Parent Boxes, Tax Credits, or Both?

4 August 2026

Michael Devereux (Oxford University Centre for Business Taxation)

Ben Lockwood (Warwick University)

Erez Yerushalmi (Birmingham City University, Business School)

Working paper | 2026-04

Patent Boxes, Tax Credits, or Both?

Michael Devereux* Ben Lockwood†
Erez Yerushalmi‡

August 4, 2026

Abstract

We analyze the relative desirability of R&D tax credits (TCs) and patent boxes (PBs) as instruments for stimulating R&D, in a setting which allows for several organizational forms within the R&D sector and where there are several market failures, all of which imply under-investment in R&D. There are two key features of the model. The first is that it is a closed economy, so the (international) profit shifting role for the PB is absent. The second is that there may be an unobservable input to R&D (e.g. managerial effort) that cannot be subsidized by a TC. The government can choose a TC, a PB, and also the main rate of CIT. We find that (i) when the unobservable input is absent, a PB should *never* be used, but a TC may be, if market failure is severe enough; (ii) when the unobservable input is present, the optimal policy depends on the need for tax revenue, as measured by the MCPF. When this is low, the TC should *never* be used, but a PB may be, but as need rises, both instruments should be used, and when it is very high, only a TC should be used.

*michael.devereux@sbs.ox.ac.uk, Oxford University Centre for Business Taxation, Said Business School.
†b.lockwood@warwick.ac.uk, Warwick University.
‡Erez.Yerushalmi@bcu.ac.uk, Birmingham City University, Business School.

1 Introduction

Governments around the world incentivize R&D and other innovative activities in the belief that such activities create positive spillovers and would therefore be under-provided from a social perspective without government support. There is a wealth of literature on many aspects of R&D - the size and extent of spillovers, the impact on productivity and growth, and the extent to which government subsidies affect the level and quality of R&D and innovation.

Within that literature, relatively little attention has been paid, either theoretically or empirically, to a comparison of possible forms of government support to the R&D sector. In the context of a corporate income tax on firm profit, there are essentially two forms of potential support.[1] One is to reduce the costs of R&D expenditure; this is often done within the corporate income tax (or personal income tax) by offering higher deductions for R&D spending compared to other forms of investment. For example, the UK currently offers an R&D tax credit of 20% of expenditure, which can be offset against the firm's corporation income tax liability.[2] An alternative approach to reducing costs is to increase allowances within the definition of taxable profit, hence reducing the overall tax liability directly. We refer throughout to this type of support as a *tax credit (TC)*; although the precise form of the support is not always technically in the form of a credit, the effect is always to give relief proportional to R&D spending.

The second form of support is to reduce the tax rate applied to the income from successful R&D activity. In this paper we refer to any subsidy which takes the form of a reduction in tax on the income arising from R&D as a *patent box (PB)*.[3] The PB approach was originally closely tied up with a rather different policy objective - to limit the extent of profit shifting by multinational enterprises (MNEs). Intangible assets, such as patents and other intellectual property, created from R&D activity tend to be highly mobile; as a result, it has been relatively easy for MNEs to locate such assets in low-tax jurisdictions so that any related income stream is subject to a low rate of tax. Largely in response to this, governments have explored reducing the tax rate on such highly mobile assets, via PBs, which apply a lower rate to income arising from such assets. In an otherwise high-tax country, the lower rate

[1]There are empirical literatures looking at the effects of one or the other forms of support separately. There is a large literature on tax credits, surveyed by Becker (2015), and including Bloom, Griffith and Reenen (2002) on aggregate data and (Bronzini and Piselli, 2016; Choi and Lee, 2017; Guceri and Liu, 2019; Hud and Hussinger, 2015) on micro data. A smaller literature addresses the impact of patent boxes on R&D, (Alstadsæter et al., 2018; Bornemann, Laplante and Osswald, 2023; Evers, Miller and Spengel, 2015; Gaessler, Hall and Harhoff, 2021; Rowe-Brown and James, 2019).

[2]The UK actually taxes the tax credit, so that at a tax rate of 25%, the credit is worth only 15% of the expenditure.

[3]These are also known as innovation or intellectual property boxes.

diminishes the incentive to shift profit.

The role of patent boxes in affecting profit shifting and the location of intangible assets has been addressed both theoretically (Haufler and Schindler (2023)) and empirically (Alstadsæter et al., 2018; Griffith, Miller and O'Connell, 2014). However, OECD guidelines for the use of patent boxes were revised as part of the OECD Base Erosion and Profit Shifting (BEPS) project in 2015, OECD (2015). This introduced a nexus rule essentially designed to allow patent boxes to apply only to income arising in a country if the R&D activity had also been undertaken in that country.[4] The result of this nexus rule is that such patent boxes are now intended primarily as a subsidy to R&D investment, rather than as a means of combating profit shifting.

In this paper, we abstract from the international profit shifting issue to focus on a more basic question: in the closed economy, what is the most effective form of support for R&D, a tax credit or a patent box? Either approach can be effective in generating additional R&D by reducing the cost of capital, or other inputs. However, we show that they differ in two key ways which create a trade-off in the government's use of each instrument.

We examine this trade-off in a very general setting with risky R&D investment, and three types of firms, which can be heterogeneous, and with positive spillovers which justify government intervention.[5] Our approach builds on the models of process innovation of Bloom, Schankerman and Reenen (2013), and Akcigit, Hanley and Stantcheva (2022) in which successful R&D investment reduces costs of production.[7] First, there are firms which engage in R&D which if successful, leads to a process innovation that reduces the cost of production in an associated intermediate firm. If the R&D is successful, this intermediate firm can then purchase a license from the R&D firm to use the innovation. Finally, there is a firm that produces a final good which it sells on the world market at a fixed price. The final good is produced from differentiated intermediate goods from each of the intermediate firms.

This setting has three forms of market failure. First, we allow the probability of success of R&D firms to depend not only on their own investment, but also positively on the investment of all other R&D firms as well, which captures positive spillovers in R&D. Second, the price at which the R&D firm licences the innovation is determined by Nash bargaining between it and the intermediate firm. As a result, the R&D firm does not capture all of the returns

[4]In practice, that could apply to both research and development activity.

[5]The empirical literature generally supports the approach taken in the theoretical model. For example, Ugur, Churchill and Luong (2020) conduct a meta-analysis of productivity effects for R&D spillovers and own-R&D investment.[6] They find heterogeneous effects, and that own R&D is necessary to build absorptive capacity and secure gains from knowledge externalities, as is the case in our model.

[7]This is opposed to the product innovation approach of Romer (1990).

associated with the R&D; part of it is also captured by the intermediate firm that purchases the license. Third, each intermediate firm has some market power due to producing a differentiated good.[8] Each of these three market failures leads to a lower R&D relative to the social optimum.[9]

We consider a government that applies a corporate income tax to profit in each firm, where all costs are deductible, so it is a cash flow tax. We assume the government can set this rate subject to an upper bound. We focus on the case where the marginal cost of public funds exceeds 1, so the government seeks to use the most cost-effective way of stimulating R&D investment. The government has two instruments to try to correct market failures: a tax credit that applies to spending by R&D firms, and a patent box that reduces the tax rate applied to income earned by R&D firms.

The first issue is one of "bang for the buck"; for a given incentive effect on R&D, the two approaches do not cost the same in terms of foregone government tax revenue. Specifically, for a given impact on the incentive to invest, we show that the tax credit has a smaller revenue cost to the government. This stems from the fact that the expected income stream for any project must be at least as high as the expected expenditure and for some projects it is greater. Hence reducing the tax on income is more expensive than subsidizing expenditure.

On the other hand, a second issue concerns the observability or verifiability of inputs to R&D activity. If one input is unobservable (such as managerial effort), then it cannot receive a tax credit. Giving a credit only to observable inputs is therefore likely to distort the choice of inputs chosen by the R&D firm, creating a deadweight cost. Optimal policy therefore must trade off these two effects of subsidizing R&D: the revenue cost and the size of the deadweight cost. The merits of the tax credit and the patent box depend on the importance of each of these two factors.

Here, we have two main results. The first, Proposition 2, follows directly from the bang for the buck argument. If all inputs are observable, and so can be given a tax credit, then it is never optimal for the government to use a PB. The government should use a tax credit if the overall positive spillover from R&D on the economy is large relative to the tax cost of the TC.

In the case where there is an unobservable input, we have a second result, Proposition 3, which is less obvious. When the optimal rate of CIT is below the upper bound, the ordering of R&D policies flips completely: it is *never* optimal for the government to use a TC. On the other hand, when the upper bound on the rate of CIT is reached, both instruments,

[8] In profit maximization, that firm therefore restricts output to charge a higher price to the final good firm.

[9] The model therefore does not incorporate the possibility of over-supply of R&D in an unregulated market, for example, through a patent race; see, for example, (Jones and Williams (1998)).

and then just the TC, should be used at the optimum. The case where the upper bound is reached can be thought of as one where the need for government tax revenue, as measured by the marginal cost of public funds, is high, so our results can be restated as follows. When the need for tax revenue is low, only a PB should be used. At an intermediate level of need, both instruments should be used, and at a high level of need, only the TC should be used.

These baseline results are derived for the case where the R&D firms are separate from the intermediate firms. We also consider an alternative organizational form where R&D and intermediate good production are done by two separate divisions of the same firm, which raises the possibility of income and cost shifting.[10] For example, the firm could exaggerate its R&D spending (for the tax credit) or income from R&D (for the patent box), by shifting income and/or costs from a non-R&D part of the overall enterprise. We assume, as is standard, that such shifting activity is costly for the firm, either because of costs of hiding such shifting from the tax authorities, or fines imposed if detected. For this organizational form, we find that our main results, Propositions 2 and 3, carry over, with some small qualifications.[11] However, simulation results show that the actual levels of tax credit and patent box support depend, inter alia, on the costs of each form of shifting.

Section 2 discusses related literature, and Section 3 motivates our theoretical approach by presenting some stylised facts about PBs and TCs. Section 4 sets out the model, and Section 5 provides our main results for the baseline organizational form where R&D firms and intermediate firms are separate entities.[12] In Section 6, we consider the alternative organizational form where the R&D and the intermediate firm merge, creating incentives for the combined firm to shift income and/or costs to attempt to take advantage of the preferential treatment of R&D. Section 7 discusses extensions, and Section 8 briefly concludes.

2 Literature Review

The two most clearly related papers to ours are Akcigit, Hanley and Stantcheva (2022) and Haufler and Schindler (2023). Our model shares many features of Akcigit, Hanley and Stantcheva (2022), namely a single economy, with intermediate firms that are heterogeneous in productivity, and multi-directional R&D spillovers across these firms. Unlike us however, they assume throughout integrated intermediate firms, that both conduct R&D and produce an intermediate input, and they do not allow for transfer pricing between these two units. They take a mechanism design approach, where a firm reports its productivity and is allo-

[10]This is akin to international profit shifting, but in a one-country context.

[11]These relate to whether the cost of shifting is a resource cost or a transfer payment to the government.

[12]In this section, we prove some propositions analytically, but to explore the full model we rely on a numerical simulation model, drawing basic parameters from the literature.

cated a level of the observable input and a transfer. This pair can be shown to be equivalent to a per unit R&D subsidy to the observable input, and a profit wedge.[13] Their model does not allow for a "baseline" level of corporate tax on the final goods firm, or equivalently sets it to zero, so the profit wedge can be regarded as a measure of the patent box (i.e. a different, possibly lower level of tax on innovating firms). Indeed, the profit wedge is generally negative in their simulations.

The R&D subsidy and a profit wedge are functions of reported firm productivity, and thus do not correspond to a profit tax and subsidy as used in reality, as they are not functions of observables. However, Akcigit, Hanley and Stantcheva (2022) show that they can be *implemented* via a transfer to the firm that is a non-linear function of current and one-period lagged profit and level of the observable input, which are observable. A calibrated model shows numerically that generally, the profit wedge (resp. R&D subsidy) will be a non-linear function of profit (resp. R&D investment). This is of course theoretically interesting, but not very policy-relevant as in practice, tax credits and patent boxes are of a simpler form, and typically are linear, as discussed in Section 3 below.

Akcigit, Hanley and Stantcheva (2022) then consider simpler policies where profit and the R&D subsidy are affine (linear with an intercept term) or other simple functions of profit and R&D investment respectively. Table V, panel B reports the case closest to our paper, where both the profit tax and R&D subsidy are constrained to be affine. Unfortunately, they do not give the values of these variables, so it is not clear whether one or both of the profit tax and the R&D subsidy are used in equilibrium. Moreover, the affine form allows for a lump-sum tax element, which is not a feature of real-world CIT systems (see Section 3 below).

To conclude, we see our paper and Akcigit, Hanley and Stantcheva (2022) as being complementary. We focus on the case where the tax instruments are constrained to be linear, but at the same time, establish general theoretical results about when each of the two instruments should be used, based on simple economic intuition. In many respects, their model is more sophisticated than ours; on the other hand, we have a more general treatment of heterogeneity.[14] Also, our simpler but still quite general structure allows us to get analytical results about when a TC or a PB should be used. Also, unlike Akcigit,

[13] The former is a tax credit, in specific rather than ad valorem, form. The latter is equal to the corporate income tax rate applied to the marginal product of the unobservable input. To make this clearer, consider a firm that has gross profit $\pi(r, l)$ from an observable input r and an unobservable input l. The costs of r, l are m, ϕ respectively. Then, the net profit of this firm, if it is subject to a CIT at rate t and with a specific subsidy of s to r, is $(1-t)\pi(r, l) - (m+s)r - \phi l$. Akcigit, Hanley and Stantcheva (2022) characterize the welfare-maximizing values of s and $\tau = t\frac{\partial \pi}{\partial l}$.

[14] In Akcigit, Hanley and Stantcheva (2022), the production function for the final good is a symmetric, and spillovers between firms are symmetric.

Hanley and Stantcheva (2022), we allow the intermediate firms to be either integrated (i.e. both conduct R&D and produce an intermediate input), or separate, so we can consider the impact of important real-world issues like hold-up, transfer pricing and cost shifting on the optimal policy.

Haufler and Schindler (2023) analyze how countries set optimal policies when both patent box regimes and R&D subsidies can be used to promote innovation. They study a multi-country setting with a MNE headquartered in each country. The MNE headquartered in country h provides an R&D input to affiliates in all other countries. They show that patent box regimes emerge endogenously under policy competition (Proposition 2 (ii)), because the patent box can be used to shift profit to the home country. They also show that with full policy coordination across countries, a patent box is never used, but a "negative" patent box i.e. a higher rate of tax on the R&D unit than on production units can arise when the government has a revenue requirement i.e. when the marginal cost of public funds is greater than unity (Proposition 1).[15]

Our model and results differ from Haufler and Schindler (2023) in a number of important respects. First we abstract from international profit shifting to focus on the fundamental drivers of tax policy to support R&D. Second, our model is significantly richer than a closed-economy version of Haufler and Schindler (2023). We allow for a population of heterogeneous firms, multi-directional R&D spillovers across firms, and a stochastic outcome to R&D investment.[16] We also study several different organizational forms: separate R&D and production firms, integrated firms with transfer pricing between R&D and production divisions, and also cost-shifting between R&D and production divisions. On top of this, importantly, we allow for a non-observable (and therefore not deductible) R&D input, as in Akcigit, Hanley and Stantcheva (2022), which is a key driver of our results.

In the case with only one input (analogous to the closed-economy version of Haufler and Schindler (2023)), we find that whether firms are independent or integrated (i.e. with or without transfer pricing), a TC always dominates a PB. Unlike Haufler and Schindler (2023), this result is not driven by tax planning costs, but a more basic intuition that it is more cost-effective to subsidize the inputs to R&D rather than the profit from R&D, due to

[15] Their intuition for these results is somewhat different to our intuitions. Specifically, they say that the only effect of a patent box "is to permit MNEs to shift profits from their producing affiliates to a tax-privileged patent box. Since this process incurs tax planning costs, profit shifting increases private after-tax income by less than it reduces global tax revenue. This can never be optimal when the welfare weight of tax revenue is equal to that of private profit income." Additionally, in the case of a revenue requirement, "The combination of a tax on the returns to R&D and a subsidy on R&D inputs gives governments the possibility to tax the economic rents in the R&D sector without distorting the MNE's decision to invest in R&D."

