Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis International Bachelor Economics and Business Economics

Optimal non-linear taxation and social insurance with ex ante moral hazard*

July 25, 2018

Author: Olivier Kooi

Student Number: 428738

Supervisor: Bas Jacobs

Second assessor: Kevin Spiritus

Abstract

Should the government include equity concerns in its health insurance policies? I study a model where individuals differ in productivity and cost of preventive care. These types are unobservable to government, introducing an equity-efficiency trade off and a moral hazard problem. Solving this model yields three new insights. First, I show that the government should differentiate effective tax rates between healthy and ill individuals. Second, I show that the government can differentiate effective tax rates through the provision of an income-dependent social insurance policy. Third, I examine the robustness of some of previous results in the literature and show they are not fully robust to a different timing of the labour supply decision.

*I would like to thank my thesis supervisor, Bas Jacobs, for his helpful comments and suggestions. All remaining errors are my own.

1 Introduction

Should the government use income-dependent health policies to redistribute income when it can also use tax and transfer policies? Governments have long intervened in the provision of healthcare, either by providing it directly or by heavily regulating insurance markets. Economists have studied the efficiency arguments for doing so. Many of the classic market failure arguments can be traced back to Arrow (1963) who explains why healthcare is unlike most other markets. Further strengthening the general case for intervention in health insurance markets were Rothschild and Stiglitz (1978), who showed that under asymmetric information insurance markets could collapse completely. Although the case for government intervention in the provision of healthcare in order to increase the efficiency of provision is clear, it is less obvious why there should be as large an equity component to this intervention as we typically see in developed countries.

One of the most significant insights of *The Theory of Public Finance* (Musgrave, 1959) was the conceptual separation between the allocative and redistributive branches of government. A government would set up two departments, which would rarely need to communicate. The allocative branch would ensure that , within an economy, all gains from trade were exploited. The redistributive branch would then implement the income distribution considered just by the government. In the context of public health insurance and redistribution this separation would imply, that the allocative branch intervenes in healthcare markets to address market failures due to moral hazard and asymmetric information. The redistributive branch can redistribute income as desired by the government through a tax system. There would be no need no need for equity considerations in the design of healthcare policies.

This separation of tasks is not what we see in practice in the design of healthcare policies. Indeed, many European policy-makers take an egalitarian viewpoint of healthcare provision, in which healthcare is provided according to need and financed according to ability to pay (Wagstaff & Van Doorslaer, 1993).

To study whether a public health insurance programme should be income-dependent when the government can also use an income tax, I extend the two-productivity type model of Stiglitz (1982) with two health types. Individuals are heterogeneous with respect to their productivity as well as their cost of effort on preventive care. Individuals are either healthy or ill. If an individual falls ill, he incurs an exogenous healthcare cost. Individuals make a discrete choice between two levels of preventive effort before they know their health state. The cost of obtaining these levels differs between workers according to their health type. After the health state is revealed workers decide on the amount of labour they supply, earn their income and consume it. The government does not observe the individual's type or labour and effort choices. It only observes the individual's production and consumption choices and their health state. The government maximises social welfare by offering a non-linear tax schedule, that can be conditioned on production level and health state. Because the government does not directly observe effort, there is a problem of moral hazard. If the government were to provide full insurance, there would be no incentive for individuals to exert any effort to remain healthy. The inability of the government to the observe productivity type implies it has to resort to distortionary taxation. This introduces a trade off between equity and efficiency which has been extensively studied in the optimal tax literature. I show how both the moral hazard problem and the equity-efficiency trade off affect optimal policy. This yields several novel insights.

First, I show how tax rates should differ with health state and productivity level. I show that optimal tax rates on the healthy individuals are lower than on the ill individuals. Intuitively, by raising the marginal tax rate of the ill above that of the healthy, the government can redistribute more income from ill high-income individuals to ill low-income individuals. Additionally, by levying a higher marginal tax rate on the unhealthy low-income individuals, the government raises more revenue from the unhealthy high-income individuals. This reduces the moral hazard problem by making it more costly to high-income individuals to fall ill. This allows the government to raise the amount of social insurance it provides, thereby raising the transfer from the healthy towards the ill.

Secondly, I show how the state-dependent tax policy described above can be implemented through the use of a non-linear income tax and an income-dependent social insurance programme. Intuitively, the government can implement the state-dependent tax schedule by offering all individuals the same non-linear income tax schedule, complemented by an income-dependent contribution to healthcare cost. By reducing the coinsurance rate of healthcare cost as income rises, the government raises the effective tax rate faced by ill individuals. Since only the ill incur healthcare cost, this effectively implements a state-dependent non-linear tax schedule.

Lastly, I shed some new light on earlier results by Rochet (1991) and Cremer and Pestieau (1996) who analyze a non-linear income tax with social insurance in a setting where the labour supply decision is made before the health state is realised. I analyze a simpler version of my model that excludes the moral hazard problem to show that the full social insurance result of Rochet (1991) and Cremer and Pestieau (1996) is mostly robust to a setting in which individuals already know their health state when they make their labour supply decision. The difference between our results is that the social insurance policy needs to be income-dependent. The intuition is that the government can generally use the health state to infer some information about an individual's productivity type. By using this information, the government can improve the equity-efficiency trade off and hence it generally makes the social insurance programme income-dependent as a way to use the information it obtains from the health state it observes, thereby improving the equity-efficiency trade off it faces.

The remainder of this paper is structured as follows. Section 2 gives an overview of the literature this paper contributes to. In section 3 I introduce and solve several versions of the main model. Subsection 3.1 considers the first-best case. Subsection 3.2 analyses the second-best without ex ante moral hazard. Subsection 3.3 considers the full model with ex ante moral hazard. Subsection 3.4 shows how the government can implement health state-dependent tax schedules through an income-dependent social insurance policy. Finally, Section 4 concludes.

2 Related Literature

This paper relates to several strands of the literature.

