

MORAL HAZARD, INCOME TAXATION, AND PROSPECT THEORY*

by

Ravi Kanbur

Cornell University

Jukka Pirttilä[#]

Labour Institute for Economic Research

and

Matti Tuomala

University of Tampere

This version: August 2007

* We are grateful to Hamish Low, Hisahiro Naito, Agnar Sandmo, Fred Schroyen, Tuomas Takalo, three anonymous referees and the editor for helpful comments. The paper has also benefited from comments by seminar participants at the Cornell/LSE/MIT Conference on Behavioral Economics, Public Economics and Development Economics, the Helsinki Centre for Economic Research and the University of Umea.

[#] Corresponding author, address: Labour Institute for Economic Research, Pitkäsillanranta 3 A, 00530 Helsinki, Finland. email: Jukka.Pirttila@labour.fi

Abstract: The standard theory of optimal income taxation under uncertainty has been developed under the assumption that individuals maximize expected utility. However, prospect theory has now been established as an alternative model of individual behaviour, with empirical support. This paper explores the theory of optimal income taxation under uncertainty when individuals behave according to the tenets of prospect theory. It is seen that many of the standard results are modified in interesting ways. The first-order approach for solving the optimisation problem is not valid over the domain of losses, and the marginal tax schedule offers full insurance around the reference consumption level. The paper also examines the implications of non-welfarist objectives under income uncertainty.

Key words: redistributive taxation, income uncertainty, moral hazard, prospect theory, loss aversion

JEL classification: D81, H21.

1 Introduction

In principal-agent models with moral hazard, the agent's income and utility depends randomly on effort. In one of his Nobel prize winning contributions, Mirrlees (1974) characterizes optimal income taxation where the government is the principal and ex ante identical individuals are the agents¹. In this contribution, as in most principal-agent analysis, expected utility theory is used as a description of agents' behaviour under uncertainty. In his Nobel Lecture, Mirrlees (1997, p. 1324) calls for closer scrutiny of this approach:

'Problems of this kind are usually analysed with the assumption that people try to maximise their expected utility. There are good reasons for thinking that may be a mistake. At least the consequences of alternative theories of decisions under uncertainty for these situations should be explored.'

One such alternative theory of decision making under uncertainty has in turn led to another Nobel prize. Prospect theory, developed by Kahneman and Tversky (1979) and modified by Tversky and Kahneman (1992), has garnered significant empirical support (see Kahneman's Nobel lecture, 2003, and Camerer and Lowenstein, 2003). In prospect theory, an individual's utility depends on how the outcome deviates from some reference point, rather than directly on the absolute value of the outcome. Individuals are loss averse, in other words, a loss leads to a larger change in welfare than a gain of a similar size. Finally, individuals may misperceive probabilities underlying the decision problem. In his review of alternative theories for expected utility theory, Starmer (2000, p.376) concludes that because of many appealing features of rank-dependent or sign-dependent models of decision-making under uncertainty, such

¹ Of course the case with no uncertainty but where individuals differ in their productivities, was also introduced and explored by Mirrlees (1971) in the more famous of his Nobel prize winning contributions. We will refer to

as prospect theory, ‘there seems good reason to push forward the task of examining what implications such models have in general economic contexts’.²

The purpose of this paper is to confront the Mirrlees project of characterizing optimal income taxation with moral hazard under uncertainty, with the Kahneman project of developing alternatives to standard expected utility theory. In the original Mirrlees’ (1974) formulation of the income tax model, workers and the government maximise workers’ expected utility over income and effort. Our purpose is to introduce elements of prospect theory into individual behaviour, in keeping with the emerging empirical consensus, to examine how optimal taxation results change with the introduction of prospect theory preferences.³

We first assume that the government respects the individual preferences, i.e. it is a ‘welfarist’ government. However, there are elements in prospect theory that may or may not be desirable from the social welfare point of view. For instance, the social planner may dislike the consumers’ tendency to be risk seeking for losses. Therefore, we also consider the case where the government is ‘non-welfarist’ (paternalistic) and its objective function may differ from that used by the individuals. This approach is relatively common in conventional public economics, and has been used recently in the behavioural public economics literature as well.⁴

this “adverse selection” case from time to time, but our focus is on the “moral hazard” case where there is uncertainty but no ex ante differentiation among individuals.

² Munro (2004) provides an analysis of welfare economics, especially the evaluation of tax reforms, under reference-dependent utility functions in a riskless choice case. In an interesting paper, Dhami and al-Nowaihi (2006) explore the consequences of prospect theory on tax evasion.

³ We became aware of the work by de Meza and Webb (2007) after having written earlier versions of this paper. The de Meza and Webb paper also explores the consequence of loss aversion and diminishing sensitivity for optimal incentive policy. Their work deals with managerial incentives, whereas our paper offers a tax policy application.

⁴ Examples of the former include Kanbur, Keen and Tuomala (1994) and Pirttilä and Tuomala (2004), while O’Donoghue and Rabin (2003), Bernheim and Rangel (2005) and McCaffery and Slemrod (2006) are examples of the latter. See Seade (1980) for seminal work.