[16] In the case of integrated firms and transfer pricing, we also allow for the transfer price paid to depend on the R&D outcome, unlike Haufler and Schindler (2023), where it is a lump-sum.

concavity of the production function. In the case with an unobserved input, we show the ranking of policy instruments is always reversed i.e. a PB is preferred to a TC, due to the fact that the latter distorts the input choice.

Our economic model is also related to a literature on the modeling of R&D in endogenous growth models (Jones (1995), Jones and Williams (1998), Jones and Williams (2000)). In particular, following the papers by Jones and co-authors, we work with a model of a final goods sector, and intermediate goods sector, and a separate R&D sector, that produces patents that can be used by the intermediate goods sector. We adopt this set-up; this allows us to consider various organizational forms i.e. types of integration between intermediate and R&D firms, as explained below.

3 Patent Boxes and Tax Credits - Some Stylised Facts

Two key features of our theoretical analysis are the following. First, we suppose that the TC and the PB are *linear* i.e. proportional to eligible costs and income respectively, at rates c, b. For R&D firms, then, the effective tax rate on income is $\tau - b$ and the effective tax relief on expenditure is $\tau + c$.[17] Second, we assume that b, c cannot be negative.

We believe that these conditions reflect actual TC and PB structures. As regards linearity of the PB and TCs, or other subsidies to R&D inputs, tere are many forms of incentives for R&D around the world, which set different rates and thresholds, as well as special regimes for different kinds of firms.[18] Typically, however, these incentives are linear for a specific type of firm. Table N1 in the Online Appendix gives details of PBs and TCs or similar schemes for all 38 countries. All of the PBs are linear, and most TCs are also linear.[19] This is very much in contrast to the case of the personal income tax, which is highly non-linear in many countries, with many brackets and rates. So, we would argue that our focus on linear

[17] Our use of the term c therefore indicates relief on expenditure over and above that provided by immediate expensing (which would imply relief at rate τ). The same effect could be achieved by offering an additional allowance of, say, A which would reduce the tax liability by $c = A\tau$. In this case, the value of c depends on which tax rate is used; but since the government chooses both A and the tax rate to be applied to the allowance, this is equivalent to simply choosing a tax credit rate, c. Details of both types of subsidy are described and analyzed in a series of OECD papers, (Gonzalez Cabral, Appelt and Hanappi, 2021; Gonzalez Cabral et al., 2023*a*,*b*,*c*).

[18] See the extensive OECD Database of Corporate Tax Statistics.

[19] As shown in Table N1, no OECD countries have a non-linear PB, although Hungary and Switzerland have an overall cap as a proportion of profit. Only two countries - Canada and France - have a clear non-linear structure for the TC. Each has only two rates - the marginal rate for higher R&D reduces to zero for high R&D expenditure in Canada, but increases from 3% to 5% in France. (Also, Netherlands has a higher reduction in labor income taxes for very low expenditures on R&D). Australia has two rates depending on the ratio of R&D to total spending, and Japan, Portugal and Spain have a higher rate where there is a substantial rise in R&D spending compared to previous years.

Table 1: Examples of R&D Support in 2024

	τ	b	c
Brazil	0.34	0	0.178
Canada	0.26	0	0.229
China	0.25	0.1	0.24
France	0.361	0.261	0.23
Germany	0.301	0	0.154
Italy	0.24	0.101	0.068
Japan	0.297	0	0.119
Netherlands	0.258	0.188	0.111
Spain	0.25	0.15	0.248
UK	0.25	0.15	0.135
USA	0.255	0	0.022

Notes: Figures are taken from, or derived from, data for 2024 from the OECD R&D Tax Incentives Database. Estimates of c are derived from OECD estimates of the "B-Index".

schemes has a strong empirical motivation.

As regards the non-negativity of the subsidy rates b, c, we are not aware of any country that sets negative rates. For example, Table 1 presents estimates of our key tax parameters, τ, b and c for 11 countries in 2024.[20] Six out of the 11 countries use a PB, reducing the overall tax rate $\tau - b$ to between 8% and 15%. All of the 11 countries have a form of R&D incentive based on expenditure - these take different forms across countries, but in all cases $c > 0$, implying that the overall treatment of R&D expenditure is more generous than immediate expensing.

4 The Model

This section sets up the model used to analyze the tax instruments available for supporting R&D expenditure. We describe in turn the position of each type of firm, the market equilibrium and the market failures that are present.

4.1 The Setting

Firms. There are three kinds of firms. At the end of the production chain is a single competitive *final good firm*, which produces an aggregate output Y from intermediate inputs $q_i,\ i = 1, ..n$. This good is sold either domestically or on the international market at a fixed

[20]Note that the estimates of τ include sub-national taxes.

price 1. The production function is a decreasing-returns function of intermediate inputs :

$$Y = \sum_{i=1}^{n} a_i (q_i)^{1-1/\varepsilon},\ 1 < \varepsilon < \infty \tag{1}$$

where $a_i > 0$. Each unit of Y is sold on the world market at a price of 1, and the final good firm takes the prices p_i as given.

Next, there are *intermediate firms* $i = 1, ..n$, each producing a differentiated intermediate input q_i from labour. The labour unit cost of the intermediate firm i is ϕ_i, and firm i sells q_i to the final good firm at price p_i.

Finally, there are *R&D firms* $i = 1, ..n$. R&D firm i invests in technology that may with some probability reduce the unit cost ϕ_i of the corresponding intermediate firm. Specifically, each R&D firm i employs both (skilled) labour l_i and the effort of the owner e_i which are then aggregated into a composite R&D input, $r_i(l_i, e_i)$. It is assumed that r_i is strictly increasing in both arguments. The key difference between the two inputs is that l_i is a cost that is observable and can be subsidized via a tax credit, whereas e_i is the private information of the owner and thus by assumption cannot be subsidized. Then, $\sigma_i = \sigma_i(r_i(l_i, e_i), r_{-i})$ is the probability that R&D firm i can produce an innovation that reduces the cost of the corresponding intermediate firm from $\phi_i = \overline{\phi}_i$ to $\underline{\phi}_i < \overline{\phi}_i$. Here, $r_{-i} = (r_j)_{j \neq i}$. Also, σ_i is assumed to be between zero and one and is strictly increasing in r_i, and non-decreasing in all r_j. This specification captures positive innovation spillovers in the R&D sector. The innovation, if it occurs, is assumed to be protected by a patent. Note that we allow for heterogeneity of firms across intermediate and R&D firms across all dimensions: a_i, ϕ_i, σ_i, and r_i can all vary with i. One of the features of our analysis is that our results largely hold without restricting this heterogeneity.

Households. There are $n + 1$ households, the n owners of the R&D firms, and a single representative household that supplies all labour inputs. The first n households have linear preferences of the form $x_i - e_i$, where x_i is consumption of the internationally traded good, and household $n + 1$ supplies labour to all firms that use it, and has quasi-linear preferences over the final good and leisure $U(x_{n+1}) + l$, where x_{n+1} is consumption of the internationally traded good. Leisure $l = l$ is equal to the total time endowment H, minus labour supplied. We will assume that in equilibrium, total labour supplied is less than the time endowment H, so the wage will be 1 in equilibrium. Finally, households $i = 1, ..n$ receive profit from the ownership of the corresponding R&D firm i, and household $n+1$ receives the profit from the other two types of firm. Together, these assumptions ensure that (i) R&D firms maximise profit net of the cost of effort; (ii) aggregate household welfare is measured by the sum of profits across all firms, net of the cost of effort.

Order of Events. The order of events, conditional on fixed taxes, is the following:

1. R&D firms choose l_i, e_i and the innovation outcome (success or failure) occurs.
2. If the innovation is successful σ_i, R&D firm i can license the cost-reducing technology to the intermediate firm i at price P_i;
3. Intermediate firms set prices p_i, conditional on ϕ_i;
4. The final good firm chooses demands q_i, having observed prices p_i.

Taxes and Profits. The baseline rate of CIT is τ.[21] The government has two additional tax instruments, b, c. Here, b is a lower rate of tax on any profits generated by an R&D firm, and is therefore a Patent Box (PB). Also, c is a tax credit (TC) on spending on the labour input l_i. Then, the post-tax profit of the R&D firm i, net of the cost of effort, is

$$\pi_i^{RD}(l_i, e_i) = \begin{cases} P_i(1-\tau+b) - l_i(1-\tau-c) - e_i & \text{with prob.} \quad \sigma_i \\ -l_i(1-\tau-c) - e_i & \text{with prob.} \quad 1-\sigma_i \end{cases} \tag{2}$$

The profits of the intermediate and final goods firms are subject only to CIT at rate τ. Motivated by Section 3, we will assume the following:

$$0 \le b \le \tau, \ 0 \le c \le 1-\tau, \ 0 \le \tau \le \overline{\tau}. \tag{3}$$

The first constraint $0 \le b \le \tau$, says that the reduction in CIT from the PB cannot lower tax on profit in the PB below zero. Similarly, $0 \le c \le 1-\tau$ says that the TC cannot lower the cost of l_i below zero. Finally, $\overline{\tau}$ is the maximum feasible CIT: this may be 100% i.e. $\overline{\tau} = 1$. This upper bound on the rate of CIT is a simple way of capturing the fact that there may be other reasons why the government cannot tax corporate profits at 100%, for example, international profit shifting or effects on small businesses.[22] The assumption of an upper bound is relaxed in Section 7.2, where we model explicitly the economic costs of a high rate of CIT. In that section, we show that the qualitative results extend to that case.

[21]We assume throughout that expenditure of all firms is tax deductible; in effect this implies that the corporate income tax is a cash flow tax, neutral for investment decisions. This helps to keep the model tractable, but has the implication that, in the absence of other incentives, raising the tax rate has no impact on economic activity. In turn this creates an incentive for the government to choose a high tax rate. An alternative approach would restrict tax deductions; for example Haufler and Schindler (2023) model the case in which investment expenditure receives no deductions.

[22]In practice, rates of CIT around the world are generally below 40%, and in many cases well below this. This may reflect a number of features not explicitly modeled here, including the negative impact of a high τ on investment in the absence of a cash flow tax, international competition for both investment and profit, potential double taxation with the personal tax system, and the creation of disincentives to use corporate form.

4.2 Equilibrium

Here, we find the equilibrium given the tax parameters (τ, b, c). As is standard, we solve the model backwards. At stage 4, the final good firm's pre-tax profit is:

$$\pi^F = \sum_i a_i (q_i)^{1-1/\varepsilon} - \sum_i p_i q_i \tag{4}$$

Note that post-tax profit is $(1-\tau)\pi_F$ and thus τ does not affect any of the final firm's decisions. Maximization of (4) with respect to q_i gives inverse demand functions for the intermediate goods as:

$$p_i = A_i \left(q_i\right)^{-1/\varepsilon}, \quad A_i = a_i \left(\frac{\varepsilon - 1}{\varepsilon}\right) \tag{5}$$

At stage 3, from (2), (5), ignoring the cost of purchasing the license from the R&D firm, firm i has pre-tax profit

$$\pi_i^I = p_i q_i - \phi_i q_i = A_i (q_i)^{1-1/\varepsilon} - \phi_i q_i \tag{6}$$

Again, post-tax profit is $(1-\tau)\pi_i^I$ and thus τ does not affect any of the firm i's decisions. So, this firm chooses q_i to maximize (6) which implies the standard mark-up equation

$$p_i = \frac{\varepsilon}{\varepsilon - 1}\phi_i \equiv \mu\phi_i$$

The second-order condition is satisfied here as marginal revenue is strictly decreasing in q_i by inspection. Also, let $\underline{\pi}_i^I, \overline{\pi}_i^I$ be the maximized values of profit in (6) for costs $\overline{\phi}_i, \underline{\phi}_i$ respectively, i.e., an innovation lowers cost to $\underline{\phi}_i$ and raises profits to $\overline{\pi}_i^I$.

At stage 2, assuming that R&D firm i has successfully innovated, there is a single buyer and seller of the license. We assume that the price P_i is determined by standard Nash bargaining (whereby the threat points are $\underline{\pi}_i^I, 0$ for the intermediate and R&D firm respectively, the latter being zero because the R&D firm's costs are sunk once the innovation has occurred). The formula for the Nash product is therefore

$$\left((1-\tau)\left(\overline{\pi}_i^I - P_i - \underline{\pi}_i^I\right)\right)^{1-\beta}\left((1-\tau+b)\,P_i\right)^{\beta} = K\left(\overline{\pi}_i^I - P_i - \underline{\pi}_i^I\right)^{1-\beta} P_i^{\beta}$$

where $K = (1-\tau)^{1-\beta}\,(1-\tau+b)^{\beta}$. Then maximization of this expression with respect to P_i yields

$$P_i = \beta\left(\overline{\pi}_i^I - \underline{\pi}_i^I\right) \tag{7}$$

This is a standard Nash bargaining solution which states that the seller gets their reservation price (in this case zero) plus some fraction $\beta \in [0,1]$ of the total surplus generated by the innovation. The latter is $\Delta_i^I \equiv \overline{\pi}_i^I - \underline{\pi}_i^I$, i.e. the increase in profit for intermediate firm i made possible by the innovation. Importantly, because the R&D firm's costs are sunk, the tax parameters τ, b, c do not appear in this expression. They are multiplicative constants (in K) that scale the objective without changing the surplus split.

This brings us to stage 1. Combining (2) and (7), the expected profit of the R&D firm is

$$E[\pi_i^{RD}(l_i, e_i)] = \sigma_i\left(r_i\left(l_i, e_i\right), r_{-i}\right)\beta\Delta_i^I\left(1-\tau+b\right) - l_i\left(1-\tau-c\right) - e_i \tag{8}$$

In order to ensure that the firm's problem is well-behaved, we will assume:

A1. *r_i is twice continuously differentiable and strictly increasing and concave in l_i, e_i.*

A2. *$\sigma_i(r_i, r_{-i})$ is twice continuously differentiable in r_i, r_{-i}, strictly increasing and concave in r_i, and non-decreasing in r_{-i}.*

A3. *l_i, e_i are not perfect substitutes i.e. r_i cannot be written as a function of a linear combination of l_i, e_i.*

To show that A1,A2,A3 can be satisfied, we consider the following running example:

$$\sigma_i = 1 - \exp\left(-\left(r_i + s\sum_{j\neq i} r_j\right)\right),\ r_i = \alpha \ln l_i + (1-\alpha)\ln e_i,\ \alpha\in(0,1),\ s\geq 0 \tag{9}$$

Note that A1, A2,A3 are all satisfied in this case. Here, α measures the relative importance of the observable input, and s measures the R&D spillover.[23]

With A1 and A2, by inspection, the R&D firm's profit $\pi_i^{RD}(l_i, e_i)$, being a linear function of σ_i, is also strictly concave in l_i, e_i.[24] Moreover, we will assume interior solutions for both inputs.[25] So, the profit-maximizing choices of l_i, e_i are fully characterized by the first-order conditions for a maximum of (8) w.r.t. l_i, e_i, which can be written:

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i}\beta\Delta_i^I = \frac{1-\tau-c}{1-\tau+b} \equiv p_l,\ \ \frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial e_i}\beta\Delta_i^I = \frac{1}{1-\tau+b} \equiv p_e \tag{10}$$

where p_l, p_e are the tax-inclusive prices of the two inputs. Note that the price of the observable input l_i is determined by both b and c, whereas the price of the unobservable input e_i is determined by b only.

