeger Publishers.

Pauly, M.V. (1986). "Taxation, Health Insurance, and Market Failure in the Medical Economy." *Journal of Economic Literature,* 24(2), 629-675.

Schlesinger, H. (1984). "Optimal Insurance for Irreplaceable Commodities." *The Journal of Risk and Insurance,* 51(1), 131-137.

Spence, M. and R. Zeckhauser. (1971). "Insurance, Information, and Individual Action." *American Economic Review,* 61(2), 380-387.

Viscusi, W.K. and W.N. Evans. (1990). "Utility Functions That Depend on Health Status: Estimates and Economic Implications." *American Economic Review,* 80(3), 353-374.

Zeckhauser, R. (1970). "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory,* 2(1), 10-26.

Zweifel, P. and F. Breyer. (1997). *Health Economics.* Oxford University Press.

Essay 3

Deductibles in Health Insurance: Pay or Pain?

Abstract

We study a health-insurance market where individuals are offered coverage against both medical expenditures and losses in income due to illness. Individuals vary in their level of innate ability and their probability of falling ill. If there is private information about the probability of illness and an individual's innate ability is sufficiently low, we find that competitive insurance contracts yield screening partly in the form of co-payment, *i.e.*, a deductible in pay, and partly in the form of reduced medical treatment, *i.e.*, a deductible in pain.

3.1 Introduction

Individuals face an inevitable risk of falling ill. Illness causes individuals to suffer a loss in income earnings and to entail expenditures on medical treatment, moreover, it causes a loss in utility *per se*. Illness thus encompasses both monetary and non-monetary losses. Traditionally, individuals are thought to hedge against potential loss in income due to permanently impaired health by holding a disability insurance, and to hedge against potential medical expenditures by holding a medical insurance. We argue, however, that disability insurance and medical insurance offer different types of coverage against the same fundamental risk, namely the *risk of falling ill*. Hence, the concept of health insurance should be expanded so as to include both types of coverage, *i.e.*, coverage against medical expenditures and coverage against loss in earnings due to impaired health. One consequence of taking this wider view of health insurance is that insurees may prefer insurance contracts offering cash compensation in part, rather than full restoration of health, if ill.[1] In this paper, we discuss how this expanded concept of health insurance affects the performance of a private health-insurance market with *asymmetric information.* In particular: if information about the probability of falling ill is private to the individuals, will low-risk individuals get both less medical treatment and less cash compensation (*i.e.*, disability payment) than they would in a world of symmetric information?

In the literature on health insurance, focus is mainly placed on insurance against medical expenditures, assuming that illness entails only monetary losses (*e.g.*, medical expenditures and temporary loss in earnings). The desire to restore health is, by and large, taken for granted.[2] When non-monetary consequences of illness are taken into account, it is assumed that utility is state dependent and that health is either not restorable or irreplaceable.[3] In this paper, we allow for both monetary and non-monetary consequences of illness without imposing the assumption that health is irreplaceable. Rather, we assume that health if ill is endogenous: poor health is improved with certainty if individuals receive medical treatment, and treatment is assumed to be divisible. Health is thus insurable, and the non-monetary consequences of illness endogenous. Our model may thus provide a bridge between models taking only monetary consequences of illness into account, and models postulating that health is irreplaceable. Flochel and Rey (2002) supplement our analysis in that they, too, study individuals' demand for health insurance when utility is a function of both consumption and health. However, they do not allow labour earnings to depend on health state, hence their analysis does not include disability insurance. Moreover, their analysis takes place in a world of symmetric information about the probability of illness.

Our study is motivated by the empirical fact that health insurance and

[1]See our companion paper, Asheim *et al.* (2003).

[2]This view has been contested by authors like Byrne and Thompson (2000) and Graboyes (2000) who argue that, when the probability of successful treatment is small, the insured may be better off with cash compensation if ill, rather than going through with the treatment.

[3]For more on irreplaceable commodities, see Cook and Graham (1977) and Schlesinger (1984).

disability insurance are integrated in a number of real-life health care systems. This is particularly prevalent in European countries where health insurance with in-kind compensation and disability insurance with cash compensation typically form parts of a public tax-financed insurance system. It seems, moreover, to be a fact that high-income individuals hold insurance contracts entitling them to more extensive and higher quality health care services than do low-income individuals. Also, disability insurance seems to be of greater importance for low-income individuals than for high-income individuals. We do not aspire to provide a definite answer to why this is so, but hope to shed some light on the question by studying how individuals with different levels of earning capabilities and different probabilities of falling ill will choose to be compensated if ill. Our analysis takes place under the assumption that insurance is supplied in a private insurance market. However, if individuals' entitlements are commensurate with their contributions, *i.e.*, no redistribution, then our findings would apply also to the design of information-constrained Pareto efficient social-insurance contracts.

In our model, individuals differ along two dimensions: ability and risk. Information about ability is assumed to be symmetrically distributed, while information about risk (*i.e.*, the probability of falling ill) is private to the individual. Some individuals are robust: they have a low probability of falling ill. Others are frail: they have a high such probability. They have identical preferences over consumption and health. Individuals can recover partially or completely from an illness if they receive partial or complete medical treatment, respectively. Their problem is to decide *ex ante* how much income to transfer between the two possible states of the world, healthy or ill, and if ill, how to allocate income between consumption and health (*i.e.*, medical treatment). The insurance contracts are thus allowed to be specified along *three dimensions*: (i) consumption if healthy, (ii) consumption if ill, and (iii) treatment if ill. A proper analysis of the market for health insurance will have to take this feature of the contracts involved into account. Our analysis thus contrasts with the text-book setting where insurance usually covers medical expenditures only and individuals differ with respect to their risk of falling ill only.

When there is asymmetric information on risk, it follows from the analysis of Rothschild and Stiglitz (1976) that contracts can be differentiated in terms of the premium paid by the insured and the level of coverage provided;

see, *e.g.*, Zweifel and Breyer (1997, chs. 5 and 6) for a health-insurance exposition. Rothschild and Stiglitz show that, under certain conditions, a separating equilibrium exists in which each insurer offers a menu of insurance contracts. Frail individuals (*i.e.*, those with a high probability of falling ill) are offered full insurance coverage, while robust individuals (with a low such probability) are offered partial coverage only. In this standard set-up, partial coverage means a reduction in the compensation paid for medical expenditures, *i.e.*, a monetary deductible. As argued above, health insurance involves three-dimensional contracts and it is, therefore, necessary to extend the Rothschild-Stiglitz analysis to such a three-dimensional case. This is what we set out to do in the subsequent analysis.

The paper is organized as follows. In Section 3.2, we outline the model and characterize insurance contracts satisfying the self-selection constraints. Individuals' choice between consumption and medical treatment if ill is shown not to change relative to a situation with symmetric information. Insurers consequently do not have to place any restrictions on how individuals allocate the insurance indemnity if ill. Our three-dimensional problem therefore reduces to one of only two dimensions: (i) consumption if healthy, and (ii) consumption if ill. In Section 3.3, analogously to Rothschild and Stiglitz (1976), we find separating contracts in which frail individuals obtain their first-best level of coverage, while robust individuals are constrained in order for insurers to induce self-selection. In Section 3.4, we study the comparative statics with respect to individuals' level of innate ability and investigate what level of medical treatment and consumption these separating contracts lead to if illness occurs. Insurers screen individuals through deductibles, and we show that robust individuals face a deductible in their level of insurance coverage. In particular, robust individuals with a sufficiently high level of innate ability will have a deductible in the form of co-payment only, *i.e.*, deductible in pay. Robust individuals with a sufficiently low level of innate ability, on the other hand, will have part of the deductible in the form of reduced treatment, *i.e.*, deductible in pain. In contrast, frail individuals are offered their first-best optimal level of insurance coverage. In particular, frail individuals with a sufficiently high level of ability obtain complete treatment if ill, while those with a sufficiently low ability obtain their optimal level of (partial) treatment and their optimal level of disability payment if ill. Our findings and their implications are discussed in Section 3.5.

3.2 The Model

We model a setting where individuals have preferences over consumption (c) and health (h). Each individual faces uncertainty with respect to her state of health. There are two such (jointly exhaustive and verifiable) states. In state 1, the individual is healthy and has a level of health normalized to 1: $h_1 = 1$. In state 2, she is ill and suffers a complete loss in health: $h_2 = 0$. Health if ill may, however, be partly or fully restored with certainty through medical treatment $t \in [0, 1]$, and health improves instantly. Medical treatment leading to full recovery (*i.e.*, $t = 1$) costs C, while treatment at rate t costs tC. Treatment is thus measured as the fraction of total cost (C) spent on treatment.[4] Consequently, if the individual receives complete medical treatment, *i.e.*, $t = 1$, health if ill is fully restored: $h_2 = 1$. If no treatment is received, *i.e.*, $t = 0$, then health equals zero: $h_2 = 0$. Health if ill is thus given by $h_2 = t$. Consumption in the two states are denoted c_1 and c_2, respectively.

The individuals know their objective probability of falling ill, which is either high or low: The probability of falling ill is π_j for type-j individuals, where $j = F, R$ denotes frail (high-risk) and robust (low-risk) individuals, respectively, and $0 < \pi_R < \pi_F < 1$. Individuals maximize the von Neumann-Morgenstern expected utility function:

$$(1 - \pi_j)u(c_1, 1) + \pi_j u(c_2, t), \tag{3.1}$$

where $u(c, h)$ is a Bernoulli utility function. We assume that $u : \mathcal{R}^2_+ \rightarrow \mathcal{R}$ is twice continuously differentiable and strictly concave, and satisfies: $\forall (c, h) \in \mathcal{R}^2_{++}$, $u_c > 0$, and $u_h > 0$, where partial derivatives are denoted by subscripts. A strictly concave utility function implies that individuals are risk averse. We also assume that $u_{ch} > 0$. Hence, in addition to being an important factor of well-being in its own right, health affects an individual's ability to enjoy consumption. Moreover, $u_c(c, h) \rightarrow \infty$ as $c \downarrow 0$ whenever $h > 0$, and $u_h(c, h) \rightarrow \infty$ as $h \downarrow 0$ whenever $c > 0$, and $u_c(c, h) \rightarrow \infty$ or $u_h(c, h) \rightarrow \infty$ as $c \downarrow 0$ and $h \downarrow 0$. Note that our assumptions on u imply normality.[5]

[4] We assume that cost of curing an illness depends on characteristics of the illness, rather than characteristics of the individuals suffering from it.

[5] What is more, if we were to measure treatment in terms of utility, then it follows from the properties of u (*i.e.*, $u_h > 0$ and $u_{hh} < 0$) that the cost of treatment would be strictly convex.

Introducing health as an argument in the utility function bears resemblance to the literature on insurance with state-dependent utility; see, *e.g.*, Zeckhauser (1970), Arrow (1974), Cook and Graham (1977), Viscusi and Evans (1990), Evans and Viscusi (1991), and Frech (1994). In these discussions, health is an unalterable characteristic of the state and they therefore fit well with the insurance being purely a disability insurance. Our formulation can be seen as filling the gap between a pure disability insurance, where the reduced health following illness is inevitable and irreversible (and thus a formulation where a state-dependent utility is appropriate), and a pure medical insurance, where the insurance coverage is used to its full extent on medical treatment in order to restore health as much as possible to its pre-illness level.[6]

Individuals' innate capacity to generate income is henceforth referred to as 'ability' and denoted by A. We choose our monetary unity so that total earnings are equal to A when in good health. Since leisure is not included in the utility function (*cf.* equation (3.1)), we implicitly assume that leisure (and thus labour supply) is constant across individuals and states. In addition, we assume that individuals, by spending tC on treatment, will generate earnings equal to tA when ill. Hence, labour earnings in state 2 are proportional to treatment t. Information about an individual's A is *symmetrically distributed.*

Individuals are risk averse and, consequently, willing to insure against the uncertainty they face. Buying insurance is the only way that an individual can transfer income across the two states. Her budget constraints in states 1 and 2 are respectively given by:

$$c_1 + P = A \tag{3.2}$$

and

$$c_2 + P + tC = tA + I, \tag{3.3}$$

where P is the total insurance premium and I the insurance benefit.

The insurance market is competitive, with risk-neutral, profit-maximizing insurers earning zero expected profits. Insurance is thus offered at an actuarially fair premium:

$$P = \pi_j I,\ j = F, R. \tag{3.4}$$

[6]See also the above-mentioned contribution by Flochel and Rey (2002).

As is standard in the insurance literature, we assume that individuals cannot buy more than one insurance contract.[7] We assume that information about which disease an individual suffers from and, consequently, the associated costs of treatment, is known by both insurer and insuree. The insurers know the proportions of frail and robust individuals, while information about each individual's risk type is *asymmetric*. To simplify, we assume that individuals can influence neither the probability of falling ill nor the costs associated with the illness, *i.e.*, there is no moral hazard.

Combining equations (3.2)-(3.4), we get:

$$(1 - \pi_j)(A - c_1) + \pi_j (tA - tC - c_2) = 0, \, j = F, \, R, \tag{3.5}$$

which gives the insurers' zero-profit condition.

3.3 Separating Equilibrium

For reasons similar to those in Rothschild and Stiglitz (1976), a (pure-strategy) Nash equilibrium, if it exists, is separating.[8] The insurers face informational constraints in the design of insurance contracts. Indeed, they face a self-selection constraint in that frail individuals may masquerade as robust individuals in order to get insurance at a lower premium. In order to induce individuals to reveal their probabilities of falling ill, insurers offer a menu of insurance contracts from which individuals can choose. Each contract is designed with a particular type of individual in mind and, since there are two risk types, only two types of contracts are offered. Individuals can be characterized by their *ex-ante* choices of consumption in the two states, as well as their levels of medical treatment if ill. Insurers thus have to design contracts in three dimensions, *i.e.*, a contract for type j is: $\{c_{1j}, c_{2j}, t_j\}$, $j = F, R$. In order to ensure that a pure-strategy equilibrium exists, we assume that there are relatively few robust individuals.

We first characterize the contract intended for *frail* individuals. As shown in the appendix, robust individuals do not wish to masquerade as frail individuals and we can, therefore, ignore the self-selection constraint

[7]For a discussion of an asymmetric-information market where consumers are allowed to have transactions with more than one firm, see Beaudry and Poitevin (1995).

[8]Rothschild and Stiglitz (1976) show that, if a Nash equilibrium exists, it is never a pooling equilibrium since pooling contracts are not robust to competition.

on robust individuals. The contract offered frail individuals constitutes the solution to the following program:

$$\max_{c_1,c_2,t} (1-\pi_F)u(c_1,1)+\pi_F u(c_2,t)$$

subject to:

$$\begin{aligned}(1-\pi_F)(A-c_1)+\pi_F(tA-tC-c_2) &= 0,\\ t &\leq 1.\end{aligned}$$

The first constraint is the insurers' zero-profit condition. The second constraint reflects that individuals cannot more than fully restore health. (In addition, there is a non-negativity constraint on t that never binds because of the assumptions we have made on u.) Let the multipliers associated with the constraints be respectively μ_F and ϕ_F, and write the Lagrangian as follows:

$$\begin{aligned}\mathcal{L} &= (1-\pi_F)u(c_1,1)+\pi_F u(c_2,t)\\ &\quad +\mu_F\left((1-\pi_F)A+\pi_F tA-(1-\pi_F)c_1-\pi_F(c_2+tC)\right)+\phi_F(1-t).\end{aligned}$$

The Lagrangian first-order necessary conditions are:

$$\frac{\partial\mathcal{L}}{\partial c_1} = (1-\pi_F)u_c(c_{1F},1)-\mu_F(1-\pi_F)\leq 0 \tag{3.6}$$

$$\frac{\partial\mathcal{L}}{\partial c_2} = \pi_F u_c(c_{2F},t_F)-\mu_F\pi_F\leq 0 \tag{3.7}$$

$$\frac{\partial\mathcal{L}}{\partial t} = \pi_F u_h(c_{2F},t_F)+\mu_F\pi_F(A-C)-\phi_F\leq 0. \tag{3.8}$$

Since c_1, c_2, and t are positive (by the properties of u), it follows from the complementary-slackness conditions that the marginal conditions will hold as equalities. From equations (3.6) and (3.7), we get:

$$u_c(c_{1F},1)=u_c(c_{2F},t_F), \tag{3.9}$$

i.e., frail individuals' marginal utility from consumption is equal across states. Combining equations (3.7) and (3.8), we find:

$$\frac{u_h(c_2,t_F)}{u_c(c_2,t_F)}+A=C+\frac{\phi_F}{\mu_F}\frac{1}{\pi_F}. \tag{3.10}$$

The left-hand side here is the *marginal willingness to pay for treatment* and is given by the sum of the marginal rate of substitution of consumption for health, $u_h(c_2, t_F)/u_c(c_2, t_F)$, and the additional earnings capacity generated by a marginal increase in treatment, A. Hence, frail individuals choose consumption and treatment if ill such that marginal willingness to pay for treatment equals the marginal cost of treatment plus the marginal imputed costs associated with the treatment constraint. The insurers' zero-profit condition obviously binds, hence $\mu_F > 0$. The marginal imputed costs incurred by restraining the individuals' level of treatment, t_F, is given by ϕ_F. According to the complementary-slackness condition, this Lagrange multiplier may take a positive or zero value. If $t_F < 1$, then $\phi_F = 0$, and it follows that:

$$\frac{u_h(c_{2F}, t_F)}{u_c(c_{2F}, t_F)} + A = C \text{ if } t_F < 1.$$

Note that there are no distortions in the contract designed for frail individuals, since self-selection constraints have no effect. The equilibrium insurance contract offered to frail individuals is, therefore, *first-best efficient.*[9]

Next, we identify the contract intended for *robust* individuals. In this case, the introduction of a self-selection constraint on frail individuals is necessary since they have an incentive to masquerade as robust individuals in order to obtain lower premium. The equilibrium contract for robust individuals solves the following program:

$$\max_{c_1, c_2, t} (1 - \pi_R)u(c_1, 1) + \pi_R u(c_2, t)$$

subject to:

$$\begin{aligned} (1-\pi_R)(A - c_1) + \pi_R(tA - tC - c_2) &= 0, \\ (1-\pi_F)u(c_1, 1) + \pi_F u(c_2, t) &\leq (1-\pi_F)u(c_{1F}, 1) + \pi_F u(c_{2F}, t_F) \\ t &\leq 1. \end{aligned}$$

The first and third constraint are as above. The second one is the self-selection constraint: Frail individuals should not wish to pretend being robust. Thus, the contract intended for robust individuals must ensure that frail individuals do not derive higher utility from choosing this contract rather than the contract intended for them. The self-selection constraint will always bind.

[9]For more on first-best contracts, see Asheim, *et al.* (2003), where the probability of falling ill is assumed to be public information.

With Lagrangian multipliers for the three constraints being denoted μ_R, λ_R, and ϕ_R, the Lagrangian is:

$$\begin{aligned}\mathcal{L} &= (1-\pi_R)u(c_1,1)+\pi_R u(c_2,t)\\ &+\mu_R\left((1-\pi_R)A+\pi_R tA-(1-\pi_R)c_1-\pi_R(c_2+tC)\right)\\ &+\lambda_R\left((1-\pi_F)u(c_{1F},1)+\pi_F u(c_{2F},t_F)\ -(1-\pi_F)u(c_1,1)-\pi_F u(c_2,t)\right)\\ &+\phi_R\left(1-t\right).\end{aligned}$$

Again, since c_1, c_2, and t are positive, it follows from the complementary-slackness conditions that the marginal conditions will hold as equalities. Thus, the first-order necessary conditions are:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial c_1} &= (1-\pi_R)u_c(c_{1R},1)-\mu_R(1-\pi_R)\\ &- \lambda_R(1-\pi_F)u_c(c_{1R},1)=0 \qquad (3.11)\\ \frac{\partial\mathcal{L}}{\partial c_2} &= \pi_R u_c(c_{2R},t_R)-\mu_R\pi_R-\lambda_R\pi_F u_c(c_{2R},t_R)=0 \qquad (3.12)\\ \frac{\partial\mathcal{L}}{\partial t} &= \pi_R u_h(c_{2R},t_R)+\mu_R\left(\pi_R A-\pi_R C\right)\\ &- \lambda_R\pi_F u_h(c_{2R},t_R)-\phi_R=0. \qquad (3.13)\end{aligned}$$

Rearranging equations (3.11) - (3.13), we get:

$$1-\frac{\mu_R}{u_c(c_{1R},1)}-\frac{(1-\pi_F)}{(1-\pi_R)}\lambda_R=0 \qquad (3.14)$$

$$1-\frac{\mu_R}{u_c(c_{2R},t_R)}-\frac{\pi_F}{\pi_R}\lambda_R=0 \qquad (3.15)$$

$$1+\frac{\mu_R(A-C)}{u_h(c_{2R},t_R)}-\frac{\pi_F}{\pi_R}\lambda_R-\frac{\phi_R}{\pi_R u_h(c_{2R},t_R)}=0. \qquad (3.16)$$

From equations (3.14) and (3.15), we have:

$$\frac{1}{u_c(c_{1R},1)}-\frac{1}{u_c(c_{2R},t_R)}=\frac{\lambda_R}{\mu_R}\frac{\pi_F-\pi_R}{\pi_R(1-\pi_R)}, \qquad (3.17)$$

which implies that marginal utility of consumption differs across states for robust individuals. In particular,

$$u_c(c_{2R},t_R)>u_c(c_{1R},1). \qquad (3.18)$$

In addition, from equations (3.15) and (3.16), we get:

$$\frac{u_h(c_2,t_R)}{u_c(c_2,t_R)}+A=C+\frac{\phi_R}{\mu_R}\frac{1}{\pi_R}. \qquad (3.19)$$

Thus, marginal willingness to pay for treatment equals marginal costs of treatment plus marginal imputed costs associated with the treatment constraint. The insurers' zero-profit condition binds, hence $\mu_R > 0$. The marginal imputed costs incurred from restraining the individuals' level of treatment, t_R, is ϕ_R. Again, according to the complementary-slackness condition, this Lagrange multiplier may take a positive or zero value. If $t_R < 1$, then $\phi_R = 0$:

$$\frac{u_h(c_2, t_R)}{u_c(c_2, t_R)} + A = C \text{ if } t_R < 1. \tag{3.20}$$

We note that the allocation of income between consumption and health if ill is *first-best efficient.* The allocation of income between consumption if well and consumption if ill is, however, *not first-best efficient* (*cf.* equation (3.18)) and, subsequently, nor is the allocation of income on consumption if healthy and treatment if ill. Thus, robust individuals are restrained in their level of insurance coverage in order to induce self-selection by the frail ones.

