

WELFARE AND OUTPUT ENHANCING MORAL HAZARD: DISABILITY BENEFITS AND ENDOGENOUS OCCUPATIONAL CHOICE

Paul Frijters

Tinbergen Institute, Free University Amsterdam, Netherlands

Michael A. Shields

Department of Economics, University of Melbourne, Australia

Abstract

One of the most prominent trends in OECD countries over the last 30 years has been the sharp increase in incidence of early retirement, and in particular the permanent take-up of disability benefits. In this paper we construct a theoretical model that shows how occupational choices, in terms of the associated health risks, made by the young can be affected by the expected provision of publicly funded disability benefits in later life. We find that because individuals are risk-averse, they take insufficient risks in the absence of insurance. Disability benefits lead to riskier aggregate behaviour, which in turn increases output and welfare at low levels of benefits, but will lead to excessive risk taking at high benefit levels to the detriment of output and welfare. We also show that the full impact of changes to the generosity of disability benefits in terms of increasing the take-up of such benefits is not immediate, but may take many years to realise because the previous career choices are largely irreversible. This time lag is consistent with the experiences of a number of countries over the last 30 years.

Keywords: Occupational Choice, Health Risks, Disability Benefits, Moral Hazard, and Output

JEL Classifications: D6, H0

1. Introduction

One of the most prominent trends in OECD countries over the last 30 years has been the sharp increase in the incidence of early retirement (Bound and Burkhauser, 1999; Diamond and Gruber, 1999; Gruber and Wise, 1998, 1999; Kerkhofs *et al.*, 1999; Burkhauser and Daly, 2002). A major component of this trend has been the large growth in the numbers receiving long-term disability benefits, which is highlighted for several countries in Table 1. This increase has been particularly pronounced for Germany and the Netherlands, where since 1980 a greater number of individuals aged 60-64 have been on disability benefits than active in the labour force. However, the figures also clearly suggest that this growth abated in the Netherlands, Sweden, the UK and the US in the last decade.

All OECD countries have some welfare provision for individuals with chronic disabilities, and these trends are placing enormous pressure on the financial viability of such welfare programs (Gruber and Wise, 1998). Consequently, a large empirical literature has investigated the relationship between economic incentives and benefit take-up, and left little doubt that individuals do react to changes in both the eligibility criteria and levels of generosity of such benefits (e.g. Autor and Duggan, 2003; Bound and Waidmann, 2002; Cullen, 2003; Harkness, 1993; Johnson and Ondrich, 1990; Kreider and Riphahn, 2000). Blondal and Scarpetta (1997), for example, calculated using pooled cross-country regressions, that removing the disincentives not to work created by such benefits would lead to an increase in the labour market participation rate of males aged 55-64 of around 10% in those countries where the financial incentives to be inactive are the most generous. It is also clearly the case that disability benefit take-up is affected by labour market opportunities (for recent contributions, see Autor and Duggan, 2003 and Black *et al.*, 2002). For example, Autor and Duggan (2003) find that the 60% increase in the share of non-elderly US adults receiving social security disability insurance income between 1984 and 2001, was partly explained by declining demand for less skilled workers and an increase in the earnings replacement ratio.

One of the remaining puzzles in this literature is that there is a long time lag between changes in benefits and the total take-up rates of such schemes (Blanchard and Wolfers, 2000; Dolado *et al.* 1996; Gruber and Wise, 1997). Consider this puzzle as it appears in Table 1: disability take-up rates increased in all countries between 1970 and 1994, whilst the relevant disability and early retirements schemes were introduced mainly at the start of this period. It appears that it took a whole generation before the full impact of the policy reforms became visible.

As one possible reason for such lags, it has been argued that some countries have an incentive in their benefit system to postpone taking up benefits even when individuals are eligible. Coile *et al.* (2002) for instance demonstrate this by the fact that around 10% of US men retiring before their 62nd birthday delay claiming for at least one year after eligibility. However, in most countries there

is no such incentive for delay, nor is 1 year sufficient to explain the decades-long lags in take-up rates.

Whilst there has been a large empirical literature investigating the economic reasons for increasing take-up rates of disability benefits by older workers in many countries, there has been a more limited theoretical literature. A popular topic in the theoretical literature has been deriving the optimal structure of disability benefits given the existence of 'tagging' (Akerlof, 1978), where social security programs can only imperfectly identify those unable to work (Diamond and Sheshinski, 1995; Boadway *et al.*, 1999; Parsons, 1996). This work to some extent also builds on the model of Diamond and Mirrlees (1978) where labour supply is affected by exogenous health, which is unobservable by government. Optimal social insurance policies are then found for one-period, two-period, and continuous-time models. In a recent paper, Westerhout (2001) extended Pissarides (1990) model of equilibrium unemployment with disability risk and disability benefits, and allows for the improper use of disability schemes by the unemployed. He concludes that disability policies that reduce the participation in disability schemes tend to increase the rate of unemployment. Only policies that lower the rate of disability shocks succeed in reducing participation in both disability and unemployment schemes.

In these papers, disability status and health risks are generally assumed to be exogenous and not a matter of choice. Consequently, the issue of why it took so long for disability rates to rise to current levels, given the introduction of most disability benefit programs in the 1970's, has not been addressed in detail. To investigate this basic puzzle, we construct a model with endogenous disability risks by linking them to consciously chosen career paths (occupations). In this model it is these early-life choices that create persistence in the composition of the current labour force and lead to long-term lags in the effect of policy. Another extension in our paper over existing models is that we allow for some productive role of risk taking. In our general equilibrium model, the wages of low-risk jobs increase when there are many other workers taking risky jobs. This implies that there is an efficient distribution of risk-taking for output maximisation. A key question is then whether risk-averse individuals take enough risks on average and whether a disability benefit scheme improves or worsens the risk-taking behaviour of the population.

The paper is structured as follows. In Section 2 we examine these issues in a basic two period general equilibrium model in which rational risk-averse individuals divide over a continuum of occupations with differential disability risks. It turns out that in the absence of a disability benefit, individuals on average take too few risks, leading to low expected utility and low output. Wage premiums for risks are then inefficiently large. Introducing a modest disability benefit scheme encourages more individuals to take the riskier jobs, which in turn reduces the wage premium of

riskier jobs. This moral hazard effect of benefits increases both expected utility and total output, though it also increases the take up rates of the benefits.

After looking at a variation of the basic model in which individuals differ according to health endowments in Section 3, we set up a calibrated version of the model in Section 4 to examine the dynamics of changes in the economy and the benefit system. We find long lags between reductions in the benefit system and reduced disability rates because of the long time length it takes for the composition of the labour force to change. We also find that reductions in the risks associated with a subset of occupations has only a limited effect on aggregate risks: when the risks in one occupation decreases, so does its associated wage premium, leading more individuals to choose other jobs that have higher risks. These general equilibrium effects reduce the advantages of reductions in risks of any particular occupation. They thereby also limit the policy relevance of merely decreasing the actual disability risks of some specific occupations. Finally, Section 5 concludes the paper.

