Journal of Public Economics 45 (1991) 271–300. North-Holland

The vanishing Harberger triangle

Hans-Werner Sinn*

University of Munich, D-8000 Munich 22, Germany

Received September 1989, revised version received December 1990

This paper shows that the double taxation of corporate dividends (or profit repatriations) implies a nucleus theory of the corporation. After the firm is set up with a small stock of original capital, it enters a phase of purely internal growth during which no dividends are paid and no shares are issued. The phase terminates when an efficient stock of capital has been accumulated and dividends are paid. During the growth phases, the tax distortion is inversely related to the tax burden and it is larger than conventional formulae for the cost of internal and external equity finance suggest.

1. The problem

One of the roots of modern tax theory lies in Harberger's (1962, 1966) problem of how the double taxation of corporate dividends affects the allocation of resources between the corporate and non-corporate sectors of the economy. Harberger's claim was that the double taxation of dividends discriminates against corporate investment and creates welfare losses by keeping too large a share of the economy's capital stock in the non-corporate sector. The larger the tax burden on dividends, the bigger the welfare loss that results.

Related results were attained in the foreign trade literature [see MacDougall (1960), Kemp (1962, 1964), and Hamada (1966)], where the allocation of capital to competing economies, rather than sectors, was studied. A typical claim of this literature is that taxes on profit repatriations discriminate against foreign investment and result in international distortions very similar to those that Harberger attributed to the double taxation of dividends in a closed economy context.

This paper reconsiders the 'Harberger problem' from an intertemporal perspective. It studies the foundation and growth of corporations in the

*I wish to thank David Bradford, Don Fullerton, Michael Hoel, Vesa Kanniainen, Mervyn King, John Shoven and the anonymous referees for useful comments. The paper originates from work begun at the University of Bergen, Norway, in 1987, and is part of the NBER's research program in International Capital Markets. The support and hospitality of these institutions is gratefully acknowledged.

presence of a dividend tax to find out whether, and if so under what circumstances, this tax creates intersectoral or international distortions.[1] The main result is that the dividend tax generates distinctive growth stages similar to those described by Penrose's (1959) nucleus theory of the corporation. To avoid the dividend tax, the firm is set up with only a small nucleus of original capital and then grows through a process of internal investment where no additional capital is injected and no dividends are paid. The phase of internal growth continues until the firm reaches a stage of maturity where it distributes its profits and stops reinvesting. During the growth phase, the dividend tax creates a distortion, but this distortion is not the same as that predicted by existing theories. It is larger than conventional cost-of-capital formulae would suggest and, in an important sense, it is negatively, rather than positively, related to the magnitude of the measured tax burden.

The traditional view of corporate taxation as formulated by Harberger has been questioned by holders of the so-called 'new' or 'trapped equity' view of corporate taxation. See King (1974a, b, 1977), Bradford (1980, 1981), Auerbach (1979, 1983), Fullerton and King (1984), and Sinn (1985) for the general argument and Hartmann (1985) and Sinn (1984) for its discussion in the context of foreign direct investment. The argument of this literature is that the dividend tax is capitalized in share prices and therefore cannot affect the firm's investment decisions. The tax is simply seen as a lump-sum levy on corporations. If true, tax reforms, whose aim is to remove the double taxation of dividends, would be superfluous. They would create windfall gains for the current owners of corporate shares, but would not improve the allocation of resources.

Unfortunately, however, the new view does not seem fully compatible with the empirical facts. As observed by Poterba and Summers (1983, 1985), who studied the effects of British tax reforms, changes in the statutory dividend tax rate did have adverse effects on the level of aggregate investment. The authors attributed their findings to the fact that the trapped equity model neglects the signalling function of dividends and they specified a model in which a reduction in share repurchases (i.e. in fact, new issues of shares) is the only marginal source of finance. While this is one possible explanation for the non-neutrality of dividend taxation, there are others. The one explored in this paper is suggested by a serious shortcoming of the trapped equity model.

Existing approaches that use this model have the common characteristic that they do not explain how equity falls into the trap. Typically, they

[1]In this paper the terms 'Harberger problem' and 'Harberger triangle' refer to the general problem of how non-uniform taxes on profit distributions affect the allocation of capital to competing uses. Given the level of abstraction used, there is no meaningful distinction between international and intersectoral allocation problems.

assume that the firm already has more than the efficient amount of equity capital at the time the investment decision is analyzed. Under these circumstances the neutrality proposition is not especially surprising. It just means that the firm retains the efficient amount of equity and distributes the remainder. The important problems of how much equity capital shareholders may wish to inject into their firms in the first place and whether the corporate stock of capital will ever reach its efficient size are unsolved.

This paper offers a solution that is consistent with Poterba and Summers' findings. It reconsiders Harberger's problem from the viewpoint of a trapped equity model, but one that starts with the process of injecting capital into the firm. In this model, dividend taxation is strongly non-neutral, implying larger (initial) distortions than Harberger's model suggests. Surprisingly, no similar model seems to exist in the literature.[2] It is true that holders of the trapped equity view typically stress that dividend taxes are distortionary to the extent that new issues of shares are a marginal source of finance. However, as far as is known, no attempt has been made to formulate an explicit intertemporal model that describes the foundation and growth of a corporation in the presence of dividend taxation, and it seems that the cost of new share issues has never been consistently derived under the assumptions of the trapped equity model.