First, it closely relates to the literature on the dual problem of income redistribution and the provision of social insurance. The papers by Blomqvist and Horn (1984), Rochet (1991) and Cremer and Pestieau (1996) consider optimal taxation and social insurance in settings where individuals differ in their ability to earn income, as in Mirrlees (1971), and in their risk of illness. Individuals choose their labour supply before they know their health state. There are no private insurance markets and the government is the only provider of insurance. The government chooses an income tax schedule and social insurance policy to maximise social welfare. The optimal amount of social insurance is generally positive even in the presence of an optimized non-linear tax schedule. In both Rochet (1991) and Cremer and Pestieau (1996) the government provides full social insurance if there is a negative correlation between risk and labour productivity. In this paper, I add to this strand of the literature in three ways. First, I show that the timing of the labour supply decision matters for the optimal social insurance policy. The literature considers cases where labour supply decisions are made before individuals know their health state. This paper considers the case when labour supply decisions are made when individuals do know their health state. I show that this timing difference matters and implies it is optimal to introduce income-dependencies in the social insurance programme. Second, I consider social insurance with a larger degree of heterogeneity. In the literature it is assumed that all individuals of a certain level of labour productivity also share the same risk of illness. This paper relaxes that assumption. Third, I introduce an ex ante moral hazard problem. In most of the literature health risk is assumed to be exogenous. In this paper I allow health risks to be endogenous and show that this breaks down the full social insurance result of Rochet (1991) and Cremer and Pestieau (1996).

This paper also relates to a second strand of the literature which studies the optimal tax policy and social insurance policy in the presence of private insurance markets (e.g. Boadway et al (2003, 2006)) and Chetty & Saez (2010)). These papers consider linear income taxes and linear social insurance policies with health insurance markets that fail due to adverse selection and moral hazard. They show that the government generally provides a non-zero, generally positive level of social insurance. The exception is Boadway et al (2003), where it is ambiguous whether social insurance functions as a tax or subsidy even when health risk and income are negatively correlated. This paper does not directly contribute to this literature as it does not model insurance markets. However, it does consider the role of moral hazard in the design of tax and social insurance policies, thereby contributing indirectly to this strand of the literature.

3 Model

Consider an economy populated by a continuum of individuals and a government. Individuals make decisions in two periods. In the first period they decide on the level of preventive effort which affects the probability they fall ill in the second period. In the second period, health states are realised and individuals make a labour supply decision. The government maximises social welfare by redistributing income between productivity types and individuals in different health states. It does so by announcing a tax and transfer policy at the start of the first period, before individuals choose preventive effort, and implementing it in the second period.

Individuals

Individuals differ in two dimensions: their productivity and the cost of preventive care. There are two distinct levels of labour productivity. An individual of type $n \in \{n_L, n_H\}$ where $n_H > n_L$ can produce z units of output by providing $\frac{z}{n}$ units of labour. Cost of preventing illness refers to the utility cost of obtaining a level of preventive care ϕ_θ. To obtain preventive care ϕ_θ ,an individual of type θ needs to provide $\frac{\phi}{\theta}$ units of effort e. Individuals face one of two levels of cost for preventive care $\theta \in \{\theta_1, \theta_2\}$ where $\theta_2 > \theta_1$.This means that the cost of preventive care is lower for the type θ_2 individual than for the type θ_1. I denote by $f_{n\theta}$ the mass of individual of productivity n that face cost of preventive care θ. The mass of all individuals is normalised to unity. The probability of being healthy is denoted as π. The likelihood of remaining healthy depends on the preventive care level an individual chooses. There are two such levels, $\pi_1 \equiv \pi(\phi_1)$ and $\pi_2 \equiv \pi(\phi_2)$, which are optimal for type θ_1 and θ_2 respectively, in the absence of government policy. This two-health level model is similar to the two-job model by Diamond and Spinnewijn (2011). Individuals choose their output level and their level of effort to maximise expected utility:

$$EU(z_n^g, z_n^b, e) = \pi(\phi_\theta) \cdot u\left(c_n^g, \frac{z_n^g}{n}\right) + [1 - \pi(\phi_\theta)] \cdot u\left(c_n^b, \frac{z_n^b}{n}\right) - H(e)$$

where superscripts $s \in \{b, g\}$ denote whether the individual is in either a good (healthy) or a bad (ill) health state. The utility-cost of exerting effort is captured by function $H(e)$ which is identical for all individuals, increasing and strictly convex. In each state, the individual derives sub-utility $u\left(c_n^s, \frac{z_n^s}{n}\right)$ from consumption c_n^s and labour supply $\frac{z_n^s}{n}$. The sub-utility function has the following form: $u\left(c_n^s, \frac{z_n^s}{n}\right) = u\left(c_n^s - v\left(\frac{z_n^s}{n}\right)\right)$. Sub-utility function $u(\cdot)$ is identical in both states and for all individuals, increasing, strictly concave and twice continuously differentiable. Disutility of production is $v\left(\frac{z_n^s}{n}\right)$, which is increasing, strictly convex and thrice continuously differentiable. This functional form for the preferences is convenient as they allow for risk aversion while eliminating income effects (Greenwood et al, 1988). Consumption and production are related through state specific budget constraints: $c_n^g = z_n^g - t^g(z_n^g)$ and $c_n^b = z_n^b - t^b(z_n^g) - C$. Where C is the exogenous cost the individual incurs if he falls ill and $t^s(z_n^s)$ is the non-linear state-dependent tax levied by the government.

The individual chooses his optimal production and effort levels in two steps. First, before he knows his health state, he chooses the level of effort on preventive care, taking into account his type and the tax schedule announced by the government. Second, after he knows his health state, he chooses the optimal amount of production and consumes the income this yields him.