[23] Additional restrictions are required to ensure that $\sigma_i > 0$ - see the Online Appendix.

[24] This implies that the expected profit of the R&D firm is always strictly positive. However, if the innovation fails, ex post, the R &D firm makes a loss.

[25] A sufficient condition for this is that both $\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i}$, $\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial e_i}$ have limits greater than $\frac{1}{(1-\overline{\tau})\beta\Delta_i^I}$ as $l_i, e_i \to 0$ respectively.

4.3 Market Failures

We now wish to compare the market equilibrium levels of l_i, e_i studied in the previous section to the social optimum levels. To do this, consider the levels of l_i, e_i that maximise welfare in the absence of any tax-raising requirement. Absent this requirement, welfare is defined as the total pre-tax profit in the economy because the final good price is fixed on the world market and thus, consumer surplus is exogenous and unaffected by domestic choices of l_i and e_i. Moreover, because final good production is additively separable across inputs, we can isolate the gross (pre-tax) profit of the final good firm from using input i:

$$\pi_i^F(\phi_i) \equiv \max_{q_i} \left\{ a_i(q_i)^{1-1/\varepsilon} - q_i\mu\phi_i \right\} \tag{11}$$

Let $\underline{\pi}_i^F < \overline{\pi}_i^F$ be the values of $\pi_i^F(\phi_i)$ at $\phi_i = \overline{\phi}_i > \underline{\phi}_i$ respectively. So, a lower ϕ_i due to innovation raises profits of the final good firm by $\Delta_i^F = \overline{\pi}_i^F - \underline{\pi}_i^F$. We therefore can define total pre-tax profit in the economy, generated by input i, without and with the innovation as

$$\underline{\pi}_i = \underline{\pi}_i^F + \underline{\pi}_i^I, \;\; \underline{\pi}_i + \Delta_i, \tag{12}$$

respectively, where $\Delta_i = \Delta_i^F + \Delta_i^I$ is the incremental total profit across intermediate firm i and the downstream firm from an innovation in sector i. These expressions do not include the costs of inputs l_i, e_i. Thus, the government's objective can be written as follows[26];

$$\Pi = \sum_{i=1}^{n} \left(\underline{\pi}_i + \sigma_i\Delta_i - l_i - e_i \right), \tag{13}$$

Note that the expected cost of the license, $P_i = \beta\Delta_i^I$ nets out of this expression because it is a pure transfer, not a real resource cost.

The socially optimal choice of l_i, e_i is therefore given by maximization of (13). It is helpful to define the *innovation spillover*:

$$s_i \equiv \sum_{j \neq i} \frac{\partial \sigma_j}{\partial r_i} \Delta_j \geq 0.$$

This is the impact of a marginal increase in the input r_i on aggregate profits in sectors $j \neq i$

[26]Note that in (13), the mark-up μ of the intermediate goods firms is included, which is also a source of distortion in the economy. Of course, this mark-up could be eliminated if the government also had access to production subsidies for all the intermediate good producers, but this is unrealistic, and we do not allow for it.

via the effect of r_i on σ_j. Then, we can write the FOC for the choice of l_i, e_i as:

$$\left(\frac{\partial \sigma_i}{\partial r_i}\Delta_i + s_i\right)\frac{\partial r_i}{\partial l_i} = 1, \ \left(\frac{\partial \sigma_i}{\partial r_i}\Delta_i + s_i\right)\frac{\partial r_i}{\partial e_i} = 1 \tag{14}$$

showing that the planner uses l_i and e_i until the marginal social benefits (i.e., the expected gain from raising σ_i plus the cross-sector spillover from s_i) equals the marginal cost of 1.

We define the overall spillover from research, denoted S_i to be the difference between the social and private return of a small increase in r_i. Using (14) and (10), this can be calculated as:

$$S_i = \underbrace{s_i}_{\text{innovation spillover}} + \underbrace{(\Delta_i - \Delta_i^I)\frac{\partial \sigma_i}{\partial r_i}}_{\text{monopoly power}} + \underbrace{(1-\beta)\Delta_i^I\frac{\partial \sigma_i}{\partial r_i}}_{\text{hold-up}} \tag{15}$$

This decomposition makes it clear that there are three market failures.

- First, and most obviously, in choosing l_i, e_i, the R&D firm ignores the innovation spillover effect of l_i, e_i on r_j and thus on σ_j, captured by s_i.

- Second, as long as $\beta < 1$, there is a hold-up problem in the choice of the R&D firm, which causes it to invest too little, as it does not appropriate all of the intermediate firm's profit increase from the innovation.

- Third, the social return to lowering the production cost of intermediate good i from $\overline{\phi}_i$ to $\underline{\phi}_i$ is the increase in overall profit Δ_i, and it is easily checked that this is strictly greater than the change in profit Δ_i^I for the intermediate goods firm i, because intermediate firm i, being a monopolist, does not capture the full additional profit the final goods firm gets from the innovation.[27]

The second and third market failures are examples of what **?** calls the *surplus appropriability problem.* For example, the intermediate goods firm cannot appropriate all of the producer surplus from the cost, which in this case, reduces its incentive to license the patent. In fact, the hold-up problem is also a surplus appropriability problem for the R&D firms, as they only get a share β of the total surplus generated by its innovation.

All three of these market failures work in the same direction. Failures one and two will directly cause the R&D firms to invest too little in the R&D inputs. Failure three will do so indirectly, because conditional on R&D firm i innovating, the downstream intermediate i firm, does not capture the full producer surplus from the innovation, meaning that even if

[27]See the Online Appendix for explicit formulae for Δ_i, Δ_i^I in terms of parameters.

R&D firm i had all the bargaining power in setting P_i i.e. $\beta = 1$, it would still under-invest, due to lack of this appropriation.[28]

5 Analysis of Tax Policy

Having set up the model, we now turn to the tax design problem for the government.

5.1 Tax Revenue and Welfare

As is standard, the government's objective function is the sum of pre-tax profits across all firms, Π, as defined in (13) above, plus $(\delta - 1)$ times the tax revenue T, where $\delta \geq 1$ is the marginal cost of public funds (MCPF). We will focus on the case where $\delta > 1$ i.e. the value of tax revenue to the government is greater than the weight it assigns to firm profit.

Total expected tax revenue is

$$T = \tau \sum_i (\underline{\pi}_i + \sigma_i \Delta_i - \sigma_i P_i) + (\tau - b) \sum_i \sigma_i P_i - (\tau + c) \sum_i l_i \tag{16}$$

where the first term is tax revenue from the final good firm and the intermediate firms i.e. all firms that are taxed at the standard rate of CIT, and the second and third terms comprise the net tax paid by the R&D firms. Re-arranging slightly, we obtain

$$T = \tau \sum_i (\underline{\pi}_i + \sigma_i \Delta_i - l_i) - b \sum_i \sigma_i P_i - c \sum_i l_i \tag{17}$$

Note that e_i does not appear in the tax base as it is unobservable and thus cannot be taxed or subsidized. Finally, combining (13) and (17), and recalling that $P_i = \beta \Delta_i^I$ and that the government objective is $W = \Pi + (\delta - 1) T$, we get:

$$\begin{aligned} W = & \sum_{i=1}^{n} (\underline{\pi}_i + \sigma_i \Delta_i - l_i - e_i) \\ & + (\delta - 1) \left(\tau \sum_i (\underline{\pi}_i + \sigma_i \Delta_i - l_i) - b \sum_i \sigma_i \beta \Delta_i^I - c \sum_i l_i \right) \end{aligned} \tag{18}$$

In what follows, we assume that W is concave in the policy variables (τ, b, c). In the numerical implementation, the model consistently produces a unique solution for all parameter

[28] Of course, if the government could subsidize the intermediate good, this third source of inefficiency could be eliminated. However, this is not very realistic, and also outside the scope of our analysis.

values and parameter combinations considered, providing strong numerical support for this assumption.

5.2 Welfare Improving Tax Reforms

We start the analysis of optimal policy by showing that there are two feasible tax reforms that always improve welfare, starting at certain values of τ, b, c. This provides intuition, as well as simple proofs, of our main results.

Proposition 1. *Assume that initially, τ, b, c are feasible i.e. satisfy (3). (i) If $e \equiv 0$ (no unobservable input) and $b > 0$, then any reform $-db < 0$, $dc = \frac{1-\tau-c}{1-\tau+b}db > 0$ increases welfare (reform 1). (ii) If $c > 0,\ \tau < \overline{\tau}$, then any reform $d\tau = db = -dc > 0$ increases welfare (reform 2).*

The intuition for these results is straightforward. It is easily checked from (10) that reform 2 leaves the input prices p_l, p_e unchanged. It is then a simple matter to check that this reform increases tax revenue. Similarly, reform 1 leaves the price of the single input l, p_l, unchanged. Also, note that σ_i is strictly concave in l_i, from A1,A2. This strict concavity implies that reform 1 strictly increases tax revenue T and thus welfare W. Intuitively, the *tax cost* of the TC is cheaper than the tax cost of the PB because it subsidizes the input, rather than the output, which is a strictly concave function of the input. It is also important to note that reforms 1 and 2 are not in contradiction, because in reform 1, the higher tax cost of increasing the PB and reducing the TC is more than offset by simultaneously increasing τ.

5.3 No Unobservable Input

With only one input, we can apply the tax reform results of Proposition 1 to characterise the optimal policy τ^*, b^*, c^*. At the optimum, neither reform 1 nor 2 can be possible. First, for reform 1 not to be possible, we cannot have $b^* > 0$. Second, for reform 2 not to be possible, we cannot have $\tau^* < \overline{\tau},\ c^* > 0$. This leaves two possibilities: (i) $b^* = 0,\ c^* = 0, \tau^* < \overline{\tau}$; (ii) $b^* = 0, \tau^* = \overline{\tau}$. But (i) cannot be optimal, because by inspection of (10), then any change in τ does not change the input price p_l which is fixed at 1. As a consequence, τ can be increased without inducing any behavioural response by the R&D firm. In turn, from (18), this must increase welfare as then tax revenue increases without changing the σ_i. So, it must be the case that (ii) holds i.e. $b^* = 0, \tau^* = \overline{\tau}$. This leaves open two possibilities; c^* could be zero or strictly positive. The following Proposition summarizes the discussion so far and provides a necessary and sufficient condition for c^* to be strictly positive.

Proposition 2. *Assume A1,A2 hold. Assume no unobservable input i.e.* $e = 0$*. Then* $\tau^* = \overline{\tau}$ *and* $b^* = 0$ *i.e. the PB should never be used to incentivise R&D investment. Also,* $c^* > 0$ *iff the following condition holds at* $b = c = 0$*:*

$$(1 + (\delta - 1)\overline{\tau}) \sum_i S_i \frac{\partial r_i}{\partial l_i} \frac{\partial l_i}{\partial c} > (\delta - 1) \sum_i l_i \tag{19}$$

The intuition for the first part of the Proposition has already been discussed in Section 5.2. The intuition for the second part is the following. The condition (19) says that the sum of total spillovers S_i across all R&D firms, weighted by the responsiveness of r_i to the tax credit i.e. $\frac{\partial r_i}{\partial l_i} \frac{\partial l_i}{\partial c}$ exceeds the deadweight cost of using the tax credit, $(\delta - 1) \sum_i l_i$. If δ is high, then the tax credit is simply too costly to use as an instrument for correcting the overall spillover and should not be used.

5.4 The General Case

In this general case, we can again use Proposition 1 to establish results. If the policy τ^*, b^*, c^* is optimal, it must be the case that reform 2 is *not* possible, i.e. we cannot have $\tau^* < \overline{\tau},\ c^* > 0$. So, it immediately follows that there are two possible cases, (i) $c^* = 0$, and (ii) $\tau^* = \overline{\tau}$. In case (ii), analytical results on b^*, c^* are difficult, so we resort to simulations, reported in Section 5.5 below, where it is shown that when $\tau^* = \overline{\tau}$ i.e. CIT is at a corner, generally, both instruments are used. In this next Proposition, we summarise case (i), and also establish conditions under which a PB is used at the optimum.[29]

Proposition 3. *Assume A1,A2 hold. Suppose that at the solution for the government's problem, there is an interior solution for* τ i.e. $\tau^* < \overline{\tau}$*. Then* $c^* = 0$*. Also,* $b^* > 0$ *iff the following condition holds at* $b = c = 0$*:*

$$(1 + (\delta - 1)\tau) \left(\sum_i S_i \frac{\partial r_i}{\partial l_i} \frac{\partial l_i}{\partial b} + \sum_i S_i' \frac{\partial r_i}{\partial e_i} \frac{\partial e_i}{\partial b} \right) > (\delta - 1) \sum_i \sigma_i \beta \Delta_i^I \tag{20}$$

where $S_i = s_i + \left(\Delta_i - \beta \Delta_i^I\right) \frac{\partial \sigma_i}{\partial r_i}$, $S_i' = s_i + \left(\Delta_i - \frac{(1-\tau)\beta\Delta_i^I}{1+(\delta-1)\tau}\right) \frac{\partial \sigma_i}{\partial r_i}$ *are the external spillovers associated with an increase in inputs* l_i, e_i *respectively.*

This proposition says that when the CIT is not used to its maximum extent, the PB

[29] A final technical point is that there is a borderline case where $\frac{\partial W}{\partial \tau} = 0$ at $\tau = \overline{\tau}$. In this case, it is also possible to prove (at the cost of considerable additional algebra) that at the optimum, $\frac{\partial W}{\partial b} > \frac{\partial W}{\partial c}$, so that $c > 0$ only if b is set at its maximum value i.e. $b = \tau$.

dominates the tax credit (TC): a TC should never be used, but the PB should be used when condition (20) holds. Condition (20) is analogous to (19), and has the same interpretation. Specifically, the right-hand side of (20) is the tax cost of using b, the LHS is the total spillover effect on welfare of b, via changes in inputs induced by b. Note that given a positive rate of CIT, $\tau > 0$, then $S_i' > S_i$ because e_i is not deductible against CIT, unlike l_i, and so the extra tax cost via deductibility is not incurred when e_i increases. It turns out that in the simulations, condition (20) is always satisfied, i.e. possibility (i) in the proposition does not occur.

Finally, we explain why regularity conditions A1, A2, A3 are needed for the result. To see this, consider first the case where inputs are perfect substitutes, violating A3. Then r_i can be written as a function of $l_i + \eta_i e_i, \eta_i > 0$. Note from (3) that for all feasible tax parameters, the observable input is weakly cheaper i.e. $p_l \leq p_e$. Assume also for simplicity that $\eta_i \leq 1, i = 1, ..n$. Then, none of the R&D firms will use the unobservable input and so we are back to the case of one input, and consequently, Proposition 2 applies and a PB should never be used.

The second case is that of perfect complements, in which case r_i can be written as a strictly concave function of $\min\{l_i, \eta_i e_i\}, \eta_i > 0$. This case clearly violates the continuous differentiability assumption in A1. Also, assume for simplicity that $\eta_i = \eta, i = 1, ..n$ so all R &D firms use the inputs in the same proportions. Then, all R &D firms again all use effectively a single input $x_i \equiv l_i = \eta e_i$ and it can be shown that Proposition 2 continues to apply. We can summarise the discussion as follows:

Proposition 4. *If the inputs l, e are perfect substitutes, i.e. r_i is a strictly concave function of* $l_i + \eta_i e_i, \eta_i > 0$, and $\eta_i \leq 1$, *then $b = 0$ i.e. the PB should never be used. If the inputs l, e are perfect complements and used in the same proportions i.e. r_i can be written as a strictly concave function of $min\{l_i, \eta e_i\}$ then $b = 0$ i.e. the PB should never be used.*

So, Propositions 3 and 4 taken together, say that a PB dominates a TC except when special conditions hold, which effectively reduce the model to one with a single R&D input.