The paper first reviews the standard, benchmark, model of optimal taxation with moral hazard under income uncertainty, in Section 2. As is common in models of this kind, we mostly focus on results based on the so-called “first-order approach”. This section therefore also examines the exact conditions under which this approach is valid, an issue that will seem to be relevant when prospect theory is introduced. Section 3 examines the validity of the first-order approach and characterises optimal tax rules when individual behaviour is described by prospect theory. Section 4 introduces non-welfarist concerns and derives results for the tax rules in this case. Section 5 concludes.

2. The standard model

Consider an economy, as in Mirrlees (1974), where the worker-consumer does not know what income (z) he or she will receive for each possible level of effort, y . In other words, income is determined both by effort and by some random element. We denote the distribution for z , given that effort y is undertaken by the worker, as $F(z, y)$, and its density as $f(z, y)$. It is assumed that $f(z, y)$ and $F(z, y)$ are continuous and continuously differentiable for all $z \in [z_0, z_1]$ and y . Moreover we assume that the support of this distribution is independent of y .

The worker-consumer chooses effort y to maximize expected utility

$$(1) \int v(x) f(z, y) dz - y,$$

where $x = z - T(z)$ is the after tax income / consumption. As in much of the literature, we concentrate on an additively separable specification of the utility function. The consumer is risk averse, hence $v' > 0$, $v'' < 0$. The first-order condition for the maximisation of (1) is

$$(2) \int v(x) f_y dz - 1 = 0.$$

The government is utilitarian and maximises (1) subject to the individual optimisation constraint (2) and the budget constraint which, for a large identical population with independent and identically distributed states of nature, can be written in the form

$$(3) \int [z - x] f(z, y) dz = 0$$

Taking multipliers α and λ for the constraints (2) and (3) respectively, the Lagrangean and the first-order condition with respect to x (pointwise optimisation) are as follows:

$$(4) L = \int \{ [v(x) + \lambda(z - x)] f(z, y) + \alpha v(x) f_y \} dz - \alpha - y$$

$$(5) 1 + \alpha g = \frac{\lambda}{v'},$$

where $g = f_y / f$ is the likelihood ratio. This approach, where incentive compatibility is modelled using equation (2), is the so-called first-order approach (FOA). Mirrlees (1975, 1999) was the first to point out that FOA is not necessarily a valid procedure in a potentially large number of cases, because it might lead to a local instead of a global optimum. Mirrlees

(1976), Rogerson (1985), Jewitt (1988) and Alvi (1997) have explored conditions for the validity of the FOA. When the utility function is separable as in our set-up, sufficient conditions are

1. Monotone likelihood ratio condition, MLRC: $\partial g / \partial z > 0$. Income is increasing stochastically in effort, that is, higher output is more likely for higher effort than lower effort.
2. Convex distribution function condition, CDFC: $F_{yy}(z, y) > 0$. This is like a diminishing returns conditions, applied to the production of information of worker's action.

Key steps for demonstrating this include showing that the after-tax income is increasing in income ($x'(z) > 0$) and that the government's maximisation problem is concave. Appendix 1 provides details of the proof. In numerical simulations, one can also find solutions that are valid but do not satisfy CDFC (See e.g. Low and Maldoom 2004).

We can now turn to the properties of the solution. Differentiation of (5) again with respect to z , substitution and reorganisation yield the shape of $x'(z)$:

$$(6) \quad x' = -\frac{\alpha(v')^2 g'}{\lambda v''}$$

Denote the coefficient of absolute risk aversion as $\delta = -(v'')/(v')$. Based on (6), the marginal tax rate ($MTR = T'(z)$) is therefore given by

$$(7) \quad MTR = 1 - x' = 1 - \frac{\alpha v' g'}{\lambda \delta}$$

This equation shows that without the need to consider incentives, i.e. when the incentive compatibility constraint is slack, $\alpha = 0$, the optimal marginal tax rate is 100%. This is the case of full insurance, which risk-averse individuals value. However, when incentives to undertake effort matter, the optimal marginal tax rate is a compromise between risk aversion and providing incentives. If the consumers become more risk averse (δ increases), the marginal tax rate increases, *ceteris paribus*. On the other hand, if effort is more tightly connected with income (g' goes up), workers' effort can be more reliably tracked, and the optimal marginal tax rate is reduced.

3. Prospect theory and moral hazard

We now turn to the new case where individual behaviour is described by prospect theory, and this is also accepted as a basis for social welfare. In prospect theory, the utility function is replaced by a value function. The key assumptions about the value function are that it ‘is (i) defined on deviations from the reference point; (ii) generally concave for gains and convex for losses; (iii) steeper for losses than for gains’ (Kahneman and Tversky 1979). The two latter properties capture the idea that individuals are loss averse. Hence, the value function takes the S-shaped value as in Figure 1.