TABLE 1: Disability Transfer Recipients Per 1,000 Active Labour Force Participants by Age

Age	1970	1975	1980	1985	1990	1994	2001
15-44							
Germany (West)	7	6	7	8	5	5	6
Netherlands	17	32	57	58	62	66	66
Sweden	18	20	19	20	21	27	32
United Kingdom	8	9	11	20	23	-	43
United States	11	17	16	20	23	38	28
45-59							
Germany	75	64	84	103	75	80	79
Netherlands	113	179	294	305	339	289	223
Sweden	66	95	99	108	116	143	172
United Kingdom	48	46	51	97	119	-	120
United States	33	68	83	71	72	96	70
Age 60-64							
Germany	419	688	1348	1291	1109	1064	1205
Netherlands	299	437	1033	1283	1987	1911	-
Sweden	229	382	382	512	577	658	658
United Kingdom	219	195	209	357	413	-	411
United States	154	265	285	254	250	294	253

Source: Aarts et al. (1996) and Buddelmeyer (2001), and a variety of sources such as the OECD for 2001.

2. A Two-Period General Equilibrium Model

Let us suppose that there is a continuum of individuals each period of measure 1. A generation of *ex ante* identical individuals lives two periods. In the first period, individuals can choose the education relevant to a particular occupation indexed by $1 \geq \theta \geq 0$ in the second period. The density of

individuals at time t choosing specialisation θ is denoted by $f_t(\theta)$ where $\int_0^1 f_t(\theta) \partial\theta = 1$. Focussing only on steady states, we will drop the time subscripts.

In the same period, individuals search production facilities. Importantly, each occupation carries a specific health risk. This is reflected in the probability of being in work, which is equal to $h(\theta)$ where $h(\theta) = 1$, $h(1) > 0$ and $\frac{\partial h}{\partial \theta} < 0$. Jobs with higher θ are by definition more risky (in terms of future health), in that they have a higher probability of disabled. A worker produces one unit of an occupation-specific good, which is an intermediary into a final good, where the production technology is CES. Total production is given by:

$$y = \left[\int_0^1 (f(\theta)h(\theta))^\gamma \partial\theta \right]^{1/\gamma} \quad (1)$$

where $f(\theta)h(\theta)$ is the total amount of the intermediary good of type θ that is produced and where $1 > \gamma > 0$. This implies that different activities are complements and that there is hence a productivity gain from a division of labour.

Individuals are assumed to be forward-looking risk-averse rational utility maximisers. Expected utility is therefore:

$$E\{U(\theta)\} = h(\theta)u((1-\tau)w(\theta)) + (1-h(\theta))u(b) \quad (2)$$

where τ is the tax rate, and $b = \frac{\tau y}{\int (1-h)f \partial\theta}$ is the level of disability benefits. Wages are set

competitively: $w(\theta) = \frac{\partial y}{\partial f(\theta)h(\theta)} = (f(\theta)h(\theta))^{\gamma-1} y^{1-\gamma}$. In equilibrium, this wage function will equal

a hedonic wage function, where $w(\theta)$ denotes the compensating wage function for the health risks associated with occupation θ . In addition, $u(\cdot)$ is convex and its derivative exists and is continuous, with $u(\infty) = \infty$. Any disutility of being on disability benefit is incorporated in the utility of having no income at all. Non-negative non-work incomes ensure that $u(\theta) > -\infty$. We initially restrict attention to cases where $b < w(\theta)$.

First, a standard argument holds that in equilibrium, $f(\theta)$ has to be continuous if $h(\theta)$ is continuous. The reason is that if $f(\theta)$ is not continuous and, say, drops at some point x , then wages will make a discontinuous jump at x . There is then a first order gain to be made for individuals just

before x to change their choice of θ to x , with only a second-order loss of finding a job. It is also the case that $f(\theta) > 0$ for every θ and every utility function $w(\theta) = \infty$ when $f(\theta) = 0$. Choosing θ will hence be preferred over choices with wages less than infinite, which ensures that $f(\theta) > 0$ for any θ . These regularities in $f(\theta)$ mean that $w(\theta)$ is also continuous when $h(\theta)$ is continuous.

Having hence checked that minimal regularity conditions apply, we can now characterise the equilibrium by noting that each choice of θ must yield the same expected utility. From (2), differentiating $E\{u(\theta)\}$ with respect to θ and setting to 0 gives the main solution equation of the model:

$$-h'(\theta)(u((1-\tau)w(\theta)) - u(0)) = (1-\tau)w'(\theta)h(\theta)u'((1-\tau)w(\theta)) \quad (3)$$

for continuity points of $h(\theta)$. This equation immediately shows that wages are increasing in the risk ($=1-h$) and hence increasing in θ . In order to judge the efficiency of the outcome, consider what would be the output maximising choice of $f(\theta)$, denoted as $f^e(\theta)$. Due to the properties of the CES-function, y is maximised when $h(\theta)w(\theta)$ is constant. Therefore, $w^e(\theta) \propto \frac{1}{h(\theta)}$ and

$\frac{\partial w^e}{\partial(1-h)} = \frac{w^e}{h}$. This corresponds to $f^e(\theta) \propto h(\theta)^{\frac{-\gamma}{\gamma-1}}$. The equilibrium of the model is now characterised in Proposition 1.

Proposition 1: *At $\tau = 0$, the model has a unique equilibrium solution $f(\theta)$, which is inefficient. Any continuous distribution of observed wages $z(w)$ can be supported as long as $Z(x)=0$ for all $0 < x < w_{\min}$.*

Proof: Existence and uniqueness is proven in the Appendix. Inefficiency can be seen by noting that we can differentiate $E\{u(\theta)\}$ with respect to h , obtaining $\frac{\partial w}{\partial(1-h)} = \frac{u(w(\theta)) - u(0)}{h(\theta)u'(w(h))} > \frac{w}{h}$ because of the risk-aversion in $u(\cdot)$. This in turn implies inefficiency of the equilibrium. The shape of observed wages $z(w)$ follows because we can write $z(x) = \frac{1}{\frac{\partial h(\theta)}{\partial \theta}} f(\theta)$ where $\theta = \arg_{\theta} w(\theta) = x$. Since

nothing bounds $\frac{w(1)}{w(0)}$ from above, any observed continuous density function that is bounded from below can then be supported by an appropriate choice of $h(\theta)$ and $u(\cdot)$.

The intuition behind existence is that the main solution equation uniquely maps $w(\theta)$ as a continuous function of $w(0)$. Conversely, this leads to a unique $f(\theta)$ and y , both continuous in $w(0)$. Since $w(0)$ is itself uniquely determined by $f(0)$ and y , there is a closing equation for which a fixed point argument shows that it has a solution for at least one $w(0)$. Uniqueness then follows because it is not possible to change $f(\theta)$ without increasing some wages and decreasing others (proven in Lemma 1 in the Appendix). Because equilibrium implies that individuals have equal utility, it cannot be the case that there is a second equilibrium in which some are better off and others strictly worse off.

The question of interest therefore is whether or not this outcome can be improved upon by introducing a disability benefit. Three general results can be obtained:

Proposition 2: (i) At 0, an increase in τ is utility, disability and output increasing. (ii) There is a critical level τ^z above which all individuals would prefer not to work and be on disability benefit where τ^z solves: $(1-\tau)w(0)=b$ and where $w(\theta)=w(0)$ and $f(\theta) \propto h(\theta)^{-1}$. (iii) There is no level of τ that yields efficiency.