5.5 Simulation Results of the Base Model

Proposition 3 identifies the optimal values of the PB and the TC in the case of an interior solution for τ. However, it does not identify their optimal values in the case in which τ is constrained. To investigate this case, we draw on numerical simulations. Our aim is two-fold: to identify when τ is more likely to be unconstrained $\tau < \bar{\tau}$ ("τ-unconstrained") or constrained at its upper level $\tau = \bar{\tau}$ ("τ-constrained"), and to identify the optimal use

Table 2: Optimal policy regimes

		τ	b	c	$\frac{\partial W}{\partial b}$	$\frac{\partial W}{\partial c}$
Regime 1: $\tau < \bar{\tau}$						
1.1	No instruments	int.	$= 0$	$= 0$	≤ 0	≤ 0
1.2	PB only	int.	$(0, \tau)$	$= 0$	$= 0$	≤ 0
1.3	PB exhausted†	int.	$= \tau$	$= 0$	≥ 0	≤ 0
Regime 2: $\tau = \bar{\tau}$						
2.1	No instruments	$\bar{\tau}$	$= 0$	$= 0$	≤ 0	≤ 0
2.2	Credit only	$\bar{\tau}$	$= 0$	> 0	≤ 0	$= 0$
2.3	PB interior, $c = 0$	$\bar{\tau}$	$(0, \bar{\tau})$	$= 0$	$= 0$	≤ 0
2.4	PB interior, $c > 0$	$\bar{\tau}$	$(0, \bar{\tau})$	> 0	$= 0$	$= 0$
2.5	PB exhausted, $c = 0$	$\bar{\tau}$	$= \bar{\tau}$	$= 0$	≥ 0	≤ 0
2.6	PB exhausted, $c > 0$	$\bar{\tau}$	$= \bar{\tau}$	> 0	≥ 0	$= 0$

Notes: For simplicity, we ignore the borderline case where $\frac{\partial W}{\partial \tau}|_{\bar{\tau}} = 0$. In this case, from Proposition 3, we know that $c = 0$. †In Regime 1.3, the constraint $b = \tau$ binds so the relevant optimality condition is the total derivative $dW/d\tau\big|_{b=\tau} = \partial W/\partial \tau + \partial W/\partial b = 0$.

of the PB and TC in the latter case. Our simulations are based on the functional forms in the running example set out in equation (9), using the computer program GAMS, and implementing the government's optimization problem as a Mixed Complementarity Problem (MCP). Table 2 is a summary of the Karush-Kuhn-Tucker (KKT) conditions for both simulated models and their regime signatures.[30]

Table 2 classifies the optimal policy outcomes into two broad regimes, according to whether τ is unconstrained or constrained. Within each of these regimes, there are sub-regimes. The first part of the table identifies *Regime 1*, which is essentially the outcomes set out in Proposition 3. In this case, the use of the PB dominates, with the TC only being used if b is at its maximum value of τ. (In our simulations, we do not find any occurrences in which the TC is used in Regime 1.) The second part of the table identifies six possible sub-regimes of *Regime 2*, the τ-constrained case. Note that, as in Proposition 2, the case of the binding upper limit for TC, $c = 1 - \tau$, is ruled out by the fact that the price of l_i cannot be zero in equilibrium. That leaves two possibilities for c: $c = 0$ and $0 < c < 1 - \tau$, along with three for b. In *Regime 2*, the binding ceiling on τ prevents the government from freely adjusting the tax rate, and the optimisation reduces to the choice of (b, c) only.

Our next step is to identify how the parameters of the model are likely to impact on which regime is relevant. In the simulation model, we do not allow for heterogeneity across

[30]This journal's Online Appendix provides a concise description of the numerical implementation and the KKT conditions underlying the MCP formulation. A complete replication package, including a detailed technical appendix, fully documented GAMS implementations of both the base case and the shifting case, and additional implementation notes, is available from the project's GitHub replication repository.

firms, and so drop the i subscripts. The structural parameters are chosen as follows. We set the two parameters from the production function of the final goods firm as $a = 5$ and $\varepsilon = 5$; this generates a mild decreasing returns to scale production function. We set the level of costs in the intermediate firm if the R&D is successful at $\phi = 0.7$ - a drop in costs of 30% per unit. We set the parameter for the R&D production function in our running example at $\alpha = 0.8$, reflecting the expectation that the unobserved factor is used less than the observed factor. In addition, in the base case we set the MCPF at $\delta = 1.3$, and the Nash bargaining parameter, $\beta = 0.4$.

We begin by exploring how variation in these parameters changes the optimal policy regime. In Figure 1, we present cases below for a range of values of two key parameters. The first is δ, the MCPF, which measures "demand" for tax revenue. The second is $1 - \beta$, which measures the intensity of the hold-up problem. From (15), $1-\beta$ is a component of the overall spillover from research activity r, so an increase in $1 - \beta$ increases the total spillover Ω. In addition, an increase in $1 - \beta$ increases the profit subject to tax in the intermediate firm. Profit in the intermediate firm is taxed as a cash flow tax - the tax base is simply income less all expenditure - and as such, does not distort its choices. A higher $1 - \beta$ therefore creates a larger incentive to raise τ to capture higher tax revenue; while this may distort R&D decisions, these effects can be offset by the use of the PB and/or the TC.

Figure 1 shows which regimes occur at the optimum policy for combinations of δ and $1-\beta$, when the maximum value of τ is set to a high value of 80%. We see that most combinations of these two parameters result in Regime 2; this is a reflection of the fact that as long as adjustments to b and c are made to keep the prices of inputs constant, τ is a lump-sum tax with a strictly positive benefit (as $\delta > 1$), so for most values of the MCPF, the government will want to increase τ to its maximum value. This is more likely when $1 - \beta$ is high, where the spillover is higher, and where more profit is captured by the intermediate firm. It is only at low values of $1 - \beta$ that there are any instances of Regime 1. These instances of Regime 1 also depend on a low value of δ, where revenue concerns are less pressing. As δ rises, then the optimal policy tends to move away from the PB. For high values of $1 - \beta$, we move from full use of the PB ($b = \tau$) in Regime 2.6 to lower values ($b < \tau$) in Regime 2.4. In these cases, $c > 0$. For low values of $1 - \beta$, we move to cases where the PB is not used at all ($b = 0$) and only the TC is used ($c > 0$) in Regime 2.2.

Figure 1: Optimal policy regimes across $(\delta, 1-\beta)$ parameter space

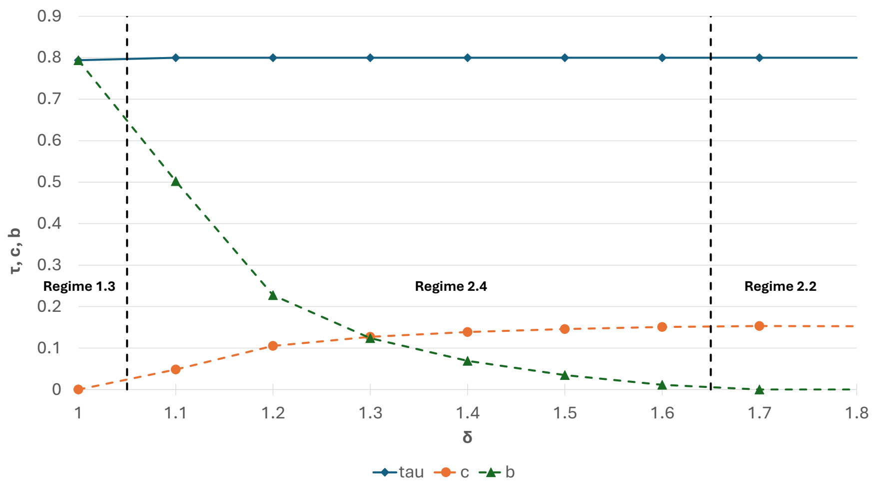
δ \ $1-\beta$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	1.3	1.3	1.3	1.3	1.3	1.3	2.6	2.6	2.6
1.1	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.6	2.6
1.2	2.2	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.6
1.3	2.2	2.2	2.2	2.4	2.4	2.4	2.4	2.4	2.6
1.4	2.2	2.2	2.2	2.2	2.4	2.4	2.4	2.4	2.6
1.5	2.2	2.2	2.2	2.2	2.2	2.4	2.4	2.4	2.4
1.6	2.2	2.2	2.2	2.2	2.2	2.4	2.4	2.4	2.4
1.7	2.2	2.2	2.2	2.2	2.2	2.2	2.4	2.4	2.4
1.8	2.2	2.2	2.2	2.2	2.2	2.2	2.4	2.4	2.4

Parameters: $\bar{\tau} = 0.8, \alpha = 0.8, \phi = 0.7, \varepsilon = 5,\ a = 5$, and loop on $1-\beta$ and δ

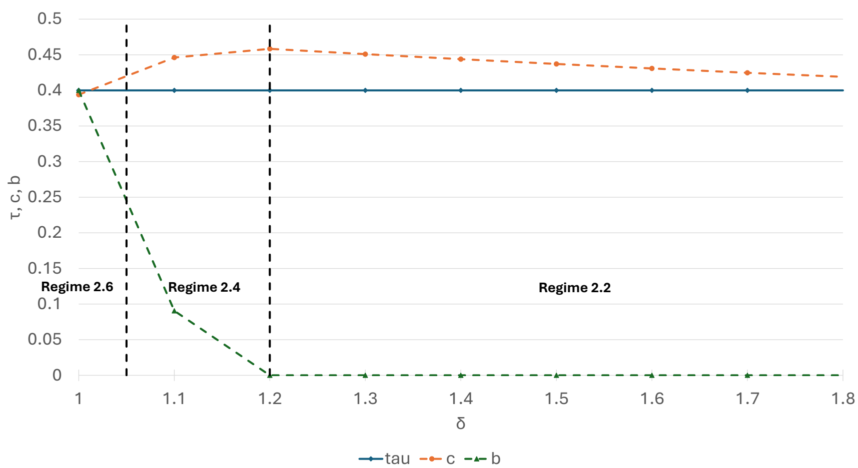
Figure 2 shows in more detail the optimal policy as the MCPF (and thus the demand for tax revenue) increases, for both high ($\bar{\tau} = 0.8$, where we effectively move down column 4 of Figure 1) and low ($\bar{\tau} = 0.4$). At $\bar{\tau} = 0.8$, we first have an instance of Regime 1.3, reflecting the first line of Figure 1, with $b = \tau$ and $c = 0$. Over a range of higher values of δ, we are in Regime 2.4, where both are used, but with a declining b, and greater use of c as δ increases. At very high values of δ, we reach Regime 2.2, where only the TC is used, and $b = 0$.

A slightly different pattern emerges for $\bar{\tau} = 0.4$, where the lower constraint on τ leads to lower revenue. In this case, additional subsidy is required beyond setting $b = \tau$ so that c is set relatively high. There is a smaller range of values for δ in Regime 2.4 with a declining b and rising c, as the effect of the lower τ bites earlier, reducing b to zero at $\delta = 1.2$. Beyond that point, in Regime 2.2, the higher δ also forces a decline in c.

Figure 2: Optimal Policy as the MCPF Increases



(a) $\bar{\tau} = 0.8$

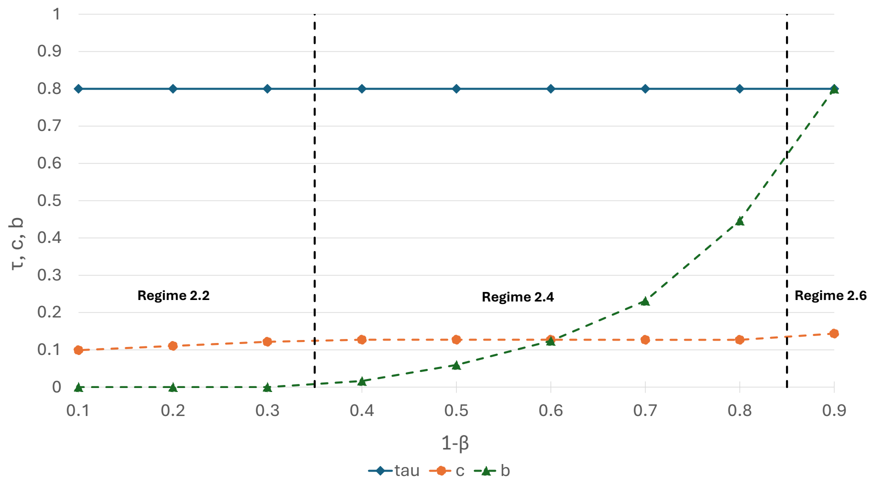


(b) $\bar{\tau} = 0.4$

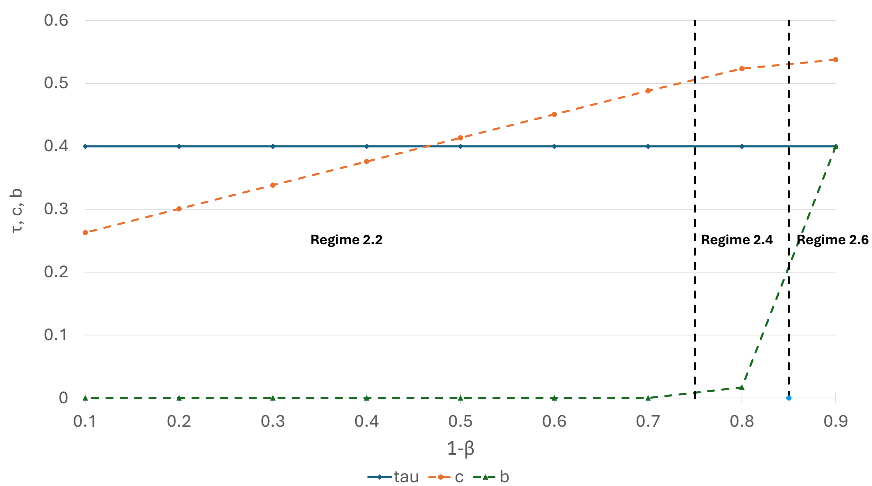
Parameters: $\alpha = 0.8, \phi = 0.7, \varepsilon = 5,\ a = 5,\ \beta$=0.4, and varying δ.

The effects of varying $1 - \beta$ are illustrated in Figure 3 for the case of $\delta = 1.3$ and both $\bar{\tau} = 0.8$ and $\bar{\tau} = 0.4$. In this case, given the relatively high value of δ, for all values of β, and both values of $\bar{\tau}$, the incentive to collect revenue from the final goods firm and intermediate firms raises τ to its maximum, resulting in Regime 2. For both values of $\bar{\tau}$, as $1-\beta$ rises, we move from Regime 2.2 with $b = 0$ and $c > 0$ and rising, to Regime 2.4 with both $b > 0$ and $c > 0$, to Regime 2.6 where b reaches its upper limit, $b = \tau$. The greater use of both b and c as $1 - \beta$ rises reflects the greater externality, and hence greater need for subsidy of R&D. For the case of the lower $\bar{\tau}$, much greater use is made of c, while b is used only at very low values of $1 - \beta$.

Figure 3: Optimal Policy as the Spillover $(1-\beta)$ Increases



(a) $\bar{\tau} = 0.8$



(b) $\bar{\tau} = 0.4$

Parameters: $\alpha = 0.8, \phi = 0.7, \varepsilon = 5,\ a = 5,\ \delta$=1.3, and varying $1-\beta$.

Despite the very significant difference in the constraint on τ in these analyses, the degree of subsidy to R&D is similar. This is measured by p_l in equation (10). This is equal to 1 in the absence of government subsidy: $b = c = 0$. With $\delta = 1.3$, $\beta = 0.4$ and $\bar{\tau} = 0.8$, the optimal use of b and c reduces p_l to 0.28. Changing $\bar{\tau}$ to 0.4 results in p_l being set optimally at 0.31.