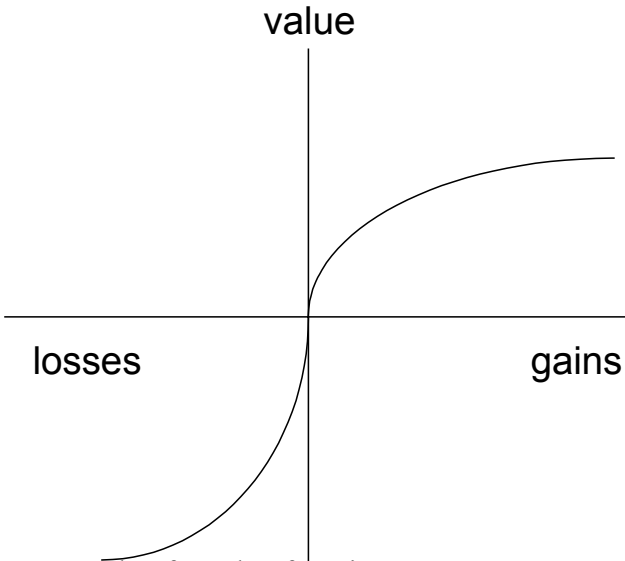


Figure 1: An example of a value function.

For simplicity, let us concentrate on the case where prospect theory is only related to the utility of income. Individuals now maximise the expectation of the value function $e(c)$

$$(1') \int e(c)f(z, y)dz - y,$$

where $c = x - \bar{x}$ denotes the change in realised income from a reference income, depicted by \bar{x} . The reference income is assumed to be exogenous.⁵ Note that the utility function is still assumed to be additively separable between effort and income.

If there was a reference point for effort level, too, within this quasi-linear model, along the lines of $\int e(c)f(z, y)dz - (y - \bar{y})$, the reference level would not change (1'). However, with a general, non-separable, formulation, $\int e(c, y - \bar{y})f(z, y)dz$, the validity of the FOA is a complicated matter even without reference levels, as shown by Alvi (1997). A possibility for loss aversion on either side of the reference level for y – the individuals could lose both from working less or more than the reference level – would add to the complexity, and therefore this issue is left for further work.

To capture the shape of the value function, we make the following assumptions about the properties of the utility function:

$$(8) (i) \quad e' > 0$$

⁵ Some of the implications of endogenising the reference point were explored in an earlier version of the paper (see Kanbur et al 2004).

$$(ii) \quad e'(-c) \geq e'(c)$$

$$(iii) \quad e'' > 0 \text{ for } c < 0, e'' < 0 \text{ for } c > 0$$

Assumption (ii) captures the principle of loss aversion: ‘losses loom larger than corresponding gains’ (Tversky and Kahneman 1992, p. 303). Assumption (iii) refers to ‘diminishing sensitivity for losses and gains’, i.e. a diminishing marginal utility for losses and diminishing marginal disutility for losses.

The specification in (ii) allows for a non-differentiability in $e(c)$ at $c=0$. In fact, much of the standard representation of prospect theory is in terms of a “kink” at zero. However, in this paper we assume that the function $e(c)$ is everywhere differentiable. This is because the purpose of this paper is to study to what extent one can utilise first-order conditions for analysing tax optima, and in particular the extent to which the standard first-order approach can be modified. This implies that we are by definition in a world of differentiable functions, thus excluding kinks. However, even without kinks the key features of Prospect Theory as stated in (i), (ii) and (iii) above can be captured with differentiable functions and, furthermore, one can with suitable functional forms and parameter values approach arbitrarily closely the description of individuals’ choices even if they were generated by a function with a kink. Nevertheless, we return to the case where the kink exists via an example in the end of this section.

We are now in a position to rewrite the government’s optimisation problem. The individual’s first-order condition can be rewritten as

$$(2') \int e(x) f_y dz - 1 = 0.^6$$

The Lagrangean and the first-order condition with respect to x (pointwise optimisation) are now as follows:

$$(4') L = \int \{ [e(c) + \lambda(z-x)] f(z, y) + \alpha e(c) f_y \} dz - \alpha - y$$

$$(5') 1 + \alpha g = \frac{\lambda}{e'}$$

We first examine whether the first-order approach is valid in this case with prospect theory preferences. For this, we need to see how consumption is related to effort. Differentiate (5') again with respect to z and reorganise to obtain the shape of $x'(z)$:

$$(6') x' = -\frac{\alpha(e')^2 g'}{\lambda e''}$$

The marginal tax rate is then written as

$$(7') MTR = 1 - x' = 1 - \frac{\alpha e' g'}{\lambda \delta_e},$$

where $\delta_e = -e''/e'$ is the coefficient of the individual attitude towards risk.

⁶ Similarly than in the standard model, the second-order condition of individual optimisation will be satisfied when the first-order approach as a whole is valid.

To provide incentives for exerting effort, x should be increasing in z . Depending on whether realised income is above or below the reference income, \bar{x} , three cases emerge.

1) For income above the reference income, $c > 0$, consumption x is indeed increasing in income z , since the value function has similar properties to the standard case of Section 2. In other words, $e = v$. The first-order approach is valid (following the arguments of equations (A.1)-(A.4), and the marginal tax rate is given by (7).

2) If $c = 0$, the right-hand side of (A.11) is not defined.