This paper rehabilitates Harberger's view that the dividend tax discriminates against corporate investment, but, in addition, it modifies and criticizes his analysis. Harberger and many of his followers have concentrated on the general equilibrium repercussions of taxation and have placed little emphasis on microeconomic considerations such as how taxes would affect the investment decisions of the firm. Frequently they have simply assumed that the corporate firm invests until the net-of-tax marginal product of capital equals the market rate of interest. This assumption is compatible with *partial* optimization given that new share issues (or reduced share repurchases) are the only marginal source of finance and that all profits resulting from an investment are distributed as dividends. However, there are at least two problems with this.

The first is that, instead of new share issues, the firm may choose *other sources of finance*. From an empirical point of view, both debt and retained profits are cheaper and much more important sources than new issues of shares. The holders of the new view have emphasized this and have derived investment conditions that typically imply lower distortions than those Harberger argued for.

[2]After writing this [see Sinn (1988)]. two important papers by King (1989) and Hines (1989) came to the author's attention. Both study the problem of setting up a firm in the presence of taxes, but they do not show the necessity of a phase of purely internal growth and do not derive the propositions of this paper.

The second problem is that, instead of using the profits from its marginal investment to pay dividends to shareholders, the firm may choose *other uses of profits*. One possibility is share repurchases. Profit-financed share repurchases can be seen as a way of avoiding the double taxation of dividends and they undermine Harberger's results for obvious reasons.[3] Another potential use of profits is internal investment. This is not only of great empirical significance in all countries, it is also suggested by theoretical considerations. In fact, it is clear that a firm would not distribute its profits immediately after it has issued new shares if, as Harberger claimed, the shares stopped being issued before the point where the marginal product of capital equals the market rate of interest. With an internal rate of return above the shareholders' discount rate, it would always pay to reinvest the profits and distribute them later. The present value of dividends, net of the dividend taxes, could be increased by postponing the distributions for as long as it takes for the process of reinvesting profits to equate the marginal product of capital and the market rate of interest. This suggests that there might be something wrong with the reasoning underlying the Harberger-type cost-of-capital formula even if it is assumed that the firm is forced by an initial shortage of retainable profits to issue new shares at the margin, cannot borrow, and cannot escape the dividend tax by repurchasing its shares.

The reinvestment of the profits generated by marginal investment projects is incompatible not only with Harberger's formula, but renders some of the formulae provided by holders of the new view also inapplicable. For example, the popular cost-of-capital formula of Fullerton and King (1984), which is a weighted average of the costs of the three alternative sources of finance, assumes that the profits from marginal investment projects are distributed. King's (1977) expressions for the cost of new share issues and retained profits, which enter this formula, are not applicable when the returns from marginal investment projects are reinvested at a rate of return above the market rate of interest. Subsection 2.5 will discuss the relationship to the cost-of-capital formulae provided by the holders of the new view in more detail and argue that these formulae may underestimate the true cost of equity capital.

Except for the exclusion of debt financing and profit-financed share repurchases, the present paper makes no assumptions about the firm's source of funds for, and the use of the profits generated by, marginal investment projects. Instead, the financial and real investment decisions are endogenously derived from the firm's optimization approach. The available sources of funds are new share issues and retained profits, and the possible uses of profits are dividends and internal investment. The exclusion of debt financing

[3]See Sinn (1990a) for an extensive discussion of the significance of share repurchases for the cost of capital.

and profit-financed share repurchases is motivated by the attempt to treat one difficulty at a time and to follow Harberger's analysis as closely as possible. Including these possibilities would weaken his case right from the outset and imply a stronger criticism of his results than the one made here.[4]

Essentially, then, the paper is a reconsideration of Harberger's problem, asking his questions and using his assumptions. It formulates an intertemporal variant of his two-sector model (or the Kemp–MacDougall two-country model), which is based on microeconomic optimization rather than arbitrarily postulated marginal conditions. The variant is slightly more complicated than the original model but it nevertheless reflects the attempt to be as simple as possible without giving up the rigor necessary to make the point. The economy has a given stock of capital for which two representative firms compete in a perfect capital market. The firms produce the same commodity, use only equity capital, and please their far-sighted owners by choosing the investment policies that maximize their market values. One of the firms is taxed, the other is tax exempt. The trapped equity property is modelled by the assumption that the government does not contribute to funds injected into the firm, but taxes all payments to shareholders that result from current or previous profits. A tax-exempt return of the original capital is allowed.[5]

For simplicity all other taxes except the dividend tax are neglected. It would be straightforward to introduce a true corporate tax on retained and distributed profits, a personal tax on interest income and dividends, and a capital gains tax. This would not affect the basic spirit of the model provided that the effective tax rate on accrued capital gains falls short of the personal tax rate.[6] Alternatively, an explicit international tax structure with home and host country taxes on parent and subsidiary profits and with alternative provisions for the taxation of repatriated earnings (deferral, exemption, crediting, etc.) could be introduced.[7] This too would be immaterial for the results as long as the overall tax burden on profit repatriations exceeds that on profits reinvested abroad.

