Rent taxation when cost monitoring is imperfect

Diderik Lund*
Department of Economics, University of Oslo

July 20, 2001

Abstract

While rent taxation in some theories is neutral, and the tax rate could be set to one hundred percent to minimize the need for distortionary taxes, this does not occur in practice. An important reason for this is the transfer incentives that would result. When cost transfers can only be imperfectly monitored, it is optimal under some conditions to combine a tax on gross revenue with a rent tax. This contributes to explaining the frequent occurrence of gross revenue taxation in actual rent tax systems, and the frequently observed tailoring of rent tax systems in response to output price changes.

Keywords: Rent taxation, transfer pricing, income shifting

JEL classification numbers: F23, H25, H26, H73
1 Introduction

Most of the literature on taxation regards a rent tax as neutral. A natural implication is that the tax rate should be increased as long as this reduces the need for other, distortionary taxes. But 100 per cent rent taxes do not occur in practice. Since significant amounts of revenue may be raised this way in many countries, it is important to consider why not.

The popularity of 100 per cent rent taxes in the theoretical literature, e.g., Stiglitz and Dasgupta (1971) and Guesnerie (1995), has to do with analytical convenience: Theoretical results are clearer when rents are always taxed away. But then one should also consider why the assumption is not satisfied in the real world.

Stiglitz and Dasgupta (1971, p. 165) write, “No government has imposed on a regular basis 100 per cent taxes on profits and the income of fixed factors, in spite of the long standing advice of economists (e.g. Henry George) of the desirability of such non-distortionary taxes.” They mention two reasons for this. One is the problem of separating out pure profits, the other the political obstacles to what may be viewed as nationalization of fixed factors. The present paper may be regarded as an elaboration of the first of these.

This paper is mainly relevant to natural resource taxation. Several countries (e.g., Australia, Denmark, the Netherlands, Norway, the U.K.) imposed taxes on oil and gas extraction in the 1970’s and 1980’s with marginal rates well above 60 percent. These systems and a proposed alternative, Garnaut and Clunies Ross (1975), had imperfect loss offset, and there has been some concern about the resulting disincentives for investment, cf. e.g. Ball and Bowers (1982). Our focus here will be different, so we assume that the rent tax is a tax on non-financial cash flows with full loss offset, cf. Brown (1948), which is neutral in standard theory.

Less standard models are needed to explain why the rate of such a tax should be set significantly below one hundred percent, even for a cash flow tax. In Haaparanta and Piekkola (1997) the reason is an entrepreneurial effort which is necessary to undertake investments. Being unobservable (or non-verifiable), this effort is not tax deductible. Although the cash flow tax formally has full loss offset, the lack of this deduction invalidates the standard
neutrality result. Observe that such a mechanism might be present in a closed, one-sector economy.

The explanation given in the present paper is complimentary to the one in Haaparanta and Piekkola. We observe that rent taxes are applied in open economies, or in one sector of a larger economy. This creates incentives to transfer revenues out of, and costs into, a high-tax sector. This is often labelled transfer pricing or, more generally, income shifting.

The monitoring by tax authorities to prevent transfers and to enforce arms-length prices is likely to be imperfect, especially on the cost side. We assume that corporations may avoid the monitoring, but only at a cost. The government may then want to introduce a tax on gross revenue, which we call a royalty, in addition to a tax on rent, even though the royalty is non-neutral in standard models. Our model shows under what conditions a royalty is desirable for the government.

Osmundsen (1995) considers the same problem, and assumes that the companies have private information about costs. He then develops a model of optimal regulation under asymmetric information based on the assumption that reported costs should play no role in determining the tax base.

At this point the present paper instead follows Gordon and MacKie-Mason (1995). Their main focus is on income shifting between corporate and personal tax bases. They introduce a convex cost function for the shifting of income. Hauffler and Schjelderup (1998) do likewise, and use this to explain the difference between marginal corporate tax rates on the revenue side and on the cost side. They argue that this explains the popularity of rate-cut-cum-base-broadening tax reforms.

In contrast to those two, the present paper considers one open economy only, with the outside world considered exogenous. This is realistic for small nations, and it also simplifies and thus allows a more detailed investigation of the optimal rent tax system. Another difference between the present paper and the other two is that they maximize the utility of a representative citizen. The present paper is motivated by situations in which
the corporation is owned by foreigners, so that no welfare weight is given to the profit. The possibility of attaching a constant weight to after-tax profits is also considered.

Sansing (1996) also considers the transfer pricing incentives created by rent taxes. Like the present paper, he assumes that transfer pricing is more easily done on the cost side than on the revenue side. His analysis is based on the comparable profits method to avoid cost transfers. This has its own problems, as comparable projects do not always exist.

The present paper does not address the desirability of rent taxes versus fixed fees. When the authorities own a natural resource at the outset, one option is to collect the resource rent through a fixed fee, perhaps determined through an auction. Another option is not to demand payment for licences, but to use resource rent taxes.

One argument for the interest in rent taxes is that the second option is also available when the ownership of the resource is private, or licences have already been granted. Another argument is that rent taxes are actually applied in many countries, in particular in the British Commonwealth, according to Garnaut and Clunies Ross (1983). Natural resource rent taxation was recommended by the Commonwealth to its member countries.