It follows from the above discussion that neither frail nor robust individuals' choice between consumption and treatment if ill is changed compared to the case of symmetric information. Consequently, our three-dimensional problem, *i.e.,* (i) consumption if healthy, (ii) consumption if ill, and (iii) treatment if ill, *reduces to one of only two dimensions*, namely that of (i) and (ii): how to allocate consumption across states. This implies that the level of medical treatment if ill does not have to be specified in the insurance contract.

Rationing of robust (low-risk) individuals as a way of separating risk-groups is, of course, in line with Rothschild and Stiglitz (1976). Frail individuals obtain their first-best allocation of consumption between the two states of the world. Robust individuals, on the other hand, are restrained in their level of insurance coverage compared to a situation with symmetric information and will have to accept a strictly positive deductible. The intriguing question is *whether this deductible is in pay or in pain, i.e.*, does the self-selection constraint restrict robust individuals' consumption if ill, their treatment if ill, or a bit of both? This is the topic of the next section.

3.4 Pay or Pain?

Individuals' decisions regarding the appropriate level of insurance coverage and the allocation of insurance indemnity if ill depend on their levels of innate ability. In a world of symmetric information, Asheim *et al.* (2003) show that individuals may *ex ante* prefer not to equalize utility across states so that $u(c_1, 1) > u(c_2, t)$. In particular, individuals with a sufficiently low level of innate ability prefer to not fully recover from an illness, but rather spend some of the indemnity on consumption. The implications of an individual's level of innate ability on her choice of insurance contract, in the present context of asymmetric information about the probability of falling ill, is discussed more closely in the following.

Proposition 1. *(i) If individuals have a high level of innate ability, in particular, if $A \geq C$, then, for a given positive level of insurance coverage, both robust and frail individuals choose complete treatment if ill: $t_R = t_F = 1$.*

(ii) If individuals have a low level of innate ability, in particular, if $A \leq \pi_R C$, then, for a given positive level of insurance coverage, both robust and frail individuals choose less than complete treatment if ill: $0 < t_R, t_F < 1$.

Proof. (i) It follows from our assumptions on u that $u_h(c_2, h_2)/u_c(c_2, h_2) > 0$. Hence, equations (3.10) and (3.19) can hold in the case when $A \geq C$ only if $\phi > 0$, which implies $t = 1$.

(ii) Note that $A \leq \pi_R C$ implies $A \leq \pi_F C$. Rewriting equation (3.5) as:

$$(1 - \pi_j) c_1 + \pi_j c_2 = (1 - \pi_j) A + \pi_j t (A - C), \; j = F, R, \qquad (3.21)$$

we see that the right-hand side is decreasing in t when $A \leq \pi_j C$. It follows from the properties of u that c_1, c_2, and t are positive. Suppose that $t = 1$. Now, the right-hand side of equation (3.21) reduces to: $A - \pi_j C$. Thus, with $A \leq \pi_j C$, there is nothing left for consumption, and the right-hand side will have to be increased through a reduction in t, that is, $t < 1$. ■

In light of this result, we assume in the subsequent analysis that individuals have one of two levels of ability: low (A_L) and high (A_H), such that $A_L \leq \pi_R C$ and $A_H \geq C$. It follows that $A_L/\pi_F < A_L/\pi_R < C \leq A_H$, since $0 \leq \pi_R < \pi_F < 1$.[10]

[10] At intermediate levels of ability, *i.e.*, where $A \in (\pi_R C, C)$, there is a possibility for cases where the constraint on treatment ($t \leq 1$) is binding for one of the risk types only. No extra insight would be gained from incorporating such hybrid situations into the analysis.

The implications of the ability level for the insurance contracts when there is asymmetric information on risk can be summarized as follows:

Proposition 2. *Equilibrium insurance contracts under asymmetric information are characterized as follows.*

(a) *Among high-ability individuals ($A_H \geq C$):*

(i) *Frail individuals face no deductibles, and receive their first-best level of insurance coverage. Their level of utility is constant across states, just like in the case of symmetric information.*

(ii) *Robust individuals are restrained in their level of insurance coverage and have to make a co-payment. Their deductible is in pay only.*

(iii) *In particular, frail individuals' marginal utility from consumption is equal across states, while that of robust individuals is not:* $0 = u_c(c_{2F}, t_F) - u_c(c_{1F}, 1) < u_c(c_{2R}, t_R) - u_c(c_{1R}, 1)$. *Both risk types choose complete medical treatment:* $t_R, t_F = 1$. *Hence,* $0 = c_{1F} - c_{2F} < c_{1R} - c_{2R}$.

(b) *Among low-ability individuals ($A_L \leq \pi_R C$):*

(i) *Frail individuals face no deductibles and achieve their first-best levels of consumption and medical treatment. However, even though not constrained in their level of insurance coverage, their utility is not equal across states; this corresponds to the case of symmetric information.*

(ii) *Robust individuals are restrained in their level of insurance coverage. They have part of the deductible in pain. Consequently, the indemnity provides for lower level of consumption and medical treatment if ill relative to a situation with symmetric information. Utility is, moreover, not equal across states.*

(iii) *In particular, both risk-types choose a lower level of consumption if ill than if healthy,* $0 < c_{1F} - c_{2F} < c_{1R} - c_{2R}$, *and choose less than complete treatment,* $0 < t_R < t_F < 1$.

Proof. We start out by establishing parts (a) and (b)(i). From the first-order conditions of the optimization problem in Section 3.2, we see that frail individuals' marginal utility from consumption is equal across states (*cf.* equation (3.9)), whereas robust individuals' marginal utility from consumption is not (*cf.* equation (3.17)). Frail individuals consequently receive their first-best level of insurance coverage, as stated in (a)(i) and (b)(i) of the Proposition, while robust individuals do not. For high-ability individuals, $t = 1$ by Proposition 1. Thus, by equation (3.18), for robust high-ability individuals, $u_c(c_{1R}, 1) < u_c(c_{2R}, t_R) = u_c(c_{2R}, 1)$ and $c_{1R} > c_{2R}$. Moreover, by part (a)(i) and Proposition 1, $u_c(c_{1F}, 1) = u_c(c_{2F}, t_F) = u_c(c_{2F}, 1)$ and $c_{1F} = c_{2F}$. This completes the proof of parts (a) and (b)(i).

Parts (b)(ii) and (b)(iii) remain to be established. For low-ability individuals (both robust and frail), $t < 1$ and $c_1 > c_2$ by Proposition 1. For robust low-ability individuals, $u_c(c_1, 1) = u_c(c_2, t)$ in first best and $u_c(c_1, 1) < u_c(c_2, t)$ under asymmetric information (*cf.* equation (3.18)). Since u is strictly concave, it follows from the zero-profit condition that c_{1R} is higher and $u(c_{2R}, t_R)$ is lower than they would have been in first-best. It now follows from the normality of c and h that both c_{2R} and t_R are lower than they would have been in first best. The consumption of frail low-ability individuals, c_{1F}, is, due to their higher cost of insurance, smaller than what robust low-ability individuals would have got in first-best, which in turn is smaller than c_{1R}, *i.e.*, $c_{1F} < c_{1R}$. It now follows from the frail low-ability individuals' self-selection constraint that $u(c_{2R}, t_R) < u(c_{2F}, t_F)$. By normality, this implies that $c_{2R} < c_{2F}$ and $t_R < t_F$. ■

We briefly restate the main results derived in the above analysis. Robust individuals are constrained in their level of insurance coverage because of the problem of self-selection. As a consequence, their allocation of income on consumption across states is not first-best efficient. Their level of consumption if ill is, indeed, distorted downwards causing consumption if ill to be less than consumption if healthy: $c_{2R} < c_{1R}$. Since robust high-ability individuals' marginal willingness to pay for treatment always exceeds marginal cost of treatment, we have that $t = 1$ for this group, entailing that their level of treatment is not altered relative to the situation of symmetric information. Indeed, health if ill is fully restored: $h_1 = h_2 = 1$. The deductible imposed on robust high-ability individuals is consequently in the form of reduced consumption if ill, *i.e.*, *deductible in pay*. Low-ability individuals' marginal willingness to pay for treatment, on the other hand, is shown to equal marginal cost of treatment only for levels of treatment less than one: $t < 1$. Hence, when faced with a reduction in the insurance indemnity, they will reduce both their level of consumption and their level of treatment, since both consumption and health are normal goods. The deductible imposed on robust low-ability individuals consequently takes the form of reduced consumption and reduced health, *i.e.*, *deductible in pay and pain.*

Considering the outcome for low-ability individuals, we note a sharp contrast between the cases of symmetric and asymmetric information. When insurers know each insuree's probability of illness, robust (*i.e.* low-risk) individuals get higher consumption and more medical treatment than frail individuals. This is turned around when information about this probability is private: In order to obtain self-selection, insurers offer robust individuals a contract that provides for lower consumption and less treatment if ill than do the contract offered frail individuals.

Note that the contracts described are *information-constrained Pareto-efficient.* Like in the Rothschild-Stiglitz model (see Crocker and Snow, 1985), the pure-strategy separating equilibrium is efficient whenever it exists.

3.5 Discussion

Analyzing a competitive health-insurance market under asymmetric information, we have identified separating insurance contracts that induce individuals to reveal information by means of deductibles. Our analysis takes place in a standard adverse-selection situation in which the insuree has more information about risk than does the insurer, the insurer offers a menu of contracts, and the insuree is restricted to buy all her insurance from the same insurer. However, the analysis deviates from the standard adverse selection situation in two related ways. Firstly, we assume that consumption and health are complements in utility. This implies that individuals may choose *not* to equalize utility across states in a world of symmetric information, and thus, that their optimal level of insurance coverage is even lower in a world of asymmetric information. Secondly, the consequences of the insured-against event are made endogenous: individuals can choose their level of recovery, and thus also their loss in income (*i.e.*, monetary loss) and utility (*i.e.*, non-monetary loss), if ill.

The *novelty* of this paper lies in the integration of medical insurance and disability insurance in a setting where adverse selection is a problem. By integrating the two types of insurance, we show that the separating scheme may involve two types of deductibles: a deductible in the form of reduced medical treatment and a monetary deductible. In fact, we find that low-ability robust individuals will have insurance contracts with both these types of deductibles, *i.e.*, deductibles in both pay (*i.e.*, in cash) and pain (*i.e.*, in kind).

Our findings may be of relevance to the practical design of health-insurance contracts. Indeed, one may observe empirically that insurance contracts specify both quantity and quality of care, rather than providing a cash compensation. Individuals will consequently have access to a predetermined level (or quality) of treatment, rather than just a cash payment. There are obviously many reasons for this, one of which being transaction costs associated with having to search for the appropriate supplier of medical treatment when ill. Thus, it is not counter-intuitive that individuals *ex ante* may find it optimal to specify their preferred level of treatment if ill, thus restricting the allocation of income when ill between consumption and health. If so, our analysis suggests that, under asymmetric information, robust individuals with low ability will achieve less treatment and less cash compensation than they would have achieved under symmetric information.

Acknowledgments

We are grateful to Vidar Christiansen and two anonymous referees of the International Journal of Health Care Finance and Economics for their valuable comments. We are also grateful for helpful comments from participants at the Nordic Health Economists' Study Group Meeting in Helsinki, and from seminar participants at Erasmus University in Rotterdam. Asheim and Nilssen gratefully acknowledge financial support for their research from the Research Council of Norway through HERO – the Health Economics Research Programme at the University of Oslo. Emblem gratefully acknowledges financial support from the SiS Programme at Agder University College.

References

Arrow, K.J. (1974). "Optimal Insurance and Generalized Deductibles." Scandinavian Actuarial Journal, 1-42; reprinted in: *Collected Papers of Kenneth J. Arrow, Vol 3: Individual Choice under Certainty and Uncertainty.* Oxford: Blackwell (1984), 212-260.

Asheim, G.B., A.W. Emblem and T. Nilssen (2003). "Health Insurance: Treatment *vs.* Compensation." Unpublished manuscript. University of Oslo and Agder University College. Available at: http://home.hia.no/~annewe/Treatment-Compensation.pdf

Beaudry, P. and M. Poitevin (1995). "Competitive Screening in Financial Markets When Borrowers Can Recontract." *Review of Economic Studies* 62, 401-423.

Byrne, M.M., and P. Thompson (2000). "Death and Dignity: Terminal Illness and the Market for Non-Treatment." *Journal of Public Economics* 76, 263-294.

Cook, P.J. and D.A. Graham (1977). "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities." *The Quarterly Journal of Economics*, 91, 143-156.

Crocker, K.J. and A. Snow (1985). "The Efficiency of Competitive Equilibria in Insurance Markets with Asymmetric Information." *Journal of Public Economics* 26, 207-219.

Evans, W.N and W.K. Viscusi (1991). "Utility-Based Measures of Health." *American Journal of Agricultural Economics* 73, 1422-1427.

Flochel, L. and B. Rey (2002). "Health Care Demand and Health Insurance." Unpublished manuscript. Groupe d'Analyse et de Théorie Economique (GATE), Université Lumière Lyon II.

Frech, H.E. III (1994). "State-Dependent Utility and the Tort System as Insurance: Strict Liability versus Negligence." *International Review of Law and Economics* 14, 261-271.

Graboyes, R.F. (2000). "Our Money or Your Life: Indemnities vs. Deductibles in Health Insurance." Working Paper 00-4, Federal Reserve Bank of Richmond.
Available at: http://www.rich.frb.org/pubs/wpapers/pdfs/wp00-4.pdf

Rothschild, M. and J.E. Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90, 629-649.

Schlesinger, H. (1984). "Optimal Insurance for Irreplaceable Commodities." *The Journal of Risk and Insurance* 51, 131-137.

Viscusi, W.K. and W.N. Evans (1990). "Utility Functions That Depend on Health Status: Estimates and Economic Implications." *American Economic Review* 80(3), 353-374.

Zeckhauser, R. (1970). "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory* 2, 10-26.

Zweifel, P. and F. Breyer (1997). *Health Economics.* Oxford University Press.

Appendix

We show here that robust individuals do not wish to masquerade as frail individuals, *i.e.*, that they will suffer a loss in expected utility if masquerading as frail:

$$(1-\pi_R)u(c_{1R},1)+\pi_R u(c_{2R},t_R) > (1-\pi_R)u(c_{1F},1)+\pi_R u(c_{2F},t_F). \quad (3.22)$$

We know from Section 3.2 that contracts for robust individuals are designed so that frail individuals are indifferent between masquerading or not:

$$(1-\pi_F)u(c_{1F},1)+\pi_F u(c_{2F},t_F) = (1-\pi_F)u(c_{1R},1)+\pi_F u(c_{2R},t_R). \quad (3.23)$$

Rewriting equation (3.23) such that:

$$\frac{(1-\pi_F)}{\pi_F}(u(c_{1R},1)-u(c_{1F},1)) = u(c_{2F},t_F)-u(c_{2R},t_R),$$

and observing that:

$$\frac{(1-\pi_R)}{\pi_R} > \frac{(1-\pi_F)}{\pi_F},$$

it follows that:

$$\frac{(1-\pi_R)}{\pi_R}(u(c_{1R},1)-u(c_{1F},1)) > u(c_{2F},t_F)-u(c_{2R},t_R).$$

Hence, inequality (3.22) holds.

Essay 4

Redistribution at the Hospital

Abstract

This paper studies redistribution by means of a public supply of medical treatment. We show that the government can redistribute income towards low-ability individuals in a world of asymmetric information by offering bundles of medical treatment and redistributive payment. If self-selection is a problem, then the separating scheme offers high-ability individuals complete treatment against a high payment, and low-ability individuals partial treatment against a low payment. In particular, the level of treatment offered low-ability individuals is distorted downwards.

4.1 Introduction

Medical treatment is but one example of a private good which in a number of countries is publicly supplied. Redistribution, along with paternalism, has been, and probably still is, a major reason why this is so. Public supply of treatment is usually financed by individuals contributing according to their (gross) income, either through general taxation or through earmarked contributions. The supply of treatment, on the other hand, is often provided according to individuals' need for treatment, rather than their contributions. Traditionally, this way of organising the financing and supply of medical treatment is thought to facilitate some kind of income redistribution towards low-income individuals. Underlying this way of reasoning is the assumption that individuals with identical medical needs prefer identical levels of medical treatment. We will, however, show that individuals with identical medical need for treatment may indeed prefer different levels of treatment and, consequently, different levels of recovery. In particular, individuals' preferred level of treatment is shown to be higher the higher their level of innate ability. This suggests that those contributing the most to the public health sector (i.e., high-income individuals) are also those utilising the services provided the most. The extent of redistribution may, therefore, not be as large as one would think.

The purpose of this paper is to identify Pareto-efficient bundles of medical treatment and payment that facilitate income redistribution towards low-ability individuals when information about individuals' level of ability (income) is private to the individuals. To simplify the analysis, we assume that this is the government's only means of redistribution, moreover, there is no private supply of medical treatment. We postulate that the government does *not* pursue a particular distribution of medical treatment (health) *per se*.[1]

There exists a fairly extensive literature on public provision of private goods as a means of redistribution in a second best world, some of which are: Blackorby and Donaldson (1988), Ireland (1990), Besley and Coate (1991), Epple and Romano (1996), Blomquist and Christiansen (1995, 1998), and Boadway, Marchand and Sato (1998).[2] The present paper is most closely related to that of Blackorby and Donaldson (1988, hereafter B-D) in that we use bundles of transfers and medical treatment as a means of redistribution. Whereas B-D study redistribution between ill and healthy individuals, i.e., redistribution *ex post*, we study redistribution between income groups *ex ante*, i.e., prior to knowing whether individuals are ill or not. Consequently, B-D study adverse selection problem in health types (i.e., whether ill or not), whereas we study adverse selection problem in ability types (i.e., whether high or low ability). B-D show that in a second-best world, publicly provided

[1] One may argue that health is a particular commodity for which the society has distinct egalitarian ambitions, i.e. that health is subject to what Tobin (1970) calls specific egalitarianism. The society may, for instance, aspire to achieve equality in health (e.g. as measured as quality adjusted life years over individuals' lifetime), or somewhat less ambitious: to reduce inequalities in health.

[2] Balestrino (1999) provides a survey of the literature on in-kind transfers in the presence of distortionary taxes.

medical treatment to the ill is 'overprovided' in the sense that their marginal willingness to pay for treatment is less than marginal cost of treatment. Here, we will show that subsidised treatment to the ill is 'underprovided,' i.e., their marginal willingness to pay is higher than marginal cost of treatment, if their level of innate ability is low. Our findings thus run counter to those of B-D, due to different informational assumptions.

The substance of our model is outlined as follows. Individuals may fall ill thus suffering a loss both in utility directly, and in ability to earn income. They can, however, buy medical treatment that restores health, and thus also ability, with certainty. Medical treatment at a given level is assumed to restore ability in the same proportion for all individuals. Benefits from treatment, and subsequently willingness to pay for treatment, is consequently higher the higher the level of innate ability. Since medical treatment is more valuable to high-ability individuals than to low-ability individuals, the government can separate the two types of individuals by offering two bundles, each specifying level of treatment and level of payment: one containing complete treatment and a high payment; type (i), and one containing partial treatment and a low payment; type (ii). If redistribution leads to a binding self-selection constraint, bundle (i) will not be distortionary, while bundle (ii) will be. In particular, the level of treatment provided in bundle (ii) will be distorted downwards, as this will be more costly to high-ability individuals than to low-ability individuals. The subsidised medical treatment allows low-ability individuals to have a higher level of consumption in the two possible states of the world (healthy or ill), and are, consequently, better off relatively to a situation without redistribution.