6 Income and Cost Shifting

We now assume that firm i is an integrated firm with two subsidiaries: an intermediate-good subsidiary, which sells the input to the final-good firm, and an R&D subsidiary, which undertakes innovation. This affects the size of the externality from R&D, since the integrated firm captures the profit of both the R&D subsidiary and the intermediate subsidiary. It also introduces two possible routes for reducing the firm's tax liability. First, the firm can shift income into the R&D subsidiary in order to benefit from the patent box rate reduction b. Second, it can relabel some costs of the intermediate subsidiary as R&D labor costs in order to benefit from the tax credit c. Hall and Reenen (2000) offered many examples of relabelling, and more recently Chen et al. (2021) documented that around 30% of reported R&D investment by Chinese firms could be due to relabeling.

Following our previous analysis, and the large literature on income-shifting, we assume that there is an arm's-length price for the license of the innovation P_i. The firm can set the price paid to the R&D subsidiary at some $P_i + d_i$, where d_i denotes the deviation from this price but at a convex cost $\frac{d_i^2}{2P_i\rho_i}$, thereby shifting income d_i from the intermediate-good subsidiary to the R&D subsidiary. Similarly, by incurring the convex cost $\frac{z_i^2}{2l_i\theta_i}$, the firm can relabel z_i of intermediate-subsidiary costs as R&D labor expenses, so that the claimed R&D labor cost becomes $l_i + z_i$.

Then, consolidated expected profits of the intermediate and R&D subsidiaries of firm i together are

$$\begin{aligned} E(\pi_i^I + \pi_i^{RD}) =& (1-\tau)(\underline{\pi}_i^I + \sigma_i\Delta_i^I - \sigma_i(P_i + d_i) + z_i) \\ &+ (1-\tau+b)\sigma_i(P_i+d_i) - (1-\tau-c)(l_i+z_i) - e_i \\ &- \sigma_i\frac{(d_i)^2}{2P_i\rho_i} - \frac{(z_i)^2}{2l_i\theta_i} \end{aligned} \tag{21}$$

where the first line gives the declared profit of the intermediate subsidiary, the second line the declared profit of the R&D subsidiary, and the last line the shifting costs.

This simplifies to:

$$\begin{aligned} E(\pi_i^I + \pi_i^{RD}) =& (1-\tau)(\underline{\pi}_i^I + \sigma_i\Delta_i^I) + \sigma_i(P_i+d_i)b + cz_i \\ &- l_i(1-\tau-c) - e_i - \sigma_i\frac{(d_i)^2}{2P_i\rho_i} - \frac{(z_i)^2}{2l_i\theta_i} \end{aligned} \tag{22}$$

Note that in this formula, z_i does not enter the normal consolidated tax base as the costs are just transferred from one subsidiary to another - the firm just gets additional profit cz_i from the tax credit.

So, from (21), the FOC for maximization of consolidated profit w.r.t. l_i, e_i, d_i and z_i are

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i}\left(\Delta_i^I(1-\tau)+(P_i+d_i)b-\frac{d_i^2}{2P_i\rho_i}\right)=1-\tau-c-\frac{(z_i)^2}{2\theta_i l_i^2} \tag{23}$$

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial e_i}\left(\Delta_i^I(1-\tau)+(P_i+d_i)b-\frac{d_i^2}{2P_i\rho_i}\right)=1 \tag{24}$$

$$d_i=P_i\rho_i b \tag{25}$$

$$z_i=l_i\theta_i c \tag{26}$$

Note that (25) says that the percentage share of the royalty payment shifted is $\frac{d_i}{P_i}=\rho_i b$ so it depends on the cost of shifting and tax benefit from doing so. Also, (26) says that the percentage share of the labor cost shifted is $\frac{z_i}{l_i}=\theta_i c$ so it also depends on the cost of shifting and tax benefit from doing so.

Substituting (25),(26) back in to (24),(23) we see:

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i}\left(\Delta_i^I(1-\tau)+bP_i+\frac{\rho_i b^2}{2}P_i\right)=1-\tau-c-\frac{\theta_i c^2}{2} \tag{27}$$

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial e_i}\left(\Delta_i^I(1-\tau)+bP_i+\frac{\rho_i b^2}{2}P_i\right)=1. \tag{28}$$

So, the effective cost of a unit of labor is $1-\tau-c-\frac{\theta_i c^2}{2}$. As the cost of labor cannot be negative, we will impose

$$1-\tau-c-\frac{\theta_i c^2}{2}\geq 0 \tag{29}$$

as an additional condition on the design of government policy. We can divide through in (27), (28) to get:

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i}=\frac{1-\tau-c-\frac{\theta_i c^2}{2}}{\Delta_i^I(1-\tau)+bP_i+\frac{\rho_i b^2}{2}P_i}=p_l^i \tag{30}$$

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial e_i}=\frac{1}{\Delta_i^I(1-\tau)+bP_i+\frac{\rho_i b^2}{2}P_i}=p_e^i \tag{31}$$

Equations (30),(31) say that the return to the inputs l_i, e_i are equal to *personalized input prices* p_l^i, p_e^i. Clearly, as long as parameters $\rho, \theta, P, \Delta^I$ differ across i, any change in b or c is going to change the personalized input prices for i, j by different amounts, so that the argument of Section 5.4 above will not carry though in general. So, we make the following assumption:

A4. $\rho_i = \rho,\ \theta_i = \theta,\ P_i = P,\ a_i = a,\ \overline{\phi}_i = 1,\ \underline{\phi}_i = \phi < 1.$

The last part, i.e., $a_i = a,\ \overline{\phi}_i = 1,\ \underline{\phi}_i = \phi < 1$ ensures $\overline{\pi}_i^I - \underline{\pi}_i^I = \Delta^I,\ \ \overline{\pi}_i - \underline{\pi}_i = \Delta$, where $\Delta > \Delta^I$. Then it is easy to check that given A3, $p_l^i = p_l, p_e^i = p_e$ in (30),(31). However, the innovation functions σ_i can still vary across R&D subsidiaries, so the model is not required to be completely symmetric.

Next, our expression for tax revenue is:

$$
\begin{aligned}
T &= \tau \sum_i (\underline{\pi} + \Delta\sigma_i - \sigma_i(P + d_i) - l_i) + (\tau - b)\sum_i \sigma_i(P + d_i) - c\sum_i (l_i + z_i) \\
&= \tau \sum_i (\underline{\pi} + \Delta\sigma_i - l_i) - \sum_i \sigma_i \left(bP + \rho P b^2\right) - \sum_i (c + c^2\theta) l_i
\end{aligned}
\tag{32}
$$

where in the second line, we use (25), (26) to substitute out d_i, z_i. The maximand for the government is therefore

$$
\begin{aligned}
W = &\sum_{i=1}^{n} \left(\underline{\pi} + \Delta\sigma_i + \xi P \rho \frac{b^2}{2}\sigma_i + \frac{\xi\theta c^2}{2} l_i\right) \\
&+ (\delta - 1)\left(\tau \sum_i (\underline{\pi} + \Delta\sigma_i - l_i) - \sum_i \sigma_i \left(bP + \rho P b^2\right) - \sum_i (c + c^2\theta) l_i\right) - \sum_i (l_i + e_i)
\end{aligned}
\tag{33}
$$

Here, ξ captures the welfare treatment of the shifting costs. If the cost of shifting is a fine paid to government, the fine costs the firm \$1 but also raises tax revenue by \$1, which is worth \$$\delta$ to the government. The net welfare effect is therefore \$$(\delta - 1)$, so in this case we set $\xi = \delta - 1$. If instead these costs are true resource costs, they reduce welfare and we set $\xi = -1$.

In (33), recall that l_i, e_i depend on p_l^i, p_e^i via (30),(31). The problem for the government is then to choose τ, b, c to maximize (33) subject to (30),(31) and (3) and (29).

We can now show that Propositions 2 and 3 carry over to the case of income and cost shifting within a unified firm when the costs of shifting are pure transfers to the governent. They also carry over with minor qualifications when the costs of shifting are a pure resource loss. The first result is the analog of Proposition 2:

Proposition 5. *Assume no unobservable input, i.e.* $e = 0$, *and that A1,A2,A3, A4 hold. If shifting costs are a pure transfer to the government* ($\xi = \delta - 1$), *then* $b^* = 0$, *i.e. the PB should never be used to incentivise R&D investment. If shifting costs are resource costs*

($\xi = -1$), the same is true if

$$\frac{\sum_i \sigma_i}{\sum_i \frac{\partial \sigma_i}{\partial l_i} l_i} > \frac{1}{\delta - 1}\left(\frac{\delta(1+2\theta)}{1+\theta} - 1\right) \tag{34}$$

i.e. the innovation function is sufficiently concave at the optimum and/or the cost of cost-shifting is relatively high (θ small).

One implication of Proposition 5 is that if cost-shifting is impossible, i.e. $\theta = 0$, then (34) always holds from A1,A2, with only income-shifting - it does not matter whether shifting costs are transfer or resource costs; in both cases, the TC dominates the PB as an incentive instrument, and hence the PB should never be used. More generally, condition (34) can be checked for any example. For instance, consider identical firms with innovation function $\sigma(l) = 1 - \exp(-l^{\alpha})$, $0 < \alpha < 1$. It is easily calculated that for this example, (34) holds if $\delta - 1 > \alpha\left(\frac{\delta(1+2\theta)}{1+\theta} - 1\right)$.[31]

This result is comparable to Proposition 2. The intuition is the following. If there is little or no cost shifting i.e. θ small, and the innovation functions are concave enough, the basic intuition of reform 1 in Section 5.2 applies. This is reinforced by the fact that income-shifting adds to the revenue cost of the PB, and also may impose real costs on the economy.

This result is also related to Haufler and Schindler (2023), Proposition 1, which considers a tax credit vs a patent box when governments coordinate, which is similar to the one-country case studied here. Their model has no cost shifting and income shifting has a real resource cost. Haufler and Schindler (2023) find that here the patent box is actually negative i.e. $b^* < 0$. This is consistent with our result that $b^* = 0$ as we constrain b to be non-negative.

Proposition 6. *Assume A1,A2,A3, A4 hold, and the cost and income shifting activities are pure transfers to the government i.e. $\xi = \delta - 1 > 0$. Moreover, suppose that at the solution for the government's problem, there is an interior solution for τ i.e. $\frac{\partial W}{\partial \tau} = 0$. Then either (i) $b, c = 0$; (ii) there is an interior solution for b i.e. $0 < b < \tau$, and $c = 0$, or (iii) $b = \tau$ and $c \geq 0$. The same result holds if the costs of shifting are a true resource cost ($\xi = -1$) and if the cost of income shifting is high enough (ρ small).*[32]

This result is comparable to Proposition 3, with the difference that it is possible that

[31] This is well-behaved: it is strictly increasing and concave in l, with $\sigma(0) = 0$ and $\sigma \to 1$ as $l \to \infty$. Then it is easily calculated that the LHS of (34) is bounded below by its value at $l = 0$. Applying L'Hopital's rule, $\lim_{l \to 0} \frac{\sigma}{\sigma' l} = \frac{1}{\alpha}$.

[32] There are other sufficient conditions for the result. In the Proof of this Proposition it is shown, for example, that if $\delta \geq \varepsilon$, the result goes through.

$c > 0$ in part (iii). Compared to Proposition 3, income shifting raises the revenue cost of increasing b. If the costs of cost shifting are sufficiently high, then using the TC to stimulate R&D may have lower revenue cost.

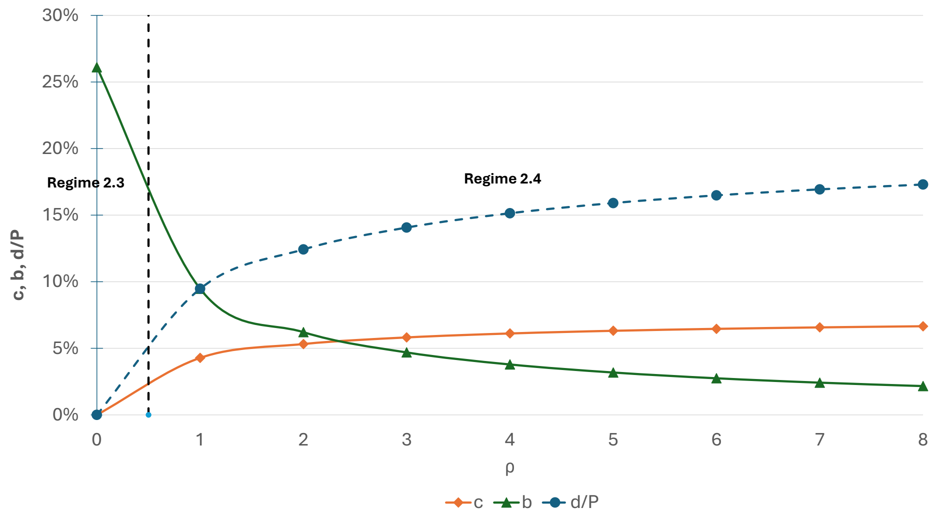
6.1 Simulation of the Shifting Model

We now present results from our numerical simulation model for the case of income and cost shifting. The possible regimes are exactly the same as shown in Table 2 for the baseline case.[33] In Figure 4 we explore the impact of a fall in the cost of income shifting, represented by a rise in ρ, setting $\theta = 0$ thereby preventing any cost shifting. In Figure 5 we explore the impact of a fall in the cost of cost shifting, represented by a rise in θ, setting $\rho = 0$ thereby preventing any income shifting. The values chosen for the other parameters are set out in the two Figures.

Figure 4 traces the increase in income shifting (d/P) as ρ increases from 0 to 8, representing a gradual fall in the cost of such shifting. Starting from a case of no income shifting, $\rho = 0$, for the set of chosen parameters, $b > 0$ and $c = 0$. As ρ rises, income shifting to take advantage of the high PB becomes more attractive. However, as a result the optimal response of the government is to reduce b in order to make income shifting less attractive. Recalling that $d/P = b\rho$, the higher is ρ the lower the b the government is forced to set. That reduces the incentives for R&D investment, which the government partially corrects by raising c, promoting the use of the TC over the PB. With the parameters and the range of values of ρ in the figure, income shifting continues to rise with ρ, even as b falls.

[33] The online supplementary appendix provides a more complete discussion of the KKT and MCP formulation, and the simulated model.

Figure 4: Optimal Policy as ρ Increases



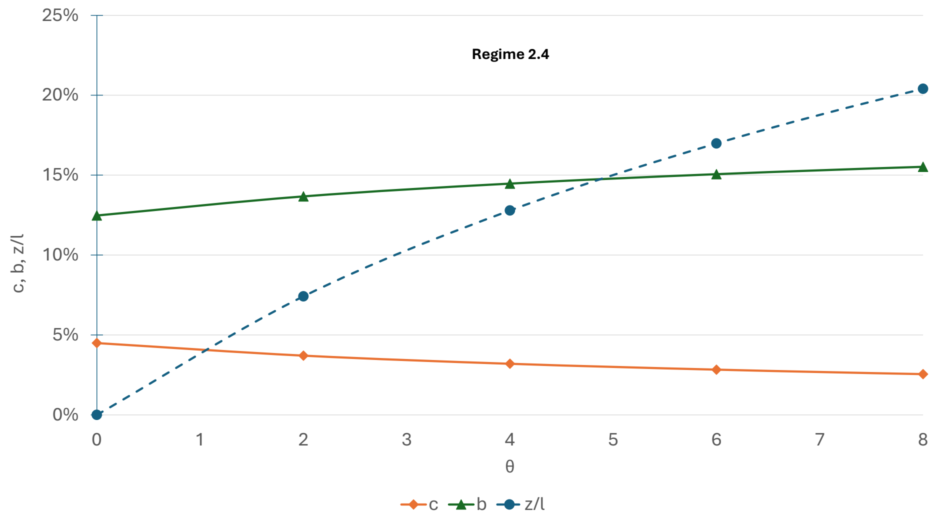
Parameters: $\alpha = 0.6, \beta = 0.5, \phi = 0.7, \varepsilon = 5,\ a = 5,\ \overline{\tau} = 0.8,\ \beta = 0.5,\ \xi = -1,\ \theta = 0,\ \delta = 1.2.$

Figure 5 presents a similar analysis for falling costs of cost shifting, as θ rises. We choose a slightly different set of parameter values for this case. That is because we would like to observe what happens to the optimal c as these costs fall, there is more cost shifting, and so the TC becomes more expensive. If $c = 0$ when there is no cost shifting, then introducing lower costs of shifting would be irrelevant as c would remain at zero. We therefore choose a set of parameters that yield $c > 0$ in the absence of shifting.[34] In this simulation, it is also the case that $b > 0$.