2 The model

The model is a partial-equilibrium model of a small open economy. For simplicity there is only one firm, which can be taken to represent one whole natural resource sector. We use a simple, static production model with a quadratic cost function. One might argue that most rent tax systems apply to natural resource extraction, and that the effects in a dynamic exhaustible resource model would be more interesting. We believe, however, that our simpler model captures important deviations from standard results on neutral taxation.

Consider a corporation which earns before-tax rents

\[ R(q) = pq - bq^2, \]  

(1)
where \( q \geq 0 \) is quantity produced, \( p > 0 \) is output price, and \( b > 0 \) is a parameter of a quadratic cost function. A factor price index is the numeraire, so \( p \) is relative to the price of inputs. We assume that \( p \) is exogenously given, both for the corporation and for the authorities.

A tax with a high rate will be levied on this rent. The corporation also has activities in another sector. This gives an incentive to income shifting. We assume that the revenue side is more easily monitored by tax authorities than the cost side. One reason is the multitude of inputs which normally go into the production of one (or a few) output(s). Another is the fact that inputs are often tailor-made, which complicates the enforcement of arms-length pricing. For simplicity we assume that no transfers are made on the revenue side. Quantity may be measured accurately, and arms-length prices may be observed in a world market.

Cost transfers may be of two kinds: Real transfers or transfer pricing. Transfer pricing consists in charging a higher price for inputs when the seller is a related company in another sector. “Related” may mean a company with the same owners, but more complicated common interests may also prevail. Although we shall not model it explicitly, we mention that the outcome of bargaining over prices is likely to be affected by tax differences.

Real transfers consist in using more input factors than would have been the case without tax differences. This may be motivated by additional non-taxed benefits, such as benefits from testing of equipment or technology or training of personnel. These benefits may be reaped in a different jurisdiction. A multinational corporation will have incentives to do testing and training in the sector in which they are deductible against the highest tax rate.

For simplicity we assume that in the corporate accounts a cost \( a \) is deducted from the rents without having any effect on output. Taxes are

\[
T = t(pq(1 - r) - bq^2 - a) + rpq = t(pq - bq^2 - a) + r(1 - t)pq, \tag{2}
\]

where \( t \) is the rate of a cash flow tax, \( r \) is the royalty rate. As long as \( t < 1 \), we can, without loss of generality, assume that the royalty is deductible in the cash flow tax, as is common in practice. The case \( t = 1 \) is discussed in section 3. As we solve the maximization
problems below, some restrictions on the tax rates will follow, and we shall prove that tax revenues are strictly positive in the optimum.

The after-tax corporate profits are

$$\Pi(q, a) = (pq - bq^2)(1 - t) - r(1 - t)pq - ca^2 + a(t - s), \quad (3)$$

where $s > 0$ is the relevant tax rate in another sector, so that the last term represents the taxes saved by the transfer (when $t > s$).

We assume that the transfer of costs is costly. These additional costs may be incurred to avoid monitoring by the authorities. In case of real transfers, there may be real transport costs, or lower efficiency in testing and training than in other sectors. As in Gordon and MacKie-Mason (1995) we assume a convex cost function. For simplicity we have specified it as quadratic, $ca^2$, where $c > 0$ is a parameter. This will depend on technology and natural conditions as well as the monitoring expertise of tax authorities. The quadratic costs could be thought of as a simplified representation of an expected penalty where the penalty in case of detection is proportional to $a$, while the probability of detection is also proportional to $a$. There are alternative assumptions in the transfer pricing literature, such as the proportional profits shifting found in Weichenrieder (1996).

For simplicity we assume, as do Hafler and Schjelderup (1998), that the costs $ca^2$ are independent of $q$, and that they are not tax deductible. For a multinational corporation they might be tax deductible in another jurisdiction, which would make no difference. The independence of $q$ is not so trivial. One could believe that it is easier to camouflage a given cost as project-related in a large project. On the other hand, a large project is likely to attract more attention from tax authorities, so a dependence on $q$ could go either way. In order not to make the model unnecessarily complicated, the dependence is left out.

The formulation in (2) and (3) covers the transfer from another sector in the same country, or from a different country, as long as the tax rate there is lower. In what follows, however, the attention is restricted to the case of transfer from a different country, so that the government does not experience increased taxes in another sector when costs are transferred.
The maximization of $\Pi$ with respect to $q$ and $a$ has the following first-order conditions:

$$\frac{\partial \Pi}{\partial q} = (1 - t)(p - 2bq) - r(1 - t)p = 0,$$

(4)

and

$$\frac{\partial \Pi}{\partial a} = t - s - 2ac = 0.$$

(5)

The second order conditions are satisfied if $t < 1$. If $t > 1$, it would be optimal for the corporation to let $q \to \infty$. We assume that the authorities want to avoid this (and the resulting negative tax revenue). Thus we assume that $t \leq 1$.

Optimal quantities are

$$q = \frac{p(1 - r)}{2b},$$

(6)

and

$$a = \frac{t - s}{2c}.$$

(7)

It is easy to show that when $t < 1$, the maximized profit will be positive.

If $t = 1$, the solution in (6) is not unique. For this $t$ value the corporation is indifferent to the choice of $q$, except for the need to create a taxable income in which to deduct $a$. It turns out below that $t = 1$ is not an uninteresting case. Thus the analysis will be carried out like this: First we assume that (6) is relevant also for $t = 1$. Later the case $t = 1$ is discussed separately.