This paper is organised as follows. In Section 4.2 we derive the model and undertake a preliminary analysis, and in Section 4.3 we derive the government's Pareto-efficient menu of contracts facilitating redistribution towards low-ability individuals. Finally, we discuss our findings in Section 4.4.

4.2 Preliminary Analysis: Allocation of Income Across States of the World.[3]

In the subsequent analysis, we describe a representative individual's *ex ante* optimisation problem. The individual has preferences over consumption (c) and health (h): $u(c,h)$. There are two possible states of health: she may with probability $(1-\pi)$ be in good health: state 1, or with probability π $(0 < \pi \leq 0,5)$ be ill: state 2.[4] The two states are jointly exhaustive and verifiable. Information on risk is symmetrically distributed. Health in state 1 is normalised to 1: $h_1 = 1$, while health in state 2 is assumed to be zero: $h_2 = 0$. Health if ill can (with certainty) be partly or fully restored if medical treatment t $(0 \leq t \leq 1)$ is utilised. Treatment leading to complete recovery, i.e., $t = 1$, has a cost of production equal to C, while treatment leading to partial recovery has a cost tC. Health if ill is henceforth represented by t: $h_2 = t$. Consumption in state 1 and 2 is denoted c_1 and c_2, respectively.

The individual is an expected utility maximiser. Hence, her preferences are represented by the following von Neumann-Morgenstern utility function:

$$(1-\pi)u(c_1, 1) + \pi u(c_2, t). \tag{4.1}$$

We assume that u is twice continuously differentiable, strictly increasing and strictly concave. Also, $u_{ch} \geq 0$, where the partial derivative is denoted by subscript. Health is thus not only an important factor of well-being in its own right, but may also affect the individual's ability to enjoy consumption. It follows that c and h are normal goods. Furthermore, $u_c(c,h) \to \infty$ as $c \downarrow 0$ whenever $h > 0$ and $u_h(c,h) \to \infty$ as $h \downarrow 0$ whenever $c > 0$. Moreover, $u_c(c,h) \to \infty$ or $u_h(c,h) \to \infty$ as $c \downarrow 0$ and $h \downarrow 0$. Hence, she strictly desires a positive level of consumption and health.

The individual's level of innate ability (productivity) is given by A. *Information about innate ability is private* to the individual. If in good health, her level of ability is equal to A, while if ill, her level of ability is equal to tA. Earnings are assumed to be proportional to ability.

By the properties of u (strict concavity) it follows that the individual is risk averse. There exists a perfectly competitive private insurance market

[3]The analysis in this section is based on a somewhat modified version of a model developed in Asheim, Emblem and Nilssen (2000).

[4]We assume that the individual can neither influence the probability of falling ill, nor the costs associated with the illness, i.e. no moral hazard.

offering cash compensation if illness occurs. The compensation can be used to cover medical expenditures and partly compensate for (permanent) loss in income due to reduced ability, e.g. in the form of a disability payment. Insurance is offered at an actuarially fair premium πI, so that πI must be paid in both states in order to have coverage equal to I if ill. To insure is the only way the individual can transfer income across the two possible states of the world. Let P represent payment for treatment. Then her budget constraint in state 1 and 2 is: $c_1 + \pi I = A$ and $c_2 + \pi I + P = tA + I$, respectively. Her *ex ante* choice of c_1 and c_2 when the budget constraint is binding in both states of the world, is thus constrained by:

$$(1 - (1 - t)\pi) A = (1 - \pi)c_1 + \pi(c_2 + P). \tag{4.2}$$

The individual's expected utility maximising choice of consumption in the two states of the world is derived by maximising equation (4.1) subject to equation (4.2). The Lagrangian is given by:

$$\begin{aligned} \mathcal{L}(c_1, c_2, \lambda; t, P, A) &= (1 - \pi)u(c_1, 1) + \pi u(c_2, t) \\ &\quad + \lambda \left[(1 - (1 - t)\pi)A - (1 - \pi)c_1 - \pi(c_2 + P)\right]. \end{aligned}$$

It follows from the first-order conditions that:

$$u_c(c_1(t, P, A), 1) = u_c(c_2(t, P, A), t) = \lambda, \tag{4.3}$$

that is; marginal utility of consumption is equal across states. The consumption demand function in each state is given by: $c_1(t, P, A)$ and $c_2(t, P, A)$. The private insurance market consequently allows her to attain her *optimal distribution of consumption across states.* In the subsequent analysis, we will therefore concentrate on characterising her preferences if ill; in particular, her marginal willingness to pay for medical treatment at the expense of consumption if ill.

The individual's indirect utility function is given by:

$$V(t, P, A) = (1 - \pi)u(c_1(t, P, A), 1) + \pi u(c_2(t, P, A), t). \tag{4.4}$$

V is strictly increasing in t, strictly decreasing in P, and strictly increasing in A. We can therefore define a curve, call it $\mathcal{P}(t, A; \bar{t}, \overline{P})$, going through $(\bar{t}, \overline{P})$ in (t, P)-space and showing combinations of t and P yielding a constant level of utility. Accordingly, the utility, $V(t, P, A)$, of an individual with ability A facing (t, P) is equal to $V(\bar{t}, \overline{P}, A)$ if and only if $P = \mathcal{P}(t, A; \bar{t}, \overline{P})$. The indifference curve is upward sloping both in t and A. Moreover,

$$\frac{\partial \mathcal{P}(t, A; \bar{t}, \overline{P})}{\partial t} = -\frac{\frac{\partial V}{\partial t}}{\frac{\partial V}{\partial P}} = -\frac{\frac{\partial \mathcal{L}}{\partial t}}{\frac{\partial \mathcal{L}}{\partial P}} = \frac{\pi(u_h(c_2, t) + \lambda A)}{\pi \lambda}$$

$$= \frac{u_h(c_2, t)}{u_c(c_2, t)} + A\,, \tag{4.5}$$

where the second equality follows from the envelope theorem, and the fourth equality is implied by the first-order condition in equation (4.3). This means that the marginal willingness to pay for treatment equals the sum of consumption and production value[5] of health.

The indifference curve $\mathcal{P}(t, A; \bar{t}, \overline{P})$ can be shown to be a strictly concave function of t. If follows that the marginal willingness to pay for treatment is positive and decreasing in t. Moreover, since $u_{cc} < 0$, $u_{ch} \geq 0$ and $\partial c_2 / \partial A > 0$, then equation (4.5) implies that:

$$\frac{\partial \mathcal{P}(\bar{t}, A; \bar{t}, \overline{P})}{\partial t \partial A} = \frac{\partial}{\partial A}\left[\frac{u_h(c_2(\bar{t}, \overline{P}, A), \bar{t})}{u_c(c_2(\bar{t}, \overline{P}, A), \bar{t})} + A\right] > 1, \tag{4.6}$$

i.e., marginal willingness to pay for treatment is *increasing* in her level of innate *ability*. Graphically, this implies that the slope of an indifference curve through any point $(\bar{t}, \overline{P})$ is increasing in A. Indifference curves are consequently steeper the higher the level of innate ability and they cross only once, i.e., *single-crossing*.[6] The single-crossing property is illustrated in Figure 4.1 for two different values of ability: $A_L < A_H$, where L and H denotes low and high ability, respectively.

Holding t and A constant, then a higher payment for treatment implies that she will have to reduce her level of consumption in state 2. Marginal willingness to pay for treatment is consequently *decreasing* in P:

$$\frac{\partial \mathcal{P}(\bar{t}, A; \bar{t}, \overline{P})}{\partial t \partial \overline{P}} = \frac{\partial}{\partial \overline{P}}\left[\frac{u_h(c_2(\bar{t}, \overline{P}, A), \bar{t})}{u_c(c_2(\bar{t}, \overline{P}, A), \bar{t})} + A\right] < 0, \tag{4.7}$$

[5]As measured by the additional earnings capacity generated by treatment.

[6]The single-crossing property corresponds to the 'Agent Monotonicity condition' in the literature on income taxation (Seade, 1982) and the 'Spence-Mirrlees condition' in the literature on screening (Macho-Stadler & Pérez-Castrillo, 1997).

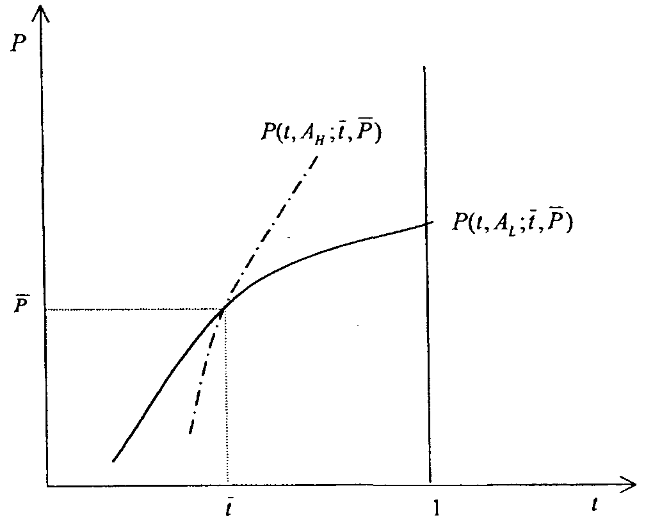

Figure 4.1: The single-crossing property.

hence, the income effect of a higher P is negative, as would be expected. It follows that the slope of the indifference curve through any point $(\bar{t}, \overline{P})$ is decreasing in payment.

The following result is useful for the main analysis. It shows that if innate ability is sufficiently high, then marginal willingness to pay for treatment exceeds cost of treatment, independently of payment. Moreover, if innate ability is sufficiently low, and payment equals cost of treatment, then marginal willingness to pay for treatment equals marginal cost only if treatment is partial.

Lemma. *If* $A \geq C$, then $\partial \mathcal{P}(t, A; t, P)/\partial t > C$ for all t and P. If $A \leq \pi C$, then $\partial \mathcal{P}(t, A; t, tC)/\partial t = C$ only if $0 < t < 1$.

Proof. Recalling that C is the production cost of complete treatment, then if $A \geq C$, it follows from equation (4.5) that the individual's marginal willingness to pay for treatment: $\partial \mathcal{P}(t, A; t, P)/\partial t$, is greater than C. Assume now that $A \leq \pi C$. If $t = 1$ and $P = C$, then we see from equation(4.2) and (4.3) that $c_1 = c_2 = A - \pi C$, implying that $t = 1$ is not feasible when $A \leq \pi C$. Moreover, if $A \downarrow \pi C$ and $t = 1$, then it follows from equation (4.2) that: $c_2 = (A - \pi C + c_1(\pi - 1))\frac{1}{\pi} \leq (A - \pi C)\frac{1}{\pi} \downarrow 0$. By the properties of

u[7], $\partial\mathcal{P}(1, A; 1, tC)/\partial t \to \pi C < C$ when $A \downarrow \pi C$. Consequently, if $A \leq \pi C$, then $\partial\mathcal{P}(t, A; t, tC)/\partial t = C$ only if $0 < t < 1$. ■

4.3 Main Analysis: Health vs. Subsidy

We now expand the analysis to include two types of individuals who are identical in all respects save their individual level of innate ability: they may either have a high ability: $A_H \geq C$, or a low ability: $A_L \leq \pi C$.[8] The number of individuals of each type is given by N_i, $i = H, L$. The government knows the proportion of each type of individuals, but cannot observe their identity. The individuals' indirect utility function is given by equation(4.4).

The government designs a menu of Pareto-efficient contracts specifying *bundles of payment and treatment*: $\{(t_L, P_L), (t_H, P_H)\}$, where (t_L, P_L) and (t_H, P_H) denotes the contract intended for low- and high-ability individuals, respectively. Contracts cannot be traded once they have been signed, moreover, medical treatment can not be supplemented. Contracts specifying combinations of treatment and payment are derived by:

$$\max_{(t_L,P_L),(t_H,P_H)} V(t_L, P_L, A_L)$$

subject to:

$$\begin{aligned} \underline{V} &\leq V(t_H, P_H, A_H) \\ V(t_H, P_H, A_L) &\leq V(t_L, P_L, A_L) \\ V(t_L, P_L, A_H) &\leq V(t_H, P_H, A_H) \\ N_L(P_L - t_L C) + N_H(P_H - t_H C) &= 0 \\ 0 &\leq t_i \leq 1, i = H, L. \end{aligned}$$

The first constraint ensures high ability individuals a certain level of utility. The second and the third constraints are the self-selection constraints. The fourth constraint is the government's balanced budget constraint, while the fifth constraint is the restriction that individuals cannot obtain more

[7] $u_c(c, h) \to \infty$ as $c \downarrow 0$ whenever $h > 0$.

[8] By assuming that they face identical risk of falling ill, we disregard questions regarding the relationship between ability and health (e.g. whether the likelihood of falling ill is correlated with the individuals' ability). Obviously, there is a relationship between health and socio-economic status, a fact that is of importance to the discussion of redistribution, yet the direction (and the strength) of causation is not straightforward.

than complete treatment (i.e., $t \leq 1$), nor 'sell' treatment (i.e., $t \geq 0$). As can be checked, the self-selection constraint and the constraint on the level of treatment are both satisfied for low-ability individuals in the subsequent analysis. It also holds that the non-negativity constraint on treatment is satisfied for high-ability individuals. Forming the Lagrangian:

$$\begin{aligned}\mathcal{L} &= V(t_L, P_L, A_L) + \lambda\left[V(t_H, P_H, A_H) - \underline{V}\right] \\ &+ \mu\left[V(t_H, P_H, A_H) - V(t_L, P_L, A_H)\right] \\ &+ \gamma\left[N_L(P_L - t_L C) + N_H(P_H - t_H C)\right] + \phi\,(1 - t_H)\,.\end{aligned}$$

The efficient t_i and P_i satisfy the conditions:

$$\frac{\partial \mathcal{L}}{\partial t_L} = \frac{\partial V(t_L, P_L, A_L)}{\partial t} - \mu\frac{\partial V(t_L, P_L, A_H)}{\partial t} - \gamma N_L C = 0 \quad (4.8)$$

$$\frac{\partial \mathcal{L}}{\partial P_L} = \frac{\partial V(t_L, P_L, A_L)}{\partial P} - \mu\frac{\partial V(t_L, P_L, A_H)}{\partial P} + \gamma N_L = 0 \quad (4.9)$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t_H} &= \lambda\frac{\partial V(t_H, P_H, A_H)}{\partial t} + \mu\frac{\partial V(t_H, P_H, A_H)}{\partial t} \\ &- \gamma N_H C - \phi = 0 \quad (4.10)\end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial P_H} = \lambda\frac{\partial V(t_H, P_H, A_H)}{\partial P} + \mu\frac{\partial V(t_H, P_H, A_H)}{\partial P} + \gamma N_H = 0 \quad (4.11)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} \geq 0, \quad \frac{\partial \mathcal{L}}{\partial \mu} \geq 0, \quad \frac{\partial \mathcal{L}}{\partial \gamma} \geq 0, \quad \frac{\partial \mathcal{L}}{\partial \phi} \geq 0.$$

For the moment, we assume that the government *has no redistribution ambitions*, and that payment reflects cost of production. In addition, we assume that the self-selection constraint on high-ability individuals is not binding (i.e., $\mu = 0$). Since, by assumption, the government has only one means of redistribution, the following provides a benchmark against which redistribution can be compared. Dividing equation (4.8) by equation (4.9) and using equation (4.5) we find:

$$\partial \mathcal{P}(t_L, A_L; t_L, t_L C)/\partial t = C,$$

hence, low-ability individuals' marginal willingness to pay for treatment equals cost of treatment. Since $A_L \leq \pi C$, it follows from the Lemma that

$0 < t_L < 1$, thus: $P_L = t_L C$. Dividing equation (4.10) by equation (4.11) and using equation (4.5) we get:

$$\partial \mathcal{P}(t_H, A_H; t_H, t_H C)/\partial t = C + \phi/\gamma N_H.$$

The government's zero-revenue constraint is assumed to hold, thus $\gamma > 0$. The marginal imputed cost incurred in restraining the individuals' level of treatment (i.e., $t_H \leq 1$) is given by ϕ. From the Lemma we know that $\phi > 0$, and thus $t_H = 1$, implying that the treatment constraint is binding. Moreover, $P_H = C$. Consequently, $\underline{V} = V(1, C, A_H)$.

The efficient bundles of treatment and payment when there is no redistribution: $\{(t_L, t_L C), (1, C)\}$, are illustrated in Figure 4.2. As can be seen, self-selection will not be a problem since high-ability individuals would suffer a loss in utility if choosing low-ability individuals' bundle (and since low-ability individuals can not afford high-ability individuals' bundle).

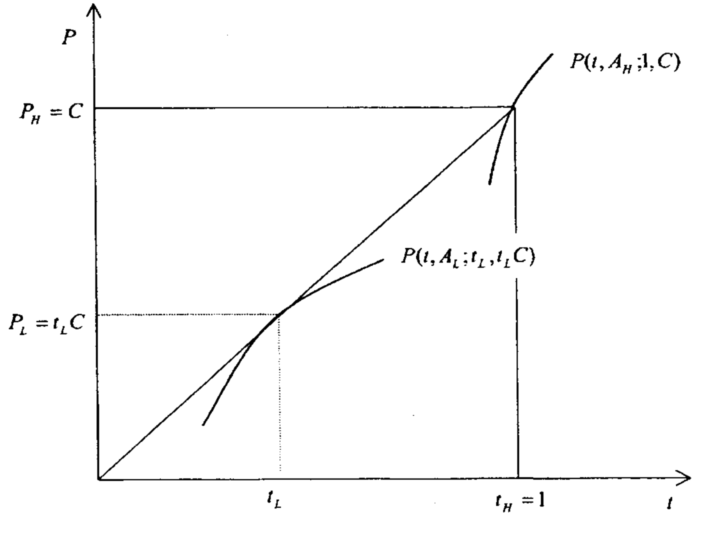

Figure 4.2: Efficient bundles of treatment and payment when there is no redistribution.

Redistribution We now assume that the government has *redistribution ambitions*, in particular, it seeks to maximise the sum of utilities, i.e., a *utilitarian welfare function*. This corresponds to $\lambda = N_H/N_L$, entailing that the weight of high-ability individuals relative to low-ability individuals corresponds solely to the numbers of individuals in each group. Utilitarianism leads to redistribution from high- to low-ability individuals if $\lambda > N_H/N_L$ in the situation without redistribution. In the subsequent, we will show that λ is indeed greater than N_H/N_L when there is no redistribution.

Since $\mu = 0$ when the government does not redistribute, it follows from equations (4.9) and (4.11) that $\lambda > N_H/N_L$ corresponds to:

$$-\partial V(t_L, t_L C, A_L)/\partial P > -\partial V(1, C, A_H)/\partial P.$$

Since, by equation (4.3), $-\partial V(t, P, A)/\partial P = -\partial \mathcal{L}/\partial P = \pi u_c(c_1, 1)$, it follows, as $u_{cc} < 0$, that λ in a situation without redistribution exceeds N_H/N_L if and only if $c_1(1, C, A_H) > c_1(t_L, t_L C, A_L)$. To show that this is the case, note that it follows from the constraint in equation (4.2) and the fact that consumption in state 2 is non-negative, that:

$$(1 - (1 - t_L)\pi) A \geq (1 - \pi) c_1(t_L, t_L C, A_L) + \pi(t_L C).$$

Hence, since $A_L \leq \pi C$, we obtain $c_1(t_L, t_L C, A_L) \leq (1 - t)\pi C$. Moreover, it follows from equation (4.3) and the constraint in equation (4.2) that high-ability individuals' level of consumption when $t_H = 1$ is given by $c_1(1, C, A_H) = c_2(1, C, A_H) = A_H - \pi C$. Recalling that $\pi \leq 0.5$, we thus see that $c_1(1, C, A_H) > c_1(t_L, t_L C, A_L)$. The level of consumption in state 1 is, in other words, higher for high-ability individuals than for low-ability individuals. Consequently, we have established that the government under utilitarianism wants to redistribute towards low-ability individuals.

Proposition 1. *If the government has a utilitarian welfare function, then income is redistributed from high-ability individuals to low-ability individuals.*

However, even if utilitarianism leads to redistribution from high- to low-ability individuals, we cannot determine without further assumptions whether redistribution is carried out to the extent that the self-selection constraint on high-ability individuals is binding. Self-selection may in fact not be a problem in the utilitarian optimum even if marginal utility of consumption is equalised across individuals and states. This is because the low-ability individuals' level of consumption and treatment in state 2 may be sufficiently low to prevent high-ability individuals from wanting to masquerade. In the following, we will study the situation where the self-selection constraint *binds*, i.e., $\mu > 0$, using the same approach to the self-selection problem as suggested by Stiglitz (1987).