Proof of (iii): Suppose that there is a τ that maximises output. This would mean that

$$\frac{\partial w}{\partial(1-h)} = \frac{u((1-\tau)w(h))-u(b)}{(1-\tau)hu'((1-\tau)w(h))} = \frac{w}{h} \text{ for any } \theta. \text{ This equation in turn can only hold for a}$$

continuum of h when $u'(\cdot)=0$ and $\tau=0$, which implies that individuals would have to be risk-neutral which contradicts the primitives of the model. The proofs of (i) and (ii) are given in the Appendix.

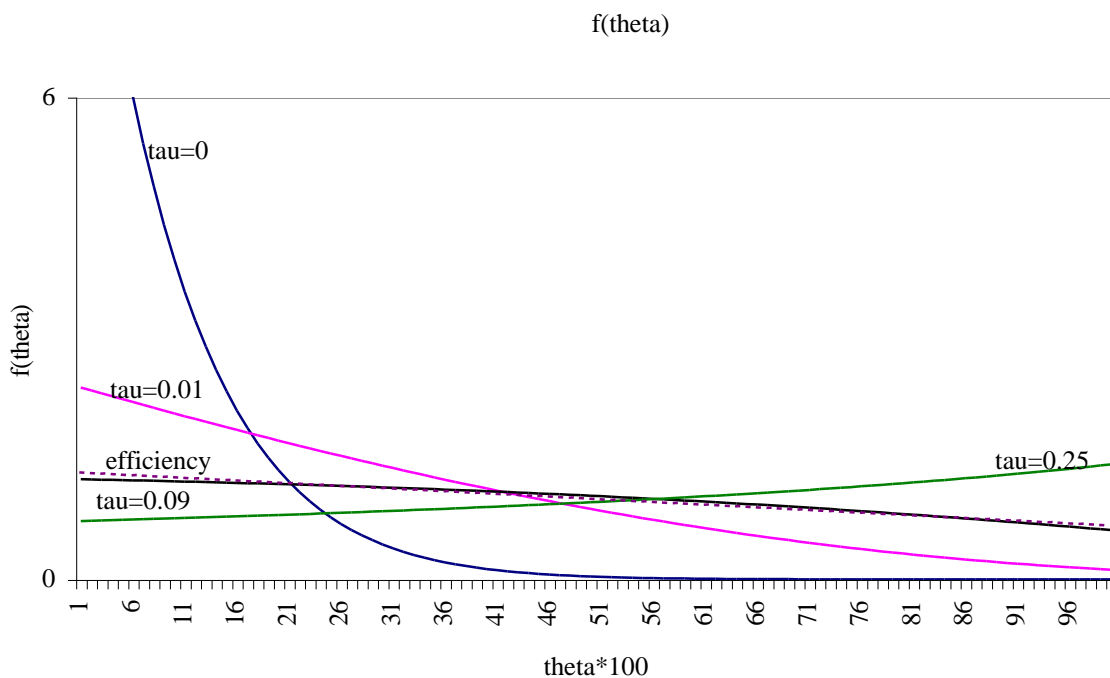
The intuition of this result is that at low levels of tax, the introduction of a benefit induces individuals to take more risks, which increases output, increases disability, and increases utility. Since the utility of each choice is the same, the utility increase following an increase in τ is *ex ante* the same for each individual. Because of the irreversibility of choosing the occupation and its associated health risk, *ex post* some individuals will not want taxation. The individuals with $\theta = 0$ for instance have no risks *ex post*, and will hence oppose any tax *ex post*, even though they have benefited from it *ex ante*.

The reasons why there is no tax level that will yield the maximum output is that *ex post* individuals want different levels of insurance: given that individuals will choose different risks, the optimal insurance should differ for different levels of risk. The *ex post* risk pooling between individuals with different risks means that those who run little risks will be over-insured and those with high risks will be under-insured. This highlights a basic difference between considering wage

distributions instead of a single wage outcome, such as Acemoglu and Shimer (1999), where there is a single tax that restores efficiency.

The key features of this model are highlighted in Figures 1, 2 and 3. Figure 1 shows the relationship between $f(\theta)$ and τ .

FIGURE 1: The Relation between Risks and the Distribution of Occupational Choices



Note: Parametric assumptions: $h(\theta)=1-0.5*\theta$, $\gamma = 0.5$, and $u(y) = y^{0.1}$.

It is clear that the distribution of risks when $\tau = 0$ is too skewed to the left in comparison to efficiency. As τ increases, the distribution becomes less tilted. At $\tau = 0.09$, the distribution is almost identical to the efficient distribution. When $\tau = \tau^Z = 0.25$, we have the limit case, where all wages are equal and hence $f(\theta) \propto \frac{1}{h(\theta)}$.

Figure 2 shows the compensating wage profile of the same tax regimes. At $\tau = 0$, wages rise quickly with risks and hence with θ . As τ increases, the wage profile becomes less skewed to the right. At $\tau = 0.09$, we get the almost efficient wage curve. In the very limit case of $\tau = 0.25$, wages are constant.

FIGURE 2: The Relation between Compensating Wages Profiles and Taxation

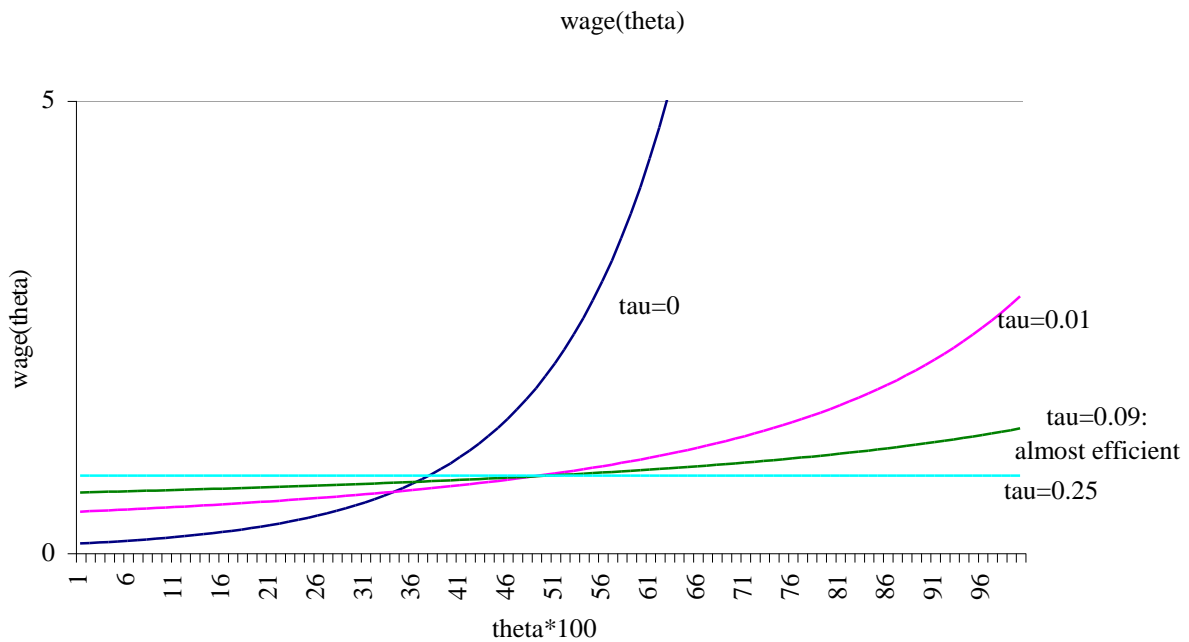
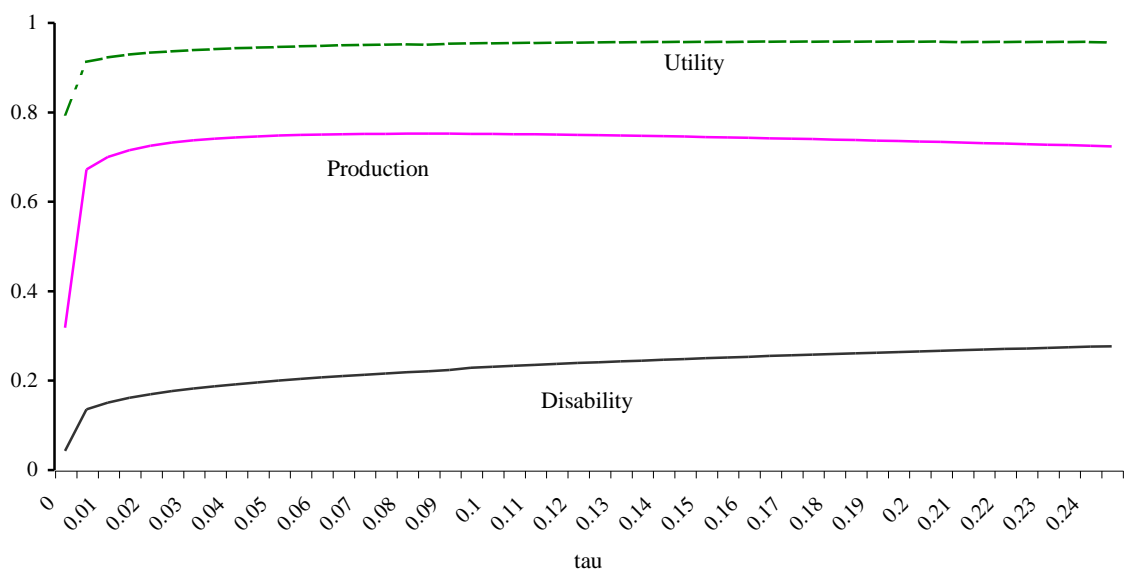