**Assumption 1** 
If $t = 1$, the corporation chooses $q$ according to equation (6).

Observe that $q$ is not affected by the rent tax rate $t$, even when there is a deductible royalty, which means that the rent tax is still neutral in this sense. However, a deductible royalty at rate $r$ does remove $q$ from its pre-tax optimum. Observe also the obvious effect that $a$ is increasing in the difference in tax rates, $t - s$.

While $s$ will be considered exogenous, it is certainly possible for the government to choose $t \leq s$. In that case the corporation would choose $a < 0$, i.e., to transfer costs away from the sector. While this is not unrealistic, one may doubt the realism of our model in one respect: The transfer cost function $ca^2$ is symmetric around zero, even though $a < 0$
means that the transfer would go in the opposite direction and the monitoring would be
done by different authorities. We shall not make a more specific formal model of this, but
shall return to the problem below.

Since the economy is small and open, taxation does not affect consumer surplus. The
welfare function depends on tax revenue and after-tax profits. Because of (partial or
full) foreign ownership of the corporation, and/or because of other distortionary taxation
(outside this model), the welfare function attaches a weight \( \lambda \in [0, 1] \) to the after-tax
profits. In this paper \( \lambda \) is considered exogenous. An extension of the model with an
edogenous \( \lambda \) would be interesting for economies in which rent tax systems supply a large
part of tax revenues.

The authorities choose tax rates \( t \) and \( r \) in order to maximize \( W = T + \lambda \Pi \) conditional
on the optimizing behavior of the corporation.

Plugging (7) and (6) into (3) and (2) gives

\[
W(t, r; \alpha, \lambda, s) = \frac{p^2(1 - r)}{4b}(t - rt + 2r) - \frac{t(t - s)}{2c} + \lambda \left[ \frac{p^2(1 - r)^2(1 - t)}{4b} + \frac{(t - s)^2}{4c} \right].
\]

(8)

Define now \( \alpha \equiv p^2c/4b \) and \( \nu \equiv (2 - \lambda)/2(1 - \lambda) \) in order to simplify the expressions. (As
\( \lambda \) goes from 0 to 1, \( \nu \) goes from 1 to infinity.) The first-order conditions are

\[
\frac{\partial W}{\partial t} = \frac{1 - \lambda}{c} \left[ \alpha(1 - r)^2 + \frac{s}{2} - \nu t \right] = 0,
\]

(9)

and

\[
\frac{\partial W}{\partial r} = \frac{2\alpha(1 - \lambda)(2\nu - t)}{c} \left[ \frac{1 - t}{2\nu - t} - r \right] = 0.
\]

(10)

Before going into details, we state the main conclusions that follow: When there is an
interior solution, its dependence on \( p, c, \) and \( b \) is intuitively reasonable: A higher output
price and a lower productive cost mean that there is more to lose from distorting production
decisions, so there will be a stronger emphasis on the tax which does not distort it. Higher
transfer costs have the same effect, since there is less to lose from allowing cost deductions.

Even the simple model presented so far has a corner solution, \( (t, r) = (1, 0) \), for some
plausible, non-trivial parameter configurations. There is also a discontinuity that may be
somewhat surprising. Thus we need to explore the solution in some detail.
Observe that when $t < s/2\nu$, then $\partial W/\partial t > 0$ always. Thus the maximizing choice of $t$ is never less than $s/2\nu$, and we can restrict the attention to $t \in [s/2\nu, 1]$.

In order to find the maximum, first solve (10) for $r$ and define the solution as $g(t; \nu)$,

$$r = \frac{(1 - t)(1 - \lambda)}{1 + (1 - t)(1 - \lambda)} = \frac{1 - t}{2\nu - t} \equiv g(t; \nu),$$

which for $t \in [0, 1]$ is a decreasing function of $t$ with values $g \in [0, 1/2]$. Equation (11) is essentially the result derived by Lund (1985) under the restriction $t = s$.

We are now able to restrict what is the interesting interval for $r$, given that $t \in [s/2\nu, 1]$. First, $r < 0$ cannot be optimal, since $r < 0$ would imply $\partial W/\partial r > 0$. Next, we have $\partial W/\partial r < 0$ if $r > g(t; \nu)$, and thus in particular for all $r > 1/2$. Thus the maximizing choice of $r$ never exceeds $1/2$. We can restrict attention to $(t, r) \in [s/2\nu, 1] \times [0, 1/2]$.

Solve (9) for $r$, use the solution which has $r < 1$, and define this as $f(t; \alpha, \nu, s)$,

$$r = 1 - \sqrt{\frac{1}{\alpha} \left( \nu t - \frac{s}{2} \right)} \equiv f(t; \alpha, \nu, s).$$

(for $t \geq s/2\nu$). From (11) and (12) we find stationary points for $W$ when

$$f(t; \alpha, \nu, s) = g(t; \nu).$$

We shall not write down the explicit analytical solutions, since they are messy solutions to third-order polynomial equations. The structure is sufficiently simple, however, to give some straightforward analytical results.

Those stationary points which occur within the set $[s/2\nu, 1] \times [0, 1/2]$, together with boundary points, comprise the set of candidates for a maximum point. Without loss of generality we expand the set for our discussion to $(t, r) \in [0, 1]^2$, since the expressions for the partial derivatives on the boundary of this set are easier to work with. For given values of $\alpha, \nu$, and $s$, the two functions $f$ and $g$ are continuous and strictly decreasing in this set, and $f$ is convex while $g$ is concave.