The effects of raising θ are as expected. Since shifting costs are lower, cost shifting (z/l) increases. To offset this, the government reduces c, given that $z/l = \theta c$. In this case, the government then relies more heavily on the PB, raising b at the same time as reducing c.

[34]Specifically, we increase α from 0.6 to 0.7, raising the (observed) labor share of R&D inputs, and making c a more effective subsidy for R&D.

Figure 5: Optimal Policy as θ Increases



Parameters: $\alpha = 0.7, \beta = 0.5, \phi = 0.7, \varepsilon = 5,\ a = 5,\ \overline{\tau} = 0.8,\ \beta = 0.5,\ \xi = -1,\ \rho = 0,\ \delta = 1.2.$

Of course, as both ρ and θ rise, then costs of both income and cost shifting fall. The effects on the optimal choice of b and c depend on the relative values of these two costs, as well as the other parameters in the model.

7 Other Extensions

7.1 Multiple Inputs

So far, we have assumed that there is only one input that is observable and can therefore be given a tax credit. However, the analysis generalizes straightforwardly to multiple observable inputs. Suppose that firm i has a vector of M observable inputs $\mathbf{l}_i = (l_i^1, ..l_i^M)$. Then we need to assume that r_i is strictly concave in $(\ \mathbf{l}_i, e_i)$. Given this, Propositions 1,2 and 3 continue to hold, although the specific conditions for $c > 0$ in Proposition 2 and $b > 0$ in Proposition 3 need to be modified. In particular, these conditions need to account for the effect of c, b on all of the observable inputs $\mathbf{l}_i$. So, for example, (19) becomes

$$(1 + (\delta - 1)\overline{\tau}) \sum_i \sum_j S_i \frac{\partial r_i}{\partial l_i^j} \frac{\partial l_i^j}{\partial c} > (\delta - 1) \sum_i \sum_j l_i^j \tag{35}$$

The condition (20) can be modified in the same way.

7.2 Incentive Costs of Raising the Rate of Corporate Income Tax

So far, we have assumed for clarity that there is an upper bound on τ, $\overline{\tau}$, that may be less than unity. This is to capture in a very stylized way that there may be negative effects on pre-tax profit and tax revenue from raising τ too high. However, the implicit assumption is that these negative effects are zero up to the threshold $\overline{\tau}$, and infinite (or at least, very high) thereafter. Here, we consider an extension where the incentive effect of τ on pre-tax profit is explicitly modeled. We do this in the simplest possible way. Suppose that the final good firm can increase profit by an additive term $\widehat{\Delta}e_F$ by exerting managerial effort e_F. The cost of this effort, which is quadratic for convenience, cannot be deducted from the tax base, so e_F maximizes $(1-\tau)\widehat{\Delta}e_F - \frac{(e_F)^2}{2}$, giving $e_F = (1-\tau)\widehat{\Delta}$. In other words, a higher corporate tax discourages effort by the manager. At optimal effort, the additional gross profit net of the cost of effort is $\widehat{\Delta}e_F - \frac{(e_F)^2}{2} = (1-\tau)(\widehat{\Delta})^2 - \frac{1}{2}(1-\tau)^2(\widehat{\Delta})^2$. The additional tax revenue is $\tau(1-\tau)(\widehat{\Delta})^2$. So, the overall addition to welfare from allowing for managerial effort is

$$(1-\tau)(\widehat{\Delta})^2 - \frac{1}{2}(1-\tau)^2(\widehat{\Delta})^2 + (\delta-1)\tau(1-\tau)\widehat{\Delta} \tag{36}$$

Now consider tax reform 2 as described in Proposition 1 i.e. $d\tau = db = -dc > 0$. Recall that the prices p_l, p_e do not change, so inputs l_i, e_i do not change either. So, the effect of this on welfare is (A.2) in the Appendix plus the derivative of the expression in (36) with respect to τ i.e.

$$dW = (\delta-1)\left(\sum_i(\underline{\pi}_i + \sigma_i\Delta_i) - \sum_i \sigma_i\beta\Delta_i^I\right)d\tau + \left(-\tau(\widehat{\Delta})^2 + (\delta-1)(1-2\tau)(\widehat{\Delta})^2\right)d\tau$$

Now define condition C to be

$$(\delta-1)\left(\sum_i(\underline{\pi}_i + \sigma_i\Delta_i) - \sum_i \sigma_i\beta\Delta_i^I\right) > (\widehat{\Delta})^2\left(\tau - (\delta-1)(1-2\tau)\right)$$

So, there are two possibilities. The first is that condition C holds at the optimal values τ^*, b^*, c^*. In this case, $c^* = 0$ at the optimum. The reason is simply that if $c^* > 0$, the reform $d\tau = db = -dc > 0$ is feasible and welfare-improving, a contradiction. So, we conclude that if C holds at the optimum, the TC should never be used. This is the extension of Proposition 3 to this case. The intuition is clear - condition C says that the welfare cost of an increase in τ on the effort incentive of the final goods manager is low (or even negative), relative to the baseline effect of the reform on tax revenue. In turn, this will be the case if $\widehat{\Delta}$ is small.

8 Conclusions

This paper has analyzed the relative cost effectiveness of R&D tax credits and PBs as instruments for stimulating R&D. We set up a model in which there are three types of market failures, all of which lead to under-investment in R&D in the absence of government intervention, and the possibility of several organisational forms, giving rise to hold-up or income and cost shifting behaviours. Assuming that the marginal cost of public funds exceeds 1, in the absence of any unobservable input to R&D firms, we find that TCs are unambiguously more tax-cost-effective than PBs in achieving the same impact on the level of R&D.

By contrast, when there is an unobservable input to R&D (e.g. managerial effort) that cannot be subsidized by a TC, we find that the optimal policy depends on the need for tax revenue, as measured by the MCPF. When this is low, the TC should *never* be used, but a PB may be, but as need rises, both instruments should be used, and when it is very high, only a TC should be used.

References

Akcigit, Ufuk, Douglas Hanley, and Stefanie Stantcheva. 2022. "Optimal Taxation and R&D Policies." *Econometrica*, 90(2): 645–684.

Alstadsæter, Annette, Salvador Barrios, Gaetan Nicodeme, Agnieszka Maria Skonieczna, and Antonio Vezzani. 2018. "Patent boxes design, patents location, and local R&D." *Economic Policy*, 33: 131–177.

Becker, Bettina. 2015. "Public R&D Policies and Private R&D Investment: A Survey of the Empirical Evdience." *Journal of Economic Surveys*, 29(5): 917–942.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica*, 81(4): 1347–1393.

Bloom, Nicholas, Rachel Griffith, and John Van Reenen. 2002. "Do R&D tax credits work? Evidence from a panel of countries 1979–1997." *Journal of Public Economics*, 85(1): 1–31.

Bornemann, Tobias, Stacie Kelley Laplante, and Benjamin Osswald. 2023. "The Effect of Intellectual Property Boxes on Innovative Activity Effective Tax Rates." *Journal of the American Tax Association*, 45(2): 7–33.

Bronzini, Raffaello, and Paolo Piselli. 2016. "The impact of R&D subsidies on firm innovation." *Research Policy*, 45(2): 442–457.

Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi. 2021. "Notching R&D Investment with Corporate Income Tax Cuts in China." *American Economic Review*, 111(7): 2065–2100.

Choi, Joonhwan, and Jaegul Lee. 2017. "Repairing the R&D market failure: Public RD subsidy and the composition of private R&D." *Research Policy*, 46(8): 1465–1478.

Evers, Lisa, Helen Miller, and Christoph Spengel. 2015. "Intellectual Property Box Regimes: Effective tax rates and tax policy considerations." *International Tax and Public Finance*, 22: 502–530.

Gaessler, Fabian, Bronwyn H. Hall, and Dietmar Harhoff. 2021. "Should there be lower taxes on patent income?" *Research Policy*, 104129: 1–15.

Gonzalez Cabral, A.C., P. O'Reilly, S. Appelt, F. Galindo-Rueda, and T. Hanappi. 2023*a*. "Design features of income-based tax incentives for RD and innovation." *OECD Taxation Working Papers*, 60.

Gonzalez Cabral, A.C., P. O'Reilly, S. Appelt, F. Galindo-Rueda, T. Hanappi, and M. Bucci. 2023*b*. "A time series perspective on income-based tax support for RD and innovation." *OECD Taxation Working Papers*, 62.

Gonzalez Cabral, A.C., S. Appelt, and T. Hanappi. 2021. "Corporate effective tax rates for R&D: The case of expenditure-based RD tax incentives." *OECD Taxation Working Papers*, 54.

Gonzalez Cabral, A.C., T. Hanappi, , S. Appelt, F. Galindo-Rueda, and P. O'Reilly. 2023*c*. "Effective tax rates for RD intangibles." *OECD Taxation Working Papers*, 63.

Griffith, Rachel, Helen Miller, and Martin O'Connell. 2014. "Ownership of intellectual property and corporate taxation." *Journal of Public Economics*, 112: 12–23.

Guceri, Irem, and Li Liu. 2019. "Effectiveness of Fiscal Incentives for R&D: Quasi-experimental Evidence." *American Economic Journal: Economic Policy*, 11(1): 266–291.

Hall, Bronwyn, and John Van Reenen. 2000. "How effective are fiscal incentives for RD? A review of the evidence." *Research Policy 29 2000 449–469*, 29: 449–469.

Haufler, Andreas, and Dirk Schindler. 2023. "Attracting profit shifting or fostering innovation? On patent boxes and R&D subsidies." *European Economic Review*, 104446: 1–28.

Hud, Martin, and Katrin Hussinger. 2015. "The impact of R&D subsidies during the crisis." *Research Policy*, 44(10): 1844–1855.

Jones, Charles, and John C Williams. 1998. "Measuring the social return to R&D." *Quarterly Journal of Economics*, 113: 1119–1135.

Jones, Charles, and John C Williams. 2000. "Too much of a good thing? The economics of investment in R&D." *Journal of Economic Growth*, 5: 65–85.

Jones, Charles I. 1995. "R & D-based models of economic growth." *Journal of political Economy*, 103(4): 759–784.

Lucking, Brian, Nicholas Bloom, and Jon Van Reenen. 2019. "Have R&D spillovers declined in the 21st century?" *Fiscal Studies*, 40(4): 561–590.

OECD. 2015. "BEPS Action 5: Agreement on Modified Nexus Approach for IP Regimes."

Romer, Paul. 1990. "Endoegenous Technical Change." *Journal of Political Economy*, 98(5): S71–S102.

Rowe-Brown, Max, and Huw James. 2019. "Patent Box Evaluation." *HMRC Evaluation Report*.

Ugur, Mehmet, Sefa Awaworyi Churchill, and Hoang M. Luong. 2020. "What do we know about RD spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power." *Research Policy*, 49(1): 103866.

Appendix

A Proofs of Propositions

Proof of Proposition 1. (i) Consider a cut in b of amount $-db < 0$, with an offsetting increase in c that leaves p_l unchanged. From (10), it is easily established that the required change in c is

$$dc = \frac{1-\tau-c}{1-\tau+b} db$$

As p_l does not change, l does not change either. So, from (17), the effect on tax revenue is simply the first-round effect

$$dT = db \sum_i \sigma_i \beta \Delta_i^I - db \frac{1-\tau-c}{1-\tau+b} \sum_i l_i$$

where the first term is the tax savings from a lower b and the second is the tax cost of a compensating increase in c. So, this tax reform increases tax revenue (and therefore welfare) iff $dT > 0$, which requires

$$\sum_i \sigma_i \beta \Delta_i^I - \frac{1-\tau-c}{1-\tau+b} \sum_i l_i > 0$$

But, using (10) to substitute out $\frac{1-\tau-c}{1-\tau+b} = \beta \Delta_i^I \frac{\partial \sigma_i}{\partial r_i} \frac{\partial r_i}{\partial l_i}$, we require:

$$\sum_i \left(\sigma_i - \frac{\partial \sigma_i}{\partial r_i} \frac{\partial r_i}{\partial l_i} l_i \right) \beta \Delta_i^I > 0 \tag{A.1}$$

Now note that σ_i is a strictly concave function of l_i by A1 and the fact that it is the only input. Thus, for any i, $\sigma_i > \frac{\partial \sigma_i}{\partial r_i} \frac{\partial r_i}{\partial l_i} l_i$ and so (A.1) must hold.

(ii) Consider a small change in policy $d\tau = db = -dc > 0$. Then from (10), the prices p_l, p_e do not change, so inputs l_i, e_i do not change either. Then, by inspection of (18), it is easily verified that

$$\begin{aligned} \frac{dW}{\delta - 1} &= d\tau \sum_i (\underline{\pi}_i + \sigma_i \Delta_i - l_i) - db \sum_i \sigma_i \beta \Delta_i^I - dc \sum_i l_i \\ &= d\tau \left(\sum_i (\underline{\pi}_i + \sigma_i \Delta_i) - \sum_i \sigma_i \beta \Delta_i^I \right) > 0 \end{aligned} \tag{A.2}$$

where the second line follows as $\beta < 1, \Delta_i > \Delta_i^I$. $\square$

Proof of Proposition 2. It remains to establish the condition for $c > 0$. If $b = c = 0$, from (10), $p_l \equiv 1$. Then, τ is non-distortionary, and so it is set at its maximum value $\overline{\tau}$. At $b = 0, \tau = \overline{\tau}$, the effect of a change in c on W is:

$$
\begin{aligned}
\frac{\partial W}{\partial c} &= \frac{\partial W}{\partial p_l}\frac{\partial p_l}{\partial c} - (\delta - 1)\sum_i l_i \qquad \text{(A.3)}\\
&= ((\delta - 1)\overline{\tau} + 1)\sum_i \left(\Delta_i \frac{\partial \sigma_i}{\partial r_i} + s_i\right)\frac{\partial r_i}{\partial p_l}\frac{\partial p_l}{\partial c} - ((\delta - 1)\overline{\tau} + 1)\sum_i \frac{\partial l_i}{\partial p_l}\frac{\partial p_l}{\partial c}\\
&- (\delta - 1)\sum_i l_i\\
&= (1 + (\delta - 1)\overline{\tau})\sum_i \left[\left(\Delta_i \frac{\partial \sigma_i}{\partial r_i} + s_i\right)\frac{\partial r_i}{\partial p_l} - \beta\Delta_i^I \frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i}\frac{\partial l_i}{\partial p_l}\right]\frac{\partial p_l}{\partial c}\\
&- (\delta - 1)\sum_i l_i\\
&= (1 + (\delta - 1)\overline{\tau})\sum_i S_i \frac{\partial r_i}{\partial l_i}\frac{\partial l_i}{\partial c} - (\delta - 1)\sum_i l_i
\end{aligned}
$$

In the second line, we have used $\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r}{\partial l_i}\beta\Delta_i^I = 1$ from (10) at $c = b = 0$. In the third line, we have used $\frac{\partial r_i}{\partial p_l}\frac{\partial p_l}{\partial c} = \frac{\partial r_i}{\partial l_i}\frac{\partial l_i}{\partial p_l}\frac{\partial p_l}{\partial c}$ as l_i is the only input that varies with p_l, and we also define $\frac{\partial l}{\partial c} = \frac{\partial l_i}{\partial p_l}\frac{\partial p_l}{\partial c}$. So, $c > 0$ iff the term in the last line of (A.3) is strictly positive. $\square$