[4]Critical discussions of the Harberger approach that allow for debt financing and other modifications of his assumptions can be found in Stiglitz (1973) and Sinn (1985, ch. 6). The neutrality of profit taxes in the presence of debt financing was first pointed out by Oberhauser (1963, p. 67).

[5]The paper should not be seen as an attempt to solve the dividend puzzle. See Poterba (1987) for an excellent discussion of this puzzle.

[6]Only one-third of the OECD countries subject capital gains realized more than one year after asset purchase to personal taxation. Moreover, even the United States, which recently increased the capital gains tax base from 40% to 100%, has an *effective* capital gains tax rate far below the personal tax rate because only realized capital gains are taxed and shareholders with long-term holding strategies pay the highest share prices, thus dominating the market.

[7]See Sinn (1990b) for an extension of the present model along these lines and Sinn (1985, pp. 172–175) for a review of tax rules applying to international capital income flows.

2. Dividend taxation and the growth of the corporation

Before the allocation problem can be discussed meaningfully, a model of the firm is needed that explains how a corporate firm is set up and how the equity capital falls into the trap. This section provides one. The next section will add an untaxed firm to analyze the Harberger problem.

2.1. A model of the firm

The firm's policy is determined by its shareholder who, in line with Fisher's separation theorem, seeks to maximize the initial market value of shares net of the original capital injected.[8] It is convenient to interpret the shareholder as a representative private household. However, it is equally possible to think of a parent company setting up and controlling its subsidiary. Let t_1 be the point in time t at which the firm's planning problem starts. The shareholder is a price-taker, looks through the corporate veil, and is endowed with perfect foresight of all variables of the model. He can borrow and lend at the going market rate of interest r, $r>0$, whose time path he takes as exogenously given. The market value of shares is therefore implicitly determined by the following arbitrage condition that requires the shareholder to be indifferent between keeping his wealth in the form of bonds or shares:

$$rM=\theta D+\dot{m}z+(\dot{z}m-Q), \quad \text{for } t>t_1. \tag{1}$$

The left-hand side of (1) is the return from selling the existing stock of shares at its market value M and investing the funds received in bonds which yield the market rate of interest r. The right-hand side of (1) measures the current return from continued shareholding. D is the gross dividend and $\theta \equiv 1-\tau$ is one minus the dividend tax rate. It is assumed that $0<\theta \leqq 1$ and that θ is a constant.[9] The next term, $\dot{m}z$, is the capital gain from existing shares where m is the price of a share and z the number of outstanding shares. The term in parentheses is the value of purchasing options for new shares issued to the existing shareholder. It is the difference between the market value of the new shares, $\dot{z}m$, and the funds Q that must be paid to the firm in exchange for the shares. When there are no purchasing options (as in the United States, for example) it can be assumed that $\dot{z}m=Q$ as the existing shareholder will not vote for a policy of diluting his shares. In general, however, there is no need to assume that $\dot{z}m$ and Q are linked to one another. It is even admissible to

[8] Note that the separation theorem is applied here to the evaluation of the cash flow of the firm as a whole rather than the cash flow generated by its investment projects. This takes account of the asymmetries in the tax treatment of dividends and equity injections.

[9] See Howitt and Sinn (1989) for an analysis of anticipated changes of the dividend tax rate in a trapped equity model with debt financing.

assume $Q > \dot{z}m = 0$ as might be appropriate in the case of a parent company financing its subsidiary.

Noting that $\dot{m}z + \dot{z}m = \dot{M}$, (1) can be transformed to $\dot{M} = -\theta D + Q + rM$ which, upon integration, gives[10]

$$M(t) = \int_t^\infty [\theta D(v) - Q(v)] \exp \int_t^v -r(u)\,du\,dv, \quad \text{for } t \geqq t_1, \tag{2}$$

plus some arbitrary constant. The constant is taken to be zero to ensure that the market value of a firm that promises never to issue new shares and never to distribute any dividends is zero. It is assumed for the derivation of (2) that the integral exists which requires that

$$\lim_{t^* \to \infty} [\theta D(t^*) - Q(t^*)] \exp \int_t^{t^*} -r(u)\,du = 0.$$

Following Harberger, it is assumed that the firm produces its output only with equity capital K. Moreover, with only small losses in generality, all commodity prices are assumed constant and normalized to unity. Physical capital equals equity capital and there is no depreciation. Under these assumptions, the firm's revenue, profit, and output can all be described by the function $f(K)$ satisfying the usual properties $f' > 0$, $f'' < 0$, $f'(0) = \infty$, and $f(0) = f'(\infty) = 0$. The dividend the firm can pay is

$$D = f(K) - \dot{K} + Q, \tag{3}$$

where $\dot{K}$ is the firm's net investment. Let K_0 be the stock of capital available from the firm's past history and K_1 the stock of equity capital reached at t_1 after the initial issue of shares.