Government

The government is assumed to maximise a utilitarian social welfare function:

$$W = \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta) \cdot u\left(c_n^g, \frac{z_n^g}{n}\right) + [1 - \pi(\phi_\theta)] \cdot u\left(c_n^b, \frac{z_n^b}{n}\right) - H(e)] \tag{1}$$

This utilitarian approach implies that the government is maximizing the sum of individual expected utilities. The drawback of this welfare function is that there is no concern for the distribution of utilities, which might be relevant in the case of health risk. However, introducing preferences over the distribution of utilities, in the the presence of risk, requires one to choose between rationality and the Pareto principle by Harsanyi's (1955) aggregation theorem. Fleurbaey (2010) describes two contenders to the approach offered here. Ex ante egalitarianism requires the government to be inequality averse over expected utilities and is advocated by Diamond (1967) among others. A drawback of ex ante egalitarianism, as explained by Fleurbaey (2010), is that a shift in the distribution of ex-post utilities to one that is more unequal would not matter to the ex ante egalitarian, provided that the distribution of expected utilities did not change. Another option, which seems particularly appealing in the case of healthcare, is ex post egalitarianism and is defended by Adler and Sanchirico (2006). The drawback of this approach is that the Pareto principle no longer applies and hence the distribution of income does not necessarily respect the individuals ex ante preferences. This is problematic in the setting of this paper, where individuals make a choice before they know the state of the world. Fleurbaey (2010) offers a solution to both of these issues by ranking individuals through an expected equally distributed equivalent approach. The resulting social welfare function is considerably more complex than the one offered here. Employing this welfare function would risk drawing the attention of this paper away from the study of a tax and social insurance schedule to the study of this welfare function. Additionally, it would frustrate comparison to the previous literature which uses the same utilitarian approach as I do here. For these two reasons I will use the utilitarian expected utility approach that is common in the literature with the side note that it would have been an interesting exercise to see whether the welfare function proposed by Fleurbaey (2010) would have altered results if should time have permitted.

The government observes the health status of all individuals. It raises welfare by redistributing income between productivity and health types and between healthy and ill individuals. In the main model the government can not condition tax policy directly on an individual's type or labour and effort choices, as these are private information. This means the government cannot use individualized lump sum taxation but has to resort to distortionary taxation, creating a trade off between equity and efficiency. The government thus sets a tax policy $t^s(z_n^s)$ where it conditions tax payments on an individual's output and health state. The governments budget constraint is then given by:

$$\sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta)t^g(z_n^g) + [1-\pi(\phi_\theta)]t^b(z_n^b)] \geq 0. \tag{2}$$

Since the government does not observe productivity or health types directly, it will also need to satisfy incentive compatibility constraints related to the unobserved characteristics. These are introduced to the analysis when they are necessary. This means that two IC constraints relating to productivity types will be introduced in section 3.2 and a further two IC constraints relating to health types will be introduced in section 3.3.

3.1 A look at the case of the first-best

Before analyzing the full model it is useful to start by considering the first-best case to illustrate the workings of the model. Therefore, I make two assumptions which I will gradually relax in the next two subsections. The first assumption is that the government can directly observe the productivity and health type of the individual. This implies the government can condition taxes on type, allowing it to use individualized lump-sum taxes rather than distortionary taxation to redistribute income. I will let go of this assumption from section 3.2 onward. The second assumption is that the effort on preventive care is directly observable to the government. This implies the government can offer a tax schedule conditional on the effort level of an individual. The ability of the government to condition taxes on effort eliminates the moral hazard problem that would occur otherwise. I will let go of this assumption in section 3.3.

In this paper I am interested in the case where it is welfare improving for the government to separate between health types. It is only in this case that moral hazard is an issue that constraints government policy. Assumption 1 ensures that it is always optimal to separate between health types, rather than letting all types pool at the low level of preventive care.

Assumption 1: *Cost of illness C is sufficiently high and cost of effort $H(\cdot)$ is sufficiently low, so that the government wishes to screen between health types in the first and second-best optimum.*

To see why this is assumption is necessary consider the case where $C = 0$. In this situation there are no gains of letting type θ_2 choose the high level of preventive care, as it does not result in a reduction in cost. Similarly if $C > 0$ and cost of effort $H(\frac{\phi_2}{\theta_2})$ is sufficiently high, a utilitarian government would allow type θ_2 to choose the low level of preventive care. These two scenarios are ruled out by assumption 1. Note that if the government can directly observe the effort and health type and it uses this information to screen between health types, it is as if health risk is exogenously determined. This implies we can express the endogenous health choice as if it were exogenous $\pi(\phi_\theta) = \pi_\theta$, which is convenient for notation. To characterize the optimal tax and transfer policy in the first-best case I solve a social planner problem. It turns out to be convenient to solve for the optimal allocation in terms of the observable variables z_n^s and (through the individual's budget constraint) c_n^s. By allocating z_n^s and c_n^s, the government maximises social welfare function (1) subject to the resource constraint of the economy. In the first-best case the social planner solves the following programme:

$$Maximize_{c,z}: \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta) \cdot u(c_n^g - v\left(\frac{z_n^g}{n}\right)) + [1-\pi(\phi_\theta)] \cdot u(c_n^b - v\left(\frac{z_n^b}{n}\right))],$$

$$Subject\, to: \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta)(z_n^g - c_n^g) + [1-\pi(\phi_\theta)](z_n^b - c_n^b)] - E,$$

where $E \equiv \sum_\theta [1-\pi(\phi_\theta)]C$ is the total cost to the government of providing healthcare.

Formulating a Lagrangian with λ as the multiplier on the resource constraint one obtains:

$$\mathcal{L} = \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta) \cdot u(c_n^g - v\left(\frac{z_n^g}{n}\right)) + [1-\pi(\phi_\theta)] \cdot u(c_n^b - v\left(\frac{z_n^b}{n}\right))]$$

$$+\lambda \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta)(z_n^g - c_n^g) + [1-\pi(\phi_\theta)](z_n^b - c_n^b)] - E.$$

Define the frequency of individuals of type n that are in a healthy state as $F_n^g \equiv \sum_\theta f_{n\theta}\pi(\phi_\theta)$, similarly for individuals in a unhealthy state $F_n^b \equiv \sum_\theta f_{n\theta}[1-\pi(\phi_\theta)]$.