Proof of Proposition 3. It remains to prove that (20) is a necessary and sufficeint condition for $b > 0$. Given that W is strictly concave in b, $b > 0$ iff $\frac{\partial W}{\partial b}\mid_{b=0} > 0$. But from (18), we have:

$$
\frac{\partial W}{\partial b}\mid_{b=0} = -(\delta - 1)\sum_i \sigma_i \beta\Delta_i^I + \sum_i \left(\frac{\partial W}{\partial l_i}\frac{\partial l_i}{\partial b} + \frac{\partial W}{\partial e_i}\frac{\partial e_i}{\partial b}\right) \qquad \text{(A.4)}
$$

Also, note that evaluating (10) at $b = c = 0$, we have

$$
\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i}\beta\Delta_i^I = 1, \;\; \frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial e_i}\beta\Delta_i^I = \frac{1}{1 - \tau} \qquad \text{(A.5)}
$$

Then again by inspection of (18),

$$
\begin{aligned}
\frac{\partial W}{\partial l_i} &= (1 + (\delta - 1)\tau)\left((\Delta_i \frac{\partial \sigma_i}{\partial r_i} + s_i)\frac{\partial r_i}{\partial l_i} - 1\right) \qquad \text{(A.6)}\\
&= (1 + (\delta - 1)\tau)\left(\Delta_i \frac{\partial \sigma_i}{\partial r_i} + s_i - \beta\Delta_i^I \frac{\partial \sigma_i}{\partial r_i}\right)\frac{\partial r_i}{\partial l_i}\\
&= (1 + (\delta - 1)\tau)\, S_i \frac{\partial r_i}{\partial l_i}
\end{aligned}
$$

where the second line follows from (A.5), and the third line from the definition of S_i In the same way, we can write;

$$\begin{aligned}\frac{\partial W}{\partial e_i} &= (1+(\delta-1)\tau)\left(\left(\Delta_i\frac{\partial \sigma_i}{\partial r_i}+s_i\right)\frac{\partial r_i}{\partial e_i}\right)-1 \\ &= (1+(\delta-1)\tau)\left(\Delta_i\frac{\partial \sigma_i}{\partial r_i}+s_i-\frac{(1-\tau)\beta\Delta_i^I}{1+(\delta-1)\tau}\frac{\partial \sigma_i}{\partial r_i}\right)\frac{\partial r_i}{\partial e_i} \\ &= (1+(\delta-1)\tau)\,S_i'\frac{\partial r_i}{\partial e_i}\end{aligned} \tag{A.7}$$

Then, substiuting (A.6), (A.7 back into (A.4), we have

$$\frac{\partial W}{\partial b}\,|_{b=0} = -(\delta-1)\sum_i \sigma_i\beta\Delta_i^I+(1+(\delta-1)\tau)\left(\sum_i S_i\frac{\partial r_i}{\partial l_i}\frac{\partial l_i}{\partial b}+\sum_i S_i'\frac{\partial r_i}{\partial e_i}\frac{\partial e_i}{\partial b}\right)$$

Setting this to be strictly positive gives (20). □

Proof of Proposition 4. With perfect complements, the firm effectively uses one input $x_i \equiv l_i = \eta e_i$ with price

$$p_x = p_l + \frac{p_e}{\eta} = \frac{1+1/\eta-\tau-c}{1-\tau+b} \tag{A.8}$$

Consider a cut in b of amount $-db<0$, with an offsetting increase in c that leaves p_x unchanged. It is easy to check that this requires $dc=\frac{1+1/\eta-\tau-c}{1-\tau+b}db$. Call this tax reform 1*. Moreover, as p_x does not change, x does not change either. So, from (17), the effect on tax revenue is

$$dT = db\sum_i \sigma_i\beta\Delta_i^I - db\frac{1+1/\eta-\tau-c}{1-\tau+b}.\sum_i l_i \tag{A.9}$$

Moreover, the first-order condition for the optimal choice of x_i is

$$\Delta_i^I\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial x_i} = \frac{1+1/\eta-\tau+b}{1-\tau-c} \tag{A.10}$$

Finally, assume σ_i is strictly concave in x_i. Then, combining (A.9), (A.10), it is easy to show that $dT>0$ i.e. this reform raises tax revenue, following the argument of Proposition 1. So, Proposition 1 (i) applies with reform 1* replacing reform 1. Also, it is easy to check that Proposition 1 (ii) continues to hold. So, the argument of Section 5.3 applies. □

Derivatives of the Welfare Functions in the Income and Cost Shifting Case. Note from (30),(31) that, given A3 and A4, l_i and e_i depend on the common input prices

$$p_l = \frac{1-\tau-c-\frac{\theta c^2}{2}}{\Delta^I(1-\tau)+bP+\frac{\rho b^2}{2}P},\ \ p_e = \frac{1}{\Delta^I(1-\tau)+bP+\frac{\rho b^2}{2}P} \tag{A.11}$$

So from (A.11), these prices depend on the tax parameters as follows:

$$\frac{\partial p_l}{\partial \tau} = \frac{p_l \Delta^I - 1}{D},\ \frac{\partial p_e}{\partial \tau} = \frac{p_e \Delta^I}{D},\ \frac{\partial p_l}{\partial b} = -\frac{p_l P(1+\rho b)}{D},\ \frac{\partial p_e}{\partial b} = -\frac{p_e P(1+\rho b)}{D},\ \frac{\partial p_l}{\partial c} = -\frac{1+\theta c}{D},\ \frac{\partial p_e}{\partial c} = 0$$

where $D \equiv \Delta^I(1-\tau) + bP + \frac{\rho b^2}{2}P$. Using these facts, it is straightforward to calculate the following derivatives of W as defined in (33):

$$\frac{\partial W}{\partial \tau} = (\delta - 1)\left(\sum_i (\underline{\pi} + \Delta\sigma_i - l_i) + A_l \frac{(p_l \Delta^I - 1)}{D} + A_e \frac{p_e \Delta^I}{D}\right) + B_l \frac{(p_l \Delta^I - 1)}{D} + B_e \frac{p_e \Delta^I}{D} \tag{A.12}$$

$$\frac{\partial W}{\partial c} = \xi c\theta \sum_i l_i + (\delta - 1)\left(-\sum_i l_i(1+2\theta c) - A_l \frac{1+\theta c}{D}\right) \tag{A.13}$$
$$- B_l \frac{1+\theta c}{D}$$

$$\frac{\partial W}{\partial b} = \xi \sum_{i=1}^{n} \sigma_i P\rho b + (\delta - 1)\left(-\sum_i \sigma_i P(1+2\rho b) - A_l \frac{p_l P(1+\rho b)}{D} - A_e \frac{p_e P(1+\rho b)}{D}\right) \tag{A.14}$$
$$- B_l \frac{p_l P(1+\rho b)}{D} - B_e \frac{p_e P(1+\rho b)}{D}$$

where A_l, A_e are the effects of p_l,p_e on tax revenue and B_l, B_e are the effects on the non-tax components of welfare:

$$A_l = \tau \sum_i \left(\Delta \frac{\partial \sigma_i}{\partial r_i} + s_i\right)\frac{\partial r_i}{\partial p_l} - P\left(b + \rho b^2\right)\sum_i \left(\frac{\partial \sigma_i}{\partial r_i} + \sum_{j\neq i}\frac{\partial \sigma_j}{\partial r_i}\right)\frac{\partial r_i}{\partial p_l} - (\tau + c + c^2\theta)\sum_i \frac{\partial l_i}{\partial p_l} \tag{A.15}$$

$$B_l = \sum_i \left((\Delta + \xi\rho P\frac{b^2}{2})\frac{\partial \sigma_i}{\partial r_i} + s_i + \xi\rho P\frac{b^2}{2}\sum_{j\neq i}\frac{\partial \sigma_j}{\partial r_i}\right)\frac{\partial r_i}{\partial p_l} + \sum_i \frac{\xi\theta c^2}{2}\frac{\partial l_i}{\partial p_l} - \sum_i \left(\frac{\partial l_i}{\partial p_l} + \frac{\partial e_i}{\partial p_l}\right) \tag{A.16}$$

$$A_e = \tau \sum_i \left(\Delta \frac{\partial \sigma_i}{\partial r_i} + s_i\right)\frac{\partial r_i}{\partial p_e} - P\left(b + \rho b^2\right)\sum_i \left(\frac{\partial \sigma_i}{\partial r_i} + \sum_{j\neq i}\frac{\partial \sigma_j}{\partial r_i}\right)\frac{\partial r_i}{\partial p_e} - (\tau + c + c^2\theta)\sum_i \frac{\partial l_i}{\partial p_e} \tag{A.17}$$

$$B_e = \sum_i \left((\Delta + \xi\rho P\frac{b^2}{2})\frac{\partial \sigma_i}{\partial r_i} + s_i + \xi\rho P\frac{b^2}{2}\sum_{j\neq i}\frac{\partial \sigma_j}{\partial r_i}\right)\frac{\partial r_i}{\partial p_e} + \sum_i \frac{\xi\theta c^2}{2}\frac{\partial l_i}{\partial p_e} - \sum_i \left(\frac{\partial l_i}{\partial p_e} + \frac{\partial e_i}{\partial p_e}\right) \tag{A.18}$$

and $s_i = \Delta \sum_{j\neq i} \frac{\partial \sigma_j}{\partial r_i}$, as before.

Proof of Proposition 5. Given A4, note from (27) that when $e_i \equiv 0$, l_i is determined by the condition

$$\frac{\partial \sigma_i}{\partial r_i}\frac{\partial r_i}{\partial l_i} \equiv \frac{\partial \sigma_i}{\partial l_i} = p_l.$$

Thus, conditional on p_l, changes in b, c, τ that affect p_e have no additional effect, i.e. $A_e = B_e = 0$. Setting $A_e = B_e = 0$ in (A.13) and (A.14) gives, after some simplification:

$$\begin{aligned}
\frac{\partial W}{\partial c}\frac{p_l P(1+\rho b)}{1+\theta c} &= \xi c\theta \frac{p_l P(1+\rho b)}{1+\theta c}\sum_i l_i + (\delta-1)\left(-\sum_i l_i(1+2\theta c)\frac{p_l P(1+\rho b)}{1+\theta c} - A_l \frac{p_l P(1+\rho b)}{D}\right) \\
-B_l \frac{p_l P(1+\rho b)}{D} &= C + \frac{\partial W}{\partial b}
\end{aligned}$$

where

$$C = ((\delta-1)\,P(1+2\rho b) - \xi P\rho b)\sum_i \sigma_i + P(1+\rho b)\left(\frac{\xi c\theta}{1+\theta c} - (\delta-1)\frac{(1+2\theta c)}{1+\theta c}\right)\sum_i \frac{\partial \sigma_i}{\partial l_i} l_i \tag{A.19}$$

There are then two cases to consider. First, set $\xi = \delta - 1$ in (A.19). After some simplification, we get:

$$C = (\delta-1)\,P(1+\rho b)\left(\sum_i \sigma_i - \sum_i \frac{\partial \sigma_i}{\partial l_i} l_i\right) > 0$$

where the second line follows from strict concavity of σ_i. As $C > 0$, there are three possibilities. First, $\frac{\partial W}{\partial b}, \frac{\partial W}{\partial c} < 0$. In this case, no PB is used i.e. $b = 0$. Second, $\frac{\partial W}{\partial c} = 0, \frac{\partial W}{\partial b} < 0$. In this case, no PB is used i.e. $b = 0$, and the TC is set at a rate such that $\frac{\partial W}{\partial c} = 0$. Third, $\frac{\partial W}{\partial c} > 0,\ \frac{\partial W}{\partial b} = 0$. In this case, the TC is fully used i.e. $c = 1 - \tau$. But then $p_l < 0$, which is inconsistent with A3. So, we conclude that $b = 0$ when $C > 0$. So, we conclude in this case, $b = 0$.

Second, set $\xi = -1$ in (A.19). After some simplification, we get:

$$C = P(1+\rho b)\left[\left(\frac{\delta(1+2\rho b)}{1+\rho b} - 1\right)\sum_i \sigma_i - \left(\frac{\delta(1+2\theta c)}{1+\theta c} - 1\right)\sum_i \frac{\partial \sigma_i}{\partial l_i} l_i\right] \tag{A.20}$$

For $x = c\theta$ or $x = b\rho$, the term $\frac{\delta(1+2x)}{1+x} - 1$ is positive and increasing in x. So, by inspection of (A.20), C is minimized over feasible values of b, c, τ at $b = 0,\ c = 1 - \tau,\ \tau = 0$. Let this value be $C_{\min}$. Then, (A.20), $C_{\min} > 0$ iff

$$\frac{\sum_i \sigma_i}{\sum_i \frac{\partial \sigma_i}{\partial l_i} l_i} > \frac{1}{\delta-1}\left(\frac{\delta(1+2\theta)}{1+\theta} - 1\right).$$

So in this case, by the argument in the first case, we also conclude $b = 0$. □

Proof of Proposition 6. From (A.14) after some rearrangement, we have:

$$\frac{1}{P(1+\rho b)}\frac{\partial W}{\partial b}=\frac{\xi\rho b-(\delta-1)\,(1+2\rho b)}{1+\rho b}\sum_i\sigma_i-\left((\delta-1)(A_l\frac{p_l}{D}+A_e\frac{p_e}{D})+B_l\frac{p_l}{D}+B_e\frac{p_e}{D}\right) \tag{A.21}$$

Also, from $\frac{\partial W}{\partial \tau}=0$, using (A.12) we can write

$$(\delta-1)A_e\frac{p_e\Delta^I}{D}+B_e\frac{p_e\Delta^I}{D}=-\,(\delta-1)\left(\sum_i\left(\underline{\pi}+\Delta\sigma_i-l_i\right)+A_l\frac{(p_l\Delta^I-1)}{D}\right)-B_l\frac{(p_l\Delta^I-1)}{D} \tag{A.22}$$

Then combining (A.21), (A.22), we get, after some manipulation:

$$\frac{\Delta^I}{P(1+\rho b)}\frac{\partial W}{\partial b}=\Delta^I\frac{\xi\rho b-(\delta-1)\,(1+2\rho b)}{1+\rho b}\sum_i\sigma_i+(\delta-1)\left(\sum_i\left(\underline{\pi}+\Delta\sigma_i-l_i\right)\right)-\left((\delta-1)(A_l\frac{1}{D})+B_l\frac{1}{D}\right) \tag{A.23}$$

Also from (A.13)

$$\frac{1}{1+\theta c}\frac{\partial W}{\partial c}=\frac{\xi c\theta-(\delta-1)(1+2\theta c)}{1+\theta c}\sum_i l_i-\left((\delta-1)(A_l\frac{1}{D})+B_l\frac{1}{D}\right) \tag{A.24}$$

So combining (A.23), (A.24), we get:

$$\frac{\Delta^I}{P(1+\rho b)}\frac{\partial W}{\partial b}-\frac{1}{1+\theta c}\frac{\partial W}{\partial c}=C \tag{A.25}$$

where

$$C=(\delta-1)\,n\underline{\pi}+\left(\Delta^I\frac{\xi\rho b-(\delta-1)\,(1+2\rho b)}{1+\rho b}+(\delta-1)\,\Delta\right)\sum_i\sigma_i-\left(\frac{\xi c\theta-(\delta-1)(1+2\theta c)}{1+\theta c}+\delta-1\right)\sum_i l_i$$

Then if $\xi=\delta-1$:

$$C=(\delta-1)\,n\underline{\pi}+(\delta-1)\,(\Delta-\Delta^I)\sum_i\sigma_i>0.$$

So, from (A.25), in this case, $\frac{\partial W}{\partial b}=0$ implies $\frac{\partial W}{\partial c}<0$, and $\frac{\partial W}{\partial c}=0$ implies $\frac{\partial W}{\partial b}>0$. Then either (i) $b=c=0$; (ii) there is an interior solution for b i.e. $0<b<\tau$, and $c=0$, or (iii) $b=\tau$ and $c\geq 0$.