The bundle intended for the low-ability individuals is found by dividing equation (4.8) by equation (4.9):

$$\frac{\mu\frac{\partial V(t_L,P_L,A_H)}{\partial t}+\gamma N_L C}{-\mu\frac{\partial V(t_L,P_L,A_H)}{\partial P}+\gamma N_L}=-\frac{\frac{\partial V(t_L,P_L,A_L)}{\partial t}}{\frac{\partial V(t_L,P_L,A_L)}{\partial P}}.$$

Defining $v \equiv \mu\frac{\partial V(t_L,P_L,A_H)}{\partial P}/\gamma N_L$ and using equation (4.5), we can rewrite the condition as:

$$\frac{\partial \mathcal{P}(t_L,A_L;t_L,P_L)}{\partial t}=C+\left(\frac{\partial \mathcal{P}(t_L,A_H;t_L,P_L)}{\partial t}-C\right)\frac{v}{v-1}.$$

From Section 4.2 we know that high-ability individuals' marginal willingness to pay for treatment is higher than that of low-ability individuals at any treatment-payment combination, so also for (t_L, P_L):

$$\partial \mathcal{P}(t_L,A_H;t_L,P_L)/\partial t > \partial \mathcal{P}(t_L,A_L;t_L,P_L)/\partial t.$$

Since $v < 0$[9], it follows that

$$C < \partial \mathcal{P}(t_L,A_L;t_L,P_L)/\partial t < \partial \mathcal{P}(t_L,A_H;t_L,P_L)/\partial t$$

in the self-selection equilibrium. Low-ability individuals are, in other words, offered a level of treatment t_L at which their marginal willingness to pay exceeds the marginal cost of production. The level of treatment is consequently *distorted downwards.* Graphically, the slope of both high- and low-ability individuals' indifference curve through the point (t_L, P_L) exceeds the slope of the isocost line.

The bundle intended for high-ability individuals is derived by diving equation (4.10) by equation (4.11), and using equation (4.5):

$$\frac{\partial \mathcal{P}(t_H,A_H;t_H,P_H)}{\partial t}=C+\frac{\phi}{\gamma N_H}.$$

By assumption, $\gamma > 0$. From the Lemma it follows that $\phi > 0$ and thus $t_H = 1$. Hence, the marginal willingness to pay for treatment exceeds the marginal cost of treatment for the non-distortionary reason that treatment cannot

[9]Because $\mu \geq 0, \gamma \geq 0$ and $\partial V_H/\partial P < 0$.

restore health beyond its original level. Therefore, the contract intended for the high-ability individuals is not distorted at the margin.[10]

Proposition 2. *If the self-selection constraint binds, then the optimal separating scheme is such that:*

(i) *High-ability individuals' bundle of treatment and payment is not distortionary.*

(ii) *Low-ability individuals' bundle of treatment and payment is distortionary. In particular, the level of medical treatment is distorted downwards since this is more costly to high-ability mimickers than to low-ability individuals.*

Consequently, if the self-selection constraint binds, the government induces individuals to reveal information by offering two bundles:

$$\{(t_L, P_L), (1, P_H)\},$$

where $0 < t_L < 1$.[11] High-ability individuals are discouraged from masquerading as low-ability individuals by restricting the level of treatment available to low-ability individuals. The attained self-selection equilibrium is illustrated in Figure 4.3.

Relating our findings to those of B-D; we have shown that when information on ability is asymmetric (and self-selection is a problem), then treatment is 'underprovided' to ill low-ability individuals, whereas B-D show that when information on health status (ill/healthy) is asymmetric, then treatment is 'overprovided' to ill individuals.

Extent of insurance coverage We apply the following terminology: by *full insurance,* we mean that $u(c_1, h_1) = u(c_2, h_2)$, i.e., that utility is constant across the two states, and by *partial insurance,* we mean that $u(c_1, h_1) > u(c_2, h_2)$, i.e., that utility is lower if ill than if healthy.

[10]This is analogous to the optimal taxation problem where the marginal tax rate faced by high-ability individuals is zero, while the marginal tax rate faced by low-ability individuals is positive (Stiglitz, 1987). This is often called a 'non-distortion at the top' property, where 'top' refers to individuals that no one would choose to masquerade as.

[11]Payment may in fact be negative if the subsidy is large, that is; individuals may not only receive medical treatment, but also a cash transfer from the hospital if ill (the cash transfer then being contingent on the individual undergoing treatment).

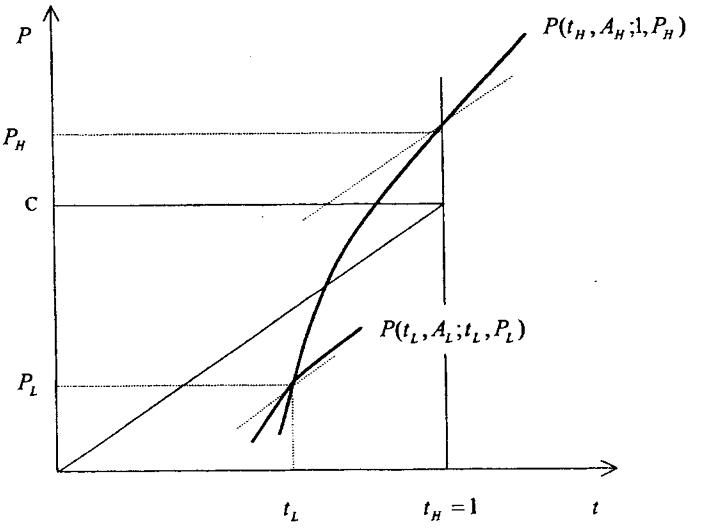

Figure 4.3: Efficient bundles of treatment and payment when self-selection is a problem.

High-ability individuals are defined by $A_H \geq C$. From the Lemma we know that $t_H = 1$, hence, high-ability individuals' level of ability will be equal across states: $h_1 = h_2 = 1$. Moreover, their level of consumption in the two states of the world will be identical: $c_1(1, P_H, A_H) = c_2(1, P_H, A_H) = A_H - \pi P_H$. It follows that $u(c_1, h_1) = u(c_2, h_2)$ and, consequently, they are *fully insured.*

Low-ability individuals are defined by $A_L \leq \pi C$. From the above analysis we know that $0 < t_L < 1$, hence, $h_1 = 1$ and $h_2 = t_L < 1$. It follows from equation (4.3) and the properties of u that $c_1(t_L, P_L, A_L) \geq c_2(t_L, P_L, A_L)$. Consequently, $u(c_1, h_1) > u(c_2, h_2)$ and the individuals are *partly insured.* When the government redistributes and self-selection is a problem, then low-ability individuals' level of medical treatment is distorted downwards. Their level of insurance coverage is hence reduced relatively to a situation without redistribution. They will, however, receive a subsidy which enables them to have a level of consumption in excess of their earnings minus the cost of treatment, and which makes them better off.

4.4 Discussion

So far, we have studied the individuals' *ex ante* decision regarding their optimal level of consumption in the two states of the world, as well as their optimal level of medical treatment given the two bundles. If individuals are rational and have perfect foresight, then their *ex post* preferred level of medical treatment will also be preferred *ex ante*. Prior to knowing which state of the world has occurred, they are therefore willing to sign a contract with the public supplier, i.e., hospital, specifying both payment and level of treatment that is to be made available if ill. Indeed, one would expect that the individuals would prefer to do so as this would prevent them from potential transaction costs associated with having to 'shop around' for the appropriate contract when ill. Such a scheme would in fact also be more in accordance with what may be observed empirically: Public supply of medical treatment financed by individuals contributing through (earmarked) contributions and supplied free at the point of delivery. The public supply of treatment may thus be thought of as an *insurance* where individuals are compensated in the form of medical treatment directly if illness occurs, i.e., *indemnity in kind*. The preceding analysis will still be valid in such a setting, the only difference is that individuals *ex ante* determine their optimal level of private and public insurance coverage.[12]

The preceding analysis is based on a highly stylised model. We assume that individuals with certainty can recover completely by consuming the appropriate level of medical treatment. Moreover, illness is presumed to be observable, hence, the private insurers do not face problems of adverse selection. We have also ignored differences in the risk of falling ill which indeed is an important reason for having a (mandatory) public health insurance (Breyer and Haufler, 2000). Possible commitment problems associated with a re-optimising government aspiring to increase treatment above the 'announced' level, is not discussed. The perhaps most striking assumption underlying this analysis is, however, the assumption that the government cannot implement income taxation. Consequently, we cannot infer whether redistribution through public pricing of contracts is more efficient than other

[12]Such a setting would be more in line with the analysis in Asheim et al., (2000) where individuals' demand for insurance against medical expenditures (i.e. in kind) and/or loss in income due to disability (i.e. in cash) are integrated. Their analysis takes place in a perfectly competitive private market with no public interference.

means of redistribution. Extension of the analysis to include also individuals' labour supply and distortionary income taxation, will be the subject of future research. We think still that our analysis provides interesting results, results that run counter to what often seems to be implicitly underlying many studies of public provision of health care; namely that individuals wish to fully recover from an illness.

Lastly, our analysis follows the tradition of neo-classical economics in that individuals' utility (or preferences) provide the foundation of the analysis. We study distribution of welfare, that is; well-being assessed in utility terms, thus, we assume that the value of medical treatment to an individual is represented by her willingness to pay for it.[13] An alternative approach would be the extra-welfarist framework in which health, and not utility, is the primary outcome of interest (Hurley, 1998). According to this approach, distributional equity (in the egalitarian concept) implies that medical treatment should be provided according to need, and not according to ability or willingness to pay.

Acknowledgments

I am grateful to Geir B. Asheim, Robin Boadway, Tore Nilssen, Agnar Sandmo and an anonymous referee of International Tax and Public Finance for their valuable comments and suggestions. I am also grateful for helpful comments from the participants at the Oslo workshop on Health Economics and the 57th Congress of the IIPF in Linz. The usual disclaimer applies. Financial support from the SiS-programme at Agder University College is gratefully acknowledged.

[13]Moreover, we disregard any merit good arguments which may justify a separation of ability to pay and treatment provided, an argument which is prevailing in the redistributional objectives in many countries.

References

Asheim, G.B., A.W. Emblem and T. Nilssen. (2000). "Health Insurance: Treatment vs. Compensation." Available at: http://www.oekonomi.uio.no/memo/memopdf/memo4000.pdf

Balestrino, A. (1999). "The Desirability of In-Kind Transfers in the Presence of Distortionary Taxes." *Journal of Economic Surveys*, 13(4), 333-354.

Besley, T. and S. Coate. (1991). "Public Provision of Private Goods and the Redistribution of Income." *The American Economic Review*, 81(4), 979-984.

Blackorby C. and D. Donaldson. (1988). "Cash versus Kind, Self-Selection, and Efficient Transfers." *The American Economic Review*, 78(4), 691-700.

Blomquist S. and V. Christiansen. (1995). "Public Provision of Private Goods as a Redistributive Device in an Optimum Income Tax Model." *Scandinavian Journal of Economics,* 97(4), 547-567.

Blomquist S. and V. Christiansen. (1998). "Price Subsidies versus Public Provision." *International Tax and Public Finance*, 5 (3), 283-306.

Boadway, R., M. Marchand and M. Sato. (1998). "Subsidies versus Public Provision of Private Goods as Instruments for Redistribution." *Scandinavian Journal of Economics*, 100(3), 545-564.

Breyer, F. and A. Haufler. (2000). "Health Care Reform: Separating Insurance from Income Redistribution." *International Tax and Public Finance*, 7(4-5), 445-461.

Epple, D. and R.E. Romano. (1996). "Public Provision of Private Goods." *Journal of Political Economy*, 104 (1), 57-84.

Hurley, J. (1998). "Welfarism, Extra-Welfarism and Evaluative Economic Analysis in the Health Sector." In M.L. Barer, T.E. Getzen and G.L. Stoddart (eds.), *Health, Health Care and Health Economics*, 373-395. John Wiley & Sons Ltd.

Ireland, N.J. (1990). "The Mix of Social and Private Provision of Goods and Services." *Journal of Public Economics*, 43(2), 201-219.

Macho-Stadler, I. and J.D. Pérez-Castrillo. (1997). *An Introduction to the Economics of Information. Incentives & Contracts.* Oxford University Press.

Seade, J. (1982). "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies*, 49(4), 637-643.

Stiglitz, J.E. (1987). "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." In A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Vol. 2, 991-1041. North-Holland.

Tobin, J. (1970). "On Limiting the Domain of Inequality." *Journal of Law and Economics*, 13(2), 263-277.

Essay 5

Able or Disabled: Optimal Income Taxation and Public Provision of Health Care

Abstract

This study examines how optimal non-linear income taxation and differentiated public provision of health care may form part of a scheme facilitating income redistribution between different ability types (high/ low), and across different health states (poor/good). Information about ability and labour supply is private. The government offers contracts in five dimension: pre- and post-tax income if in good health, and pre- and post-tax income, as well as medical treatment, if in poor health. The derived information constrained scheme facilitates income redistribution towards low-ability individuals. In addition, it provides insurance against medical expenditures and loss in income due to disability. The scheme is shown to entail an downward distortion in low-ability individuals' labour supply in both health states, and a downward distortion in their level of treatment if in poor health.

5.1 Introduction

Health care is in a number of countries publicly supplied and financed through taxation, either in the form of general income taxation or specific (earmarked) contributions. For instance, in the United Kingdom and the Scandinavian countries, there exists a tax financed universal compulsory public insurance against a wide range of health care expenditures, such as primary care and hospital care. The motivation for, and justification of, a public health care supply differs among countries, and over time. Often, distributional (equity) concerns are put forward as the most important rationale, but also arguments of paternalism, externalities and market failure may be advanced.[1]

In this paper, we do not aspire to explain *why* health care is publicly supplied. Rather, we acknowledge that a public supply exists. We assume that the government's ambition is (more) equal *income distribution* among individuals. Hence, it pursues a general distributional motive and not a particular distribution of health, or health care, *per se*. Redistribution by means of individualized lump-sum taxation is not possible due to private information about ability (productivity) and labour supply. Our objective is to study *how* health care provision and non-linear income taxation may form part of an information constrained Pareto-efficient income redistribution scheme. Integrating what is often considered to be two separate policy instruments, *i.e.*, non-linear income taxation and public health care provision, is motivated by the fact that the two are interrelated. Indeed, the tax base (*i.e.*, income) is dependent on individuals' health state, and the monetary benefits from health improvements are dependent on the tax burden.

We assume that health is stochastic: individuals may either be healthy or sick. If sick, health may be fully or partly restored (with certainty) if the appropriate level of care is provided. Individuals' health, and thus productivity, is consequently endogenous to the model. The government's problem is to *ex ante* design contracts facilitating income redistribution between individuals with different levels of innate ability, and across different health states. Hence, contracts are designed *prior* to knowing individuals' health states (which is verifiable *ex post*). The similarity between the tax/provision scheme and an optimal *insurance* scheme under asymmetric information will be readily apparent. Indeed, the 'ex ante view' enables us to highlight the in-

[1] Indeed, paternalism was a central argument when introducing the Law of Health Insurance (Sykeforsikringsloven) in Norway 1906/07: *"Thi all Erfaring viser det, at Menneskene tænker ikke- iallefald ikke alle...paa den Dag imorgen...og da bliver det nødvændigt, at Samfundet, udover hvilket tilslut Skaden gaar, sørger for, at Individet bliver forpligtet til at sørge for den Dag imorgen."* Gunnar Knudsen om sykeforsikringsloven, 1906/07 as cited in Seip (1984).

surance aspect of the tax/provision scheme: the non-linear income tax does not only provide an instrument for achieving (more) equity, but also an instrument for insurance when health, and thus also earnings, are stochastic (Mazur, 1989). The tax/provision scheme provides coverage against medical expenditures, *i.e.*, a medical insurance with in-kind compensation, and against (permanent) loss in income due to illness, *i.e.*, a disability insurance with cash compensation.[2] Indeed, we show that the provision of health care constitutes a *screening* device, that is: it induces self-selection. The income tax may, hence, constitute an 'insurance premium', nevertheless separated from individual health risks or benefits.

In the literature on optimal income taxation, it is well established that asymmetric information about characteristics on which the taxation is to be based, restricts the range of policies available to the government and the design of these. Stiglitz (1982, 1987) provides a comprehensive analysis of the implications of informational constraints on income taxation. Nichols and Zeckhauser (1982), Blackorby and Donaldson (1988), Besley and Coate (1991), and Boadway and Marchand (1995) show that in-kind provision may serve as a screening mechanism and thus enhance the target efficiency of a distributional policy. Indeed, Blomquist and Christiansen (1998) show that if the self-selection constraint is binding, then public provision of private goods may strictly Pareto dominate the optimal income tax optimum. The good in question should, though, not be re-tradeable or supplementable, and it should be a substitute for leisure (Blomquist and Christiansen, 1995). Health care is often mentioned as an example of a private good that may be suitable for public provision. However, there are only few studies that actually scrutinize public provision of health care, some of which are Blackorby and Donaldson (1988) and Anderberg (2001). Their adverse selection problem is in health status, *i.e.*, whether individuals are sick (infirm) or healthy (firm). Hence, their analysis takes place after health status, and thus ability to work, is determined, *i.e.*, *ex post*. They assume, moreover, that sick (infirm) individuals cannot work, hence they do not consider the consequences of the tax/provision scheme on sick individuals' trade-off between consumption and leisure.[3] Our analysis is similar to that of Anderberg (2001) in

[2]Empirically, insurance against medical expenditures and loss in income due to permanent disability, *i.e.*, a disability payment, form part of a social insurance in a number of countries such as Norway and the United Kingdom.

[3]Blackorby and Donaldson (1988) assume that sick and healthy individuals have dif-

that the government's provision of health care is contingent on earnings, *i.e.*, non-uniform provision. Also, there is *no* private supply of health care, and neither resale nor supplementing is possible.[4]

Our model deviates from previous studies in several respects. First, we characterize a policy of in-kind provision when the policy is to be designed *prior* to knowing individuals' health status, *i.e.*, *prior* to knowing whether an individual is sick or not. Second, the government faces adverse selection in ability (high/low) and not in health status (sick/well). Third, we allow sick individuals to (partly or fully) restore their health, and thus also their productivity in the labour market, by undergoing medical treatment. Hence, productivity if ill is *endogenous*. Fourth, individuals can participate in the labour market even if health is not fully restored. This allows us to study the consequences of a tax/provision scheme on sick individuals' incentives to participate in the labour market. Lastly, the government faces also a moral hazard problem: since labour supply is assumed endogenous and not observable, individuals can influence the potential loss in income due to illness.[5]

The outline of this paper is as follows. In Section 5.2 we derive the model and show that individuals with identical medical needs do *not* prefer identical levels of health care. Indeed, individuals with high levels of inherent ability prefer to undergo more medical treatment than do individuals with a low such ability. In Section 5.3 we describe an information constrained efficient scheme in which public supply of health care and non-linear income taxation are integrated. Since the benefit per euro of utilizing health care services increases with productivity (ability), the government uses different

ferent preferences for health care, and that the government's ambition is to provide care to the sick. Anderberg (2001) study income redistribution from healthy (firm) individuals towards sick (infirm) individuals.

[4]In Norway, health care services are mainly publicly provided and only a few services are supplied by private institutions (and then usually in agreement with the government).

[5]This may be referred to as *ex post* moral hazard since we focus on individuals' behavior after illness has occurred. In the literature on health insurance, *ex post* moral hazard usually refers to a situation in which individuals may influence the cost of treatment, *i.e.*, the financial consequences of illness. Here, we focus on a situation in which individuals influence the financial consequences of illness, *not* through their choice of treatment (cheap/expensive), but through their choice of leisure (more/less). Also, we assume that health state is verifiable, thereby ignoring another type of possible *ex post* moral hazard problems: individuals having incentives to falsely claiming to be sick when indeed they are not.

levels of health care combined with different bundles of earnings and taxes to separate high-ability individuals from low-ability individuals. The optimal separating scheme entails a *downward* distortion in low-ability individuals' level of treatment (*i.e.*, health) and labour supply. High-ability individuals, on the other hand, are provided with complete treatment if ill, and are required to spend more time in the labour market in both states. In Section 5.4 we elaborate on the implications of a public provision of treatment with respect to redistribution. Our findings are discussed in Section 5.5.

5.2 The Model

Individuals are assumed to have identical preferences over consumption (c), health (h) and leisure (l). Moreover, preferences over consumption and health are assumed to be additive separable from leisure:

$$U(c,h,l) = u(c,h) + v(l),$$

where U will provide input to individuals' von Neumann-Morgenstern expected utility function in the below. While health is a multidimensional concept, we assume that health is represented by the variable h. Also, consumption c is a vector of consumption goods. The utility function is assumed to be continuously differentiable and it satisfies: $u_c; u_h; v_l > 0$, and $u_{cc}; u_{hh}; v_{ll} < 0$, where the subscripts denote the partial derivatives. The utility function is, thus, strictly increasing and strictly concave. Also, we assume that $u_c(0,h) = \infty$, $u_h(c,0) = \infty$, $u_c(0,0) = \infty$, $u_h(0,0) = \infty$, and $v_l(0) = \infty$. From this it follows that individuals strictly desire a positive level of consumption, health and leisure. Consumption, health and leisure are, moreover, assumed to be normal goods. We make the additional assumption that health and consumption are complements in utility: $u_{ch} > 0$. Thus, individuals derive utility from health *per se*, in addition, health influences their marginal utility from consumption: $u_{ch} > 0$.[6] In the literature on health economics, this assumption is often made about severe illnesses (Evans and Viscusi, 1991).[7]

[6]Since the partial derivatives of the utility function with respect to health and consumption are both assumed continuous, then according to Young's theorem: $u_{ch} = u_{hc}$, *i.e.*, the cross partial derivatives are identical.