Figure 3 highlights the relationship between τ on the one hand and utility, production and disability. It is clear that each rises quite rapidly for very low levels of τ . Production peaks quickly (i.e. $\tau=0.09$) and then slowly levels off. Utility peaks much later at $\tau=0.22$, and disability increases until the limit of $\tau=0.25$.

FIGURE 3: The Effect of Taxation on Utility, Production and Disability Benefit Take-Up



3. Including Heterogeneous Health Endowments

As an additional exercise we explore how individuals with different health endowments behave in this economy. To this extent, suppose that individuals have an innate health endowment $1 > q > 0$ that is drawn from a differentiable population distribution $G(q)$. The lower q , the better the innate health of some is. This health endowment can be interpreted as an efficiency unit where the probability of having a job is equal $h(\theta^*q)$. This means we can interpret q as the innate propensity of an individual to be unhealthy.

An immediate implication is that:

Proposition 3: Wages compensate for risks and those with higher innate health choose the riskier occupations, i.e. $q_a > q_b$ then $\theta_a < \theta_b$.

Proof: Suppose we have an equilibrium wage function $w(\theta)$. We now only have to look at the minimum compensation $w^{mc}(\theta, q)$ someone with innate health q would need to prefer occupation θ^* over occupation 0:

$$u((1-\tau)w^{mc}(q, \theta)) = u(b) + \frac{\{u((1-\tau)w(0)) - u(b)\}}{h(q\theta)} \quad (4)$$

which implicitly defines the minimum compensation $w^{mc}(q, \theta)$ as an increasing function of $\frac{\{u((1-\tau)w(0)) - u(b)\}}{h(q\theta)}$. This term is increasing in q , which implies the minimum compensation for accepting an occupation is increasing with q . Therefore, those who have the highest innate health (low q) will at lower compensation accept higher θ occupations. This means it cannot be the case that those with higher health accept occupations with lower θ .

This proposition hence shows that jobs with 'objectively' the highest health risks (θ) will be taken up by those individuals whose health is least affected by these risks (low q). Thus there is a self-selection into occupational type on the basis of health endowments.

4. Dynamics: A Calibrated OLG Model

In order to investigate the dynamics of the two-period model we perform the following calibrations. Assume that there are only two occupations: one with a high health risk (e.g. a coal-miner) and one with a low health risk (e.g. an academic). Newborn, ex ante identical, individuals choose their future occupation. After 20 periods they die. The population grows at rate n , which means that a

new cohort then replaces the cohort that dies, which is a fraction $(1+n)^{20}$ bigger.¹ A fraction β_s of the cohort born at time s will choose low-risk future occupations, and a fraction $(1-\beta_s)$ choose high-risk future occupations. Those with low-risk future occupations have job-finding rates equal to λ^L and job-destruction rates equal to δ^L . Those with high-risk (often specialised) future occupations have job-finding rates equal to $\lambda^H = \lambda^L$ and job-destruction rates equal to $\delta^H > \delta^L$. The timing is such that at the beginning of the period those individuals working the previous period can become disabled, after which they can never work again. The unemployed search for jobs during the period, whereby production takes place at the end of each period. Individuals are assumed to have rational expectations and have standard per-period utility functions: $u(x) = \ln(x+A)$. Because we calibrate working lives, each period corresponds to about 2.5 years. Taking standard estimates for the discount rate from empirical studies (e.g. Frijters and Van der Klaauw, 2001), we take ρ to be equal to 10% per period. There is no mortality before period 20.

An individual i born at time s hence maximises:

$$E\{U_i\} = \sum_{t=s}^{s+T} P_t(\text{working} | \theta_i, s) \ln[A + (1-\tau_t)w_t(\theta_i)] + (1-P_t(\text{working}, \theta_i))u(A+b_t) \quad (5)$$

where $P_t(\text{working} | \theta_{i,s})$ denotes the probability that someone with choice $\theta_i \in \{H, L\}$ who is born at time s works at time t . This probability follows the following Markov-chain rule:

$$\begin{aligned} P_t(\text{working} | \theta_i, s) &= (1-\delta_{i,t})P_t(\text{working} | \theta_i, s) + \lambda_{i,t} * NU_{s,t} \\ NU_{s,t} &= (1-\lambda_{i,t-1})NU_{s,t} \end{aligned} \quad (6)$$

where $NU_{s,t}$ is the proportion of those born at time s who are still unemployed at the beginning of period t . By definition, $NU_{s,s} = 1$ and $P_t(\text{working} | \theta_{i,s}) = 0$ for $t < s$.

Total production equals:

$$Y_t = \{(X_t^H)^\gamma + (X_t^L)^\gamma\}^{1/\gamma} \quad (7)$$

where $X_t^H = \sum_{s=t-T}^{s=t} P_t(\text{working} | H, s)$ denotes the total measure of individuals working in high-risk occupations at time t and X_t^L denotes the total measure of individuals working in low-risk

¹ Because of the constant returns to scale production function, population growth will not substantially affect steady state relations. Population growth will however serve to 'dampen out' fluctuations.

occupations at time t . Wages are competitive: $w_t(H) = \frac{\partial Y_t}{\partial X_t^H}$ and $w_t(L) = \frac{\partial Y_t}{\partial X_t^L}$. Benefits solve the budget constraint of the government: benefits are equal to total amount of taxes collected divided by the number of unemployed plus disabled. Individuals have rational expectations and hence correctly predict future wages, benefits, and working probabilities.