The other case is $\xi=-1$. Here,

$$C=(\delta-1)\,n\underline{\pi}+\left(\Delta^I\left(1-\frac{\delta(1+2\rho b)}{1+\rho b}\right)+(\delta-1)\,\Delta\right)\sum_i\sigma_i+\delta\left(\frac{c\theta}{1+c\theta}\right)\sum_i l_i \tag{A.26}$$

So, by inspection of (A.26), if the cost of income shifting is infinite, i.e. $\rho = 0$, then $C > 0$. Recall that $\Delta = (1+m)\Delta^I$, and $m = \frac{\varepsilon}{\varepsilon - 1} > 0$, where m is the mark-up of the intermediate firms. Substituting these into (A.26) yields:

$$\begin{aligned} \Delta^I \left(1 - \frac{\delta(1+2\rho b)}{1+\rho b}\right) + (\delta - 1)\,\Delta &= \Delta^I \left(1 - \delta - \frac{\delta \rho b}{1+\rho b}\right) + (\delta - 1)\,(1+m)\Delta^I \\ &= \Delta^I \left((\delta - 1)\, m - \delta \frac{\rho b}{1+\rho b}\right) \qquad \text{(A.27)} \\ &> \Delta^I \left((\delta - 1)\, \frac{\varepsilon}{\varepsilon - 1} - \delta\right) \end{aligned}$$

where the final inequality follows from $\frac{\rho b}{1+\rho b} < 1$.

So, combining (A.26), (A.27), we see that if $\delta \geq \varepsilon$, $C > 0$. $\square$

Online Appendix

Table N1: Rate Structure of Patent Boxes and R&D tax credits, OECD, 2026

Country	(a) Patent Box — multiple rates by profit level?	(b) R&D tax credit — multiple rates by expenditure level?	Source
Australia	No Patent Box (proposed patent box never legislated)	Yes — large-firm premium tiered by R&D intensity (8.5 / 16.5 pts) and rate steps down above A$150m cap	ATO
Austria	No Patent Box	No — research premium flat 14% (refundable)	PwC WWTS
Belgium	No — innovation deduction 85% → ∼3.75%, flat	No — R&D investment credit flat rate	Tax Foundation
Canada	No federal Patent Box (Quebec has a provincial one)	Yes — 35% up to expenditure limit (C$6m, Budget 2025), 15% above; CCPC-conditioned	MNP
Chile	No Patent Box	No — R&D credit flat (∼35%), with cap	PwC WWTS
Colombia	No Patent Box	No — R&D / innovation credit flat	PwC WWTS
Costa Rica	No Patent Box	No R&D tax credit	PwC WWTS
Czech Republic	No Patent Box	Allowance, not credit — flat super-deduction, no ceilings	PwC WWTS
Denmark	No Patent Box	No — deduction + flat 22% credit for loss-making R&D (capped)	PwC WWTS
Estonia	No Patent Box	No R&D credit (distributed-profits CIT)	PwC WWTS
Finland	No Patent Box	Deduction, not credit — general + incremental additional deduction	PwC WWTS

continued on next page

Table N1 continued from previous page

Country	(a) Patent Box — multiple rates by profit level?	(b) R&D tax credit — multiple rates by expenditure level?	Source
France	No — reduced 10% IP rate, flat	Yes — CIR 30% up to €100m, 5% above	service-public.fr
Germany	No Patent Box	No — Forschungszulage flat 25% (SME uplift; base capped)	PwC WWTS
Greece	Limited — patent-income exemption, flat	Super-deduction, not credit (enhanced, flat)	PwC WWTS
Hungary	No — royalty deduction, flat rate; deduction capped at 50% of pre-tax profit (cap-induced marginal effect)	No — R&D credit / deductions flat	PwC WWTS
Iceland	No Patent Box	No — reimbursable credit flat (∼20%), with expenditure caps	PwC WWTS
Ireland	No — Knowledge Development Box 10%, flat	No — R&D credit flat 30%	PwC WWTS
Israel	No — Preferred Tech Enterprise 12% (7.5% dev-zone) vs Special Preferred 6%, by ≥ NIS 10bn group revenue	Grants-based; a "special credit" rate applies above ∼NIS 1.05bn R&D threshold (verify)	PwC WWTS
Italy	No Patent Box (repealed 2021 → 110% super-deduction)	Credit rates vary by activity type + ceilings (not by expenditure level as such)	PwC WWTS
Japan	No — 2025 innovation box, flat 30% deduction	Yes — general-type rate varies with change in R&D + intensity premium	PwC WWTS
Korea	Limited — SME IP-income relief (size-conditioned)	Size-conditioned (SME vs large) + incremental option → effectively multiple rates	PwC WWTS
Latvia	No Patent Box	No R&D credit (distributed-profits CIT); enhanced deduction	PwC WWTS
Lithuania	No — 5% IP rate, flat	Super-deduction (300%), not credit	PwC WWTS

continued on next page

Table N1 continued from previous page

Country	(a) Patent Box — multiple rates by profit level?	(b) R&D tax credit — multiple rates by expenditure level?	Source
Luxembourg	No — 80% exemption → ∼4.99%, flat	No R&D credit (grants / subsidies)	PwC WWTS
Mexico	No Patent Box	Incremental only — 30% on the increase over prior 3-yr average	PwC WWTS
Netherlands	No — innovation box 9%, flat	Yes — 36% on first €380k, 16% above	PwC WWTS
New Zealand	No Patent Box	No — R&DTI flat 15% (capped)	PwC WWTS
Norway	No Patent Box	No — SkatteFUNN flat 19% (cost caps)	PwC WWTS
Poland	No — Patent Box 5%, flat	Super-deduction, not credit	PwC WWTS
Portugal	No — patent box 85% exemption, flat	Yes — SIFIDE base 32.5% + 50% incremental on increase over 2-yr average	PwC WWTS
Slovak Republic	No — patent / software 50% exemption, flat	Super-deduction, not credit	PwC WWTS
Slovenia	No Patent Box	Allowance, not credit (flat)	PwC WWTS
Spain	No — ∼60% income reduction, flat	Yes — base 25% + 42% on excess over 2-yr average (+17% staff, 8% assets)	PwC WWTS
Sweden	No Patent Box	Payroll relief, not credit (flat, capped)	PwC WWTS
Switzerland	No headline tiering — up to 90% relief (cantonal); overall relief capped ∼70% of taxable income (cap-induced marginal effect)	Super-deduction (up to 150%, cantonal), not credit	Tax Foundation

continued on next page

Table N1 continued from previous page

Country	(a) Patent Box — multiple rates by profit level?	(b) R&D tax credit — multiple rates by expenditure level?	Source
Türkiye	No — tech-zone exemptions, flat	Allowance / zone incentives, not a classic credit	PwC WWTS
United Kingdom	No — Patent Box 10%, flat	Merged RDEC flat 20%; intensity-based enhanced support (ERIS) for loss-making R&D-intensive SMEs	PwC WWTS
United States	No — FDII export-IP deduction, flat effective rate	No — §41 flat 20% regular (incremental base) / 14% ASC	IRS

Shading: green = rate genuinely tiered by the absolute level of profit/expenditure; amber = cap-induced marginal effect rather than a headline rate schedule.

Details of the Simulations

This Online Appendix provides a description of the numerical implementation used in the paper. A complete replication package, including a detailed technical appendix, fully documented GAMS implementations of both the base case and the shifting case, and additional implementation notes, is available from the project's GitHub replication repository.

In the simulations, we assume all firms are identical and for simplicity, we set $s = 0$. Instead of n identical firms, we set $n = 1$ so there is one representative R&D firm and one intermediate firm. The innovation technology used in the simulations is from the Example i.e. (9), with $s = 0$ imposed i.e.

$$\sigma = 1 - \exp(-r), \qquad r = \alpha \ln l + (1-\alpha) \ln e, \qquad 0 < \alpha < 1. \tag{N.1}$$

N1 Solving for $\Delta^I, \Delta^F, \Delta$

The first step is to solve for $\Delta^I, \Delta^F, \Delta$ in terms of the underlying model parameters $\varepsilon, a, \underline{\phi}, \overline{\phi}$. First, intermediate firms face inverse demand

$$p_i = A_i q_i^{-1/\varepsilon}, \qquad A_i = a_i \left(\frac{\varepsilon - 1}{\varepsilon}\right), \qquad \varepsilon > 1, \tag{N.2}$$

and maximise pre-tax profits

$$\pi_i^I = p_i q_i - \phi_i q_i. \tag{N.3}$$

Profit maximisation yields the constant-markup pricing rule

$$p_i = \frac{\varepsilon}{\varepsilon - 1} \phi_i. \tag{N.4}$$

Substituting the equilibrium quantity into profits gives

$$\pi_i^I = Z a_i^{\varepsilon} \phi_i^{1-\varepsilon}, \qquad Z \equiv \varepsilon^{-2\varepsilon} (\varepsilon - 1)^{2\varepsilon - 1}. \tag{N.5}$$

Now w.l.o.g, set $\underline{\phi} = \phi, \overline{\phi} = 1$. Then pre- and post-innovation intermediate profits are

$$\underline{\pi}^I = Z a^{\varepsilon}, \qquad \overline{\pi}^I = Z a^{\varepsilon} \phi^{1-\varepsilon}. \tag{N.6}$$

The gain in intermediate firm profit from innovation is therefore

$$\Delta^I = \overline{\pi}^I - \underline{\pi}^I = Z a^{\varepsilon} \left(\phi^{1-\varepsilon} - 1\right) > 0. \tag{N.7}$$

Final-good profits are proportional to intermediate profits:

$$\pi_i^F = m\pi_i^I, \qquad m = \frac{\varepsilon}{\varepsilon - 1}. \tag{N.8}$$

Therefore, the gain in final-good profits and the total gain from innovation are

$$\Delta^F = m\Delta^I, \qquad \Delta = \Delta^I + \Delta^F = (1+m)\Delta^I. \tag{N.9}$$

Note that these solutions do not depend on the functional form of the innovation technology in (N.1), or the organizational form (base case or shifting case).

N2 Base Model: Solving for l, e, σ

Here, we solve for l, e, σ as functions of the tax parameters for the representative R&D firm in the base case without income shifting. In this case, using (N.1), the first-order conditions for l, e (10) can be written as

$$\beta\Delta^I \exp(-r)\frac{\alpha}{l} = p_l, \qquad \beta\Delta^I \exp(-r)\frac{1-\alpha}{e} = p_e. \tag{N.10}$$

Solving these two equations gives the closed-form input demands

$$l = \left(\frac{\alpha\beta\Delta^I}{p_l}\right)^{\frac{2-\alpha}{2}} \left(\frac{p_e}{(1-\alpha)\beta\Delta^I}\right)^{\frac{1-\alpha}{2}}, \tag{N.11}$$

$$e = \left(\frac{p_l}{\alpha\beta\Delta^I}\right)^{\frac{\alpha}{2}} \left(\frac{(1-\alpha)\beta\Delta^I}{p_e}\right)^{\frac{1+\alpha}{2}}. \tag{N.12}$$

Using the definitions of p_l and p_e from ( ), these can be written directly in terms of (τ, c, b):

$$l = (1-\tau+b)^{1/2}(1-\tau-c)^{(\alpha-2)/2}(\beta\Delta^I)^{1/2}\alpha^{(2-\alpha)/2}(1-\alpha)^{(\alpha-1)/2}, \tag{N.13}$$

$$e = (1-\tau+b)^{1/2}(1-\tau-c)^{\alpha/2}(\beta\Delta^I)^{1/2}\alpha^{-\alpha/2}(1-\alpha)^{(1+\alpha)/2}. \tag{N.14}$$

Substiuting these into (N.1), we can write the innovation probability as a function of the tax paramaters τ, b, c:

$$\sigma = 1 - (1-\tau+b)^{-1/2}(1-\tau-c)^{\alpha/2}(\beta\Delta^I)^{-1/2}\alpha^{-\alpha/2}(1-\alpha)^{(\alpha-1)/2}. \tag{N.15}$$

N3 Base Model: Welfare Function and Tax Design Problem

With symmetry and a single intermediate and R&D firm, welfare (11) takes on a very simple form

$$W = \underline{\pi} + \sigma\Delta - l - e + (\delta - 1)\left[\tau(\underline{\pi} + \sigma\Delta - l) - b\sigma\beta\Delta^I - cl\right]. \tag{N.16}$$

The problem for the government is then to choose tax parameters τ, b, c to maximise (N.16) subject to (N.13), (N.14), (N.15) and the feasibility constraints (3). This problem is solved by using the computer progam GAMS, and implementing the government's optimisation problem as a Mixed Complementarity Problem (MCP).

N4 The Model with Shifting: Solving for l, e, σ

Imposing firm symmetry, no cross-firm spillovers $s_i = 0$, and a single representative firms, from (30),(30), the effective input prices faced by the firm are

$$p_l = \frac{1 - \tau - c - \frac{\theta c^2}{2}}{D}, \qquad p_e = \frac{1}{D} \tag{N.17}$$

Here, D is the return to an innovation i.e.

$$D = \Delta^I(1-\tau) + bP + \frac{\rho b^2}{2}P = \Delta^I\left[(1-\tau) + \beta b + \frac{\beta\rho b^2}{2}\right]. \tag{N.18}$$

where in the second equality, we use $P = \beta\Delta^I$.

The functional form of the innovation technology is (N.1) as before. So, the R&D firm's first-order conditions are

$$\exp(-r)\frac{\alpha}{l} = p_l, \qquad \exp(-r)\frac{1-\alpha}{e} = p_e. \tag{N.19}$$

Since these conditions require $p_l, p_e > 0$, the numerical implementation imposes the feasibility condition $D > 0$ and

$$N_1 \equiv 1 - \tau - c - \frac{\theta c^2}{2} > 0.$$

The latter condition is tighter than the simple upper-bound constraint $c \leq 1 - \tau$, and ensures that the input demands are well defined.

Solving the first-order conditions (N.19) gives the closed-form input demands

$$l = \alpha^{\frac{2-\alpha}{2}}(1-\alpha)^{-\frac{1-\alpha}{2}} p_l^{-\frac{2-\alpha}{2}} p_e^{\frac{1-\alpha}{2}}, \; e = \alpha^{-\frac{\alpha}{2}}(1-\alpha)^{\frac{1+\alpha}{2}} p_l^{\frac{\alpha}{2}} p_e^{-\frac{1+\alpha}{2}}. \tag{N.20}$$

Substituting (N.20) into (N.1), we get

$$\sigma = 1 - p_l^{\frac{\alpha}{2}} p_e^{\frac{1-\alpha}{2}} \alpha^{\frac{\alpha}{2}} (1-\alpha)^{\frac{1-\alpha}{2}}. \tag{N.21}$$

N5 Shifting Model: Welfare Function and Tax Design Problem

Imposing firm symmetry, no cross-firm spillovers $s_i = 0$, and a single representative firms, the welfare function (33) simplifies to:

$$\begin{aligned} W =& \underline{\pi} + \Delta\sigma - (l + e) + \xi \left[\sigma \frac{P\rho b^2}{2} + \frac{\theta c^2}{2} l \right] \\ &+ (\delta - 1) \left[\tau(\underline{\pi} + \Delta\sigma - l) - \sigma P(b + \rho b^2) - (c + \theta c^2) l \right]. \end{aligned} \tag{N.22}$$

The government's problem is therefore to choose the policy vector (τ, b, c) to maximize (N.22), subject to (N.20), (N.21), and the feasibility constraints (3) and the conditions $D > 0, N_1 > 0$. This problem is solved by using the computer program GAMS, and implementing the government's optimisation problem as a Mixed Complementarity Problem (MCP).