[7]Minor illnesses are, on the other hand, often argued to cause marginal utility from income to increase (Evans and Viscusi, 1991).

Individuals face an identical risk of falling ill thus suffering a loss in health. There is a probability $(1-\pi)$ that they are in good health, *i.e.*, state 1, and a probability π that they are sick, *i.e.*, state 2, where $0 \leq \pi \leq 1$. The two states are mutually exhaustive and verifiable. Risk is identical and stochastically independent, thus we disregard collective risk. Also, individuals can not influence the probability of illness, *i.e.*, there is no *ex ante* moral hazard. Health in state 1 is normalised to 1: $h_1 = 1$. If sick, individuals are assumed to suffer a complete loss in health: $h_2 = 0$. Health if sick may, however, be partly or fully restored with certainty if individuals undergo medical treatment: $t \in [0, 1]$. Thus, treatment is assumed to be indefinitely divisible. The cost of treatment leading to complete recovery is given by C and is *public information.* Treatment t is measured as the rate of total treatment cost spent on treatment. Thus, individuals will by spending tC on treatment recover by rate t. Health in state 2 is henceforth measured by the fraction of C spent on treatment, *i.e.*, $h_2 = t$. We assume that medical treatment cannot be resold, and there is no supplementing in a private market.

We assume that individuals maximize an expected utility function of the von Neumann-Morgenstern type:

$$E\{U\} = (1-\pi)\left[u(c_1, 1) + v(l_1)\right] + \pi\left[u(c_2, t) + v(l_2)\right]. \tag{5.1}$$

As follows by the assumptions on $U(c, h, l)$, individuals are risk averse towards consumption, health and leisure. Moreover, as stated in the above, for levels of health less than perfect (*i.e.*, $h_2 < 1$), we do not regard health as an income-equivalent. In the literature on health insurance, such as Zeckhauser (1970), Arrow (1974), Viscusi and Evans (1990), utility functions may assume different shapes depending on the individuals' health state (good/poor). Traditionally, health status is incorporated by indexing the utility function by the state of the world (Evans and Viscusi, 1991). Underlying this is an implicit assumption that health is not restorable, or that illness as such entails pain and suffering which cannot be compensated for. In this paper, however, we let health if ill be an *alterable* and endogenous characteristic, indeed, health if ill can be fully restored. Utility from consumption and health if in state 2 may hence be identical to that if in state 1 provided that health if sick is fully restored, *i.e.*, $t = 1$.[8] We assume that

[8]Geometrically, one can depict the traditional formulation of state-dependent prefer-

treatment does not involve pain and suffering, hence, restitution for pain is not necessary. Also, since health states are assumed to be verifiable, medical treatment is provided to sick individuals only.

Individuals are assumed to *ex ante* be identical in all respects, save their level of inherent ability. Moreover, we assume that production is linear in labour supply and that individuals are perfect substitutes in production. Labour units are normalized so that individuals' inherent level of ability is reflected in their wage rate: w. If in good health, *i.e.,* in state 1, then individuals will by spending $(1 - l_1)$ hours in the labour market generate earnings equal to $w(1 - l_1)$. Labour supply is assumed strictly positive and less than one: $0 < (1 - l_i) < 1$. If sick, *i.e.,* in state 2, and receiving a level of medical treatment equal to t, then they will by spending $(1 - l_2)$ hours in the labour market generate earnings equal to $tw(1 - l_2)$. Productivity if sick is thus given by the product tw and is *endogenous.* Information about w and l_i is *private* to the individuals. By assumption, health care is free at the point of delivery, thus net benefits from treatment if ill is given by the total productivity during the $(1 - l_2)$ hours of work: $tw(1 - l_2)$. Naturally, the higher the level of innate ability (w), the higher the benefits from treatment. Gross labour earnings, henceforth referred to as earnings, are defined as:

$$E_1 \equiv w(1 - l_1),$$

if well and

$$E_2 \equiv tw(1 - l_2)$$

if sick. In order to simplify, we assume that individuals are *not* able to transfer income across states through a private insurance market.

ences as two curves in (c, u)-space: one depicting $u(c, 1)$ and one depicting $u(c, 0)$. The two curves are two sections of the overall utility function $u(c, h)$. (For more on this, see Hirshleifer and Riley, 1992). Hence, it is easy to see that if health is not restorable (*i.e.,* $h_2 = 0$), then identical marginal utility from consumption across states is achieved at different levels of consumption: $c_2 < c_1$. Here, however, we allow health to be improvable. Thus, we can depict a number of such curves between the 'lower' curve giving$u(c, 0)$ and the 'upper' curve giving $u(c, 1)$, since $h_2 = t \in (0, 1]$. Thus, if individuals choose a $t = 1$ if sick, then equal marginal utility from consumption across states implies equal level of consumption: $c_1 = c_2$. Correspondingly, if $0 < t < 1$, then $c_2 < c_1$.

Normalizing the price of consumption to 1, we can write individuals' after-tax budget constraint in each of the two states as:

$$c_1 = E_1 - T(E_1) \tag{5.2}$$

and

$$c_2 = E_2 - T(E_2) \tag{5.3}$$

where T denotes earnings tax.

We elaborate somewhat on individuals' labour supply. Medical treatment if ill improves productivity by a fraction t. Improved productivity has both an income and substitution effect on labour supply; on the one hand, it implies that leisure becomes relatively more expensive thus leading to a decrease in the demand for leisure (*i.e.*, the substitution effect on $(1 - l_2)$ is positive), on the other hand, it causes earnings to increase, thus making more leisure affordable (*i.e.*, the income effect on $(1 - l_2)$ is negative). Since leisure is assumed to be a normal good, then labour supply may increase (decrease) in t (and w) if the substitution effect dominates (is inferior to) the income effect. In order to simplify the subsequent analysis, we assume that labour supply is strictly increasing in w and t, and strictly decreasing in T. Thus, the labour supply function is *upward sloping* in both $((1-l_i), w)$ and $((1 - l_2), t)$-space.[9]

We can now rewrite the von Neumann-Morgenstern expected utility function (*cf.* equation (5.1)) of a representative individual in terms of observable variables:

$$V(E, c, t) = (1 - \pi)\left[u(c_1, 1) + v(1 - E_1/w)\right] + \pi\left[u(c_2, t) + v(1 - E_2/tw)\right], \tag{5.4}$$

where $V(E, c, t)$ will differ among individuals since ability (w) differs. As follows from the budget constraints, individuals' level of consumption in each of the two states of the world is equal to net income (*i.e.*, earnings minus tax), thus we let c_i, $i = 1, 2$, represent net income in state 1 and 2, respectively. The expected utility function $V(E, c, t)$ is increasing in consumption (c) and treatment (t) and decreasing in earnings (E).

[9] In the literature on income taxation, the basic model postulates that labour supply at low wages increases if wage increases, whereas labour supply at high wages decreases if wage increases (Atkinson and Stiglitz, 1987). The labour supply curve is thus backward bending at high wages.

We now set out to characterize the utility function as given in equation (5.4). First, the marginal rate of substitution between consumption and earnings (mrs_{c_i,E_i}) at a given level of health, is given by:

$$mrs_{c_1,E_1} \equiv \frac{dc_1}{dE_1} |_{\overline{V}} = -\frac{\frac{\partial V}{\partial E_1}}{\frac{\partial V}{\partial c_1}} = \frac{1}{w} \frac{v_l(1 - \frac{E_1}{w})}{u_c(c_1, 1)} > 0, \tag{5.5}$$

if healthy, and:

$$mrs_{c_2,E_2} \equiv \frac{dc_2}{dE_2} |_{\overline{V}} = -\frac{\frac{\partial V}{\partial E_2}}{\frac{\partial V}{\partial c_2}} = \frac{1}{tw} \frac{v_l(1 - \frac{E_2}{tw})}{u_{c_2}(c_2, t)} > 0 \tag{5.6}$$

if sick. Consumption and leisure are assumed to be normal goods, thus mrs_{c_i,E_i} is decreasing in ability (w). This is shown for state 2:

$$\frac{\partial}{\partial w}(mrs_{c_2,E_2}) = \left[\frac{v_{ll}(1 - \frac{E_2}{tw})}{u_c(c_2, t)} \frac{E_2}{tw} - \frac{v_l(1 - \frac{E_2}{tw})}{u_c(c_2, t)} \right] \frac{1}{w^2} < 0, \tag{5.7}$$

since, by the properties of u and v: $u_c; v_l > 0$ and $v_{ll} < 0$. This is an *agent monotonicity property* which may be referred to as a 'Mirrlees-Spence' single-crossing property.[10] From this it follows that the slope of an indifference curve through any point $(\overline{E}_i, \overline{c_i})$, $i = 1, 2$, is steeper for individuals with ability w_L than that of individuals with ability w_H, if $w_L < w_H$. This is illustrated in Figure 5.1. The monotonicity property entails that any two indifference curves of individuals with different ability cross only once, *i.e.*, *single-crossing*.

Note that the marginal rate of substitution between consumption and earnings (mrs_{c_2,E_2}) is decreasing in treatment:

$$\begin{aligned} &\frac{\partial}{\partial t}(mrs_{c_2,E_2}) \\ &= \frac{1}{tw} \left[-\frac{1}{t} \frac{v_l(1 - \frac{E_2}{tw})}{u_c(c_2, t)} + \frac{\frac{E_2}{t^2 w} v_{ll}(1 - \frac{E_2}{tw}) u_c(c_2, t) - v_l(1 - \frac{E_2}{tw}) u_{ch}(c_2, t)}{(u_c(c_2, t))^2} \right] < 0, \end{aligned} \tag{5.8}$$

since, by the properties of u and v: $u_c; u_{ch}; v_l > 0$ and $v_{ll} < 0$. From this it follows that individuals are less concerned (in ordinal terms) about leisure

[10] The intuition for this result is that low-ability individuals will spend relatively more time in the labour market than do high-ability individuals in order to generate the given level of earnings. Hence, low-ability individuals' marginal rate of substitution of leisure for consumption is higher than that of high-ability individuals.

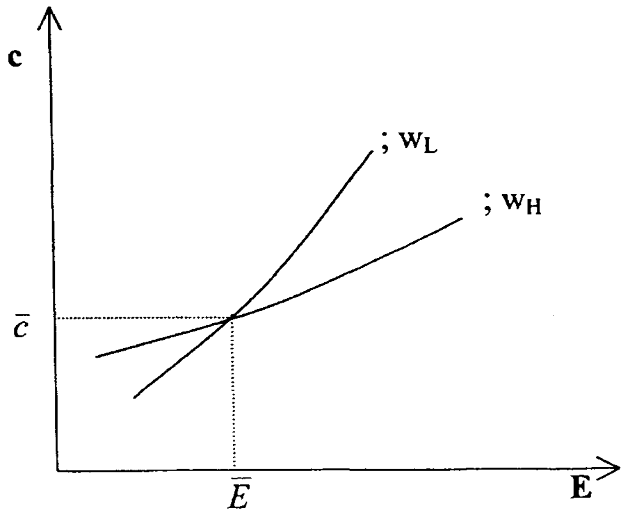

Figure 5.1: Earnings and consumption.

than consumption when health is good. Geometrically, this implies that the slope of an indifference curve through $(\overline{E}_2, \bar{c}_2)$ in (E_2, c_2)-space is less steep for health h_{2H} than for h_{2L}, when $h_{2L} < h_{2H}$.

Second, the slope of a curve through $(\bar{t}, \overline{E}_2)$ in (t, E_2)-space showing combinations of treatment and earnings yielding a constant level of utility is given by:

$$mwpt \equiv \frac{dE_2}{dt} |_{\overline{V}} = -\frac{\frac{\partial V}{\partial t}}{\frac{\partial V}{\partial E_2}} = \frac{u_{h_2}(c_2, t)}{v_{l_2}(1 - E_2/tw)} tw + \frac{E_2}{t}. \tag{5.9}$$

This is the individuals' *marginal willingness to pay for treatment* ($mwpt$) and is given by the sum of marginal willingness to pay for health: $(u_{h_2}/v_{l_2})tw$, *i.e.*, the consumption and leisure value of treatment, and marginal earnings: E_2/t, *i.e.*, the production value of treatment. As follows from equation (5.9), $mwpt$ is increasing in ability:

$$\frac{\partial}{\partial w}(mwpt) = \frac{u_{h_2}(c_2, t)}{v_{l_2}(1 - E_2/tw)} \left[\frac{-v_{ll}(1 - E_2/tw)}{v_l(1 - E_2/tw)} \frac{E_2}{w} + t \right] > 0. \tag{5.10}$$

Geometrically, this *monotonicity property* is illustrated by the slope of any pair of two indifference curves in (t, E_2)-space going through $(\bar{t}, \overline{E_2})$ being steeper the higher the level of ability. This is illustrated in Figure 5.2 for $w_L < w_H$.

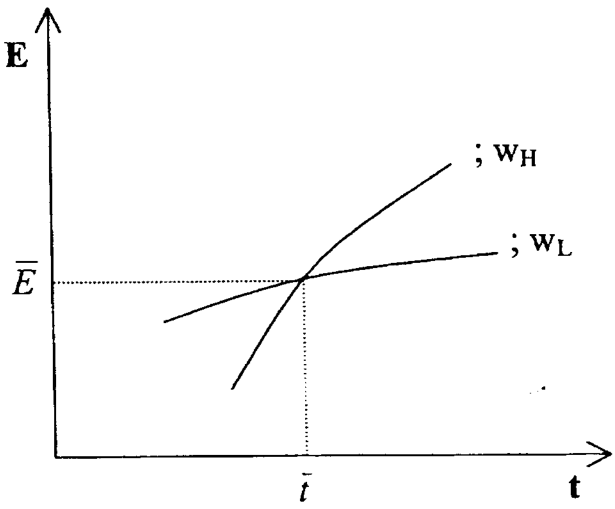

Figure 5.2: Medical treatment and earnings.

Third, the slope of an indifference curve through $(\bar{t}, \overline{c_2})$ in (t, c_2)-space is given by:

$$mwpt_{c_2,t} \equiv (-\frac{dc_2}{dt}) \mid_{\overline{V}} = \frac{\frac{\partial V}{\partial t}}{\frac{\partial V}{\partial c_2}} = \frac{u_{h_2}(c_2, t)}{u_{c_2}(c_2, t)} + \frac{v_{l_2}(1 - E_2/tw)}{u_{c_2}(c_2, t)} \frac{E_2}{t^2 w}. \quad (5.11)$$

Also this reflects the individuals' marginal willingness to pay for treatment, this time in terms of foregone consumption. From equation (5.11) we see that $mwpt_{c_2,t}$ is given by the sum of marginal willingness to pay for health in terms of consumption (*i.e.*, u_{h_2}/u_{c_2}) and the 'returns' on labour, as measured by the marginal willingness to pay for leisure in terms of consumption (v_{l_2}/u_{c_2}) times the 'effective' labour hours $(E_2/t^2w = (1 - l_2)/t)$. We can now identify a third monotonicity property: the slope of an indifference curve through $(\bar{t}, \overline{c_2})$ in (t, c_2)-space is decreasing in ability:

$$\frac{\partial}{\partial w}(mwpt_{c_2,t}) = \frac{E_2}{(tw)^2}\left[\frac{v_{ll}(1 - E_2/tw)}{u_{c_2}(c_2, t)}\frac{E_2}{tw} - \frac{v_l(1 - E_2/tw)}{u_c(c_2, t)}\right] < 0. \quad (5.12)$$

Naturally, if individualized lump-sum taxation had been possible, then marginal willingness to pay for treatment as given by equations (5.9) and (5.11) would be identical. This is so because in a first-best situation, the marginal rate of substitution between consumption and leisure is set equal to the wage rate, *i.e.*, $v_{l_2}/u_{c_2} = tw$, implying that $(dE_2/dt) \mid_{\overline{V}} = -(dc_2/dt) \mid_{\overline{V}}$.

Forth, we note that the slope of an curve in (E_1, E_2)-space showing combinations of gross earnings in state 1 and 2, yielding a constant level of utility is given by:

$$mrs_{E_2,E_1} = (-\frac{dE_2}{dE_1}) \mid_{\overline{V}} = \frac{\frac{\partial V}{\partial E_1}}{\frac{\partial V}{\partial E_2}} = \frac{(1-\pi)}{\pi} \frac{v_l(1-\frac{E_1}{w})t}{v_l(1-\frac{E_2}{tw})} > 0. \tag{5.13}$$

Recall that, by assumption, labour supply is strictly increasing in ability[11], thus:

$$\frac{\partial}{\partial w}(mrs_{E_2,E_1}) > 0. \tag{5.14}$$

We elaborate somewhat on the implications of this assumption in the subsequent. If $h_1 = h_2 = 1$, and $\overline{E}_1 = \overline{E}_2$, then leisure is constant across states: $l_1 = l_2$. It thus follows that $mrs_{E_2,E_1}$is equal to the ratio of state probabilities, *i.e.*, $mrs_{E_2,E_1} = \frac{(1-\pi)}{\pi}$. For $0 < \bar{t} < 1$ (*i.e.*, $h_2 < 1$), then individuals will spend relatively more labour time if sick than if well in order to generate the required level of earnings: $\overline{E}_1 = \overline{E}_2$. Hence, marginal utility from leisure is lower if well than if sick and, subsequently, $mrs_{E_2,E_1} < \frac{(1-\pi)}{\pi}$ when $0 < \bar{t} < 1$. We will now consider mrs_{E_2,E_1} for two levels of ability: $w_L < w_H$, when treatment is given by: $0 < \bar{t} < 1$. For a given level of $\overline{E}_1$ and $\overline{E}_2$, then if $\overline{E}_2$ increases and $\overline{E}_1$ decreases so as to keep expected utility constant, then a sick individual with ability w_H will have to increase her labour supply with less than do a sick individual with ability w_L in order to generate the required increase in earnings. Hence, the increase in marginal utility from leisure if sick is less for the individual with ability w_H. Correspondingly, a decrease in E_1 implies that a healthy individual will enjoy more leisure, hence marginal utility from leisure decreases, and more so for an individual with ability w_H than for an individual with w_L. Thus, the increase in earnings if ill (E_2) required to compensate for a decrease in earnings if healthy (E_1) is less for individuals with low ability than for those with high ability. Hence, the slope of an indifference curve in (E_1,E_2)-space

[11] Without this assumption the slope of the indifference curve would be indeterminate: $\frac{\partial}{\partial w}(mrs_{E_2 E_1}) = \frac{(1-\pi)}{\pi} \frac{E_1 t v_{ll}(1-\frac{E_1}{w})v_l(1-\frac{E_2}{tw}) - E_2 v_{ll}(1-\frac{E_2}{tw})v_l(1-\frac{E_1}{w})}{w^2(v_l(1-\frac{E_2}{tw})t)^2} \lesseqgtr 0$.

through a given point $(\overline{E}_1,\overline{E}_2)$ is steeper for w_H than for w_L. (Correspondingly, the slope of a curve in (l_1,l_2)-space showing combinations of l_1 and l_2 attainable at different combinations of (E_1,E_2) yielding a constant level of utility, is less steep for individuals with ability w_L than that of individuals with ability w_H.)

Note also that the marginal rate of substitution of earnings across states is increasing in treatment:

$$\frac{\partial}{\partial t}(mrs_{E_2,E_1}) = \frac{(1-\pi)}{\pi}\left[\frac{v_l(1-\frac{E_1}{w})}{v_l(1-\frac{E_2}{tw})} - t\frac{v_l(1-\frac{E_1}{w})v_{ll}(1-\frac{E_2}{tw})\frac{E_2}{t^2 w}}{(v_l(1-\frac{E_2}{tw}))^2}\right] > 0.$$

Finally, we note that the marginal rate of substitution between consumption if sick for consumption if well is given by:

$$mrs_{c_2,c_1} \equiv (-\frac{dc_2}{dc_1}) \mid_{\overline{V}} = \frac{\frac{\partial V}{\partial c_1}}{\frac{\partial V}{\partial c_2}} = \frac{1-\pi}{\pi}\frac{u_c(c_1,1)}{u_c(c_2,t)} > 0. \qquad (5.15)$$

Moreover,

$$\frac{\partial}{\partial t}(mrs_{c_2 c_1}) = \frac{1-\pi}{\pi}\left[\frac{-u_c(c_1,1)u_{ch}(c_2,t)}{(u_c(c_2,t))^2}\right] < 0,$$

since, by assumption, $u_{ch} > 0$. Thus, the slope of an indifference curve in (c_1,c_2)-space through $(\overline{c}_1,\overline{c}_2)$ is less steep for a high level of treatment than for a low such level.