The first order conditions with respect to production are:

$$\frac{\partial \mathcal{L}}{\partial z_n^s} = -u_n^{s\prime} v'\left(\frac{z_n^s}{n}\right)\frac{1}{n} + \lambda = 0, \tag{3}$$

the first order conditions with respect to consumption are:

$$\frac{\partial \mathcal{L}}{\partial c_n^s} = u_n^{s\prime} - \lambda = 0, \tag{4}$$

where $u_n^{s\prime} \equiv u'(c_n^s - v\left(\frac{z_n^s}{n}\right))$ to shorten the notation. Define an individual's social welfare weight in money units as $\gamma_n^s \equiv \frac{u_n^{s\prime}}{\lambda}$. This is the social value of a marginal increase in the income of an individual with productivity n in health state s. Equation (4) implies that $\gamma_n^s = \frac{u_n^{s\prime}}{\lambda} = 1$ for all individuals and hence there is no inequality in utility levels. This implies that there is full redistribution between productivity types and health states. Combining equation (3) and (4) shows that $v'\left(\frac{z_n^s}{n}\right)\frac{1}{n} = 1$ which implies that the labour supply decision is not distorted in either health state. Note that an undistorted labour supply and equalized marginal utilities do not imply that consumption is also equalized. Individuals with a high level of productivity supply more labour and are compensated through a higher level of consumption.

The first-best allocation can be implemented through a Musgravian separation of tasks (Musgrave, 1956). The equity and efficiency branches of government would not have to interact to implement the allocation found above. Because the government is well informed, it does not have to condition taxes on observed actions and there is no trade off between equity and efficiency. The branch of government concerned with equity can redistribute income through the use of individualized lump-sum taxation and therefore it does not have to communicate with the efficiency branch about distortions in the supply of labour. In the next section I show that this separation of tasks is no longer possible when we reduce the amount of information that is available to the government.

3.2 Second-best analysis without moral hazard

In this subsection I consider an intermediate version of the full model. I let go of the assumption that the government can observe productivity and health type directly. I do introduce another assumption, which will allow for comparison of the results with those of Rochet (1991) and Cremer and Pestieau (1996) (RCP). I assume that individuals do not choose their effort level to maximise utility but simply choose to obtain their designated level of preventive care. That is, an individual with low cost of care always chooses the high level of effort and an individual with high cost of care chooses the low level. Since individuals always choose their designated level of care, there is no moral hazard. In this section we thus model health risk as exogenous, as it is in RCP. Note that we do retain the unobserved heterogeneity in health. Furthermore, I study a second-best problem without moral hazard. Since the government does not directly observe productivity it has to condition its tax schedule on output, which it does observe. I will thus introduce two incentive compatibility (IC) constraints, one for each health state, to screen between high and low-productivity types. If these IC constraints are not satisfied, it is in the interest of the high-productivity worker to mimic the low-productivity worker. In this setting the problem of the government is equivalent to solving this programme:

$$Maximise_{c,z}: \sum_{n,\theta} f_{n\theta}[\pi_\theta \cdot u(c_n^g - v\left(\frac{z_n^g}{n}\right) + [1-\pi_\theta] \cdot u(c_n^b - v\left(\frac{z_n^b}{n}\right)],$$

$$Subject\,to: \sum_{n,\theta} f_{n\theta}[\pi_\theta(z_n^g - c_n^g) + [1-\pi_\theta](z_n^b - c_n^b)] - E,$$

$$u(c_H^g - v\left(\frac{z_H^g}{n_H}\right)) \geq u(c_L^g - v\left(\frac{z_L^g}{n_H}\right)), \tag{5}$$

$$u(c_H^b - v\left(\frac{z_H^b}{n_H}\right)) \geq u(c_L^b - v\left(\frac{z_L^b}{n_H}\right)), \tag{6}$$

where equation (5) and (6) are the IC constraints. They imply that, high-productivity type obtains a higher level of utility by choosing its designated level of production than it would if it were to choose the level of production of the low-productivity worker. The Lagrangian with η^s as the Lagrange multiplier for the corresponding IC constraint is:

$$\mathcal{L} = \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta) \cdot u(c_n^g - v\left(\frac{z_n^g}{n}\right) + (1-\pi(\phi_\theta)) \cdot u(c_n^b - v\left(\frac{z_n^b}{n}\right))]$$

$$+\lambda \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta)(z_n^g - c_n^g) + (1-\pi(\phi_\theta))(z_n^b - c_n^b)] - E$$

$$+\sum_s \eta^s[u(c_H^s - v\left(\frac{z_H^s}{n_H}\right)) - u(c_L^s - v\left(\frac{z_L^s}{n_H}\right))].$$

The first order conditions with respect to production for respectively the high and low-productivity are:

$$\frac{\partial \mathcal{L}}{\partial z_H^s} = -F_H^s {u_H^s}'[v'\left(\frac{z_H^s}{n_H}\right)\frac{1}{n_H}] + \lambda F_H^s - \eta^s {u_H^s}'[v'\left(\frac{z_H^s}{n_H}\right)\frac{1}{n_H}] = 0, \tag{7}$$

$$\frac{\partial \mathcal{L}}{\partial z_L^s} = -F_L^s {u_L^s}'[v'\left(\frac{z_L^s}{n_L}\right)\frac{1}{n_L}] + \lambda F_L^s + \eta^s {u_L^s}'[v'\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H}] = 0. \tag{8}$$

The first order conditions with respect to consumption for respectively the high and low-productivity type are:

$$\frac{\partial \mathcal{L}}{\partial c_H^s} = F_H^s {u_H^s}' - \lambda F_H^s + \eta^s {u_H^s}' = 0 \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial c_L^b} = F_L^b {u_L^b}' - \lambda F_L^b - \eta^b {u_L^b}' = 0 \tag{10}$$

Combining the first order conditions for the high-productivity type, (7) and (9), yields a production distortion of zero at the top in both states:

$$v'\left(\frac{z_H^s}{n_H}\right)\frac{1}{n_H} = 1.$$

This condition states that the high-productivity types are not taxed at their marginal production level. This is intuitive, as marginal taxation at production level z raises revenue from individuals above level z but distorts production at level z. Since there are no individuals producing more than the high-productivity type, the government raises no revenue by taxing them at the margin but it does distort production. Therefore the government optimally does not levy a tax at the marginal level of production of the high-productivity type. Next, by combining the first order conditions of the low-productivity type we can characterize the optimal marginal tax rates of the low-productivity type:

$$\gamma_L^s [v'\left(\frac{z_L^s}{n_L}\right)\frac{1}{n_L} - v'\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H}] = 1 + v'\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H}. \tag{11}$$