3) If the income is below the reference income, $c < 0$, the right-hand side of (6'), i.e. $x'(z)$, is non-negative only if $\alpha < 0$, since $e'' > 0$. Using a similar procedure as in Appendix 1, one can determine the sign of α from

$$(9) \quad \begin{aligned} \frac{\alpha}{\lambda} \int e f_y dz &= \text{cov}(e, \frac{1}{e'}) \\ \frac{\alpha}{\lambda} &= \text{cov}(e, \frac{1}{e'}) \end{aligned} ,$$

where the second line follows from (2'). Equation (9) is a counterpart of earlier equation (A.3). Now e and e' covary in the *same* directions for $c < 0$, we necessarily have $\alpha \leq 0$. However, the only case where the covariance is zero is when x is constant irrespective of income. But then the worker has no incentives to provide positive effort. Therefore, to induce effort, $\alpha < 0$.

However, if $\alpha < 0$, a relaxation in the incentive constraint reduces the social welfare determined by (4'). For a meaningful government optimisation problem this cannot hold. Therefore, one must conclude that the FOA is not valid for income below the reference level.

Proposition 1 summarises the discussion above.

***Proposition 1.** In a moral hazard tax problem, when the individuals' decision making is based on prospect theory, the FOA is valid if income is above the reference point. When income is below the reference point, the FOA is not valid.*

The optimal solution is therefore non-continuous. For consumption above the reference level, the marginal tax rate is given by (7'). For income below the reference point, a randomised schedule is optimal. To see this, recall that the convexity of the value function implies that the consumer is in fact risk loving, if income is below the reference point. It is therefore conceivable why conditions needed for an *insurance* scheme are then not valid. The point that randomisation may be desirable in moral hazard context is not new. Holmström (1979) and Arnott and Stiglitz (1988) have shown, however, that randomisation is never optimal for a standard, concave, moral hazard problem as in Section 2. Rather, it can become optimal in more complicated situations (Arnott and Stiglitz 1988).⁷ The optimality of randomisation depends on the magnitude of diminishing sensitivity for losses. It may be the case that for stakes on the scale involved in income taxation, diminishing sensitivity for losses may not hold. In this case, the tax rate could remain continuous.

⁷ Of course, the question remains how randomisation can be implemented in real world tax policy. One of the ideas presented in this context is lax control of tax evasion. Another example is related to poverty traps due to

The next step is to characterise the tax schedule in more detail. Usually loss aversion is introduced by a kink at the reference consumption. This is not necessary a major issue. We can always approximate arbitrary well any kinked value function.⁸ If $-e''$ is very large in a neighbourhood of the reference consumption, we can approximate a kinked function by a differentiable one. In an approximation e' would fall continuously in an interval $[\bar{x}, \bar{x} + \varepsilon]$ and e'' would approach $-\infty$. From (6') we see that when e'' is very large x' is close to zero. Let denote $x(z^*) = \bar{x}$ and $x(\hat{z}) = \bar{x} + \varepsilon$. Since $e'(x)$ is constant over the short interval it implies $e'(\bar{x})g(z^*) \approx e'(\bar{x} + \varepsilon)g(\hat{z})$. This shows that consumption $x(z)$ is constant at \bar{x} over the interval $[z^*, \hat{z}]$. The net income is insensitive to gross income over this interval. In other words, the marginal tax rate is 100% in that range. Beyond \hat{z} consumption is monotonically increasing in z . Combining this discussion and the point about randomisation, one can deduce the following.

Proposition 2. *In a moral hazard tax problem, when the individuals' decision making is based on prospect theory, the tax schedule is characterised by*

- *100% marginal tax rate around some interval at the reference point*
- *Optimal combination of incentives and insurance above the reference point (as in the standard model)*
- *A randomised schedule between a minimum consumption and the reference consumption below the reference point.*

the interaction of social benefits and taxation. Then for low incomes, a minor change in before-tax income can generate a large variation in after-tax income.

⁸ We are very grateful to a referee for this point.

The intuition for this result is that around the reference point, the individual is extremely risk averse, and therefore it is locally optimal to provide full insurance. In the area below the reference point the individual prefers a gamble, whereas above the reference point, the determinants of the tax schedule are similar to the standard model with expected utility.

An example

It seems that it is difficult to characterise the tax schedule in more detail within the general model. We therefore consider the following example, where the utility function and the distribution function are assumed to take certain functional forms. We now also explicitly allow for the kink in the utility function.

Suppose first that the utility function is

$$(10a) \quad e(x) = \frac{(x - \bar{x})^{1-\beta}}{1-\beta}, \quad \text{if } (x - \bar{x}) \geq 0$$

$$(10b) \quad e(x) = -h \frac{(\bar{x} - x)^{1-\beta}}{1-\beta}, \quad \text{if } (x - \bar{x}) < 0.$$

The form in (10a) is of CRRA form and similar to what has been used in simulations of the moral hazard model, such as those in Tuomala (1990). The function in (10b) is otherwise similar, but it includes h , which is a loss aversion parameter. The function also satisfies the conditions that $e'' < 0$ above the reference point and $e'' > 0$ below the reference point.