Before we proceed with the main analysis, the following result is useful.

Lemma. *If individuals' level of inherent ability (w) is such that they for any level of labour supply, $(1-l_2) > 0$, can generate earnings greater than or equal to the marginal cost of treatment (C), then their marginal willingness to pay for treatment exceeds marginal cost of treatment, i.e., $(u_{h_2}/v_{l_2})tw + E_2/t > C$, for all t and $T(E_2)$. If, however, individuals' level of inherent ability (w) is such that they for any level of labour supply, $(1-l_2) > 0$, can generate earnings less than or equal to the expected cost of treatment (πC), then their marginal willingness to pay for treatment is equal to marginal cost of treatment, i.e., $(u_{h_2}/v_{l_2})tw + E_2/t = C$, only if $0 < t < 1$.*

Proof. Recall that $0 < l_2 < 1$, and that C is the production cost of complete treatment. If $E_2 = tw(1-l_2) \geq C$ for any level of labour supply, $(1-l_2) > 0$, it follows from equation (5.9) that marginal willingness to pay for treatment is greater than C. Now, assume that $t = 1$, $E_2 = w(1 - l_2) \leq \pi C$, and $T(E_2) = \pi C$. It then follows from the budget constraint in state 2 (*cf.* equation (5.3)) that $c_2 = E_2 - \pi C$. By the properties of u, c_2 is positive, but in order for c_2 to be positive, treatment will have to be reduced below 1, *i.e.*, $0 < t < 1$. ∎

Henceforth, healthy individuals with a level of inherent ability (earnings capacity) enabling them to generate earnings higher than, or equal to, cost of treatment at any level of labour supply, *i.e.*, $w(1 - l_i) > C$ where $i = 1, 2$, are referred to as *high-ability individuals*. Healthy individuals with a level of inherent ability enabling them to generate earnings less than, or equal to, expected cost of treatment at any level of labour supply, *i.e.*, $w(1 - l_2) \leq \pi C$, are referred to as *low-ability individuals*. High- and low-ability individuals are denoted by subscripts H and L, respectively.

5.3 Optimal Non-Linear Income Taxation and Public Supply of Medical Treatment

We confine our analysis to that of characterising *information constrained Pareto-efficient contracts*. Indeed, we do not explicitly formulate the redistributional ambitions of the government, but assume that the government has a concave social welfare function reflecting its non-negative aversion to inequality.

The government pursues a policy of *ex ante* income redistribution towards low-ability individuals. As stated earlier, information on earnings and probability of illness (π) is public, while information on inherent ability and labour supply is private. The government knows, however, that a fraction $(1 - p)$ of the individuals has a high level of inherent ability: w_H, and that a fraction p has a low level of ability: w_L. The government's problem is to design mechanisms that induce individuals to reveal their private information. It has two policy instruments at its disposal: income taxation and health care provision.

In order to induce individuals to self-select, the government specifies two types of contracts, each contract offering specific bundles of earnings: $E = (E_{1j}, E_{2j})$, consumption: $c = (c_{1j}, c_{2j})$, and medical treatment if ill: t_j, where $j = H, L$. Hence, individuals can choose from contacts in five dimensions. The contracts intended for low- and high-ability individuals are respectively given by: $\{E_{1L}, E_{2L}, c_{1L}, c_{2L}, t_L\}$, and $\{E_{1H}, E_{2H}, c_{1H}, c_{2H}, t_H\}$. The contracts are designed so as to facilitate redistribution and to induce individuals to self-select.

Individuals' expected utility function in terms of observable variables is given in equation (5.4). In order to simplify, we do not specify the expected utility function, but simply write $V_j(E_{1j}, E_{2j}, c_{1j}, c_{2j}, t_j)$, $j = H, L$. Naturally, the first-order partial derivatives of the utility function follow directly from equation (5.4).

The government's problem is to:

$$\max_{E_{1j}, E_{2j}, c_{1j}, c_{2j}, t_j} V_L(E_{1L}, E_{2L}, c_{1L}, c_{2L}, t_L),\ j = H, L$$

subject to:

$$\begin{aligned}
\underline{V}_H &\leq V_H(E_{1H}, E_{2H}, c_{1H}, c_{2H}, t_H) \\
V_H(E_{1L}, E_{2L}, c_{1L}, c_{2L}, t_L) &\leq V_H(E_{1H}, E_{2H}, c_{1H}, c_{2H}, t_H) \\
V_L(E_{1H}, E_{2H}, c_{1H}, c_{2H}, t_H) &\leq V_L(E_{1L}, E_{2L}, c_{1L}, c_{2L}, t_L) \\
0 &\leq t_j \leq 1,\ j = H, L
\end{aligned}$$

and

$$\begin{aligned}
&(1-p)\left[(1-\pi)(E_{1H} - c_{1H}) + \pi(E_{2H} - c_{2H} - t_H C)\right] \\
&+p\left[(1-\pi)(E_{1L} - c_{1L}) + \pi(E_{2L} - c_{2L} - t_L C)\right] = 0.
\end{aligned}$$

The first constraint ensures high-ability individuals a minimum level of utility. The second and third constraints are the self-selection constraints: individuals should not derive higher utility from choosing a different contract than that intended for them. The fourth constraint reflects the fact that individuals cannot more than fully recover from illness and the fifth constraint gives the government's budget constraint.

We assume that the government has no revenue requirements exceeding the sum of expected cost of treatment (*i.e.*, $\pi C(t_L + t_H)$) and expected income transfers between the different types of individuals and between health states. Moreover, we focus on what in the literature is considered to be the 'normal' case, namely that there is redistribution towards low-ability individuals, and, thus, that high-ability individuals are the potential mimickers.

Thus, the self-selection constraint and the constraint on the level of treatment are both satisfied for low-ability individuals.[12] As is standard in the literature on optimal taxation, we assume that the maximization problem has a solution and that the bundles derived from the first-order conditions are globally optimal.

Writing the Lagrangian:

$$\begin{aligned} \mathcal{L} \;=\; & V_L(E_{1L}, E_{2L}, c_{1L}, c_{2L}, t_L) + \lambda \left[V_H(E_{1H}, E_{2H}, c_{1H}, c_{2H}, t_H) - \underline{V}_H \right] \\ & + \mu \left[V_H(E_{1H}, E_{2H}, c_{1H}, c_{2H}, t_H) - V_H(E_{1L}, E_{2L}, c_{1L}, c_{2L}, t_L) \right] \\ & + \gamma \left[\begin{array}{c} (1-p)\left[(1-\pi)(E_{1H} - c_{1H}) + \pi(E_{2H} - c_{2H} - t_H C)\right] \\ + p\left[(1-\pi)(E_{1L} - c_{1L}) + \pi(E_{2L} - c_{2L} - t_L C)\right] \end{array} \right] \\ & + \phi \left[1 - t_H\right]. \end{aligned}$$

Since E_{1j}, E_{2j}, c_{1j}, c_{2j}, and t_j are all positive, it follows from the complementary slackness conditions[13] that the marginal conditions will hold as equalities.

The first-order conditions (FOCs) for the low-ability individuals' contract are:

$$\frac{\partial \mathcal{L}}{\partial E_{1L}} = \frac{\partial V_L}{\partial E_{1L}} - \mu \frac{\partial V_H}{\partial E_{1L}} + \gamma p (1-\pi) = 0 \tag{5.16}$$

$$\frac{\partial \mathcal{L}}{\partial E_{2L}} = \frac{\partial V_L}{\partial E_{2L}} - \mu \frac{\partial V_H}{\partial E_{2L}} + \gamma p \pi = 0 \tag{5.17}$$

$$\frac{\partial \mathcal{L}}{\partial c_{1L}} = \frac{\partial V_L}{\partial c_{1L}} - \mu \frac{\partial V_H}{\partial c_{1L}} - \gamma p (1-\pi) = 0 \tag{5.18}$$

$$\frac{\partial \mathcal{L}}{\partial c_{2L}} = \frac{\partial V_L}{\partial c_{2L}} - \mu \frac{\partial V_H}{\partial c_{2L}} - \gamma p \pi = 0 \tag{5.19}$$

$$\frac{\partial \mathcal{L}}{\partial t_L} = \frac{\partial V_L}{\partial t_L} - \mu \frac{\partial V_H}{\partial t_L} - \gamma p \pi C = 0 \tag{5.20}$$

[12] Indeed, the two self-selection constraints can not both be binding because of the single-crossing property.

[13] $t_j \frac{\partial \mathcal{L}}{\partial t_j}; E_{1j} \frac{\partial \mathcal{L}}{\partial E_{1j}}; E_{2j} \frac{\partial \mathcal{L}}{\partial E_{2j}}; c_{1j} \frac{\partial \mathcal{L}}{\partial c_{1j}}; c_{2j} \frac{\partial \mathcal{L}}{\partial c_{2j}} = 0$, and $\lambda \frac{\partial \mathcal{L}}{\partial \lambda}; \mu \frac{\partial \mathcal{L}}{\partial \mu}; \gamma \frac{\partial \mathcal{L}}{\partial \gamma}; \phi \frac{\partial \mathcal{L}}{\partial \phi} = 0$.

and the FOCs for high-ability individuals:

$$\frac{\partial \mathcal{L}}{\partial E_{1H}} = \lambda \frac{\partial V_H}{\partial E_{1H}} + \mu \frac{\partial V_H}{\partial E_{1H}} + \gamma(1-p)(1-\pi) = 0 \quad (5.21)$$

$$\frac{\partial \mathcal{L}}{\partial E_{2H}} = \lambda \frac{\partial V_H}{\partial E_{2H}} + \mu \frac{\partial V_H}{\partial E_{2H}} + \gamma(1-p)\pi = 0 \quad (5.22)$$

$$\frac{\partial \mathcal{L}}{\partial c_{1H}} = \lambda \frac{\partial V_H}{\partial c_{1H}} + \mu \frac{\partial V_H}{\partial c_{1H}} - \gamma(1-p)(1-\pi) = 0 \quad (5.23)$$

$$\frac{\partial \mathcal{L}}{\partial c_{2H}} = \lambda \frac{\partial V_H}{\partial c_{2H}} + \mu \frac{\partial V_H}{\partial c_{2H}} - \gamma(1-p)\pi = 0 \quad (5.24)$$

$$\frac{\partial \mathcal{L}}{\partial t_H} = \lambda \frac{\partial V_H}{\partial t_H} + \mu \frac{\partial V_H}{\partial t_H} - \gamma(1-p)\pi C - \phi = 0. \quad (5.25)$$

As in Stiglitz (1982, 1987), we define the implicit marginal tax rate for individual $j = H, L$ in state 1 by $T'(E_{1j}) = 1 - mrs^j_{c_{1j},E_{1j}}$, where $mrs^j_{c_{1j},E_{1j}} \equiv \frac{\partial V_j/\partial E_{1j}}{\partial V_j/\partial c_{1j}} = \frac{1}{w_j}(\frac{v_l(1-E_{1j}/w_j)}{u_c(c_{1j},1)})$, *cf.* equation (5.5). Correspondingly, the marginal tax rate in state 2 is defined by: $T'(E_{2j}) = 1 - mrs^j_{c_{2j},E_{2j}}$, where $mrs^j_{c_{2j},E_{2j}} \equiv \frac{1}{t_j w_j}(\frac{v_l(1-E_{21j}/t_j w_j)}{u_c(c_{2j},t_j)})$, *cf.* equation (5.6). The marginal tax rate hence *differs across health states.*

5.3.1 Self-Selection is not a Problem

For now, we assume that redistribution is *not* carried out to an extent at which the self-selection constraint on high-ability individuals is binding, *i.e.*, $\mu = 0$. The following hence constitutes a benchmark.

We start out by identifying the contract intended for low-ability individuals. The bundle of earnings and consumption to be made available if well is derived by dividing equation (5.16) by equation (5.18), and substituting in from equation (5.5):

$$-\frac{\frac{\partial V_L}{\partial E_{1L}}}{\frac{\partial V_L}{\partial c_{1L}}} = 1 \iff \frac{1}{w_L} \frac{v_{l_{1L}}(1-\frac{E_1}{w})}{u_{c1_L}(c_1, 1)} = 1.$$

Correspondingly, the bundle of earnings and consumption to be made available if sick is found by dividing equation (5.17) by equation (5.19) and substituting in from equation (5.6):

$$-\frac{\frac{\partial V_L}{\partial E_{2L}}}{\frac{\partial V_L}{\partial c_{2L}}} = 1 \iff \frac{1}{t_L w_L} \frac{v_{l_{2L}}(1-\frac{E_2}{tw})}{u_{c_{2L}}(c_2, t)} = 1.$$

Hence, the bundle of earnings and consumption intended for low-ability individuals entails a *zero marginal tax rate* in both states of the world, *i.e.*, $T'(E_{iL}) = 0$. (If, however, they generate earnings exceeding E_{iL}, they face a 100 percent marginal tax rate, $i = 1, 2$.) The marginal rate of substitution between consumption and leisure in each of the two states is, consequently, equal to the wage rate in that state, *i.e.*, $v_{l_{1L}}/u_{c_{1L}} = w_L$ and $v_{l_{2L}}/u_{c_{2L}} = t_L w_L$.

The level of medical treatment to be made available if ill, is found by dividing equation (5.20) by equation (5.17). Substituting for *mwpt*, *cf.* equation (5.9), gives:

$$\frac{u_{h_{2L}}(c_{2L}, t_L)}{v_{l_{2L}}(1 - \frac{E_{2L}}{t_L w_L})} t_L w_L + \frac{E_{2L}}{t_L} = C. \qquad (5.26)$$

Hence, the contract specifies a level of medical treatment if ill such that their marginal willingness to pay for treatment is equal to marginal cost of treatment. Note that marginal cost is also total cost since we have defined $0 < t \leq 1$. As follows from the Lemma, low-ability individuals' marginal willingness to pay for treatment is equal to cost of treatment for $0 < \widehat{t}_L < 1$ only. Thus, their health (productivity) if ill is *partly restored.*

The allocation of earnings across states is derived by dividing equation (5.16) by equation (5.17). From this we see that the marginal rate of substitution between earnings in state 1 and 2 equals the ratio of state probabilities, *i.e.*, $mrs_{E_2,E_1} = (1 - \pi)/\pi$. Substituting from equation (5.13) we thus have that:

$$v_{l_{1L}}(1 - \frac{E_{1L}}{w_L}) = \frac{1}{t_L} v_{l_{2L}}(1 - \frac{E_{2L}}{t_L w_L}), \qquad (5.27)$$

i.e., marginal utility from leisure in state 1 is equal to marginal utility from leisure in state 2 per unit of recovery in state 2. Since the contract entails $0 < \widehat{t}_L < 1$, it follows from equation (5.27) that $l_1 < l_2$, that is, *leisure is not equalized across states.* Subsequently, *earnings if healthy are higher than if ill:* $\widehat{E}_{1L} > \widehat{E}_{2L}$. Geometrically, this is illustrated in (E_1,E_2)-space by an indifference curve with slope equal to the state probabilities through ($\widehat{E}_{1L}, \widehat{E}_{2L}$). The bundle of earnings is located to the right of the 45° 'certainty line'.[14]

[14] Alternatively, we can illustrate this in (l_1, l_2)-space by an indifference with slope equal to the state probabilities through the point ($\widehat{l}_{1L}, \widehat{l}_{2L}$). The ($\widehat{l}_1$,$\widehat{l}_2$) bundle is located to the left of the 'certainty line'. (Naturally, if $t = 1$, then the optimal bundle of leisure across states would be located at the 'certainty line').

Lastly, dividing equation (5.18) by equation (5.19) and substituting from equation (5.15), we find that the level of disposable income (*i.e.*, consumption) is set so as to equalize marginal utility from consumption across states: $u_{c_{1L}} = u_{c_{2L}}$. Since $0 < \widehat{t}_L < 1$ and, by assumption, $u_{ch} > 0$, then the level of consumption is higher if in good health than if sick, *i.e.*, $\widehat{c}_{1L} > \widehat{c}_{2L}$.

We now describe the properties of the contract intended for ***high-ability individuals***. We start out by characterizing their bundle of earnings and consumption. Dividing equation (5.21) by equation (5.23), and equation (5.22) by equation (5.24), and substituting from equations (5.5) and (5.6), we find that the bundle of earnings and consumption if well and if sick, are respectively given by:

$$\frac{1}{w_H}\frac{v_{l_{1H}}(1-\frac{E_{1H}}{w_H})}{u_{c_{1H}}(c_1,1)} = 1, \tag{5.28}$$

$$\frac{1}{t_H w_H}\frac{v_{l_{2H}}(1-\frac{E_{2H}}{t_H w_H})}{u_{c_{2H}}(c_2,t_H)} = 1. \tag{5.29}$$

Consequently, also high-ability individuals face a *zero marginal tax rate*: $T'(E_{iH}) = 0$, $i = 1,2$. Earnings in excess of E_{iH} are, however, subject to a 100 percent marginal tax rate.

Next, dividing equation (5.25) by equation (5.22) and substituting from equation (5.9), we find that the bundle of treatment and earnings intended for high-ability individuals is such that:

$$\frac{u_{h_2}(c_2,t_H)}{v_{l_2}(1-E_{2H}/t_H w_H)}t_H w_H + \frac{E_2}{t_H} = C + \frac{\phi}{\gamma(1-p)}.$$

The government's budget constraint binds, *i.e.*, $\gamma > 0$. The marginal imputed cost incurred in restraining high-ability individuals' level of treatment is ϕ. According to the complementary-slackness condition, the Lagrange multiplier may take a positive or zero value. Since, by definition, high-ability individuals for all levels of t generate earnings exceeding cost of treatment, *i.e.*, $E_2/t \geq C$, it follows that $\widehat{t}_H = 1$ and $\phi > 0$. The contract thus entitles high-ability individuals to *complete medical treatment if ill.*

Moreover, dividing equation (5.21) by equation (5.22) and substituting from equation (5.13), we find that earnings are allocated across states so that individuals' labour supply is not distorted, that is: $v_{l_{1H}} = v_{l_{2H}}\frac{1}{t_H}$. From this it follows that since $\widehat{t}_H = 1$, the level of *earnings (and thus leisure) is constant across states*: $\widehat{E}_{1H} = \widehat{E}_{2H} = w_H(1 - l_H)$. Geometrically, the bundle intended for high-ability individuals is located on the certainty line, reflecting that $\widehat{E}_{2H} = \widehat{E}_{1H}$. Also, dividing equation (5.23) by equation (5.24) and substituting from equation (5.15), gives: $u_{c_{1H}}(c_1, 1) = u_{c_{2H}}(c_2, t_H)$. Thus, marginal utility from consumption is equal across states and, since $\widehat{t}_H = 1$, the level of *consumption (net income) is also constant across states*: *i.e.*, $\widehat{c}_{1H} = \widehat{c}_{2H}$.

Our findings are summarized in the following Proposition.

Proposition 1. *When self-selection is not a problem, the tax/provision scheme does not induce any distortions in the allocation of income on consumption, leisure or treatment, neither within, nor between, the two possible states of the world. In particular:*

1. *The contract intended for low-ability individuals:*

   (a) *entitles them to a cash transfer in both states of the world, i.e.,* $\widehat{c}_{iL} > \widehat{E}_{iL}$, *and partial medical treatment if ill:* $0 < \widehat{t}_L < 1$. *They are thus, partly disabled if in state 2. The cash transfers allow them to enjoy a higher level of consumption, treatment and leisure relatively to a situation without redistribution. Indeed, the cash transfer if ill constitutes partly a compensation for loss in earnings due to reduced productivity, i.e., a disability payment, and partly a cash redistributive transfer.*

   (b) *implies higher level of earnings and consumption if well than if ill:* $\widehat{E}_{1L} > \widehat{E}_{2L}$ *and* $\widehat{c}_{1L} > \widehat{c}_{2L}$. *Consequently, they enjoy a higher level of leisure if sick than if well:* $l_{2L} > l_{1L}$. *The transfer is thus less in the form of consumption and more in the form of leisure if ill than if well.*

   (c) *allows them to enjoy a higher level of expected utility relative to a situation without redistribution.*

   (d) *provides insurance against both loss in income caused by illness and medical expenditures (at the specified level of treatment:* $\widehat{t}_L$*).*

2. *The contract intended for high-ability individuals:*

   (a) *entitles them to complete medical treatment if ill:* $\widehat{t}_H = 1$. *A positive average tax is imposed in each state:* $\widehat{c}_{iH} < \widehat{E}_{iH}$. *Consequently, they have to spend more time in the labour market than they would in a situation without redistribution.*

(b) *provides for a constant level of earnings and consumption across states.*

(c) *allows for a constant level of utility across states:* $u(\widehat{c}_{1H}, 1) + v(l_{1H}) = u(\widehat{c}_{2H}, \widehat{t}_H) + v(l_{2H})$.