We first show some baseline calculations of the steady state, where we have constructed two different baseline economies. The key variables $\{\gamma, \tau, n, \lambda^L, \lambda^H, \delta^H, \delta^L\}$ are calibrated on statistics from the US and the EU. In line with Ljungqvist and Sargent (1998), we do not allow job-destruction rates or basic production processes to differ. We have hence set $\{\gamma, \lambda^H, \lambda^L, \delta^H, \delta^L\}$ to be equal for both economies, in order to allow for proper comparisons of the dynamics in the shock experiments. Also, the parameters were set such that the level of specialists is much higher in one economy, in order to be assured that our risk-arguments have some relevance for the two economies. The statistics we took into account are average job-finding rates, disability rates, population growth rates in the last 30 years, job-destruction rates, and level of IB. The underlying data is from the OECD.

Table 2: Baseline Calibrations for a High Tax and Low Tax Economy

	European Union	USA
Specific Inputs		
n per year	0.01	0.02
τ	0.15	0.02
Outcomes		
β	0.58	0.79
$w^G * (1 - \tau)$	1.43	1.34
$w^S * (1 - \tau)$	2.44	3.66
Benefits	0.63	0.26
Average production per unit of work	1.44	1.46
Disability plus unemployment	0.23	0.11
Utility	3.63	3.04

Note: Common inputs: $A=0.1$; ρ per year=0.05; $\gamma=0.5$; $\lambda=0.9$; $\delta^G=0$; $\delta^L=0.1$.

In the 'EU' baseline model, taxes and disability levels are much higher than in the 'US' baseline model but welfare is slightly higher in the 'EU'. In the 'US', output (per hour input), population

growth, and wage-differentials are higher (i.e. GDP per person is much higher in the US than in the EU, yet this is largely due to more units of inputs). The sensitivity to taxes is quite large: even though basic production possibilities are set equal for both economies, the differences in the taxes translate to big differences in disability. Output maximising taxes in these economies would be 0.05. Given the assumptions used in these calibrations, the `US' would have too low taxes and the `EU' too high. We do not want to argue that these are convincing numbers, but merely that our simple model can fit an array of observed differences between economies.

We now turn to the effect of unanticipated shocks. Starting from steady state, we perturb the economy at $t=0$. We then assume that expectations instantaneously realign themselves i.e. after the shock all individuals know what will happen next. We look at two different shocks:

1.A permanent biased search shock. At $t = 0$, δ^S is halved.

2.A welfare system change. At $t = 0$, τ is halved.

The results for (1) are shown in Figures 4 and 5, and the results for (2) are shown in Figures 6 and 7.

FIGURE 4: The response of the 'USA' to decreased risks for specialists

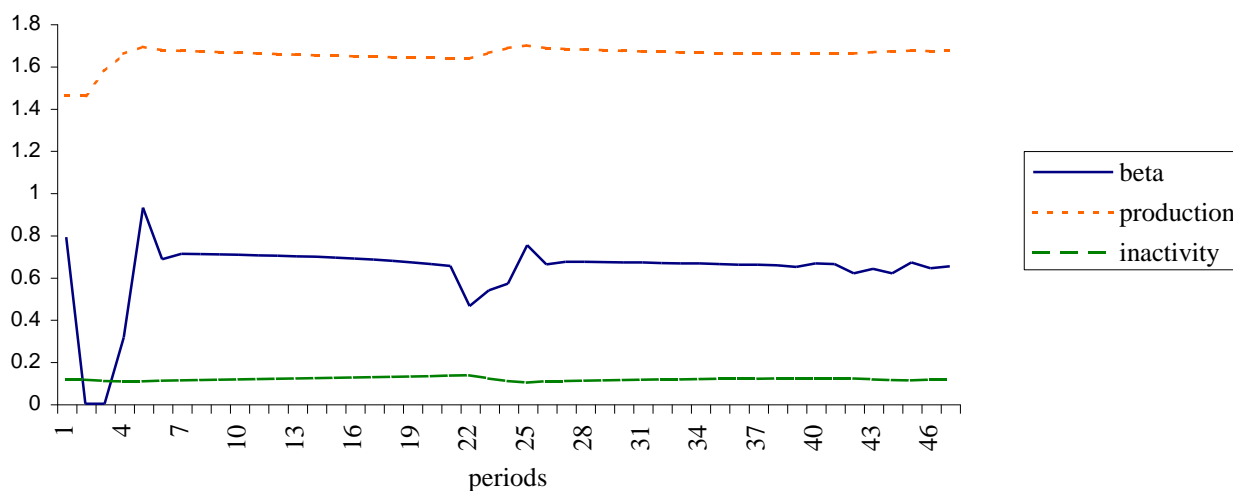
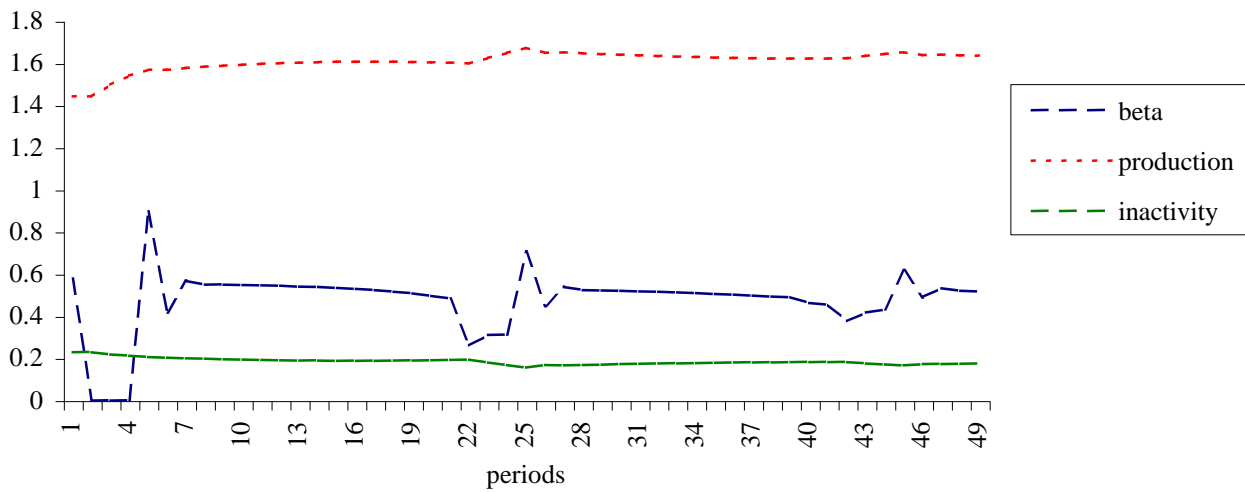


FIGURE 5: The response of the 'EU' to decreased risks for specialists



A permanent biased search shock has stronger effects on the 'US' than on the 'EU', simply because the proportion of specialists at the time of the shock is much higher in the 'EU'. The risks associated with being a specialist fall, which encourages more specialists. As an immediate effect, entire cohorts become specialists in the 'US' and the 'EU'. This causes a ripple effect for the subsequent periods that slowly dies down. In the long run, the proportion of specialists goes up in both economies and production increases. Disability levels change very little in the 'US', even though the risk for any individual job has decreased. In the 'EU', there is an initial drop in disability as fewer specialists become inactive. However, in the longer run the increased proportion of specialists, who still run higher risks than the generalists, leads to disability levels that are only about 25% lower, even though risks for any particular occupation have been halved. These general equilibrium effects thus reduce the long-term impact of any (policy induced) change in the risk level of a subset of occupations.