Then the firm's problem can be expressed as

$$\max_{\{\dot{K}, Q, K_1\}} M(t_1) - K_1 \tag{4}$$

$$\text{s.t. } K_1 \geqq K_0 = 0,$$

$$D \geqq 0,$$

$$Q \geqq 0,$$

where K is the state variable and $\dot{K}$, Q, and K_1 are the controls. The three

[10]For a comparison of this and alternative, but equivalent, market value functions that can be constructed by way of integrating the differential equation, see Sinn (1985, pp. 63–65).

constraints implicitly capture the trapped equity assumption that the government participates in profit distributions but not in what the shareholder injects into his corporation. For the time being, the trapped equity assumption is made in the extreme form that it is impossible for the firm to pay cash to its shareholder that is not taxed as dividends, i.e. that $Q \geqq 0$ for all $K \geqq 0$. This assumption will be relaxed in section 4 to allow the firm to return its original share capital. A similar remark applies to the assumption $K_0 = 0$.

2.2. The optimality conditions

The problem of the firm can be solved by using Pontryagin's Maximum Principle. Using (3), and associating a co-state variable q (Tobin's q) with K and Kuhn–Tucker multipliers μ_D and μ_Q with the flow constraints, the current value Hamiltonian of problem (4) can be written as:

$$\mathscr{H} = (\theta + \mu_D)[f(K) - \dot{K} + Q] + q\dot{K} - Q(1 - \mu_Q).$$

From $\partial\mathscr{H}/\partial\dot{K} = 0$ it follows that

$$q = \theta + \mu_D \tag{5}$$

and from $\partial\mathscr{H}/\partial Q = 0$ that

$$\mu_D + \mu_Q = \tau. \tag{6}$$

Both equations together imply

$$q = 1 - \mu_Q. \tag{7}$$

Because of (5), the canonical equation $\dot{q} - rq = -\partial\mathscr{H}/\partial K$ can be transformed to

$$f'(K) + \frac{\dot{q}}{q} = r. \tag{8}$$

The Kuhn–Tucker conditions of the problem are

$$\mu_Q Q = 0, \quad \mu_Q \geqq 0, \quad Q \geqq 0, \tag{9}$$

$$\mu_D D = 0, \quad \mu_D \geqq 0, \quad D \geqq 0. \tag{10}$$

The firm's starting condition is $\partial M(t_1)/\partial K_1 - 1 = 0$ which, as $\partial M(t)/\partial K(t) \equiv q(t)$ holds by definition, implies that

$$q(t_1) = 1. \tag{11}$$

Finally, the transversality condition is

$$\lim_{t\to\infty} q(t)K(t) \exp \int_{t_1}^{t} -r(v)\,dv = 0. \tag{12}$$

Notice that, because of (6), (9), and (10), the firm cannot simultaneously issue new shares and pay dividends. Instead, at any point in time after t_1, it must either be the case that $(Q>0; D=\mu_Q=0)$, that $(Q=D=0; \mu_Q, \mu_D \geqq 0)$, or that $(Q=\mu_D=0; D>0)$. Together with the initial condition, this implies that the following activity phases are available. The names of these phases anticipate properties yet to be derived.

Phase Ia $(K_1 \geqq 0; t=t_1)$. Phase Ia refers to the starting point where the original stock of equity K_1 may be injected. According to (11), this phase is characterized by

$$q(t_1) = 1.$$

Phase Ib $(Q>0,\ D=\mu_Q=0;\ t>t_1)$. Phase Ib is a phase of continuing equity injections after the time of foundation. During this phase, $\mu_Q=0$ and hence (7) implies that

$$q=1, \quad \dot{q}=0.$$

It therefore follows from (8) that

$$r = f'(K). \tag{13}$$

Phase II $(Q=D=0;\ \mu_Q, \mu_D \geqq 0;\ t>t_1)$. If the firm neither issues new shares nor pays out any dividends, then from (8):

$$\dot{q} = q[r - f'(K)], \tag{14}$$

and from (3):

$$\dot{K} = f(K). \tag{15}$$

Phase III $(D>0,\ Q=\mu_D=0;\ t>t_1)$. For a firm that pays dividends, (5) and $\mu_D=0$ indicate that

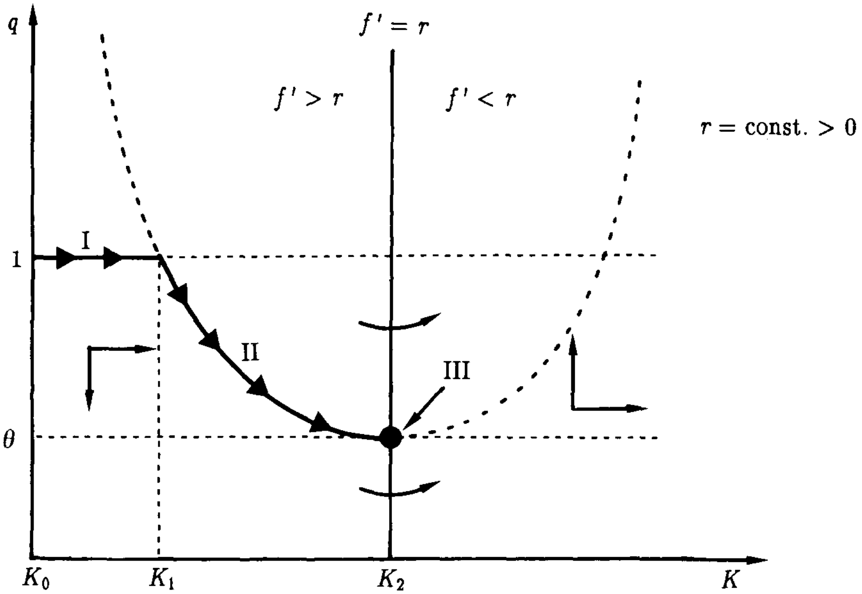

Fig. 1. The optimal growth path under the dividend tax.