(d) *provides insurance against medical expenditures (at the specified level of treatment:* $\widehat{t}_H$*).*

We illustrate the information constrained efficient bundles of treatment and earnings in Figure 5.3. The bundle intended for high-ability individuals implies a high level of earnings and complete treatment, while that of low-ability individuals implies a low level of earnings and less than complete treatment.

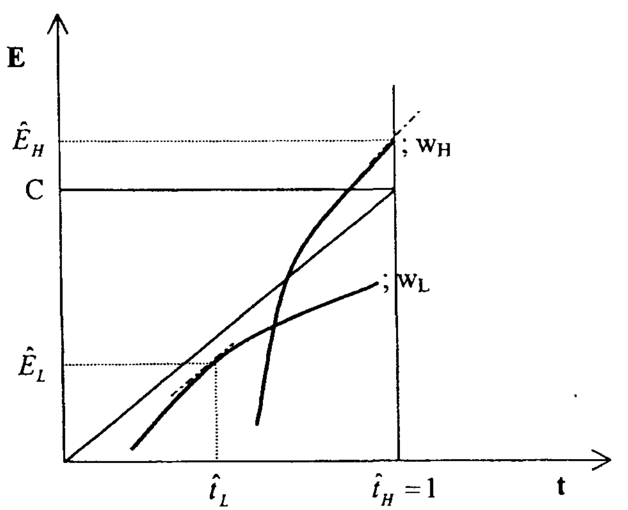

Figure 5.3: Efficient bundles of treatment and earnings.

The efficient bundles of consumption and earnings provided to high- and low-ability individuals are illustrated in Figure 5.4. Self-selection is not a problem since high-ability individuals would suffer a loss in utility by choosing the bundle intended for low-ability individuals (and low-ability individuals would not wish to choose high-ability individuals' bundle since they, in order to achieve the level of earnings ($\widehat{E}_{iH}$), would have to supply an insurmountable level of labour hours). As shown, at the designated bundles, the slope of the indifference curve is equal to 1 for both ability types. The bundles are, *i.e.*, *first-best efficient.* Consequently, the government is able to achieve *first-best efficient redistribution* by means of its integrated scheme of non-linear income taxation and health care provision.

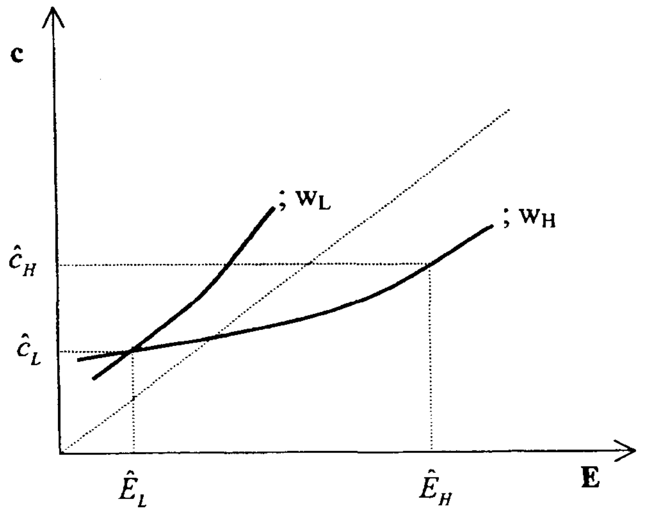

Figure 5.4: Efficient bundles of earnings and consumption.

So far, we have identified a pair of information constrained Pareto efficient bundles of earnings, consumption and treatment that facilitates redistribution towards low-ability individuals, at the given level of utility designated high-ability individuals: $\underline{V}_H$. Now, if we assume that the government's ambition is to not only construct a policy of Pareto-efficient income redistribution, but in particular, to maximize a *utilitarian welfare function*, *i.e.*, $\lambda = (1-p)/p$, then it would pursue an income distribution at which expected marginal utility from income is equal among individuals.[15] Here, individuals are assumed to be identical in all respects save their inherent level of ability, and, moreover, marginal utility from consumption is increasing in health ($u_{ch} > 0$). Then, even if marginal utility from consumption would be equalized among high- and low-ability individuals, their level of consumption would *not* be equal. This is so because their level of treat-

[15]Individual sovereignty is presumed not to be violated, hence willingness to pay has relevance.

ment (*i.e.*, health) if ill is different. Moreover, the level of leisure induced by low-ability individuals' contract is high, and may be conceived as excessive by a potential mimicking high-ability individuals. (This is so because the more productive individuals would spend even less labour time in order to generate the required level of earnings.) Hence, we cannot exclude the possibility that the contracts maximizing a utilitarian welfare function indeed prevent high-ability individuals from mimicking, *i.e.*, that self-selection is not a problem.

The government may, consequently, by means of its supply of treatment achieve redistribution towards low-ability individuals *without* causing the familiar efficiency loss. In particular, its supply of treatment enables it to reduce the level of utility required by high-ability individuals in order not to mimic (*i.e.*, $\underline{V}_H$). This is so since $mrs^m_{c,E}$ is decreasing in treatment. Thus, a differentiated provision of treatment allows the government to slacken the self-selection constraint relative to a situation in which treatment is not used as a redistributive instrument.

5.3.2 Self-Selection is a Problem

We now proceed to a situation in which the *self-selection constraint is binding, i.e.,* $\mu > 0$. Again, we start out by deriving the contract intended for low-ability individuals.

The bundle of earnings and consumption in state 1 is found by dividing equation (5.16) by equation (5.18):

$$\frac{\frac{\partial V_L}{\partial E_{1L}}}{\frac{\partial V_L}{\partial c_{1L}}} = \frac{\mu\frac{\partial V_H}{\partial E_{1L}} - \gamma p(1-\pi)}{\mu\frac{\partial V_H}{\partial c_{1L}} + \gamma p(1-\pi)}$$

$$\Leftrightarrow \quad mrs_{c_{1L},E_{1L}} = mrs^m_{c_{1L},E_{1L}} + (1 - mrs^m_{c_{1L},E_{1L}})\frac{k}{1+k},$$

where $mrs_{c_{1L},E_{1L}}$ follows from equation (5.5), $mrs^m_{c_{1L},E_{1L}} \equiv (-\frac{\partial V_H}{\partial E_{1L}}/\frac{\partial V_H}{\partial c_{1L}}) > 0$, $k \equiv \gamma p(1-\pi)/\mu\frac{\partial V_H}{\partial c_{1L}} > 0$, and superscript m denotes 'mimicker'. From equation (5.7) we know that the marginal rate of substitution between earnings and consumption is decreasing in ability, hence $mrs^m_{c_{1L},E_{1L}} < mrs_{c_{1L},E_{1L}}$. Subsequently, as follows from the above condition, $mrs^m_{c_{1L},E_{1L}} < 1$ and $mrs_{c_{1L},E_{1L}} < 1$, thus: $mrs^m_{c_{1L},E_{1L}} < mrs_{c_{1L},E_{1L}} < 1$. The contract thus entails a positive marginal tax rate: $T'(E_{1L}) = 1 - mrs_{c_{1L},E_{1L}} > 0$, implying that $v_{l_{1L}}/u_{c_{1L}} < w_L$.

Analogously, we find that the designated bundle of earnings and consumption if in state 2, entails a positive marginal tax rate: $T'(E_{2L}) = 1 - mrs_{c_{2L},E_{2L}} > 0$, implying that $v_{l_{2L}}/u_{c_{2L}} < t_L w_L$. Low-ability individuals are, therefore, faced with a *positive marginal tax rate in both states.* Indeed, their (implicit) marginal tax rate is between zero and 100 percent. The distortion originates from the problem of self-selection; the contract intended for low-ability individuals is distorted so as to reduce high-ability individuals' incentive to mimic.

Geometrically, this corresponds to a situation in which the slope of low-ability individuals' indifference curve going through the allotted bundle of $(\widetilde{E}_{2L}, \widetilde{c}_{2L})$ in (E_{2L}, c_{2L})-space is less than 1. As illustrated in Figure 5.5, the allotted bundle of $(\widetilde{E}_{2L}, \widetilde{c}_{2L})$ entails a negative total tax payment, *i.e.*, $\widetilde{E}_{2L} < \widetilde{c}_{2L}$.

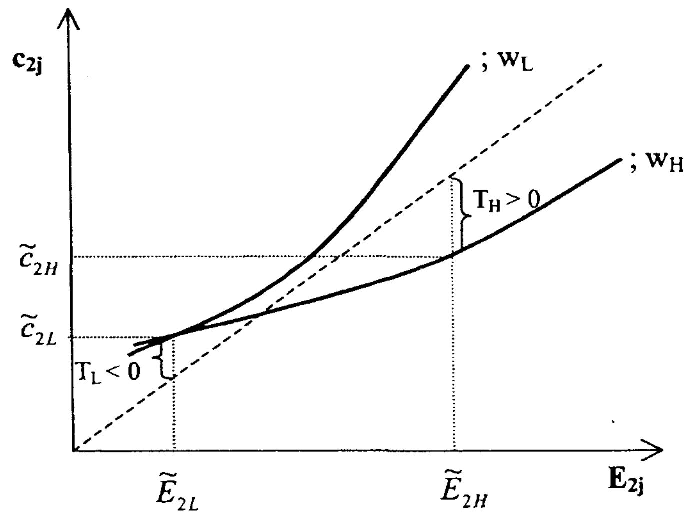

Figure 5.5: Information-constrained efficient bundles of earnings and consumption.

The bundle of earnings and treatment if ill is derived by dividing equation (5.20) by equation (5.17):

$$-\frac{\frac{\partial V_L}{\partial t_L}}{\frac{\partial V_L}{\partial E_{2L}}} = \frac{-\mu\frac{\partial V_H}{\partial t_L} - \gamma p \pi C}{\mu\frac{\partial V_H}{\partial E_{2L}} - \gamma p \pi}.$$

The left hand side gives low-ability individuals' marginal willingness to pay for treatment ($mwpt^L$), *cf.* equation (5.9). We define mimicking high-ability individuals' *mwpt* by: $mwpt^m \equiv -(\frac{\partial V_H}{\partial t_L}/\frac{\partial V_H}{\partial E_{2L}})$ and $\beta \equiv -\gamma p\pi/\mu\frac{\partial V_H}{\partial E_{2L}}$. The above condition can then be rewritten as:

$$mwpt^L = mwpt^m - (mwpt^m - C)\frac{\beta}{1+\beta}. \tag{5.30}$$

From equation (5.10) we know that marginal willingness to pay for treatment is increasing in the level of ability. Hence, $mwpt^m > mwpt^L$. Since $\beta > 0$[16], it follows from the above expression that $C < mwpt^L < mwpt^m$ (recalling that C denotes the marginal production cost of treatment). The optimal level of treatment provided to low-ability individuals is such that both mimickers' and low-ability individuals' *mwpt* exceeds cost of treatment. Indeed, the level of treatment, $\widetilde{t}_L$, is distorted *downwards* in order to discourage mimicking. Restricting t_L induces a first-order effect on the utility of high-ability individuals, while the effect on low-ability individuals is second-order. Moreover, the level of earnings if ill ($\widetilde{E}_{2L}$) is distorted downwards, implying that individuals' labour supply is reduced. The treatment/earnings bundle intended for low-ability individuals thus entails '*under-provision*' of treatment and induces '*over-consumption*' of leisure if ill. The information constrained efficient bundle of treatment and earnings is illustrated in Figure 5.6. Of course, since individuals face a positive marginal tax rate, their *mwpt* is reduced relatively to a situation without distortionary taxation.

The allocation of earnings across states is found by dividing equation (5.16) by equation (5.17):

$$\frac{\frac{\partial V_L}{\partial E_{1L}}}{\frac{\partial V_L}{\partial E_{2L}}} = \frac{\mu\frac{\partial V_H}{\partial E_{1L}} - \gamma p(1-\pi)}{\mu\frac{\partial V_H}{\partial E_{2L}} - \gamma p\pi}. \tag{5.31}$$

Again, substituting from equation (5.13) and defining $mrs^m_{E_1E_2} \equiv \frac{\partial V_H}{\partial E_{1L}}/\frac{\partial V_H}{\partial E_{2L}}$ and $g \equiv \gamma p/\mu\frac{\partial V_H}{\partial E_{2L}}$, we can rewrite this condition as:

$$mrs^L_{E_1E_2} = mrs^m_{E_1E_2} + g\left(\frac{(mrs^m_{E_1E_2}+1)\pi - 1}{1-g\pi}\right).$$

By assumption, mrs_{E_2,E_1} is increasing in ability (w), *cf.* equation (5.14). Hence, at the given level of treatment ($\widetilde{t}_L$) and at the specified level of earnings ($\widetilde{E}_{1L}, \widetilde{E}_{2L}$), then $mrs^m_{E_1E_2} > mrs^L_{E_1E_2}$. Since marginal utility from

[16]Since $\mu \geq 0, \gamma \geq 0$, and $\partial V_H/\partial E_{2L} < 0$.

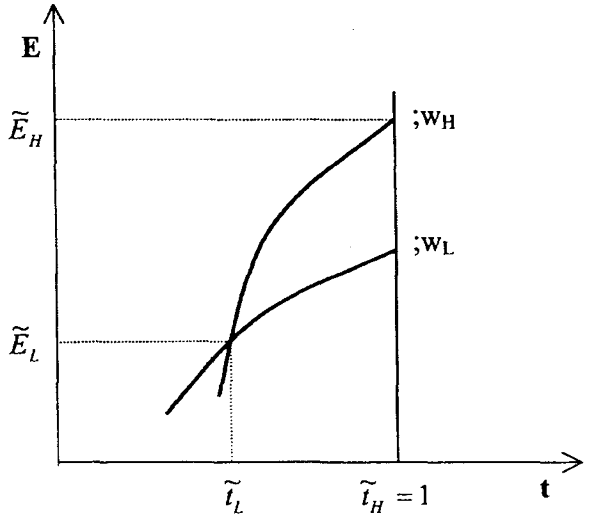

Figure 5.6: Information-constrained efficient bundles of treatment and earnings.

earnings is negative, *ceteris paribus*, then $g < 0$. It follows that in order for $mrs^m_{E_1E_2} > mrs^L_{E_1E_2}$ then $mrs^m_{E_1E_2} > (1-\pi)/\pi > mrs^L_{E_1E_2}$. Geometrically, this means that the pair of earnings-points in (E_{1L}, E_{2L})-space is below the certainty line (at which $mrs^L_{E_1E_2}$ equals the ratio of state probabilities). Hence, the contract provides *less than complete smoothing of earnings* across states, *i.e.*, $\widetilde{E}_{2L} < \widetilde{E}_{1L}$.

Additionally, as can be seen from equation (5.27), the bundle implies that low-ability individuals' marginal rate of substitution of leisure if healthy for leisure if ill, is greater than the rate or recovery (*i.e.*, treatment): $\frac{v_{l_{2L}}}{v_{l_{1L}}} > \widetilde{t}_L$. The contract intended for low-ability individuals thus *distorts the allocation of leisure across states.* Indeed, the level of earnings specified in the contract entails *excessive leisure time* in both states.

The allocation of consumption across states is found by dividing equation (5.18) by equation (5.19):

$$\frac{\frac{\partial V_L}{\partial c_{1L}}}{\frac{\partial V_L}{\partial c_{2L}}} = \frac{\mu\frac{\partial V_H}{\partial c_{1L}} - \gamma p(1-\pi)}{\mu\frac{\partial V_H}{\partial c_{2L}} - \gamma p\pi}.$$

Substituting from equation (5.15) and defining $mrs^m_{c_1c_2} \equiv \frac{\partial V_H}{\partial c_{1L}} / \frac{\partial V_H}{\partial c_{2L}}$ and $e \equiv \gamma p / \mu \frac{\partial V_H}{\partial c_{2L}}$, we get:

$$mrs^L_{c_1c_2} = \frac{mrs^m_{c_1c_2} - (1-\pi)e}{1-\pi e}$$

$$\iff mrs^L_{c_1c_2} = mrs^m_{c_1c_2} + e\left(\frac{(1+mrs^m_{c_1c_2})\pi - 1}{1-e\pi}\right).$$

As follows from the first-order derivative of equation (5.4), $mrs^L_{c_1c_2} = (1-\pi)u_{c_{1L}}/\pi u_{c_{2L}}$. Marginal utility from consumption is, by assumption, independent of leisure, *i.e.*, $u_{cl} = 0$. Thus, at the given level of treatment $\widetilde{t}_L$, low-ability and mimicking individuals' expected marginal utility from consumption is identical: $mrs^L_{c_1c_2} = mrs^m_{c_1c_2}$. Consequently, the bundle of net income across states $(\widetilde{c}_{1L},\widetilde{c}_{2L})$, implies that $mrs^L_{c_1c_2} = mrs^m_{c_1c_2} = (1-\pi)/\pi$. Geometrically, this corresponds to a situation in which an indifference curve through $(\widetilde{c}_{1L},\widetilde{c}_{2L})$ in (c_{1L},c_{2L})-space is tangent to a line showing the ratio of state probabilities. As illustrated in Figure 5.7, the bundle of net income is located to the right of the 'certainty line'. This is so because $h_{2L} = \widetilde{t}_L < 1 = h_{1L}$ and, since marginal utility from consumption is increasing in t, (*i.e.*, $u_{ch} > 0$), then $\widetilde{c}_{2L} < \widetilde{c}_{1L}$.

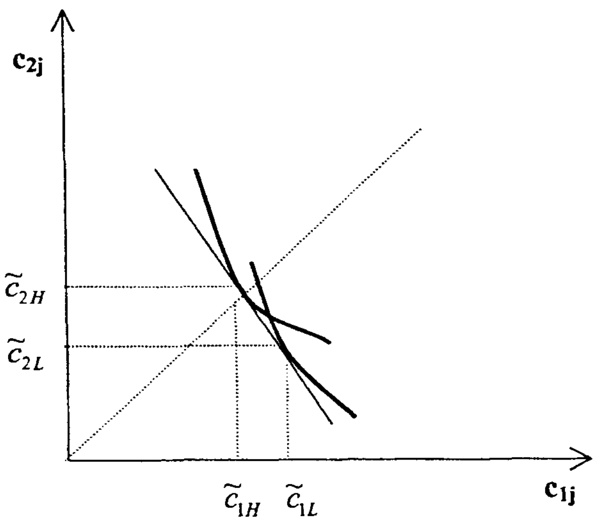

Figure 5.7: Consumption across states.

We now set out to characterize the contract intended for high-ability individuals. Their bundle of earnings and medical treatment is found by dividing equation (5.25) by equation (5.22):

$$mwpt_{E_2,t} \equiv \left(-\frac{\frac{\partial V_H}{\partial t_H}}{\frac{\partial V_H}{\partial E_{2H}}}\right) = C + \frac{\phi}{\gamma(1-p)\pi}. \tag{5.32}$$

High-ability individuals' marginal willingness to pay for treatment exceeds cost of treatment by $(\phi/\gamma(1-p)\pi)$ (indeed, they are constrained in their level of treatment since, by definition, $t \leq 1$). High-ability individuals are thus provided with *complete treatment*: $\widetilde{t}_H = 1$. This is first-best efficient. Their bundle of earnings and consumption if in state 2 is found by dividing equation (5.22) by equation (5.24). This gives that the marginal rate of substitution of earnings and consumption in each state is equal to 1, hence, $v_{l_{iH}}/u_{c_{iH}} = w_H$ (as $\widetilde{t}_H = 1$). *I.e.*, high-ability individuals face a *zero marginal tax rate*: $T'(E_{iH}) = 0$, $i = 1, 2$, and, thus, there is no distortion in their assigned bundle of consumption and earnings.[17]

Additionally, dividing equation (5.23) by equation (5.24), we find that their level of consumption across states is not distorted: $u_{c_{1H}}/u_{c_{2H}} = (1-\pi)/\pi$. By dividing equation (5.21) by equation (5.22), we find that high-ability individuals' earnings are constant across states: $\widetilde{E}_{1H} = \widetilde{E}_{2H}$. Consequently, their allocation of leisure across states is first-best efficient: $v_{l_{1H}} = v_{l_{2H}}$ for $\widetilde{t}_H = 1$.

Our findings are summarized in the following Proposition.

Proposition 2. *Information constrained efficient contracts when self-selection is a problem.*

1. *The contract intended for low-ability individuals:*

   (a) *imposes a distortion in their allocation of income on consumption and leisure across states, and in the allocation of income between consumption and treatment if ill.*

   (b) *entitles them to less than complete recovery if ill, i.e.,* $0 < \widetilde{t}_L < 1$. *Indeed, their level of treatment is distorted downwards. They are, thus, partly 'disabled' if in state 2.*

[17]Geometrically, the slope of the indifference curve in (E_{iH}, c_{iH})-space going through the bundle $(\overline{E}_{iH}, \overline{c}_{iH})$ is equal to one.

(c) *implies a non-zero implicit marginal tax rate on earnings:* $T'(E_{iL}) \neq 0$ in both states. *Their average tax payment is negative:* $\widetilde{E}_{iL} < \widetilde{c}_{iL}$.