We shall interpret $(1 + v'\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H})$ as the marginal tax rate in state s at output level z_L^s. This condition shows that the government distorts the labour supply of the low-productivity type when it has to condition taxes on output. This distortion of the labour supply of the low-productivity type makes it more costly for the high-productivity type to mimic the low-productivity type, this allows the government to increase its redistribution towards the low-productivity type. In this model, the amount of insurance the government provides depends on the covariance between health risk and income. To see why this is the case, note that, in the optimum the government is indifferent to redistributing an additional euro from the healthy to the ill. If the government raises an additional euro from individuals in state s, their consumption is lowered by $\frac{1}{F^s}$. Raising the income of all individuals in group s by one euro raises welfare by $F_H^s {u_H^s}' + F_L^s {u_L^s}'$. Together this implies:

$$\frac{F_H^g w_H^g + F_L^g w_L^g}{F^g} = \frac{F_H^b w_H^b + F_L^b w_L^b}{F^b} = 1$$

where $F^s \equiv \sum\limits_n F_n^s$, is the share of individuals in state s and the marginal cost of public funds is one. This condition shows that, in the optimum, the government equalizes average social welfare in each health state. Equalizing average social welfare between health states does not generally imply full public insurance against health risk. The reason is that health state reveals information about productivity type if $Cov(n, \pi) \neq 0$. The government uses this information to improve the equity-efficiency trade off by not providing full public insurance. It uses health states to, on average, redistribute more income to the low-productivity types. In the special case that $Cov(n, \pi) = 0$, the government cannot use health states to improve the equity-efficiency trade off. The equity and the insurance problem are in this case unrelated. The government then provides full social insurance as individuals are risk averse. If $Cov(n, \pi) > 0$ then the government provides social insurance in excess of full insurance. By providing insurance in excess of full insurance the government is able to, on average, redistribute more income to low-productivity types as these are more likely to be ill. Empirically, this is the most relevant case as high-income individuals also tend to be healthier (Ettner, 1996).

If $Cov(n, \pi) < 0$ the amount of social insurance is ambiguous. As the covariance becomes more negative the government decreases the amount of social insurance it provides. If the high-productivity individuals are those that are ill and and health cost are not too large the government might use social insurance as a net tax on the ill in order to redistribute from high-productivity individuals to low-productivity individuals. These observations are formalized in proposition 1:

Proposition 1: *In the second-best optimum without moral hazard the government equalizes average social welfare weights in each health state. The optimal amount of social insurance depends on the sign of $Cov(n, \pi)$:*

1. *If $Cov(n, \pi) = 0$ the government provides full social insurance.*
2. *If $Cov(n, \pi) > 0$ the government provides social insurance in excess of full insurance.*
3. *If $Cov(n, \pi) < 0$ the sign of social insurance is ambiguous.*

This proposition agrees with some the results of RCP and clarifies some of the driving forces. They find that if $Cov(n, \pi) \geq 0$ the government provides full social insurance. The reason that they do not find that the government provides in excess of full social insurance is that they assume the government cannot offer more than full social insurance. It is clear that the desire of government to provide social insurance is partially driven by the incentive to use social insurance to redistribute income without distorting the labour market by redistributing on the health margin. By employing the information that the health state gives about the likelihood that an individual is of a low-productivity type, the government can, on average, improve redistribution towards low-productivity type without additional labour supply distortions.

3.3 Second-best analysis with moral hazard

Now we move to the full model. We relax the assumption made in the previous subsection that individuals always choose their designated level of preventive effort. We thus no longer assume that individuals with low cost of effort always choose the high level of preventive care. This allows for a problem of moral hazard. Suppose that the government provides full social insurance, then there is no incentive for an individual to exert any effort to reduce the chance of falling ill. Assumption 1 implies that it is optimal for the government to prevent individuals with a low cost of effort from choosing a low level of preventive care. Therefore the government wants to screen between individuals of different health types. It needs to be rewarding to both high and low-productivity types with a low cost of effort to choose the high level of preventive care. Thus the two IC constraints take the following form:

$$u(c_H^g - v\left(\frac{z_H^g}{n_H}\right)) - u(c_H^b - v\left(\frac{z_H^b}{n_H}\right)) \geq \frac{H(\frac{\phi_2}{\theta_2}) - H(\frac{\phi_1}{\theta_2})}{\pi_H - \pi_L}, \tag{12}$$

$$u(c_L^g - v\left(\frac{z_L^g}{n_L}\right)) - u(c_L^b - v\left(\frac{z_L^b}{n_L}\right)) \geq \frac{H(\frac{\phi_2}{\theta_2}) - H(\frac{\phi_1}{\theta_2})}{\pi_H - \pi_L}. \tag{13}$$

In the optimum, only one of the IC constraints can bind, given that individuals have to make a discrete choice of effort on preventive care. If one of the constraints binds and the other one is satisfied, then the tax policy is admissible. The constraint that will bind is the constraint on the high-productivity side. To see why, consider a case where the government provides full insurance through a transfer T^g on the healthy, that funds a transfer T^b on the ill. Budget balance requires $F^g T^g = F^b T^b$.

Full insurance then implies:

$$\triangle V_H \equiv u(c_H^g - v\left(\frac{z_H^g}{n_H}\right) - T^g) - u(c_H^b - v\left(\frac{z_H^b}{n_H}\right) + T^b) = 0,$$

$$\triangle V_L \equiv u(c_L^g - v\left(\frac{z_L^g}{n_L}\right) - T^g) - u(c_L^b - v\left(\frac{z_L^b}{n_L}\right) + T^b) = 0.$$

When there is full insurance IC constraints (12) and (13) are not satisfied. Next, we decrease T^g and T^b until the both of the IC constraints are satisfied. To see which one will be binding, consider a decrease in T^g and T^b while requiring budget balance:

$$-\frac{\partial \triangle V_H}{\partial T^g} = u'(c_H^g - v\left(\frac{z_H^g}{n_H}\right) - T^g) + u'(c_H^b - v\left(\frac{z_H^b}{n_H}\right) + T^b)\frac{F^b}{F^g} > 0,$$

$$-\frac{\partial \triangle V_L}{\partial T^g} = u'(c_L^g - v\left(\frac{z_L^g}{n_L}\right) - T^g) + u'(c_H^b - v\left(\frac{z_L^b}{n_L}\right) + T^b)\frac{F^b}{F^g} > 0.$$

Remember that marginal utility of income is larger for the low-productivity type: $u_L^{b\prime} > u_H^b$ and $u_L^{g\prime} > u_H^g$. This implies that the utility loss of non-insurance is larger for the low-productivity type: $-\frac{\partial \triangle V_L}{\partial T} > -\frac{\partial \triangle V_H}{\partial T}$. Given that insurance is less beneficial to the high-productivity type it is the IC constraint on this type that will bind.