FIGURE 6: The response of the 'USA' to a halving of tax rates

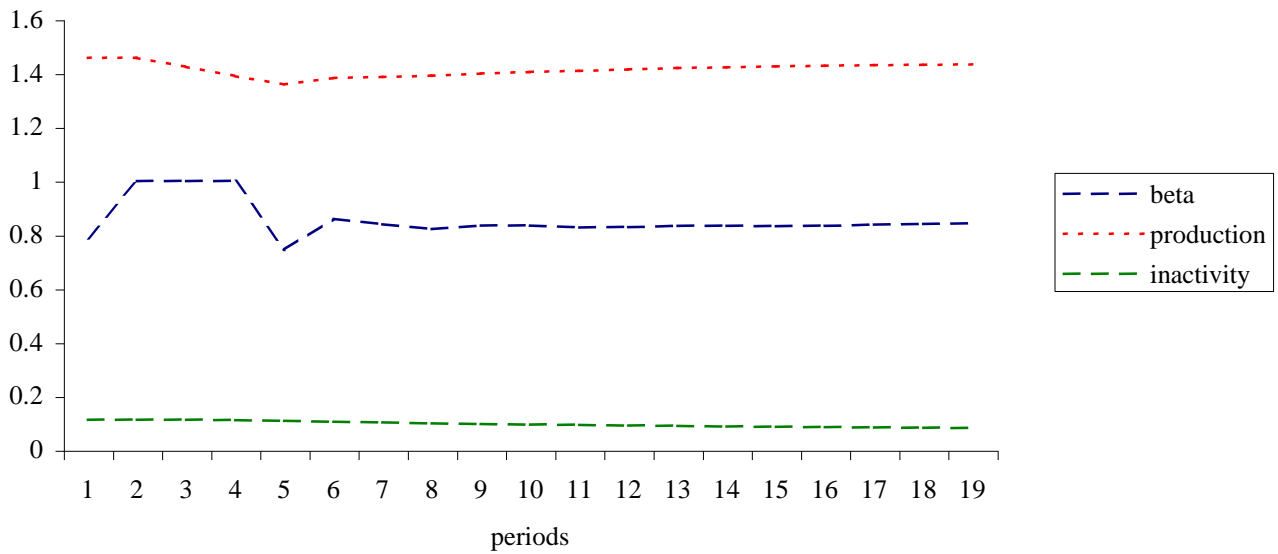
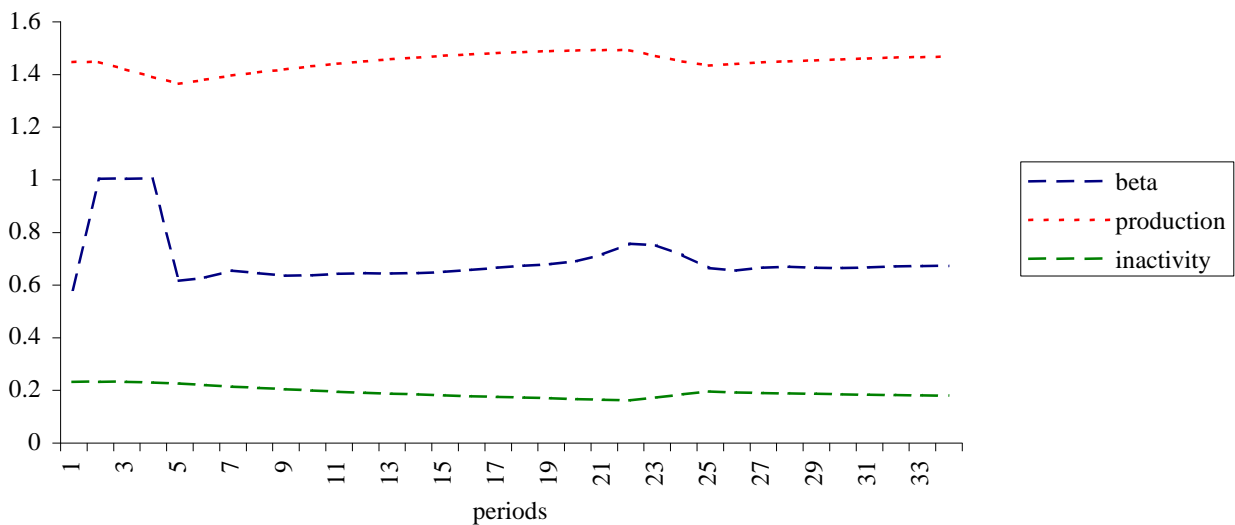


FIGURE 7: The response of the 'USA' to halving the tax rate



The welfare system change increases the incentives to become generalist. Because in the 'US', the proportion of generalists was already very high, the long-run effect of lowering τ on disability and β is small: production and disability rates decrease very slightly. All that really changes is a large increase in the wages of specialists (not shown). There are bigger effects in the 'EU', where the changes in β are larger, leading to long-lasting ripples. The effect on production and output is very delayed, however, simply because it takes time for new cohorts of generalists to come through the education system and substantially change the composition of the labour force: it takes about 20 periods (≈ 50 years) before production and disability levels have reached their new steady state. The long-term effects are that production increases and disability decreases because of the reduced

numbers of specialists in the economy. These predictions seem to mirror the sluggishness with which disability rates have been found to react to changes in the level of benefits, and indeed to other welfare changes (Blanchard and Wolfers, 2000; Dolado *et al.*, 1996; Gruber and Wise, 1997).

5. Conclusions

One of the most prominent trends in OECD countries over the last 30 years has been the sharp increase in incidence of early retirement, and in particular the permanent take-up of disability benefits. In this paper we construct a theoretical model that shows how occupational choices, in terms of the associated health risks, made by the young can be affected by the expected provision of publicly funded disability benefits in later life. Modest disability benefits are found to lead to higher welfare and output because risk-averse individuals take too few risks in the absence of any insurance. Too high benefits lead to inefficiently high-risk taking, compressed wage distributions and lower welfare. These results were shown to hold in a general equilibrium setting where the wages of all occupations were jointly determined. We have also extended the model to allow for heterogeneous health endowments. We show that occupations with the highest health risks will be chosen by those individuals whose health is least affected by these risks. Thus there is likely to be self-selection into occupational type on the basis of health endowments.