(d) *entails a low level of earnings in both states. This implies a downward distortion in their labour supply in both states.*

(e) *gross earnings and disposable income if sick are less than that if healthy.*

(f) *implies that utility is not constant across states:* $u(\widetilde{c}_{1L}, 1) + v(1 - \widetilde{E}_{1L}/w_L) > u(\widetilde{c}_{2L}, \widetilde{t}_L) + v(1 - \widetilde{E}_{2L}/\widetilde{t}_L w_L)$. *Hence, despite receiving a net transfer from the government, they will suffer a loss in utility if ill.*

2. *The contract intended for high-ability individuals:*

(a) *does not introduce any distortions in the allocation of income on consumption and leisure across states, or in the allocation of income between consumption and treatment if ill.*

(b) *entitles them to complete recovery, i.e.,* $\widetilde{t}_H = 1$, *and obliges them to generate high earnings. Thus, if ill, they are still 'able'.*

(c) *implies a zero implicit marginal tax rate, i.e.,* $T'(E_{iH}) = 0$, *but a positive average tax, i.e.,* $\widetilde{E}_{iH} > \widetilde{c}_{iH}$. *They are, thus, 'net contributors' to the public sector.*

(d) *earnings are identical across states,* $\widetilde{E}_{1H} = \widetilde{E}_{2H}$, *as is net income,* $\widetilde{c}_{1H} = \widetilde{c}_{2H}$.

(e) *provides for constant utility across states:*
$u(\widetilde{c}_{1H}, 1) + v(1 - \widetilde{E}_{1H}/w_H) = u(\widetilde{c}_{2H}, 1) + v(1 - \widetilde{E}_{2H}/w_H)$.

Hence, the information constrained Pareto-efficient contracts are given by $\{\widetilde{E}_{1L}, \widetilde{E}_{2L}, \widetilde{c}_{1L}, \widetilde{c}_{2L}, \widetilde{t}_L\}$ where $\widetilde{E}_{2L} < \widetilde{E}_{1L}$, $\widetilde{c}_{2L} < \widetilde{c}_{1L}$, $\widetilde{t}_L < 1$, and $\{\widetilde{E}_{1H}, \widetilde{E}_{2H}, \widetilde{c}_{1H}, \widetilde{c}_{2H}, \widetilde{t}_H\}$ where $\widetilde{E}_{1H} = \widetilde{E}_{2H}$, $\widetilde{c}_{1H} = \widetilde{c}_{2H}$ and $\widetilde{t}_H = 1$.

It follows that the efficiency conditions are distorted for only one of the ability types.[18] Stiglitz (1987) refers to this as a property of Pareto efficient

[18]Our finding resembles the results found in the literature on asymmetric information in insurance markets where high- and low-risk individuals are offered different contracts and where low-risk individuals are constrained in the extent of coverage offered, see for example Rothschild and Stiglitz (1976) who address the private market mechanism.

tax structures in a second best world. Notice that the efficient contracts entail that high-ability individuals can fully restore their ability if ill, *i.e.*, they are *'able'*, while low-ability individuals can restore health only partly, thus they are (partly) *'disabled'*.

The derived contracts induce individuals to self-select. Indeed, high-ability individuals trade off leisure time to 'signal' their ability type in order to receive complete medical treatment. Public provision of health care hence provides an instrument for screening. What are the consequences of integrating income taxation and differentiated public provision for the extent of income redistribution possible, and how does this compare with alternative redistributional schemes? This is the subject of the following.

5.4 Medical Treatment as a Screening Device

In the literature on optimal non-linear taxation under asymmetric information, it is well established that public provision of private goods may enhance redistribution. The good in question is traditionally assumed to be a consumption good for which individuals' preferences are non-separable from leisure. It is established that when the good in question enters the utility function as a consumption good only, then separability implies that in-kind transfers (or commodity taxation) is redundant. In this paper, we have shown that public provision may still have a role to play even if preferences are separable in consumption and leisure. This is so because the publicly provided good, *i.e.*, health, alters individuals' productivity in the labour market if ill, moreover, it affects individuals' marginal utility from consumption of other goods.[19] We will in the following discuss more explicitly how public provision of treatment enhances redistribution when information is asymmetric. In particular, we describe what will be the effect of reducing the available level of medical treatment on the extent of redistribution possible. The argument is along the lines of those of Boadway and Keen (1993) and Edwards *et al.*, (1994).

From equation (5.10) we know that marginal willingness to pay for treat-

[19] Provision of goods influencing individuals productivity has also been studied by Boadway and Marchand (1995). They analyze uniform public provision of education assuming that individuals derive utility from education indirectly through its effect on individuals' wage rate. Moreover, they assume that preferences are non-separable in consumption and leisure.

ment is higher for a mimicker than a low-ability individual:

$$mwpt^L < mwpt^m.$$

If now the level of treatment intended for low-ability individuals (t_L) is reduced infinitesimally below the efficient level as derived in Section 5.3.1, then mimickers' earnings liability decreases by $mwpt^L$, which is less than the reduction in $mwpt^m$, *i.e.*, their valuation of the loss in treatment. Thus, reducing treatment has a first-order effect on high-ability individuals' *mwpt* while the effect on low-ability individuals' *mwpt* is second-order. Mimicking is consequently made less attractive, implying that the self-selection constraint is relaxed. Thus, it is possible to change the optimal income taxation so as to improve the welfare of low-ability individuals, or in other words, reduce the reservation utility of high-ability individuals ($\underline{V}_H$). Analogously to Boadway and Keen's (1993) analysis of public goods, the rule for optimal provision involves '*under-provision*' (relatively to the Samuelson rule) since mimickers' *mwpt* is greater than that of low-ability individuals:

$$mwpt^L > C \text{ when } mwpt^L < mwpt^m.$$

This is in fact what we derived in Section 5.3.2, *cf.* equation (5.30), namely that low-ability individuals are restrained in the level of treatment available. Since, by assumption, high-ability individuals' *mwpt* exceeds cost of full treatment (*cf.* Lemma), then the government can subtract (some) of their 'consumer surplus' by restricting the level of treatment available to low-ability individuals.

Proposition 3. *Restraining the level of treatment available to sick low-ability individuals enables the government to slacken the self-selection constraint. In particular, it:*

- *enables it to deduce higher tax revenues from high ability individuals by compelling them to generate a high level of earnings and by constraining their level of consumption in both states.*
- *allows it to subtract more of high-ability individuals' 'consumer-surplus' from treatment.*
- *facilitates larger transfers towards low-ability individuals in both states.*

We have established that the slope of an indifference curve in (c_2, E_2)-space is decreasing in t, *cf.* equation (5.9). Hence, the slope of a mimicker's indifference curve in (c_2, E_2)-space changes relative to that of a 'non-deceiving' high-ability individual. This is so because the bundle of earnings and consumption intended for low-ability individuals is specified for a particular level of treatment, namely $\widetilde{t}_L$. Consequently, a mimicker will not retrieve her full productivity if ill and will, *ceteris paribus*, have to spend relatively more time in the labour market in order to generate earnings E_{2L}^m than she would for a higher t_L. Thus, her level of utility is lower. Also, low-ability individuals' level of utility is higher when the level of treatment t_L is reduced since more redistribution is carried out. Low-ability individuals' designated earnings are lower, implying more leisure time and, thus, higher $v(l_{iL})$. Their designated level of consumption if ill $(\widetilde{c}_{2L})$ is, however, lower due to the reduction in treatment (recalling that $u_{ch} > 0$), thus $u(\widetilde{c}_{2L}, \widetilde{t}_L)$ is lower. The increase in utility from more leisure, $v(l_{2L})$, more than compensates for the loss in utility from lower health and consumption if ill. Indeed, as follows from the utility function, individuals are more concerned (in ordinal terms) about leisure than consumption when health is poor than when health is good. In other words, the marginal rate of substitution of leisure for consumption, $\partial mrs_{lc}/\partial h$, decreases if health decreases. Consequently, low-ability individuals' expected utility V_L increases.[20]

As follows from the above, a scheme of uniform public provision of treatment allows for less redistribution than does a scheme of differentiated provision, since the incentive to mimic is stronger. Consequently, $\underline{V}_H$ is higher when treatment is uniform than if treatment is differentiated (*cf.* Section 5.3.1) which again is higher than if treatment is restrained (*cf.* Section 5.3.2).

Proposition 4. *Public provision of a uniform level of medical treatment does not Pareto dominate a scheme with differentiated public provision.*

Geometrically, this is illustrated by high-ability individuals' indifference curve in (c_2, E_2)-space showing the minimum level of utility required in order

[20] Our findings runs counter to those of Boadway and Marchand (1995) who show that the self-selection constraint may be slacken by compelling individuals to '*over-consume*' education which is uniformly provided. In our model of health care provision, the self-selection constraint is relaxed by forcing low-ability individuals to '*under-consume*' treatment.

not to mimic, is more flat when the level of treatment intended for low-ability individuals is restrained. Moreover, it is located above the corresponding curve in a scheme with uniform provision. Low-ability individuals' level of utility is, consequently, lower when treatment is uniformly provided. Thus, low-ability individuals are relatively better off when they are constrained in their level of treatment if ill. This finding may seem counterintuitive: low-ability individuals' expected utility is higher in a scheme where health is poorer. The reason for this result is as follows. The income transfer allows low-ability individuals to reduce their labour supply and, thus, to increase utility from leisure if ill to an extent at which the loss in utility due to lower health and consumption is more than compensated for.

5.4.1 The Utility Possibilities Frontier

Individuals are assumed to have strictly concave utility functions. It may thus be reasonable to infer that the (*ex ante*) utility possibilities frontier (*upf*) is also concave. The *upf* may in a situation without asymmetric information, henceforth called first-best, be depicted as in Figure 5.8. The distribution of utility among the two types of individuals *prior* to redistribution is indicated by point H on the frontier. In a world of symmetric information, redistribution may take the form of lump sum transfers, hence, any point on the *upf* is attainable. In the present context, however, the government faces problems in identifying individuals. If the self-selection constraint on high-ability individuals does not bind, then the government can through its menu of contracts as described in Section 5.3.1, change the distribution of utility from point H to point K on the first-best frontier. The second-best efficient allocation consequently coincides with the first-best efficient allocation. If, on the other hand, the self-selection constraint does bind, then the redistributional policy imposes a distortion. Indeed, low-ability individuals are constrained in their level of treatment and earnings in order to restrain high-ability individuals from mimicking. When self-selection is a problem, we thus have the familiar equity-efficiency trade-off. For a small degree of redistribution, the distortion, caused by the self-selection mechanism, will have a second-order effect on low-ability individuals. Naturally, the larger the extent of redistribution, the more 'severe' is the distortion.

Since, by assumption, individuals have smooth and convex indifference curves, we can for different levels of $\underline{V}^H$ (and for a given λ) depict a smooth second-best frontier that coincides with the first-best frontier over a range (from H to K), but then diverges from the first-best frontier and is located inside the first-best frontier. The second-best *upf*, illustrated by the heavily drawn curve, will not intersect the 45° line, but rather be located to the right of the 45° line. The intuition for this is that high-ability individuals in a self-selection equilibrium will, by choosing the contract intended for low-ability individuals, derive a higher level of utility in both states of the world, than do low-ability individuals (since, at the given level of treatment and earnings, they can enjoy a higher level of leisure in both states of the world). The separating bundles (as illustrated by point S) can also be thought to describe the maximum level of redistribution achievable, that is; it represents a 'maxi-min' solution.

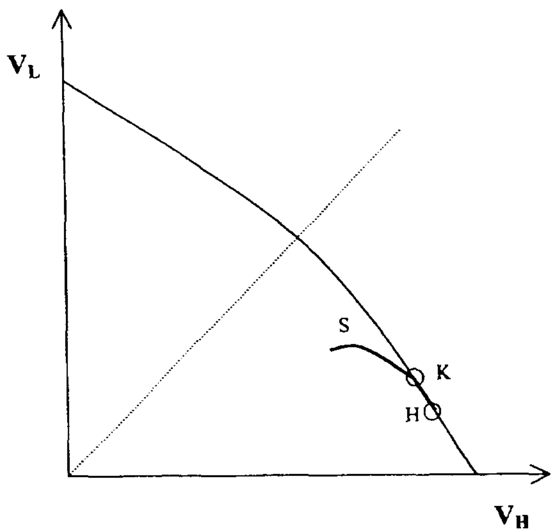

Figure 5.8: The utility possibilities frontier.

5.5 Discussion

We have shown that a public supply of health care may be justified on the grounds of *efficiency*. The derived contracts may also be justified on the grounds of *ex post redistributional justice*. The argument is as follows. Individuals may face uncertainty, not only with respect to their state of health, but also with respect to their inherent level of ability. If this is so, then the preceding analysis can be considered to take place in the 'interim'. Contracts would now be offered *prior* to individuals knowing their innate ability, hence, they would have to make their decision behind a 'veil of ignorance'. In this situation, individuals would prefer to maximize their *ex ante* average utility, which is indeed the utilitarian objective.[21]

The government may pursue its redistributional ambitions without reference to the state of the world if insurance markets are complete. However, perfect insurance markets do not exists. Reasons for this is asymmetric information leading to problems of adverse selection and moral hazard, transaction costs, *etc.* Redistribution may, hence, be motivated by a desire for (social) insurance. As Varian (1980, p. 51) puts it:

> "Indeed, the fact that redistributive taxation helps to insure against individual risk is a common justification for redistributive programs."

In the literature on redistributive taxation focus may, according to Varian (1980), be placed on (i) the equity effect of altering the income distribution, (ii) the efficiency effects from reducing incentives, and (iii) the insurance effects from reducing the variance in income streams. The insurance aspect of redistribution of income has, however, received less attention than for instance equity and efficiency aspects of taxation.

In this paper, we assume that private insurance markets do not exist and we show that non-linear income taxes and public provision of health care facilitates redistribution both between different income groups and across different health states. The tax/provision scheme thus constitutes an instrument for insurance. Indeed, the derived redistributive insurance scheme defines prices (taxes) and quantities (level of treatment).

Our analysis is based on a stylized model in which the government does not pursue a particular distribution of health *per se*. However, one may argue that health is a 'particular' commodity for which the society has specific egalitarian ambitions, *i.e.*, that health is subject to what Tobin (1970)

[21]Utilitarianism may also be supported by a different set of arguments. According to Roemer (1996), Harsanyi introduced in his 1955 paper on "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility", three axioms for social choice under uncertainty, and showed that the social welfare function is given by the sum of individuals' expected utilities.

calls specific egalitarianism. Furthermore, our analysis ignores possible correlation between inherent ability and the probability of falling ill. It seems natural to extend the analysis to include also differences in the probability of illness. This will be the subject of future research. We still think that our analysis may offer a contribution to the literature on the public provision of private goods in which health care is perhaps the most outstanding example.

Acknowledgments

I am grateful to Vidar Christiansen and Agnar Sandmo for their valuable comments and suggestions. Helpful comments from seminar participants at Programme for Health Economics in Bergen (HEB), and seminar participants at Norwegian School of Economics and Business Administration, are much appreciated.

References

Anderberg, D. (2001). "Social Insurance with In-kind Provision of Private Goods." *Scandinavian Journal of Economics,* 103(1), 41-61.

Arrow, K.J. (1974). "Optimal Insurance and Generalized Deductibles." Scandinavian Actuarial Journal, 1-42; reprinted in: *Collected Papers of Kenneth J. Arrow, Vol 3: Individual Choice under Certainty and Uncertainty.* Oxford: Blackwell (1984), 212-260.

Atkinson, A.B. (1987). "Income Maintenance and Social Insurance." In A.J. Auerbach and M. Feldstein (eds.) *Handbook of Public Economics*, 2, 779-908. North-Holland, Amsterdam.

Atkinson, A.B. and J.E. Stiglitz. (1987). *Lectures on Public Economics.* McGraw-Hill, Singapore.

Besley, T. and S. Coate. (1991). "Public Provision of Private Goods and The Redistribution of Income." *The American Economic Review*, 81(4), 979-984.

Blackorby, C. and D. Donaldson. (1988). "Cash Versus Kind, Self-Selection, and Efficient Transfers." *The American Economic Review*, 78(4), 691-700.

Blomquist, S. and V. Christiansen. (1995). "Public Provision of Private Goods as a Redistributive Device in an Optimum Income Tax Model." *Scandinavian Journal of Economics*, 97 (84), 547-567.

Blomquist, S. and V. Christiansen. (1998). "Price Subsidies Versus Public Provision." *International Tax and Public Finance*, 5(3), 283-306.

Blomquist, S. and V. Christiansen. (1998). "Topping Up or Opting Out? The Optimal Design of Public Provision Schemes." *International Economic Review*, 39(2), 399-411.

Blomqvist, Å. and H. Horn. (1984). "Public Health Insurance and Optimal Income Taxation." *Journal of Public Economics*, 24(3), 353-371.

Boadway, R. and M. Keen. (1993). "Public Goods, Self-Selection and Optimal Income Taxation." *International Economic Review*, 34(3), 463-478.

Boadway, R. and M. Marchand. (1995). "The Use of Public Expenditures for Redistributive Purposes." *Oxford Economic Papers*, 47(1), 45-59.

Boadway, R., M. Marchand and M. Sato. (1998). "Subsidies versus Public Provision of Private Goods as Instruments for Redistribution." *Scandinavian Journal of Economics*, 100(3), 539-564.

Breyer, F. and A. Haufler. (2000) "Health Care Reform: Separating Insurance from Income Redistribution." *International Tax and Public Finance*, 7(4-5), 445-461.

Calcott, P. (2000). "Health Care Evaluation, Utilitarianism and Distortionary Taxes." *Journal of Health Economics*, 19(5), 719-730.

Christiansen, V. (1998). "Public Provision and Distortionary Taxes: Two Second Best Regimes." Paper prepared for Symposium in honour of Agnar Sandmo, NHH 9-10 January, 1998.

Cremer, H. and F. Gahvari. (1997). "In-Kind Transfers, Self-selection and Optimal Tax Policy." *European Economic Review*, 41(1), 97-114.

Diamond, P.A. and J.A. Mirrlees. (1971). "Optimal Taxation and Public Production, I: Production Efficiency." *The American Economic Review*, 61(1), 8-27.

Edwards, J. M. Keen and M. Tuomala. (1994). "Income Tax, Commodity Taxes and Public Good Provision: A Brief Guide." *FinanzArchiv*, 51(4), 472-487.

Evans, W.N. and W.K. Viscusi. (1991). "Utility-Based Measures of Health." *American Journal of Agricultural Economics*, 73 (5), 1422-1427.

Hirshleifer, J. and J.G. Riley. (1992). *The Analytics of Uncertainty and Information.* Cambridge Surveys of Economic Literature. Cambridge University Press.

Mazur, M.J. (1989). "Optimal Linear Taxation with Stochastic Incomes." *Public Finance*, 44 (1), 31-50.

Myles, G.D. (1995). *Public Economics.* Cambridge University Press.

Nichols, A.L. and R. J. Zeckhauser. (1982). "Targeting Transfers through Restrictions on Recipients." *The American Economic Review*, 72(2), 372-377.

Petretto, A. (1999). "Optimal Social Health Insurance with Supplementary Private Insurance." *Journal of Health Economics*, 18(6), 727-45.

Rochet, J-C. (1991). "Incentives, Redistribution and Social Insurance." *The Geneva Papers on Risk and Insurance Theory*, 16(2), 143-65.

Roemer, J.E. (1996). *Theories of Distributive Justice.* Harvard University Press.

Rothschild, M. and J.E. Stiglitz. (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics,* 90(4), 629-649.

Seade, J. (1982). "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies,* 49(4), 637-643.

Seip, A.L. (1984). *Sosialhjelpstaten blir til : Norsk Sosialpolitikk 1740-1920.* Gyldendal, Oslo.

Sen, A. (2000). "Social Justice and the Distribution of Income." In A.B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution*, 1, 59-85. Elsevier Science B.V.

Stiglitz, J.E. (1982). "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics*, 17(2), 213-240.

Stiglitz, J.E. (1987). "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." In: A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics,* 2, 991-1041. North Holland.

Stiglitz, J.E. (1977). "Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market." *The Review of Economic Studies*, 44(3), 407-430.

Tillmann, G. (1989). "Equity, Incentives, and Taxation." Lecture Notes in Economics and Mathematical Systems. Managing editors: M. Beckmann and W. Krelle. Springer-Verlag.

Tobin, J. (1970). "On Limiting the Domain of Inequality." *Journal of Law and Economics*, 13(2), 263-277.

Varian, H.R. (1980). "Redistributive Taxation as Social Insurance." *Journal of Public Economics*, 14(1), 49-68.

Viscusi, W.K. and W.N. Evans. (1990). "Utility Functions That Depend on Health Status: Estimates and Economic Implications." *American Economic Review*, 80 (3), 353-375.

Zeckhauser, R. (1970). "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory*, 2(1), 10-26.