In the following we shall focus on the dependence of the maximum point on $\alpha$, while the discussion is valid for different values of $\nu$ and $s$. The focus on variation in $\alpha$ is due to the high variability of $p$, in particular for many natural resource industries.
Figure 1 shows an example of graphs of $f$ and $g$ in a $(t, r)$ diagram for $\alpha = 0.89, \nu = 1,$ and $s = 0.3$. The value $\nu = 1$ corresponds to $\lambda = 0$, which is relevant when the corporation is wholly owned by foreigners. The value $s = 0.3$ is chosen as an approximation to an average corporate income tax rate across countries of the world.

In the diagram there are two intersections between the two curves, i.e., two stationary points. The second-order conditions for a maximum depend on the parameters $\alpha$ and $\nu$, and are tedious to work out. Instead we can determine the character of the stationary points by considering the arrows, which indicate the directions of increasing $W$ within the five regions delimited by the graphs and the boundaries. It is clear that the north-western stationary point is a local maximum and gives the optimal tax rate mix. Both rates are strictly positive.

Next we ask what happens to the optimal mix if the output price changes (or if the parameter $b$ or $c$ changes). This means that $\alpha \equiv p^2 c / 4b$ changes. For a given $\nu$ (i.e., a given $\lambda$), if we change $\alpha$, the graph of the $g$ function is unaltered, while the graph of $f$ is changed. Its intersection with the horizontal $r = 1$ line is fixed, however, at $t = s / 2\nu$, which is 0.15 in the diagram. A higher $\alpha$ lifts the rest of the curve and makes it flatter (i.e., a reduced $|\partial f / \partial t|$). There is a slightly higher $\alpha$ for which there is just one intersection between the curves (a tangency), and for even higher $\alpha$ values, there is none. The graph of $f$ “leaves” the set $(t, r) \in [0, 1]^2$ through the vertical $t = 1$ line when $\alpha > \nu - s/2$. An example is given in figure 2 for $\alpha = 1.3, \nu = 1.0$.

The graph of $f$ “leaves” through the horizontal axis when $\alpha < \nu - s/2$. Then there is only one intersection with the graph of $g$ within the set. An example is given in figure 3 for $\alpha = 0.5, \nu = 1.0$. 

...
For high values of $\alpha$ (as in figure 2) the maximum $W$ within the set $[0, 1]^2$ is not found at a stationary point, but as a corner solution, $(t, r) = (1, 0)$. This occurs when there is no intersection between the curves, but may also occur for slightly lower $\alpha$ values.

In the appendix we prove the following:

**Lemma**  
Keep fixed values of $\lambda \in [0, 1)$ and $s \in [0, 1)$, and adopt Assumption 1. Within the set $(t, r) \in [0, 1]^2$ the function $W(t, r; \alpha, \lambda, s)$, as defined in (8), has an interior maximum for low values of $\alpha$. This occurs in the interval of low $\alpha$ values for which there is only one stationary point in $[0, 1]^2$, but also in an adjacent interval of higher $\alpha$ values for which there are two stationary points. For even higher values of $\alpha$ the maximum occurs at the corner $(t, r) = (1, 0)$. When there is an interior global maximum, this point has a $t$ value which is strictly increasing in $\alpha$ and an $r$ value which is strictly decreasing in $\alpha$.

From the Lemma follows the following:

**Proposition 1**  
Assume some fixed values $\lambda \in [0, 1)$ and $s \in [0, 1)$. The optimal rent tax rate is strictly increasing, and the optimal royalty rate is strictly decreasing, as we consider these partial changes: (i) The output price $p$ is increasing, from zero to some upper limit, (ii) the parameter $c$ of the transfer cost function is increasing, from zero to some upper limit, and (iii) the parameter $b$ of the productive cost function is decreasing, from infinity to some lower limit. Above these two upper limits for $p$ and $c$, and below the lower limit for $b$, the optimal rent tax rate is one hundred percent, while the optimal royalty rate is zero, given Assumption 1.

The intuition for the interior solution is that there is a trade-off between maximization of the rent tax base, and thus tax revenue, on one hand, and the need to reduce incentives for transfer pricing on the other. The solution is a compromise.

Perhaps the corner solution for high $\alpha$ values is more surprising. In a second-best situation, one could believe that all instruments would be used to some extent. A closer look at the first-order conditions may give some insight. Equation (10) has no relation to
the transfer pricing problem. It gives the optimal royalty for any given $t \in [0, 1)$ even if $c \to \infty$ so that the transfer pricing problem disappears.

Equation (9), however, can be rewritten

$$\frac{\partial W}{\partial t} = \frac{(1 - \lambda)p^2(1 - r)^2}{4b} + \frac{s(1 - \lambda) - t(2 - \lambda)}{2c} = 0. \quad (14)$$

If there is no transfer pricing problem, $c \to \infty$, the second term vanishes, and $\partial W/\partial t$ is always positive. In this case the corner solution $(t, r) = (0, 1)$ is clearly optimal. But when transfer pricing becomes a problem, the second term appears. It is negative when $s < t$, but $p^2/b$ can clearly be so large that $\partial W/\partial t$ is everywhere positive in the region below and just above the graph of $g$. This is the situation illustrated in figure 3.