$$q=\theta, \quad \dot{q}=0$$

which, because of (8), implies that

$$f'(K)=r. \tag{16}$$

2.3. The optimal growth path

The optimal growth path of the firm is a combination of the four phases that satisfies the transversality condition (12) and the Maximum Principle's general requirement that there be no jumps in the state and co-state variables. Assuming for the time being that the market rate of interest is a positive constant[11] for $t \geqq t_1$, the growth path can be uniquely determined in (q, K) space as illustrated in fig. 1.

The position of the vertical line in this diagram characterizes the equity level K_2 implicitly defined by the laissez faire condition $f'(K_2)=r$. During Phase III, the firm is on this vertical at $q=\theta$. It distributes all its profits since $\dot{K}=Q=0$ and it can stay in this phase for ever since the transversality condition (12) is satisfied.

[11]In the next section, this assumption will be relaxed and the time path of r will be endogenously determined by the conditions of a market equilibrium.

Phase III is the phase on which the new view of corporate taxation concentrates. The shadow price of equity, q, is θ rather than one, since θ is the shareholders' net dividend forgone if the firm decides to retain and invest one additional unit of profit. The low value of the shadow price just compensates for the tax on the returns which an additional unit of equity capital will generate and explains why, despite the tax, the firm follows the laissez faire investment rule $f'(K)=r$. Let t_2 be the point in time at which Phase III begins.

Clearly, Phase III cannot be a starting phase. The firm first has to raise enough equity capital to get there. A potential candidate for explaining how to reach Phase III is Phase II, for the two differential equations (14) and (15) define a set of possible paths in (q,K) space one of which intersects the vertical at $q=\theta$. The slope of these paths is given by

$$\mathrm{d}q/\mathrm{d}K=\dot{q}/\dot{K}=q[r-f'(K)]/f(K) \quad \text{(Phase II)}. \tag{17}$$

As $f''<0$ and $\dot{K}>0$ for $K>0$, the slope is negative in the region to the left of the vertical, zero on the vertical, and positive in the region to the right of it. The arrows in fig. 1 indicate the possible movements. It is assumed that the path leading to the steady-state point (θ, K_2) intersects the horizontal line of height $q=1$ at some strictly positive value of K; call this value K_1^*. If there is no such intersection point, then the dividend tax prevents the firm from being founded.[12]

Notice that Phase II cannot be a final phase. If it were, the firm would never again pay dividends. Yet it is clear from (2) that this cannot be an optimum since, at any point in time t^* during this phase and given the then available stock of equity $K(t^*)$, the firm could increase its market value $M(t^*)$ by paying out all future profits as dividends and keeping the stock of equity constant. The existence of such a possibility would violate the Bellman Principle on which the Maximum Principle is built. Neither can Phase II be an initial phase. Starting with Phase II means starting with a value of q above one. Given the possibility of injecting equity capital at a price of one from outside the firm, this cannot be optimal.

Potential candidates for an initial phase are Phases Ia and Ib. It can easily be seen, however, that Phase Ib does not exist. During Phase Ib, $q=1$ and the firm issues new shares without distributing any dividends. In the diagram, this means that there is a horizontal movement to the right ($\dot{K}>0$) with q being kept constant at the level of unity. Such a horizontal movement clearly implies that condition (13) cannot be met, for this condition and the

[12]The discussion paper [Sinn (1988, appendix)] from which this article was derived contains a proof that the ordinate is an asymptote to all possible paths, i.e. $q(K)\to\infty$ for $K\to 0$, if the production elasticity of capital is bounded away from unity.

assumption $\dot{r}=0$ imply that $\dot{K}=\dot{r}/f''=0$. The firm would have to satisfy the laissez faire condition $f'=r$ if it continued to issue new shares, but it cannot.

Thus, only Phase Ia remains. Like Phase Ib, Phase Ia requires that $q=1$. However, since this phase merely refers to the starting point $t=t_1$, the flow condition (13) is not required. The firm issues a sufficient number of shares to reach the Phase II path in one step when it is founded: $K_1=K_1^*$. Because of the non-existence of Phase Ib, the 'a' is dropped in the remainder of this paper and all references to Phase I are now meant to refer to Phase Ia.

With Phase I as the necessary starting phase and Phase III as the only available final one, the question is whether Phase II is needed at all. It might be tempting to believe that the optimal strategy involves issuing enough shares to reach K_2 during Phase I and then to start with Phase III ($t_2=t_1$). However, such a direct move between the two phases would require a forbidden jump in the co-state variable q from 1 to θ. To avoid the infinitely large capital losses on the last unit of funds injected into the firm, a rational investor injects an amount of capital smaller than K_2. In fact, as will be clarified in subsection 2.5, the threat of capital losses will even make it wise to start with a smaller stock of capital than Harberger's formula for the cost of new share issues suggests.