Consider first the area above the reference point. Substituting (10a) in the first-order condition (5') yields

$$(11) \quad x = \left(\frac{1}{\lambda} + \frac{\mu}{\lambda} g\right)^{(1/\beta)} + \bar{x}$$

Suppose now that the distribution function is a gamma distribution,

$f(z, y) = \{b/\Gamma(r)\}(bz)^{r-1} y^{-r} e^{-(bz/y)}$. Then it can be shown that the likelihood ratio has the following form $g = f_y/f = \frac{bz}{y^2} + \frac{1-r}{y}$. The likelihood ratio is also linear in z . Assuming

$A + Bz > 0$, consumption is determined by

$$(12) \quad x(z) = (A + Bz)^{(1/\beta)} + \bar{x},$$

where $A = \left(\frac{1}{\lambda} - \frac{\alpha s}{\lambda y}\right)$, $B = \frac{\alpha b}{\lambda y^2}$, and $s = r - 1$. In the example below, $r = 2.75$ and $b = 3$. If g is linear in z (as in gamma and exponential distributions), the tax schedule will have a declining marginal tax rate if $\beta < 1$ and increasing marginal tax rate if $\beta > 1$. These results are the same as in the standard model. In calculations we have $\beta = 1.2$.

We now turn to the implications of (10b). Combining it with the first-order condition (5') implies

$$(12') \quad x(z) = \bar{x} - h(A + Bz)^{(1/\beta)}$$

This equation implies that x is decreasing in $z \in (z_0, z^*)$, where z_0 is the smallest z and $x = \bar{x}$ at $z = z^*$). Combining this information with the message of Proposition 2, the government should offer a randomised schedule between the reference consumption level and the consumption level targeted for the smallest possible z . However, in this example the consumption level offered at lower end is just equal to the reference consumption. Therefore, using these functional forms, the optimal contract offers full insurance below the reference point, as illustrated in Figure 2. Needless to say, this is not a general result, but it does suggest that it can be optimal to lengthen the flat part of the consumption schedule downwards from reference consumption.

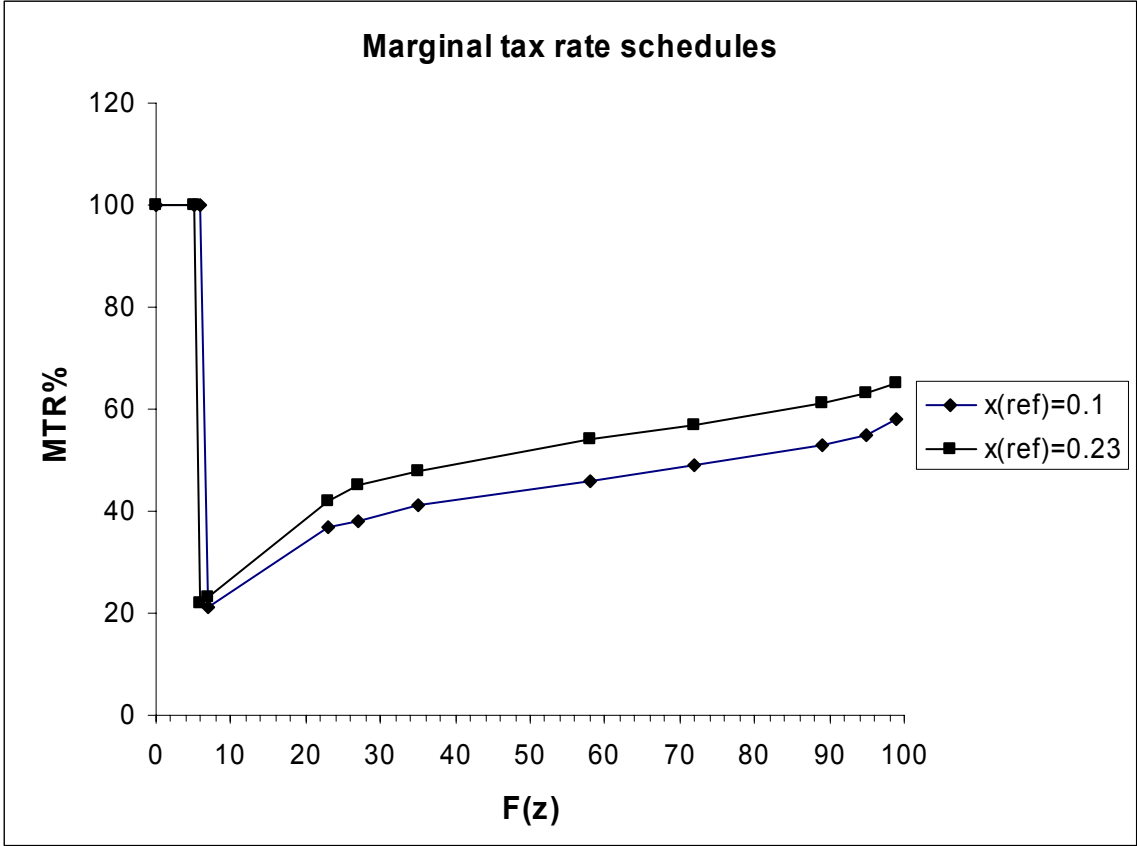


Figure 2

As we might expect the optimum marginal tax rate increase with reference consumption. In Figure 2 we also see that the shape of the optimal income tax schedule becomes more pro

gressive when \bar{x} increases. Our calculations showed that the flat segment covers about 4 percent of population with $\bar{x}=0.1$ (29 percent of mean consumption) and 3 percent with $\bar{x}=0.23$ (49 percent of mean income). The mean x and z also increase with \bar{x} .