The main policy-relevant prediction of the model is that the composition of the labour force reacts sluggishly to changes in circumstances because of the irreversibility of the early life choices of older cohorts. This could be a partial explanation for why it took several decades for generous disability benefit schemes to lead to the very high levels of take-up rates in several OECD countries and would mean that take-up rates would only decrease slowly if generosity were to be reduced. Dynamically, we would expect that new cohorts would be affected sooner by changes in the generosity of benefits, because new cohorts are still in a position to make changes to their career choices. In summary, whilst we have proposed a model that we believe is useful in understanding these dynamic relationships, there still remains a large amount of future research which needs to be undertaken into the dynamic reaction of cohorts to changing expectations about different welfare systems.

References

- Aarts, L., Burkhauser, R. and de Jong, P (eds) (1996). *Curing the Dutch Disease: An International Perspective on Disability on Disability Policy Reform*. International Studies on Social Security, vol. 1. Aldershot: UK.
- Acemoglu, D. and Shimer, R. (1999). Efficient unemployment insurance. *Journal of Political Economy*, 107, pp. 893-928.
- Akerlof, G. (1978). The economics of 'tagging' as applied to the optimal income tax, welfare programs, and manpower planning'. *American Economic Review*, 68, pp. 8-19.
- Autor, D. and Duggan, M. (2003). The rise in the disability rolls and the decline in unemployment. *Quarterly Journal of Economics*, CXVIII, pp. 157-205.
- Black, D., Daniel, K. and Sanders, S. (2002). The impact of economic conditions on participation in disability programs: Evidence from the coal boom and bust. *American Economic Review*, 92, pp. 27-51.
- Blanchard, O. and Wolfers, J. (2000). The role of shocks and institutions in the rise of European unemployment. *Economic Journal*, 110, pp. C1-C33.
- Blondal, S. and Scarpetta, S. (1997). Early retirement in OECD countries: The role of social security systems. *OECD Economic Studies*, 29, pp. 1-27.
- Boadway, R., Marceau, N. and Sato, M. (1999). Agency and the design of welfare systems. *Journal of Public Economics*, 73, pp. 1-30.
- Bound, J. and Burkhauser, R. (1999). Economic analysis of transfer programs targeted on people with disabilities. In Ashenfelter, O. and Card, D. (eds.). *Handbook of labor economics*. Volume 3C. *Handbooks in Economics*, vol. 5. Amsterdam; New York and Oxford: Elsevier Science, North-Holland, 1999, pages 3417-3528.
- Bound, J. and Waidmann, T. (2002). Accounting for recent declines in employment rates among working-aged men and women with disabilities. *Journal of Human Resources*, 37, pp. 231-250.
- Buddelmeyer, H. (2001). Re-employment dynamics of disabled workers. IZA Discussion Paper no. 269, March.
- Burkhauser, R. and Daly, M. (2002). Policy watch: US disability policy in a changing environment. *Journal of Economic Perspectives*, 16, pp. 213-224.
- Coile, C., Daimond, P., Gruber, J. and Josten, A. (2002). Delays in claiming social security benefits. *Journal of Public Economics*, 84, pp. 357-385.
- Cullen, J. (2003). The impact of fiscal incentives on student disability rates. *Journal of Public Economics*, 87, pp. 1557-1589.
- Diamond, P. and Mirrlees, J. (1978). A model of social insurance with variable retirement. *Journal of Public Economics*, 10, pp. 295-336.

- Diamond, P. and Gruber, J. (1999). Social security and retirement in the US. In Gruber, J. and Wise, D. (eds.) *Social Security and Retirement around the World*. University of Chicago press: Chicago.
- Diamond, P. and Sheshinski, E. (1995). Economic aspects of optimal disability benefits. *Journal of Public Economics*, 57, pp. 1-23.
- Gruber, J. and Wise, D. (1998). Social security and retirement: An international comparison. *American Economic Review, Papers and Proceedings*, 88, p. 158-163.
- Harkness, J. (1993). Labour force participation by disabled males in Canada. *Canadian Journal of Economics*, 26, pp. 878-889.
- Johnson, W. and Ondrich, J. (1990). The duration of post-injury absences from work. *Review of Economics and Statistics*, 72, pp. 578-586.
- Kerkhofs, M., Lindeboom, M. and Theeuwes, J. (1999). Retirement, financial incentives and health. *Labour Economics*, 6, pp. 203-227.
- Kreider, B. and Riphahn, R. (2000). Explaining applications to the US disability System: A semiparametric approach. *Journal of Human Resources*, 35, pp. 82-115.
- Ljungqvist, L. and Sargent, T. (1998). The European unemployment dilemma. *Journal of Political Economy*, 106, pp. 514-550.
- Parsons, D. (1996). Imperfect 'tagging' in social insurance programs. *Journal of Public Economics*, 62, pp. 183-207.
- Pissarides, C. (1990). *Equilibrium Unemployment Theory*. Blackwell: Oxford.
- Westerhout, E. (2001). Disability risk, disability benefits, and equilibrium unemployment. *International Tax and Public Finance*, 8, pp. 219-243.

Appendix

We begin by highlighting a useful result::

Lemma 1: When going from an initial distribution $F_1(\theta)$ to a new distribution $F_2(\theta)$ with a continuous CES, wages will be higher for at least one θ and wages will be lower for at least one θ .

Proof: Consider the level $\theta^* = \arg_{\theta} \max \left\{ \frac{f_2(\theta)}{f_1(\theta)} \right\}$. If $F_2 \neq F_1$, it has to hold that $\frac{f_2(\theta^*)}{f_1(\theta^*)} > 1$.

Therefore:

$$\begin{aligned} \frac{w_2(\theta^*)}{w_1(\theta^*)} &= \frac{(h(\theta^*)f_2(\theta^*))^{\gamma-1} \left[\int (h(\theta)f_2(\theta))^{\gamma} \partial\theta \right]^{\frac{1}{\gamma-1}}}{(h(\theta^*)f_1(\theta^*))^{\gamma-1} \left[\int h(\theta)f_1(\theta)^{\gamma} \partial\theta \right]^{\frac{1}{\gamma-1}}} \\ &= \left[\frac{f_2(\theta^*)}{f_1(\theta^*)} \right]^{\gamma-1} \left[\frac{\int (h(\theta)f_2(\theta))^{\gamma} \partial\theta}{\int (h(\theta)f_1(\theta))^{\gamma} \partial\theta} \right]^{\frac{1}{\gamma-1}} \end{aligned} \quad (\text{A1})$$

Hence:

$$\frac{w_2(\theta^*)}{w_1(\theta^*)} < \left[\frac{f_2(\theta^*)}{f_1(\theta^*)} \right]^{\gamma-1} \left[\frac{f_2(\theta^*)}{f_1(\theta^*)} \right]^{\frac{1}{\gamma-1}} < 1 \quad (\text{A2})$$

which implies that $w_2(\theta^*) < w_1(\theta^*)$. Using the same argument for $\theta^* = \arg \min \left\{ \frac{f_2(\theta)}{f_1(\theta)} \right\}$, we can see

that there is also a θ^* for which $w_2(\theta^*) > w_1(\theta^*)$. When $f(\theta)$ is continuous, this strict inequality implies that any change in $f(\theta)$ means that a whole range of wages must decrease and another range must increase.