Since the two effects in $\partial W/\partial t$ are additive, there can be parameter configurations for which their sum is positive for all relevant $t, r$, but other configurations for which the first-order condition is satisfied. This explains the occurrence of the corner solution.

Next we show that tax revenues are positive at an interior solution. Together with the fact that after-tax profits are positive, this is important because it implies that the qualitative properties of the interior solutions would not change if a small fixed cost were introduced into the model.

**Proposition 2** When the optimal tax rates are interior, $t < 1, r > 0$, both taxes give strictly positive revenue.

The proof is in the appendix.

One might suspect that the interval of $\alpha$ values for which an interior solution is optimal, is less realistic in many cases, because it might imply $t < s$ (cf. the discussion above). This is not the case. It is certainly true that the model, taken literally, implies an optimal $t < s$ as $\alpha$ becomes low enough. Specifically, the optimal $t$ approaches $s/2\nu < s$ as $\alpha$ goes to zero. However, for any configuration of $\nu$ (i.e., $\lambda$) and $s$ there will be an interval of $\alpha$ values which gives interior solutions with $t > s$. This is shown in the following Proposition, which also points out the condition for an important discontinuity.
Proposition 3  

For any values of $\lambda \in [0, 1)$ and $s \in [0, 1)$, there exist positive values of $\alpha$ for which $W$ has an interior global maximum in the set $(t,r) \in [0,1]^2$ for which $t > s$. The optimum $(t,r)$ is discontinuous in $\alpha$ when $\lambda < (\sqrt{9-8s} - 1)/(\sqrt{9-8s} + 1)$. In that case the optimum makes a discrete jump from an interior maximum to the corner $(t,r) = (1,0)$ as $\alpha$ increases, given Assumption 1.

The proof is in the appendix. As an illustration, when $s = 0.3$, the condition for a discontinuity is $\lambda < 0.44$. If $\lambda$ exceeds this, the $f$ and $g$ curves are flatter, and the optimum point moves smoothly to the corner as $\alpha$ increases. The intuition behind the result seems to be that a higher $\lambda$ gives a higher penalty for distorting production decisions, thus a reluctance to jump to a high $r$ when output prices fall.

3 The corner solution

As noted above, the solution $(t,r) = (1,0)$ should not be taken too literally. We had to impose the somewhat unrealistic Assumption 1 to arrive at this solution. This weakness, due to the corporation’s indifference to $q$ when $t = 1$, is inherent also in the standard theory. For some parameter configurations our model exhibits exactly the same weakness. Just like in the case without transfer pricing, we can only suggest some rent tax rate close to one hundred percent as the practical solution in this situation.

We cannot predict the corporation’s choice in case a hundred percent rent tax is introduced. But the model suggests one reason why a positive $q$ will nevertheless be chosen: Otherwise there will be no tax base in which to deduct the “cost” $a$. If this is really a camouflaged transfer, one could imagine that there should be a taxable income of some size in order for the camouflage to work. We do not offer a more elaborate model of this.

In the presentation of the tax system, we suggested that it is insubstantial whether a royalty is deductible in the base for the rent tax or not. This is true as long as $t < 1$. In that region the corporation reacts similarly to a deductible royalty at rate $r$ or a non-deductible royalty at rate $\theta \equiv r(1-t)$, as can be easily seen from (3). The authorities may
obtain the same maximum \( W \) from using \((t, \theta)\) as instruments as from using \((t, r)\), as can be seen from (2).

For \( t = 1 \), however, a non-deductible royalty \( \theta \) is a different instrument from a deductible royalty \( r \), since any deductible royalty would be completely refunded by the deduction, which is not the case for a non-deductible royalty. Assumption 1 does not work, since \( q = p(1 - t - \theta)/2b(1 - t) \), which is optimal for \( t < 1 \), is no longer well defined for \( t = 1 \). If we disregard the need for taxable income for deduction of \( a \), we find: Since \( \partial \Pi / \partial q = (1 - t)(p - 2bq) - \theta p \), the corporation is indifferent to \( q \) if \( (t, \theta) = (1, 0) \). If \( t = 1, \theta > 0 \), however, the corporation would want \( q = 0 \). But the benefit from deducting \( a \) raises doubts about these conclusions. Thus the corner solution, \((t, \theta) = (1, 0)\), is equally unrealistic if the royalty is non-deductible.

4 Discussion

Proposition 1 shows that the model may work well as a descriptive model: While many other models prescribe neutral cash flow taxes and tax rates which do not change when prices change, this model explains why authorities may want to use royalties and may want to change tax rates when prices change.

The introduction of cost transfers and a convex cost function attached to these transfers makes it intuitively reasonable that a deviation from a cash-flow tax may be optimal. The higher marginal tax rate on revenues than the rate at which costs are deductible, counteracts the transfers. But the results of the model are more specific than this: An analytical solution is found, and more importantly, it is shown that under some conditions the cash flow tax is optimal after all. There may also be a discontinuity as the output price changes. This contributes to explaining why some countries have not imposed a tax on gross revenues while other countries have, and why some countries introduce or abolish such taxes at significant rates from time to time.
The model is of course highly stylized. Many existing tax systems are more complicated, with depreciation schemes and less than full loss offset. But they often have a marginal tax rate on revenues which is higher than the marginal tax rate at which costs can be deducted. The problem of monitoring cost transfers may be a reason for such a feature.