The phase of internal growth (II) is a necessary link between Phase I and Phase III, and the only continuous transition between the phases is one that satisfies the following pattern. During Phase I, $q=1$ and new shares are issued until the desired stock of original capital K_1 is reached. Then the firm drifts along the curved Phase II path towards the steady state point ($q=\theta$, $K=K_2$) accumulating a surplus reserve of amount K_2-K_1. Once there ($t=t_2$), it is in Phase III and stays there for ever.

It is important to realize that, unlike many other intertemporal models, the steady state is reached in finite time. As $Q=D=0$ during Phase II, it follows from (3) or (15) that the increase in K per period is positive and bounded away from zero, and, in fact, even the speed of increase is increasing:

$$\dot{K}(t)=f[K(t)]\geqq f(K_1)>0,$$

$$\ddot{K}(t)=f'[K(t)]\dot{K}(t)>0, \quad \text{for all } t_1\leqq t\leqq t_2 \quad \text{(Phase II)}.$$

This clearly implies that $t_2<\infty$, i.e. that the steady-state stock of surplus reserves K_2-K_1 is accumulated in finite time.

The following proposition summarizes these findings. It describes the optimal growth path of a corporation that is subject to dividend taxation.

Proposition 1. When the firm is founded, new shares are issued to generate some equity to start with. The starting stock is smaller than the one at which the marginal product of capital equals the market rate of interest. After the *foundation, a phase of internal growth follows during which the firm neither issues new shares nor distributes any profits. This phase terminates in finite time when sufficient surplus reserves have been accumulated to equate the marginal product of capital with the market rate of interest. The firm will then stop growing, issue no shares, and distribute all its profits.*

While Proposition 1 refers to the qualitative aspects of the growth path when there is a dividend tax, it does not clarify the role of the dividend tax itself. To understand this role, note that the size of the dividend tax rate neither affects the initial value of q, nor the set of Period II paths compatible with (14) and (15), nor the steady state value of equity, K_2. The only thing that is affected is the steady-state value of $q, q(t_2)=\theta$. This value singles out the optimal path during Period II and determines both the length of this period and the size of the original capital K_1. Obviously, the higher τ, the lower θ and q, the lower K_1, and the longer the time span that must elapse before the missing amount of surplus reserves, K_2-K_1, is accumulated. When there is no dividend tax, then $q(t_2)=q(t_1)=1$, and Phase II is not needed to avoid a jump in q. Shareholders inject a sufficient amount of original capital to reach the steady-state value K_2 in one step.

Again, this is summarized in a proposition.

Proposition 2. The phase of internal growth is longer and the starting stock of equity smaller the higher the tax rate on dividends. Without the dividend tax, there is no such phase. All equity is then generated through equity injections when the firm is founded.

2.4. The nucleus theory of the corporation

While Propositions 1 and 2 are meant to prepare the ground for a discussion of the Harberger problem, they may be interesting in their own right. They show that the policy of maximizing the rate of internal growth and minimizing dividend payments that has been so graphically described by Penrose (1959) and others does not have to be explained by a divergence between manager and shareholder interests. The high burden of dividend taxes can also be an explanation. It is particularly confirmative in this context to hear what Barlow and Wender (1955, ch. 11) and Penrose (1956, pp. 227–229) say about the growth of foreign affiliates of U.S. corporations. The typical pattern of growth these authors observed was that, when founded, the affiliates were given only a nucleus of equity capital and then had to grow by themselves. Only mature affiliates that had reached their desired size were expected to distribute profits. In the light of the fact that cross-border profit distributions are frequently subject to international double and triple taxation, this observation is not at all surprising.

Similar evidence has recently been provided by studies of Fazzari, Hubbard and Petersen (1988) and Hines and Hubbard (1989). This literature shows that, for firms with low payout rates, investment and cash flow are tightly linked, and that there are indeed many firms that fall into this category. While the observation has been made for U.S. firms in general, an extreme situation seems to prevail with U.S. multinational subsidiaries. As Hines and Hubbard report, in 1984 more than 80% of these subsidiaries did not pay any dividends to their parent companies. Although one should be careful with a generalization of this figure, it is a clear indication that the phase of purely internal growth where all profits are invested is not a theoretical artifact with little practical meaning, but a potentially important empirical phenomenon in cases where the trapped equity model applies and high dividend taxes are raised.[13]

2.5. *Relationship to the new view and the old view*

Propositions 1 and 2 complement the new view of corporate taxation. They confirm this view for Phase III and they show that this phase will indeed be reached. In Phase III, profit retentions are a *potential* marginal source of finance and the profit from marginal investment projects is paid out as dividends. As predicted by holders of the new view, the marginal product of capital equals the market rate of interest.

There is less agreement, though, for Phases I and II. As is well known, cost-of-capital expressions of the Fullerton–King type say that, when there is only a dividend tax, the cost of capital is r for a firm that relies exclusively on retained earnings and r/θ for one that relies exclusively on new issues of shares. The latter value is the same as the one Harberger assumed in his calculations and indeed it is the most frequently used value for the cost of equity finance in the literature on tax distortions. It is this value that characterizes the so-called 'old' view of corporate taxation. By way of contrast, in Phase II, where retained earnings are the only marginal source of finance, the cost of capital exceeds r, and the cost of capital is not, in general, equal to r/θ when new share issues are the marginal source of finance (Phase I). In fact, under mild conditions, it turns out to be higher than this value.