As is well known, in the expected utility version, one can approximate the first best with an extreme punishment for the worst possible outcome and full insurance for all other outcomes. Usually the possibility for this ‘capital punishment’ is assumed away, but its existence is still an undesirable outcome of the basic moral hazard model, as Mirrlees (1997) has pointed out. It is interesting that our example provides an escape route from this problem by providing full insurance for low values of gross income. In this sense, prospect theory helps the motivation to work with the otherwise standard version of moral hazard.

4. Prospect theory non-welfarism

It is not necessarily clear that the social planner ought to accept all facets of prospect theory when forming its social objectives. There is, for example, some evidence according to which people underestimate their ability to cope with negative life-events, for instance income losses (see the discussion in Frey and Stutzer (2004). Loewenstein et al (2002) provide theoretical reasoning for this behaviour based on projection bias. The projection bias can therefore imply that the utility function governing individuals’ long-term welfare is different from that of their short-term welfare.⁹ But to the extent loss aversion is real, it should, of course, be respected when evaluating social welfare.

⁹ Some forms of prospect theory also include biases in evaluating probabilities. These biases provide further reasons for paternalism, akin to Sandmo (1983).

Perhaps a stronger case for paternalism could be built on the idea that the government is not willing to accept risk loving over the domain of losses. The society may want to restrict gambles on stakes involved in income taxation. We therefore also consider the case, where the government's and the individual's objective functions differ. The individual still maximises the same value function (e) as in the previous section, but the government's objective function is globally concave, as in expected utility model, and denoted by v . Because of the concavity, it does not also involve any approximated kink near the reference point.

The Lagrangean and the first-order conditions can now be written as

$$(4'') \quad L = \int \{ [v(x) + \lambda(z - x)]f(z, y) + \alpha e(x)f_y \} dz - \alpha - y$$

$$(5'') \quad v'f + \alpha e'f_y - \lambda f = 0$$

To determine how consumption is related to income, differentiate (5'') with respect to z to get

$$(6'') \quad x' = \frac{\alpha e' g'}{v' \delta_v + \delta_e (\lambda - v')}$$

where $\delta_v = -v''/v'$ is the social coefficient of absolute risk aversion and $\delta_e = -e''/e'$ is the coefficient of the individual attitude towards risk. For $c > 0$, the individual is risk averse, and δ_e is defined to be positive. However, for $c < 0$, because of diminishing marginal disutility

for losses the individual is in fact risk loving, and $\delta_e < 0$. The term $\lambda - v'$ is positive because of (5''). The marginal tax rate is

$$(7'') \quad MTR = 1 - x' = 1 - \frac{\alpha e' g'}{v' \delta_v + \delta_e (\lambda - v')}$$

Let us now examine the conditions for the validity of the first-order approach. Appendix 2 demonstrates that the same conditions, namely monotone likelihood ratio condition (MLRC) and the convexity of the distribution function condition (CDFC) still constitute part of the sufficient conditions. In addition, consumption x should be increasing in income z . For $c > 0$, this is always the case because then $\delta_e \geq 0$, and the denominator in (6') and the term at the right of (6'') are positive.

If $c < 0$, the denominator in (6'') may be either positive or negative, depending on the strength of diminishing marginal disutility for losses. Then the FOA remains valid if the individuals' risk loving is sufficiently smaller than the social coefficient of risk aversion, i.e. $v' \delta_v > -\delta_e (\lambda - v')$. However, if the individual is sufficiently risk loving, $v' \delta_v < -\delta_e (\lambda - v')$, the denominator and the right-hand side of (6'') are negative, i.e. consumption would be decreasing in income (effort). This violates the validity of the first-order approach. Therefore, the following proposition holds.

Proposition 3. *In a non-welfarist moral hazard tax problem, when the individuals' decision making is based on prospect theory, the FOA is valid if the government's risk aversion sufficiently outweighs individuals' diminishing sensitivity for losses or income is above the reference point. When income is below the reference point and individuals'*

diminishing sensitivity for losses sufficiently outweighs the government's risk aversion, the FOA is not valid.

The introduction of non-welfarism, therefore, brings the interesting point that if the government's willingness to override loss seeking behaviour is sufficiently strong, one avoids randomisation over the domain of losses. Because of the potential complexities of carrying out randomisation in the real world, this may be a desirable outcome from paternalism.

The optimal tax rate still features full insurance around the reference point. To see, this notice that around the reference consumption level, the individual is very risk averse, and therefore $\delta_e = -e''/e'$ is very large. This implies the marginal tax rate approaches 100%. Therefore it seems that the point about a small section of full insurance remains valid even if the government's objective function does not directly include loss aversion. Even though the government is non-welfarist, it must take individuals' preferences into account through the incentive compatibility constraint.

5. Conclusion

This study analysed a model of optimal non-linear income taxation under income uncertainty, along the lines of Mirrlees (1974). While the standard model is based on expected utility, in our work the preferences are modelled along the lines of prospect theory, as developed by Kahneman and Tversky (1979).