The optimal tax policy then solves the following programme:

$$Maximise_{c,z}: \sum_{n,\theta} f_{n\theta}[\pi_\theta \cdot u(c_n^g - v\left(\frac{z_n^g}{n}\right) + (1-\pi_\theta)\cdot u(c_n^b - v\left(\frac{z_n^b}{n}\right))],$$

$$Subject\, to: \sum_{n,\theta} f_{n\theta}[\pi_\theta(z_n^g - c_n^g) + (1-\pi_\theta)(z_n^b - c_n^b)] - E,$$

$$u(c_H^g - v\left(\frac{z_H^g}{n_H}\right)) \geq u(c_L^g - v\left(\frac{z_L^g}{n_H}\right)),$$

$$u(c_H^b - v\left(\frac{z_H^b}{n_H}\right)) \geq u(c_L^b - v\left(\frac{z_L^b}{n_H}\right)),$$

$$u(c_H^g - v\left(\frac{z_H^g}{n_H}\right)) - u(c_H^b - v\left(\frac{z_H^b}{n_H}\right)) \geq \frac{\phi(\frac{\phi_H}{\theta_H}) - \phi(\frac{\phi_H}{\theta_H})}{\pi_H - \pi_L}.$$

Forming a Lagrangian with μ as the multiplier on IC constraint (12) we have

$$\mathcal{L} = \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta)\cdot u(c_n^g - v\left(\frac{z_n^g}{n}\right) + (1-\pi(\phi_\theta))\cdot u(c_n^b - v\left(\frac{z_n^b}{n}\right))]+$$

$$+\lambda \sum_{n,\theta} f_{n\theta}[\pi(\phi_\theta)(z_n^g - c_n^g) + (1-\pi(\phi_\theta))(z_n^b - c_n^b)] - E$$

$$+\sum_s \eta^s[u(c_H^s - v\left(\frac{z_H^s}{n_H}\right)) - u(c_L^s - v\left(\frac{z_L^s}{n_H}\right))] + \mu[u(c_H^g - v\left(\frac{z_H^g}{n_H}\right)) - u(c_H^b - v\left(\frac{z_H^b}{n_H}\right)) - \frac{\phi(\frac{\phi_H}{\theta_H}) - \phi(\frac{\phi_L}{\theta_H})}{\pi_H - \pi_L}].$$

The first order conditions with respect to production are:

$$\frac{\partial \mathcal{L}}{\partial z_H^g} = -F_H^g {u_H^g}'[v'\left(\frac{z_H^g}{n_H}\right)\frac{1}{n_H}] + \lambda F_H^g - \eta^g {u_H^g}'[v'\left(\frac{z_H^g}{n_H}\right)\frac{1}{n_H}] - \mu {u_H^g}'[v'\left(\frac{z_H^g}{n_H}\right)\frac{1}{n_H}] = 0$$

$$\frac{\partial \mathcal{L}}{\partial z_H^b} = -F_H^b {u_H^b}'[v'\left(\frac{z_H^b}{n_H}\right)\frac{1}{n_H}] + \lambda F_H^b - \eta^b {u_H^b}'[v'\left(\frac{z_H^b}{n_H}\right)\frac{1}{n_H}] + \mu {u_H^b}'[v'\left(\frac{z_H^b}{n_H}\right)\frac{1}{n_H}] = 0$$

$$\frac{\partial \mathcal{L}}{\partial z_L^g} = -F_L^g {u_L^g}'[v'\left(\frac{z_L^g}{n_L}\right)\frac{1}{n_L}] + \lambda F_L^g + \eta^g {u_L^g}'[v'\left(\frac{z_L^g}{n_H}\right)\frac{1}{n_H}] = 0$$

$$\frac{\partial \mathcal{L}}{\partial z_L^b} = -F_L^b u_L^{b\prime}[v'\left(\frac{z_L^b}{n_L}\right)\frac{1}{n_L}] + \lambda F_L^b + \eta^b u_L^{b\prime}[v'\left(\frac{z_L^b}{n_H}\right)\frac{1}{n_H}] = 0$$

The first order conditions with respect to consumption are:

$$\frac{\partial \mathcal{L}}{\partial c_H^g} = F_H^g u_H^{g\prime} - \lambda F_H^g + \eta^g u_H^{g\prime} + \mu u_H^{g\prime} = 0$$

$$\frac{\partial \mathcal{L}}{\partial c_H^b} = F_H^b u_H^{b\prime} - \lambda F_H^b + \eta^b u_H^{b\prime} - \mu u_H^{b\prime} = 0$$

$$\frac{\partial \mathcal{L}}{\partial c_L^g} = F_L^g u_L^{g\prime} - \lambda F_L^g - \eta^g u_L^{g\prime} = 0$$

$$\frac{\partial \mathcal{L}}{\partial c_L^b} = F_L^b u_L^{b\prime} - \lambda F_L^b - \eta^b u_L^{b\prime} = 0$$

Next, if we combine the first order conditions with respect to production and consumption of the high-productivity type we obtain:

$$v'\left(\frac{z_H^s}{n_H}\right)\frac{1}{n_H} = 1$$

This condition shows that the high-productivity type is not taxed at its marginal level of production. This implies that the introduction of ex ante moral hazard does not affect the zero at the top result. The government does not use marginal taxation on the high-productivity type to ensure that IC constraints (12) and (13) are satisfied. Rather, it uses lump sum taxation conditional on health state to ensure that it is in the interest of low cost health types to choose the high level of preventive care. Next, we examine how the government uses marginal taxation on the low-productivity types. By combining the first order conditions of the low-productivity types we can characterize the tax schedule of the government.