Assume that the pure profit or rent that the concave production function

[13]A further observation that supports the trapped equity model in general and which has been made by many authors involves the relatively infrequent occurrence of new share issues. For example, in the period from 1980 to 1985, on average 67.8% of gross investment by U.S. non-financial corporations was internally financed and 31.0% was debt financed, but only 1.2% was financed by share issues. [Calculated from *Survey of Current Business*, Volumes 57 (July 1977, p. 26n.), 61 (1981, special supplement, p. 10), 63 (July 1983, p. 30), 66 (July 1986, p. 33); and *Federal Reserve Bulletin*, Volumes 55 (November 1969, p. A 71.4), 60 (October 1974, p. A 59.4), 64 (June 1978, p. 433), 65 (December 1979, p. A 44)]. The actual figures may be somewhat different for other countries, but their tendency clearly describes a general empirical phenomenon. In developed economies, corporations are self-perpetuating enterprises that rarely rely on equity injections from the household sector but generate most of their equity capital internally.

$f(K)$ implies is the return to a hidden fixed factor of production and that output is linearly homogeneous in K and this factor. Let α be the hidden factor's partial production elasticity (the share of rents), β the partial production elasticity of K (the share of capital), and σ the Hicksian elasticity of substitution between the two factors. Then a sufficient condition for the cost of new share issues to exceed the conventional value r/θ is

$$0<\tau<\frac{\alpha/\beta}{\sigma}. \tag{18}$$

A proof, which also covers the extension of the model provided in section 3, is contained in the appendix.

Interpreting α and β as income shares, these results can be summarized in Proposition 3.

Proposition 3. Cost-of-capital formulae of the Fullerton–King variety are not applicable to immature firms that prefer to reinvest their profits. These formulae definitely underestimate the true cost of retained earnings and they (as well as the Harberger formula) may underestimate the true cost of new share issues. With an arbitrary dividend tax rate $(0<\tau<1)$, *a sufficient condition for the underestimaton of the cost of new share issues is that pure corporate rents do not fall short of the return to capital* $(\alpha \geqq \beta)$ *and that the demand for capital is not more elastic than in the Cobb–Douglas case* $(\sigma \leqq 1)$. *Another sufficient condition is simply that the dividend tax rate is small enough.*

As mentioned in the Introduction, the reason for the revision of the cost-of-capital formulae is the reinvestment of profits generated by marginal investment projects. While the Fullerton–King methodology allows for alternative sources of equity finance, it assumes invariably that the marginal profit from equity-financed investment is distributed to shareholders. This assumption is admissible when retained earnings are the source of finance and the firm is mature. Even if the firm retains its marginal profit, the cost of capital can be calculated as if the profit were distributed since the shareholders are indifferent between dividends and retentions. Retentions generate a present value of future dividends that just equals the value of the dividends forgone. Things are different, though, for young firms that have investment projects yielding a rate of return above the market rate of interest. For these firms retentions dominate dividends strictly and so it does make a difference whether the returns from marginal investment projects are retained or distributed. In fact, the reinvestment of profits reduces q, the marginal value of equity, and this reduction is a capital loss that increases the firm's cost of capital beyond the value which simple arbitrage conditions are able to predict.[14]

A useful study in the firm's cost of equity capital that also allows for a change in q is that of Edwards and Keen (1984). These authors calculate cost-of-capital formulae for situations where adjacent periods are characterized by different combinations of new share issues and retained profits. However, they do not allow for a phase of internal growth and, except for Phase III, their formulae are not applicable to the present model. Their result (p. 214), that the dividend tax does not affect the cost of capital 'whenever the marginal source of funds is the same in two adjacent periods', cannot be confirmed. During Phase II, retained profits are the marginal source of funds in all adjacent 'periods', but nevertheless the dividend tax is able to drive a wedge between the marginal product of capital and the market rate of interest. The cited statement is only true in Phase III where retained profits are a potential marginal source of funds *and* dividends are paid.

The assumption that marginal profits are distributed is particularly misleading when the firm is forced to issue new shares because it does not have enough retainable profits. Economists sharing the 'new' view sometimes argue that the Harberger–Fullerton–King formula, $f(K)\theta = r$, is correct in cases where the firm is forced to rely on equity injections from the household sector. However, these economists overlook the fact that the underlying assumption, namely that the marginal profits are being distributed, conflicts with the goal of market value maximization. If the firm were indeed to obey the Harberger–Fullerton–King formula when choosing its investment policy, it could increase its market value by reinvesting the profits at the internal rate of return $f'(K)$ rather than by distributing them to the shareholders. The shareholders have access to the market rate r, but the firm could reinvest at a rate of return that exceeds this market rate initially by the amount $\tau f'(K)$ and loses its lead only gradually with the passage of time. Obviously the formula $f'(K)\theta = r$ implies a financial behavior other than the one from which it is derived. Correcting this inconsistency by allowing the use of profits to be determined endogenously when new share issues are chosen as a marginal source of finance leads to a different, and typically higher, value for the cost of capital. Table 1 summarizes the relationship between the source of funds, the use of profits, and the cost of capital.