As does most of the literature in the area, we focused on the so-called first-order approach for solving the optimisation problem. The introduction of prospect theory implies that the first-

order approach is valid only in the area above the reference point. In the area clearly below the reference point, the individuals are risk lovers because of diminishing sensitivity for losses, and the optimal policy is to offer a randomised schedule between a minimum payment and the consumption at the reference level. At a small interval around the reference, the individuals are extremely risk averse owing to loss aversion, and the marginal tax rate is 100 per cent. Finally, in the area clearly above the reference point, the optimal policy resembles a standard compromise between incentives and insurance.

We also dealt with the possibility that the government does not accept all aspects of prospect theory as a basis for social welfare, in other words, we introduced non-welfarist preferences. For instance, the society may not necessarily approve of risk loving behaviour in the domain of losses. In this case, it was shown that the first-order approach remains valid if the risk aversion by the social planner outweighs the diminishing sensitivity for losses of the individual. This is potentially desirable feature if one wants to avoid randomisation. Finally, we considered an example that created full insurance below the reference point. This is interesting, because in this case one avoids an undesirable feature of the standard moral hazard model with expected utility, namely the possibility that the government can achieve the first best by imposing an extreme punishment for the lowest possible realisation of income.

These findings have potentially important implications. For income taxation models, the optimality of full insurance around or below the reference point suggests that high marginal tax rates at low incomes may not necessarily be as harmful as it is sometimes claimed. In fact, such as system supports organising a minimum income level through e.g. social assistance. For research in information economics in general, the results of this paper show that it may

become worthwhile to assess the robustness of other results, derived using expected utility maximisation, to assumptions that are more in line with findings in behavioural economics.

Appendix 1: Sufficient conditions for FOA in the benchmark model

First, as an intermediate step, it must be checked that consumption is increasing in income / effort, i.e. $x'(z)$ is positive. The right-hand side of the first-order condition in (5) is increasing in $z(y)$, since $v'' < 0$. The left-hand side of (5) is increasing with $z(y)$, provided that $\alpha > 0$, since $g' > 0$ when we assume that the MLRC (monotone likelihood ratio condition) holds.

Following Jewitt (1988) and Laffont and Martimort (2002) we now show that α is indeed positive. Dividing (5) by λ , multiplying it with f and integrating over the support $[\underline{z}, \bar{z}]$ yields

$$(A.1) \quad \frac{1}{\lambda} = \int \frac{1}{v'} f dz$$

since $\int f_y dz = 0$. Using (5) again gives

$$(A.2) \quad \frac{\alpha}{\lambda} g = \frac{1}{v'} - \int \frac{1}{v'} f dz$$

Multiply both sides by $v'f$ and integrate over the support $[\underline{z}, \bar{z}]$ to get

$$(A.3) \quad \frac{\alpha}{\lambda} \int v f_y dz = \text{cov}(v, \frac{1}{v'})$$

$$\frac{\alpha}{\lambda} = \text{cov}(v, \frac{1}{v'})$$

where $\text{cov}()$ denotes the covariance operator. The second line follows from the incentive constraint (2). Since v and v' covary in opposite directions, we necessarily have $\alpha \geq 0$. However, the only case where the covariance is zero is when x is constant irrespective of income. But then the worker has no incentives to provide positive effort. Therefore, to induce effort, $\alpha > 0$.

Finally, it remains to be shown that expected utility is concave. For this, rewrite the expected utility as follows:

$$\int v(x) f(z, y) dz - y$$

$$(A.4) = [vF(z, y)]_{\underline{z}}^{\bar{z}} - \int v' x' F(z, y) dz - y$$

$$= v(x(\bar{z})) - \int v' x' F(z, y) dz - y$$

where the second line follows from integration by parts and the third from using the property that $F(\underline{z}, y) = 0$ and $F(\bar{z}, y) = 1$. From the last row, since v' and x' are positive, the expected utility is concave in y if $F_{yy} > 0$. This property is called the convexity of the distribution function condition (CDFC). Therefore, the FOA is a valid strategy given that MLRP and CDFC hold.

Appendix 2: Sufficient conditions for FOA under non-welfarism

In the main text we show that for $x' > 0$, $v'' + \alpha e''g < 0$. The next step is to show that $\alpha > 0$.

Rearrange (5'') to get

$$(A.5) \quad \frac{1}{\lambda} + \frac{\alpha e'}{\lambda v'} g = \frac{1}{v'}$$

Integrating over the support $[\underline{z}, \bar{z}]$ yields

$$(A.6) \quad \frac{1}{\lambda} = \int \frac{1}{v'} f dz$$

and the first-order condition can be written as:

$$(A.7) \quad \frac{\alpha e'}{\lambda v'} g = \frac{1}{v'} - \int \frac{1}{v'} f dz$$

By multiplying with v and integrating over the support $[\underline{z}, \bar{z}]$ one obtains

$$(A.8) \quad \frac{\alpha}{\lambda} \int \frac{e'v}{v'} g = \text{cov}(v, \frac{1}{v'})$$

The right-hand side is always non-negative. So is the term multiplying α on the left. Therefore we necessarily have $\alpha \geq 0$. Again, to induce a positive effort level, α cannot be equal to

zero. Hence, α is positive. It remains to show that the objective function is concave. This has actually been already shown above in equation (A.4).