As noted, the corner solution should be interpreted with particular care. The main significance here is that for some parameter values, the existence of cost transfers at a convex cost is not sufficient to explain the departure from 100 per cent rent taxes.

Our analytical results were obtained by the use of quadratic cost functions. While this made the derivation of results easier, there is a cost in loss of generality. With more general functional forms, the loci of $\partial W/\partial t = 0$ and $\partial W/\partial r = 0$ would be less tractable. There might be more intersections and discontinuities. It would be difficult to investigate the importance of discontinuities and corner solutions, while the quadratic functions give a clear example of how this works. Anyhow, the basic messages of the model would remain: Depending on parameters, either a corner or interior solutions may be optimal. These solutions may be discontinuous in the exogenous variables. It should also be noted that our model has $\Pi > 0$ and $W > 0$ at the optimum for all parameter values. This means that the introduction of a small fixed cost will not change the results.

The simple objective function for the authorities is realistic, with $\lambda = 0$, for many small nations without the capital and knowledge to develop their own natural resources. Foreign firms operate, but their after-tax profits have no welfare weight. Another reason, besides simplicity, for the choice of objective function is that we wanted a model which has 100 per cent rates as the optimum if cost transfers become prohibitively expensive. This occurs in the model when $c$ is large. If instead a concave utility of profits had been added to the objective function, with small after-tax profits resulting in very high marginal utility, this would in itself have ruled out tax rates approaching 100 per cent. Our question has been whether convex transfer costs, $ca^2$, are sufficient to explain that a 100 per cent cash flow tax is not used. We were looking for a minimal departure from a model implying 100
per cent tax rates, and we have been able to show how this departure depends on some parameters of our model.

In the part of Haufler and Schjelderup (1998) which is parallel to the present paper, they claim that that “a cash-flow tax that leaves investment decisions undistorted cannot be optimal in a setting where multinational firms engage in transfer pricing” (p. 12). In our model it can, depending on the parameters and prices. Our welfare function is not concave in after-tax profits, while theirs is, and this explains the difference.

The fact that the optimal tax rates depend on the fraction of foreign ownership in the sector, could lead to testable implications for cross-country studies. The fraction is observable in many countries, either because of general disclosure rules for corporations, or through particular disclosure rules for licensees in the natural resource sectors.

It should be noted that the model does not capture the trade-off between rent taxation systems on one hand and auctions or other fixed fee systems on the other. If there are serious problems with transfer pricing, this is an argument for fixed fees, since there will then (ideally) be no need for the high marginal rent tax rates. On the other hand, if transfer pricing problems are due to a weak government, it is conceivable that there will also be a perception of political risk which restricts the willingness to pay much up front for natural resource licences.

We regard the model more as a descriptive than as a normative model. If we were to advise tax authorities, we might recommend something different from a cash flow tax, but the idea of letting tax rates depend on the output price (relative to the input price) seems to be flawed, or at least too simplistic. In a more realistic multi-period model, one would run into the problem that output prices change, and the authorities might want to change tax rates over time. If firms realize this, there will be a dynamic game situation. With uncertainty also, the problem is difficult to solve. We leave this for future research.

An interesting analogy to the problem analyzed here is the problem of a minority shareholder in protection against dilution of profits by a majority shareholder. The analysis suggests that the minority shareholder might be interested in handing in some of the shares
in exchange for a claim to a fraction of gross revenue. This will be the case if monitoring is more difficult on the cost side, and cost transfers are costly to the majority shareholder.

5 Conclusion

It may be optimal for governments to use a combination of a cash flow tax (called a rent tax) and a tax on gross production value (called a royalty) in order to maximize tax revenue in the presence of costly cost transfers. This is a possible explanation for the frequent occurrence of royalties in practice, in spite of the well-known non-neutrality of such taxes. But a cash flow tax alone will be optimal in some situations. In particular this happens when the output price is high enough.

Furthermore, both the optimal royalty rate and the optimal rent tax rate depend on, among other things, the output price. This is one possible explanation for the frequent occurrence of tailoring of rent tax systems when large price changes occur. The optimal tax rates may be discontinuous functions of the output price. The optimal tax rates also depend on the marginal cost of public funds and/or the foreign ownership fraction in the sector, which could lead to testable implications.

The normative conclusions are less obvious. A tax system which needs tailoring in case prices change, will have draw-backs which are not analyzed here. A more elaborate model is needed to discuss what an optimal tax system is.

Acknowledgements

A Appendix

A.1 Proof of Lemma

The maximum of $W(t,r;\alpha,\lambda,s)$ for $(t,r) \in [0,1]^2$, and its dependence on $\alpha$.

For fixed $(\alpha,\lambda,s)$, $W(t,r;\alpha,\lambda,s)$ is twice continuously differentiable everywhere in $[0,1]^2$. Thus it has at least one maximum there, and the maximum or maxima occur(s) at boundary or stationary points. Stationary points occur when $f(t;\alpha,\nu,s) = g(t;\nu)$.