Proof of Proposition 1: Uniqueness of $w(\theta)$. First, we note that the differential equation:

$$-h'(\theta)(u(w(\theta)) - u(0)) = (1 - \tau)w'(\theta)h(\theta)u'(w(\theta)) \quad (\text{A3})$$

defining the equilibrium $w(\theta)$ is well-behaved in the sense that $w(\theta)$ is uniquely determined by

$w(0)$. Also, $\frac{\partial w(\theta)}{\partial w(0)}$ is continuous and bigger than 0. In turn, the equation $\frac{w(\theta)}{w(0)} = \frac{(h(\theta)f(\theta))^{\gamma-1}}{(h(0)f(0))^{\gamma-1}}$

means that $\frac{w(\theta)}{w(0)}$ uniquely determines $\frac{f(\theta)}{f(0)}$ and $f(0)$ is then solved by $\int f d\theta = 1$. The function $f(\theta, w(\theta))$ is also continuous in $w(\theta)$. By implication of $w(\theta)$ being continuous in $w(0)$, the implicit function $f(\theta, w(0))$ must therefore also be continuous in $w(0)$.

Finally, equilibrium requires that the level $w(0)$ also solves:

$$(h(0)f(0, w(0)))^{\gamma-1} \left[\int (h(\theta)f(\theta, w(0)))^{\gamma} \partial\theta \right]^{\frac{1}{\gamma}-1} = w(0) \quad (\text{A4})$$

For $w(0) \downarrow 0$, we first can note that at $\lim_{w(\theta) \downarrow 0}$ the defining condition $-h'(\theta)(w(w(\theta)) - u(0)) = (1 - \tau)w'(\theta)h(\theta)u'(w(\theta))$ reduces to $\frac{w'(\theta)}{w(\theta)} = \frac{-h'(\theta)}{h(\theta)}$. Translating this into $f(\theta)$ implies that the left-hand side of the above expression will converge to some positive number. For $w(0) \uparrow \infty$, we can note that in the limit $u'(w(\theta))$ becomes constant, which in turn pins down the left-hand side of the expression above to a finite number. Because of the continuity of $f(\theta, w(0))$ the fixed-point theorem hence applies and there must be at least some level $w(0)$ for which the condition is satisfied.

Considering uniqueness, suppose that there are two wage functions w_1 and w_2 , that are an equilibrium. Without loss of generality, take $w_2(0) > w_1(0)$. It then has to be the case for w_2 all will prefer $\theta = 0$ above any other level θ unless $w_2(\theta) > w_1(\theta)$ for any $\theta > 0$. If this does not hold, there is a positive mass that will choose 0 and there will be no mass choosing a quality slightly above 0. This in turn would mean that $w_2(0) = 0$, which cannot be the case. Hence, there can only be a second equilibrium if $w_2(\theta) > w_1(\theta)$ for all $\theta > 0$. As shown in Lemma 1, this is an impossibility. If there is an equilibrium, it therefore has to be unique.

In this proof it is not necessary to assume that $w'(\theta)$ or $f(\theta)$ are continuous. This means that we can include cases where $h(\theta)$ is not continuous. In such cases, we can apply the same proof, but simply note that $w(\theta)$ is then defined by the more general requirement that $E\{U(\theta)\} - E(U(0)) = 0$. Note that in the text, all the formulas are given for the continuous case for ease of exposition.

Proof of Proposition 2:

Our strategy for proving i) is to prove that when taxation increases from 0 to some arbitrarily small ε , that there will be a $1 > \theta^* > 0$ for which $w(\theta)$ remains constant. Then we show that for all

$\theta < \theta^*$, $w(\theta)$ has increased. For all other θ , $w(\theta)$ has decreased. Because $w'(h)$ will still be greater than $\frac{w}{h}$, we can then write the changes in $w(\theta)$ as a succession of production-increasing changes, which establishes production improvements. For θ^* we then show that utility has increased, which implies that it has to have increased for all choices.

Looking at $\frac{d \frac{dw}{d(1-h)}}{d\tau}$ there holds:

$$\begin{aligned} \frac{d \frac{dw}{d(1-h)}}{d\tau} = & \frac{1}{(1-\tau)hu'((1-\tau)w(h))} * [\{-w(h)u'((1-\tau)w(h)) - \frac{db}{d\tau}u'(b)\} + \\ & \{u((1-\tau)w(h)) - u(b)\} * \{ \frac{1}{1-\tau} + \frac{w(h)u''((1-\tau)w(h))}{u'((1-\tau)w(h))} \}] \end{aligned} \quad (A4)$$

Using that $\frac{db}{d\tau} |_{\tau=0} = \frac{\int hfw d\theta}{\int (1-h)fd\theta} > \frac{h(0)w(0)}{\int (1-h)fd\theta} > \frac{h(0)w(0)}{1-h(1)} > w(0)$ and that

$u((1-\tau)w(\theta)) - u(b) < ((1-\tau)w(\theta) - b)u'(b)$ we hence know that $\frac{d \frac{dw}{d(1-h)}}{d\tau} < 0$ at $\theta = 0$ for τ close

to 0. This in turn establishes that the whole wage function must have changed. We denote the changed wage function as $w(\theta, \varepsilon)$. Using the argument as in Lemma 1, we thus know that there has to be a whole range of θ for which $w(\theta)$ has increased. Because of continuity, this also means that there will be a $1 > \theta^* > 0$ for which $w(\theta, \varepsilon) = w(\theta)$. Since $w(\theta, \varepsilon)$ equalizes utility, this also implies that $sign\{w(\theta, \varepsilon) - w(\theta)\} = sign\{\theta^* - \theta\}$: because the importance of the risk of disability has decreased, taking more risks must have become relatively more attractive. This in turn means that

$\frac{d(f(\theta, \varepsilon) - f(\theta))}{d\theta} > 0$. For small ε we can write

$y(\tau = \varepsilon) - y(\tau = 0) = \int \{f(\theta, \varepsilon) - f(\theta)\}h(\theta)w(\theta)d\theta$. Now, because

$\frac{dw}{d(1-h)} = \frac{u((1-\tau)w(h)) - u(b)}{(1-\tau)hu'((1-\tau)w(h))}$ is continuous in τ , $\frac{dw}{d(1-h)}$ will still be greater than $\frac{w}{h}$ for very

small τ . Hence we can still use that $\frac{dh(\theta)w(\theta)}{d\theta} > 0$, implying that $y(\tau = \varepsilon) - y(\tau = 0) > 0$.

Using that $\frac{dw}{d(1-h)} \Big|_{h=1/\frac{d\frac{dw}{d(1-h)}}{d\tau}}^{\tau=0, h=1} < -1$ and that average wages have increased, we also know

that $h(\theta^*) < (1-\varepsilon)$ and thereby that $w(0, \varepsilon) > (1+\varepsilon)w(0)$. This in turn means that the wage increase at $\theta = 0$ more than offsets the tax increase, which establishes a utility increase for all θ .

(ii) First, we simply define τ^Z as the τ that solves $(1-\tau)w(0, \tau) = b(\tau)$. For this level of τ ,

$\frac{dw}{d(1-h)}$ has to equal 0 because otherwise individuals could increase their expected utility by

changing θ . Hence $w(\theta) = w(0)$ and $f(\theta) \propto h(\theta)^{-1}$. Also, $\frac{d\frac{dw}{d(1-h)}}{d\tau} \Big|_{\tau=\tau^Z} > 0$. This then implies that

if $\tau > \tau^Z$, that $\frac{dw}{d(1-h)} < 0$ and that individuals would want to take as much risk of disability as

possible, which in turn means that they would not want to work at all.