References

- Alvi, E. (1997) 'First-Order Approach to the Principal-Agent Problems: A Generalization', *The Geneva Papers on Risk and Insurance Theory* 22, 59-65.
- Arnott, R. and J.E. Stiglitz (1988) 'Randomization with asymmetric information', *RAND Journal of Economics* 19, 344-362.
- Bernheim, E.J. and A. Rangel (2005) 'Behavioral public economics: Welfare and policy analysis with non-standard decision makers', NBER Working Paper 11518.
- Camerer, C.F. and G. Loewenstein (2003) 'Behavioral economics: Past, Present, Future' in C.F. Camerer, G. Loewenstein and M. Rabin (eds.) *Advances in Behavioral Economics*. Princeton: University Press.
- De Meza, D. and D.C. Webb (2007) 'Incentive design under loss aversion', *Journal of the European Economic Association* 5, 66-92.
- Dhami, S. and A. al-Nowaihi (2006) 'Why do people pay taxes? An explanation based on loss aversion and overweighting of low probabilities', mimeo, University of Leicester.
- Frey, B. and A. Stutzer (2004) 'Economic consequences of miscalculating utility', Institute for Empirical Research in Economics Working Paper No. 218, University of Zurich.
- Holmström, B. (1979) 'Moral hazard and observability', *Bell Journal of Economics* 10, 74-91.
- Jewitt, I. (1988), Justifying the first order approach to principal-agent problems, *Econometrica* 56, 1177-90.

- Kahneman, D. (2003) 'Maps of bounded rationality: Psychology for behavioral economics', *American Economic Review* 93, 1449-1475.
- Kahneman, D. and A. Tversky (1979) 'Prospect theory: An analysis of decision under risk' *Econometrica* 47, 263-281.
- Kanbur, R., Keen, M. and Tuomala, M. (1994) 'Labor supply and targeting in poverty alleviation programs', *The World Bank Economic Review* 8, 191-211.
- Kanbur, R., J. Pirttilä and M. Tuomala (2004) 'Moral hazard, income taxation and prospect theory', Tampere Economic Working Papers Net Series No. 30.
- Laffont, J.-J. and D. Martimort (2002) *The theory of incentives. The principal-agent model*. Princeton University Press.
- Loewenstein, G., T. O'Donoghue and M. Rabin (2002) 'Projection Bias in predicting future utility', Carnegie Mellon University, mimeo.
- Low, H. and Maldoom, D. (2004) 'Optimal taxation, prudence and risk-sharing', *Journal of Public Economics* 88, 443-464.
- Mirrlees, J.A. (1971) 'An exploration in the theory of optimal income taxation', *Review of Economic Studies* 38, 175-208.
- Mirrlees, J.A. (1974), 'Notes on welfare economics, information and uncertainty', in Balch, McFadden and Wu (Eds.), *Essays on Economic Behaviour under Uncertainty*, Amsterdam: North Holland.
- Mirrlees, J.A (1975,1999) 'The Theory of Moral Hazard and Unobservable Behaviour. Part I', *Review of Economic Studies* 66, 3-22.
- Mirrlees, J.A. (1976) 'The optimal structure of authority and incentives within an organization', *Bell Journal of Economics* 7, 105-31.

- Mirrlees, J.A. (1997) 'Information and incentives: The economics of carrot and sticks', *The Economic Journal* 107, 1311-1329.
- McCaffery, E.J. and J. Slemrod (Eds.) (2006) *Behavioral Public Finance*, New York: Russell Sage Foundation.
- Munro, A. (2004) 'Public Policy and Bounded Rationality', University of East Anglia, mimeo.
- O'Donoghue, T. and M. Rabin (2003) 'Studying optimal paternalism, illustrated by a model of sin taxes', *American Economic Review* 93, 186-191.
- Pirttilä, J. and M. Tuomala (2004) 'Poverty alleviation and tax policy', *European Economic Review* 48, 1075-1090.
- Rogerson, W. (1985) 'The First-Order Approach to Principal-Agent Problems', *Econometrica* 53, 1357-67.
- Sandmo, A. (1983) 'Ex post welfare economics and the theory of merit goods', *Economica* 50, 19-33.
- Seade, J. (1980) 'Optimal non-linear policies for non-utilitarian motives', in D. Collard, R. Lecomber and M. Slater (eds.) *Income distribution: the limits to redistribution*, Bristol: Scientechica.
- Starmer, C. (2000) 'Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk', *Journal of Economic Literature* 38, 332-382.
- Tuomala, M. (1990) *Optimal income tax and redistribution*, Oxford: Clarendon Press.
- Tuomala, M. (1984) 'Optimal degree of progressivity under income uncertainty', *Scandinavian Journal of Economics*, 87, 184-93.

Tversky, A. and D. Kahneman (1992) 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty* 5, 297-323.

Varian, H. (1980), 'Redistributive taxation as social insurance', *Journal of Public Economics* 14, 49-68.