The unhealthy type is taxed at:

$$\gamma_L^b[v'\left(\frac{z_L^b}{n_L}\right)\frac{1}{n_L} - v'\left(\frac{z_L^b}{n_H}\right)\frac{1}{n_H}] = 1 + v'\left(\frac{z_L^b}{n_H}\right)\frac{1}{n_H} \tag{14}$$

The healthy type is taxed at:

$$\gamma_L^g[v'\left(\frac{z_L^g}{n_L}\right)\frac{1}{n_L} - v'\left(\frac{z_L^g}{n_H}\right)\frac{1}{n_H}] = 1 + v'\left(\frac{z_L^g}{n_H}\right)\frac{1}{n_H} \tag{15}$$

Where we shall again interpret $[1+v'\left(\frac{z_L^g}{n_H}\right)\frac{1}{n_H}]$ as the marginal tax rate on individual in state s at production level z_L^s. This expression shows that the labour supply distortion differs between low-productivity individuals in different health states, because their social welfare weight differs. Once we know in which health state individuals have a higher social welfare weight, we know which individual face a higher marginal tax rate. It follows from IC constraint (13) that it is the individual in the bad health state who has a higher social welfare weight: $\gamma_L^b > \gamma_L^g$. In the *Appendix* I show that labour supply distortions increase as the welfare weight increases. This implies that low-productivity individuals in a bad health state face a higher marginal tax rate. This is stated formally in proposition 2:

Proposition 2: *In the second-best optimum with moral hazard, the government places a lower marginal tax rate on the healthy low-productivity type than on the unhealthy low-productivity type.*

To see why this makes sense consider a situation in which low-productivity individuals in both health states face identical marginal tax rates. Suppose that the government now starts raising the marginal tax rate on the ill. This makes it more costly for the ill high-productivity individual to mimic the ill low-productivity individual. This allows the government to raise transfers to the ill low-productivity individual financed by taxes on the ill high-productivity individual. Thus, by differentiating marginal tax rates, the government can redistribute more income between ill individuals, thereby improving social welfare. Differentiating marginal tax rates also allows the government to increase the aggregate amount of social insurance that it provides. By raising marginal tax rates, the government is able to collect more taxes from ill high-productivity individuals. This makes it more costly to high-productivity individuals to fall ill, thereby relaxing IC constraints (12). This allows the government to transfer more income from the healthy to the ill individuals and thus provides more insurance and raising social welfare.

3.4 Implementation through social insurance

In this subsection I consider how the optimal policies found in subsections 3.2 and 3.3 can be implemented through a social insurance and tax policy. There are few governments which differentiate between the ill and the healthy through tax policy. Rather, this is typically done through social insurance policies. I show that the government could offer all individuals the same tax schedule, while differentiating effective marginal tax rates between the healthy and the ill, through an income-dependent social insurance policy. I analyze the case where all individuals face the marginal tax rate of the healthy type $t^{g\prime}(\cdot)$ but ill individuals face an additional tax wedge through an income-dependent social insurance policy. The total effective marginal tax rate of unhealthy low-productivity individuals then is:

$$t^{g\prime}(z_L^g) + I'(z_L^b) = t^{b\prime}(z_L^b) \tag{16}$$

Where $I'(\cdot)$ should be interpreted as the marginal change in the social insurance payment the individual receives as income rises. We can then use the identity in equation (16) and the optimal tax expressions in equations (14) and (15) to characterize the optimal change in the social insurance payment as income increases:

$$I'(z_L^b) = t^{b\prime}(z_L^b) - t^{g\prime}(z_L^g) = \gamma_L^b[[v'\left(\frac{z_L^b}{n_L}\right)\frac{1}{n_L}] - [v'\left(\frac{z_L^b}{n_H}\right)\frac{1}{n_H}]] - \gamma_L^g[[v'\left(\frac{z_L^g}{n_L}\right)\frac{1}{n_L}] - [v'\left(\frac{z_L^g}{n_H}\right)\frac{1}{n_H}]]$$

This expression shows that the optimal income-dependent social insurance policy is equal to the difference in marginal tax rates under the state-dependent tax rate implementation. We can also think of the income-dependent social insurance policy as an income-dependent coinsurance rate on health cost C. The optimal coinsurance rate can be characterized as follows:

$$\alpha(z_L^b)C = t^{b\prime}(z_L^b) - t^{g\prime}(z_L^g).$$

Where $\alpha(\cdot) \equiv \frac{I'(z_L^b)}{C}$. The optimal coinsurance rate can also implement the state-dependent tax policy through a common tax schedule for all individuals, combined with an income-dependent coinsurance rate.

4 Concluding remarks

The design of an optimal tax system depends on many factors that affect incentives. This paper considered two of these factors: endogenous health risks and labour supply decisions. I attempted to understand whether heterogeneity in productivity and health risk warranted a policy that differs between healthy and ill individuals. The main contribution of this paper is that it shows that the government does want to differentiate in its policy. By differentiating, the government is able to redistribute more income and provide more social insurance. By increasing the marginal tax rate on the ill, the government is able to redistribute more income from the ill high-productivity individuals to the ill low-productivity individuals. Additionally, by increasing the amount of social insurance the government is able to raise the transfers from the healthy to the ill individuals. In subsection 3.4 I showed that instead of offering different marginal tax rates to the healthy and the ill the government can offer all individuals the same tax schedule and differentiate effective taxes through the use of an income-dependent social insurance policy. To the best of my knowledge, it is an original result in the literature that studies non-linear taxation and social insurance, that the government should use an income-dependent social insurance policy, even if it has access to non-linear income tax.