It can be shown that the necessary and sufficient second-order conditions for a local maximum are satisfied if and only if

$$t < 2\nu - 2\frac{\alpha}{\nu}(1-r)^2.$$  \hfill (A.1)

However, all we need in order to establish whether a stationary point is a local maximum or not, is the signs of the first-order partial derivatives in the point’s neighborhood. These are easily established from equations (9) and (10), and are indicated in figures 1–3.

Let $\alpha^*(\nu,s)$ be the value of $\alpha$ which gives tangency between the graphs of $f(t;\alpha,\nu,s)$ and $g(t;\nu)$, i.e., there is exactly one $t$ which satisfies $f(t;\alpha^*,\nu,s) = g(t;\nu)$. A formula for $\alpha^*$ is given in the proof of Proposition 3. Observe that when $\alpha < \nu - s/2$, the intersection of the graph of $f$ with the horizontal axis is $t = (\alpha + s/2)/\nu$, which approaches 1 as $\alpha$ approaches $\nu - s/2$. We consider three separate intervals for $\alpha$:

(a): $\alpha < \nu - s/2$. The only stationary point is a local maximum, cf. figure 3. The three candidates on the boundary are excluded since $\partial W(0,1/2\nu;\alpha,\nu,s)/\partial t > 0$, $\partial W((\alpha + s/2)/\nu,0;\alpha,\nu,s)/\partial r > 0$, and $\partial W(s/2\nu,1;\alpha,\nu,s)/\partial r < 0$. This implies that the global maximum is the stationary point.

(b): $\nu - s/2 \leq \alpha < \alpha^*(\nu,s)$. There are two stationary points, both along the graph of the decreasing $g(t;\nu)$. The situation is illustrated in figure 1, in which the arrows show the directions of increasing $W$ values in the various areas. It is clearly possible to find paths from the south-eastern to the north-western stationary point (— paths in
the area enclosed by the graphs of \( f \) and \( g \) —) with a monotonous increase in \( W \).

Thus the south-eastern point is not a local maximum, while the north-western is.

Consider the three candidates on the boundary: Two of them can be excluded since \( \partial W(0, 1/2\nu; \alpha, \nu, s)/\partial t > 0 \), and \( \partial W(s/2\nu, 1; \alpha, \nu, s)/\partial r < 0 \). The remaining candidate, \((1, 0)\), cannot be ruled out, however. Numerical examination for the case \( \nu = 1 \) shows that for \( \alpha \) values close to \( \nu - s/2 \), the interior candidate is the global maximum, while for greater \( \alpha \) values, closer to \( \alpha^* \), the corner point is the global maximum. The Lemma does not rely on this numerical examination, however. A closer examination of the occurrence of a non-continuity is given in Proposition 3.

\((c): \alpha > \alpha^*(\nu, s)\). There are no stationary points. On the boundary, two candidates can be excluded since \( \partial W(0, 1/2\nu; \alpha, \nu, s)/\partial t > 0 \), and \( \partial W(s/2\nu, 1; \alpha, \nu, s)/\partial r < 0 \). Thus \((1, 0)\) is the global maximum point.

This suffices to prove the Lemma, since the effect of changes in \( \alpha \) on the global maximum when it is interior, follows via the change in the graph of \( f \), while the graph of \( g \) is fixed.

### A.2 Proof of Proposition 2

The revenue from taxes, even from the rent tax alone, is always strictly positive at an interior optimum: After plugging (7) and (6) into (2), we find the base for the rent tax,

\[
(1 - r)pq - bq^2 - a = \frac{p^2(1 - r)^2}{4b} - \frac{t - s}{2c}.
\]

This is strictly positive if and only if

\[
t < \frac{p^2c(1 - r)^2}{2b} + s.
\]

Consider the possibility that an interior solution \((t, r) \in (0, 1)^2\) had instead

\[
t \geq \frac{p^2c(1 - r)^2}{2b} + s.
\]

We can show that for such high values of \( t \), one has \( \partial W/\partial t \) strictly negative: Equation (9) implies

\[
\frac{\partial W}{\partial t} \geq 0 \Leftrightarrow t \leq \frac{1}{2\nu} \left[ \frac{p^2c(1 - r)^2}{2b} + s \right] < \frac{p^2c(1 - r)^2}{2b} + s.
\]
For values of $t$ not satisfying (A.3), one could therefore not have a stationary point of $W$, and thus no interior solution, which proves the proposition.

### A.3 Proof of Proposition 3

At any given $t$ value, consider the $\alpha$ value for which a stationary point of $W$ occurs:

$$f = g \iff \frac{1 - t}{2\nu - t} = 1 - \sqrt{\frac{1}{\alpha} \left(\nu t - s\right)}$$

(A.6)

which can be solved for

$$\alpha = \frac{(2\nu - t)^2(2\nu t - s)}{2(2\nu - 1)^2} \equiv \varphi(t; \nu, s),$$

(A.7)

in which we have defined a new function $\varphi$.

The graph of $\varphi$ is shown in figure 4 for $s = 0.3$. The solid curve has $\nu = 1$, while the dashed curve has $\nu = 1.6$. If $W$ were defined over all positive $t, r$, there could be one, two or three stationary points for each value of $\alpha$.