My results contrast with the results of Rochet (1991) and Cremer and Pestieau (1996). They show that the government provides full, non-income-dependent social insurance if $Cov(n, \pi) > 0$. The difference between our results is driven by two factors: the presence of moral hazard in my model and the difference in the timing of the labour supply decision. Due to moral hazard, as in Boadway et al (2003), the government does not provide full social insurance. The presence of moral hazard makes it costly to the government to provide full social insurance, as individuals will reduce their effort to avoid falling ill. As to the second difference between our results, that the social insurance programme is income-dependent, this is driven by the different timing of the labour supply decision. In this paper, the individual chooses his labour supply while knowing his health state. This means that the government can use the health state to differentiate effective taxes allowing, it to improve its equity-efficiency trade off and provide more social insurance. In Rochet (1991) and Cremer and Pestieau (1996), individuals choose their labour supply before they know their health state. Thus, the government is not able to use health states to improve the equity-efficiency trade off. And hence in their model, there is no reason to introduce an income-dependent social insurance policy. Both labour supply models are empirically relevant. The model in this paper is most appropriate to study the provision of social insurance to workers that face health risks during their working life. That is, to workers that are able to alter their labour supply based on their realised health state. The model of Rochet (1991) and Cremer and Pestieau (1996) is most appropriate to model social insurance for individuals that face health risks when they are retired, that is for individuals who are unable to adjust labour supply based on realised health state.

Although this paper establishes that the optimal social insurance programme is income-dependent, I do not quantify how rapidly healthcare support should be lowered as income increases. This is an empirical question. A recent literature in public economics develops models that allow for quantitative analysis of optimal policy. It does so by deriving sufficient statistics that allow for quantitative results based on empirical elasticities estimates (e.g. Saez 2001, Chetty 2009). Formulating a sufficient statistics model with non-linear taxation and moral hazard would allow for study of the quantitative policy implications of the problem that I analyzed in this paper.

5 Literature

Adler, M. D., & Sanchirico, C. W. (2006). Inequality and uncertainty: Theory and legal applications. *U. Pa. L. Rev.*, 155, 279.

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

Blomqvist, Å., & Horn, H. (1984). Public health insurance and optimal income taxation. *Journal of Public Economics*, 24(3), 353-371.

Breyer, F., & Haufler, A. (2000). Health care reform: Separating insurance from income redistribution. *International Tax and Public Finance*, 7(4-5), 445-461.

Boadway, R., Leite-Monteiro, M., Marchand, M., & Pestieau, P. (2003). Social insurance and redistribution. *Public Finance and Public Policy in the New Century, MIT Press, Cambridge*, MA, 333, 358.

Boadway, R., Leite-Monteiro, M., Marchand, M., & Pestieau, P. (2006). Social insurance and redistribution with moral hazard and adverse selection. *The Scandinavian Journal of Economics*, 108(2), 279-298.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ.,* 1(1), 451-488.

Chetty, R., & Saez, E. (2010). Optimal taxation and social insurance with endogenous private insurance. *American Economic Journal: Economic Policy*, 2(2), 85-114.

Cremer, H., & Pestieau, P. (1996). Redistributive taxation and social insurance. *International Tax and Public Finance*, 3(3), 281-295.

Diamond, P. A. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *The Journal of Political Economy*, 75(5), 765.

Diamond, P., & Spinnewijn, J. (2011). Capital income taxes with heterogeneous discount rates. *American Economic Journal: Economic Policy,* 3(4), 52-76.

Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of health economics*, 15(1), 67-85.

Fleurbaey, M. (2010). Assessing risky social situations. *Journal of Political Economy*, 118(4), 649-680.

Greenwood, J., Hercowitz, Z., & Huffman, G. W. (1988). Investment, capacity utilization, and the real business cycle. *The American Economic Review*, 402-417.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, 63(4), 309-321.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review* *of economic studies*, 38(2), 175-208.

Musgrave, R. A. (1959). Theory of public finance; a study in public economy.

Rochet, J. C. (1991). Incentives, redistribution and social insurance. *The Geneva Papers on risk and insurance theory*, 16(2), 143-165.

Rothschild, M., & Stiglitz, J. (1978). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. In *Uncertainty in economics* (pp. 257-280).

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The review of economic studies*, 68(1), 205-229.

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of public economics*, 17(2), 213-240.

Wagstaff, A., & Van Doorslaer, E. (1993). Equity in the finance and delivery of health care: concepts and definitions. *Equity in the finance and delivery of health care: an international perspective*, 7-19.

Appendix

In this Appendix I derive the relationship between welfare weight γ_L^s and production level z_L^s. Equations (14) and (15) show that the optimal tax rates are characterized by the following implicit function:

$$G(b,z) = \gamma_L^s[v'\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H} - v'\left(\frac{z_L^s}{n_L}\right)\frac{1}{n_L}] + v'\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H} - 1 = 0$$

The relationship between welfare weight γ_L^s and production level z_L^sdepends on the sign of $\frac{dz_L^b}{dw_L^b}$. The implicit function theorem yields:

$$\frac{dz_L^b}{d\gamma_L^b} = \frac{\partial G/\partial w}{\partial G/\partial z} = -\frac{[v'\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H} - v'\left(\frac{z_L^s}{n_L}\right)\frac{1}{n_L}]}{(b-1)v''\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H^2} - bv''\left(\frac{z_L^s}{n_L}\right)\frac{1}{n_L^2}}$$

The sign of this expressions depends on the sign of the numerator and the denominator. To see that the numerator is negative, remember that $v(\cdot)$ is a strictly convex function and $n_H > n_L$. This implies that $v'\left(\frac{z_L^s}{n_H}\right) < v'\left(\frac{z_L^s}{n_L}\right)$ and hence the numerator is negative.

To see that the denominator is negative interpret it as a function $X(n_H; n_L) \equiv (b-1)v''\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H^2} - bv''\left(\frac{z_L^s}{n_L}\right)\frac{1}{n_L^2}$ and consider the case where $n_H = n_L$. I now use a Taylor expansion to show that this is negative and decreases as n_H and n_L are further apart, the Taylor expansion yields:

$$-v''\left(\frac{z_L^s}{n_H}\right)\frac{1}{n_H^2}-v'''\left(\frac{z_L^s}{n_H}\right)\frac{z_L^s}{n_H^4}(n_H-n_L)<0$$

Strict convexity implies that the expression is negative at $n_H = n_L$ since $v''(\cdot) > 0$. Note that the strict convexity assumption can only hold everywhere with certainty if $v'''(\cdot) \geq 0$. This implies that $v'''\left(\frac{z_L^s}{n_H}\right)\frac{z_L^s}{n_H^4}(n_H-n_L)$ and hence the denominator is more negative as n_H and $n_L$are further apart. Since both the numerator and denominator are negative it follows that $\frac{dz_L^b}{dw_L^b}$ is negative as well.