(INSERT FIGURE 4 HERE)

The third-degree polynomial $\varphi(t; \nu, s)$ is positive when $t > s/2\nu$, and is increasing in a $t$ interval starting at this value. In order to find the local maximum denoted $(t^*, \alpha^*)$, consider the stationary points of $\varphi$. They are given by

$$t = \frac{8\nu^2 + s \pm (4\nu^2 - s)}{6\nu}.$$  

The lower of these, with the minus sign, is a local maximum for $\varphi$, while the higher is a local minimum. The local maximum corresponds to the tangency of $f$ and $g$ in the $(t, r)$ diagram, and it has the $t$ value

$$t^* = \frac{8\nu^2 + s - (4\nu^2 - s)}{6\nu} = \frac{2}{3\nu} + \frac{s}{3\nu}.$$  

(A.8)

Plugging (A.8) into (A.7) gives

$$\alpha^* = \frac{(4\nu^2 - s)^3}{54\nu^2(2\nu - 1)^2}.$$
Observe that (since $\nu \geq 1$,) $t^* < 1$ if and only if $\nu < (3 + \sqrt{9 - 8s})/4$ (which corresponds to $\lambda < (\sqrt{9 - 8s} - 1)/(\sqrt{9 - 8s} + 1)$). Call this case I, and let case II be a situation with $\nu \geq (3 + \sqrt{9 - 8s})/4$. Clearly, the solid curve in figure 4 illustrates case I, while the dashed curve illustrates case II.

**Case I, $\nu < (3 + \sqrt{9 - 8s})/4$**

Figure 4 illustrates that for $\alpha \in (0, \nu - s/2)$, there is only one stationary point of $W$ for $(t, r) \in [0, 1]^2$. Moreover, there are two stationary points for the interval $\alpha \in [\nu - s/2, \alpha^*)$. The lower of these $t$ values corresponds to the north-western stationary point in figure 1. This means that $\varphi(t; \nu, s)$ is increasing at the interior maximum.

In case I, if $\alpha$ is increased from $\alpha = \nu - s/2$, then the $t$ value at the local maximum will increase continuously (according to the inverse of the $\varphi$ function) until it reaches the value $t^*$ for $\alpha = \alpha^*$. When $\alpha$ exceeds $\alpha^*$, the (global) maximum is at the corner $(t, r) = (1, 0)$. Thus there is a discontinuity in case I. Actually, numerical examination shows that the global maximum shifts from the interior maximum to the corner for some $\alpha$ between $\nu - s/2$ and $\alpha^*$. Even though we do not give a formula for this critical $\alpha$ value, it is clear that there is a discontinuity for some $\alpha \in (\nu - s/2, \alpha^*)$.

Proposition 3 claims that the jump occurs for $t > s$. We show this by showing that $\varphi(s; \nu, s) < \nu - s/2$: Assume the contrary, that

$$\varphi(s; \nu, s) = \frac{(2\nu - s)^2s}{2(2\nu - 1)} > \nu - \frac{s}{2}.$$  

This implies

$$1 + s > 2\nu,$$

which is impossible for $s < 1$, $\nu \geq 1$.

Thus there is an interval of $\alpha$ values for which there is an interior maximum for $W$ in $[0, 1]^2$ which has $t > s$.  

---

20
Case II, \( \nu \geq (3 + \sqrt{9-8s})/4 \)

If \( \nu = (3 + \sqrt{9-8s})/4 \), then for \( \alpha = \nu - s/2 \), a tangency between \( f \) and \( g \) occurs exactly at \((t, r) = (1, 0)\). If \( \nu > (3 + \sqrt{9-8s})/4 \), then the tangency between \( f \) and \( g \) occurs for \( t = t^* > 1 \). Within the set of our interest, \((t, r) \in [0, 1]^2\), we either have the situation shown in figure 2 (if \( \alpha > \nu - s/2 \)), and in that case, the corner is the global maximum for \( W \). Or we have the situation shown in figure 3 (if \( \alpha \leq \nu - s/2 \)), and in that case, there is an interior global maximum. Clearly the \((t, r)\) which maximizes \( W \) is now continuous in \( \alpha \): When \( \alpha \) goes from just above \( \nu - s/2 \) to just below, the global maximum starts to move along the \( g \) curve, but without any jump.

References


Lund, Diderik: Compromising on rent tax neutrality to ensure economizing behavior. Memorandum no. 15/85, Dept. of Economics, University of Oslo, Norway, 1985.


Figure 1: Maximizing $W$ by choice of $t$ and $r$: Graphs of $g(t;\nu)$ (thickly drawn) and $f(t;\alpha,\nu,s)$ when there are two intersections, $\alpha = 0.89, \nu = 1.0, s = 0.3$.

Figure 2: Maximizing $W$ by choice of $t$ and $r$: Graphs of $g(t;\nu)$ (thickly drawn) and $f(t;\alpha,\nu,s)$ when there is no intersection, $\alpha = 1.3, \nu = 1.0, s = 0.3$.

Figure 3: Maximizing $W$ by choice of $t$ and $r$: Graphs of $g(t;\nu)$ (thickly drawn) and $f(t;\alpha,\nu,s)$ when there is only one intersection, $\alpha = 0.5, \nu = 1.0, s = 0.3$.

Figure 4: Dependence of stationary points on $\alpha$: Graph of $\varphi(t;\nu,s)$ when $s = 0.3$. The solid curve has $\nu = 1.0$, and the indicated values $(t^*,\alpha^*)$, $s/2\nu$, and $\nu - s/2$ correspond to this case. The dashed curve has $\nu = 1.6$. 