[14]According to eq. (14), the cost of capital is $r - \dot{q}/q$. It exceeds r since $\dot{q}/q < 0$. Note that the decline of q does not mean that the value of equity itself is being reduced. It follows from (1) that, during Phase II where $D = Q = 0$, $(\dot{m}z + \dot{z}m)/M = \dot{M}/M = r$. Thus, the market value of equity grows at a rate that equals the market rate of interest. The co-state variable q is the slope of a concave curve in (M, K) space that depicts the market value of the firm as a function of its stock of equity capital. During Phase II, there is an upward movement along this curve during which the slope is declining but M is increasing. The movement comes to a halt when the slope equals θ.

Table 1

The cost of equity finance.

Source of funds \ Use of profits	Dividends	Retentions
New issues of shares	r/θ (impossible)	$>r/\theta$[a] (Phase I)
Retained earnings	r (Phase III)	$>r$ (Phase II)

[a]Cf. sufficiency condition (18).

A corollary of the results reported in table 1 is that the traditional view of the significance of deferral needs to be revised. The standard argument is that the possibility of deferring the dividend tax by reinvesting the profits and distributing them later reduces the firm's cost of capital.[15] This argument is wrong, because it neglects the fact that a preferential treatment of retained earnings increases the opportunity cost of new equity injections. This opportunity cost must be added to, and not subtracted from, the cost of capital in the absence of deferral, r/θ. A firm that enjoys the advantage of deferral is founded with a lower stock of equity capital than one whose retained profits are subjected to the same tax rate as its dividends are.

The possibility of, and preference for, generating equity capital through profit retentions makes it wise to start with only a small nucleus of equity when the firm is founded and it eliminates the need for equity injections thereafter. It implies an extended period of internal growth during which the cost of capital is higher than the Edwards–Keen and Fullerton–King formulae predict, but it also implies that the firm will eventually reach a stage of maturity where the cost of capital is lower than Harberger assumed.

3. The Harberger problem

3.1. The traditional view

Consider now the Harberger problem more closely. Suppose there are two sectors, X and Y, competing for a given aggregate stock of capital $\bar{K}$:

$$\bar{K} = K_X + K_Y = \text{const.} \tag{19}$$

Sector X is subjected to dividend taxation and sector Y is untaxed. The sectors produce the same commodity, but are endowed with specific production functions $f_X(K_X)$ and $f_Y(K_Y)$.

In a narrow interpretation of the Harberger problem, X can be identified

[15]For a more extensive discussion of the deferral problem, see Sinn (1990b).

with the corporate and Y with the non-corporate sector. However, for the purposes of this paper, the two sectors can also be seen as two countries competing for the world capital stock $\bar{K}$, where country X imposes a source tax on distributed profits and country Y has no tax or a pure Schanz–Haig–Simons tax combined with the residence principle for the taxation of foreign capital income.[16]

Assume that, before t_1, there was only a non-corporate sector, Y, but at t_1 the corporate sector X is 'invented', the new production function being $f_X(\cdot)$. One may think of the corporation as a new form of organizing a firm which increases the efficiency of production and induces the government to impose a dividend tax in order to participate in the rents this form can be expected to generate. Alternatively, assume there is initially only one country, Y, which uses all the capital, and that, at t_1, another country, X, which has no capital, opens its borders and imposes a source tax to participate in the returns it can offer to international investors. In the light of the current stage of world history the sudden appearance of a capital–poor country is certainly not a strange assumption.

An efficient allocation of capital that maximizes aggregate output would require an immediate transfer of capital from Y to X until both sectors' marginal products of capital are equated: $f'_X(K_X)=f'_Y(K_Y)$. However, with a dividend tax this may not happen.

According to the Harberger–Kemp–MacDougall theory, the taxed sector invests up to the point where $\theta f'_X(K_X)=r$ and the untaxed sector retains capital so that $f'_Y(K_Y)=r$. In a capital market equilibrium the available stock of capital will therefore be allocated to the two sectors such that $\theta f'_X(K_X)= f'_Y(K_Y)$. This means that a constant wedge the size $\tau f'_X$ is being driven between the two marginal products of capital and that there is a permanent welfare loss in terms of a level of output that is less than is possible under the given aggregate stock of capital and the given set of investment opportunities.

The lower part of fig. 2 illustrates the Harberger–Kemp–MacDougall equilibrium. The downward-sloping and upward-sloping curves are the marginal product curves of the two sectors. The employment of capital in the taxed sector is measured from left to right and in the untaxed sector from right to left. The distance between the two verticals is the total amount of capital, $\bar{K}$, that is available. The stock of capital is optimally allocated to the two sectors when $K_X=DF$ and $K_Y=FG$, for then aggregate output, the area below the two curves, is maximized. However, the distorted allocation

[16]In fact, the assumption that the two sectors produce the same commodity fits even better to the Kemp–MacDougall foreign trade model than to the Harberger model. Harberger assumed that the two sectors were not allowed to produce the same commodities. However, nothing essential is lost if this assumption is relaxed as done here. For other examples of this simplified interpretation, see Gravelle and Kotlikoff (1989) and Sinn (1985, ch. 6).

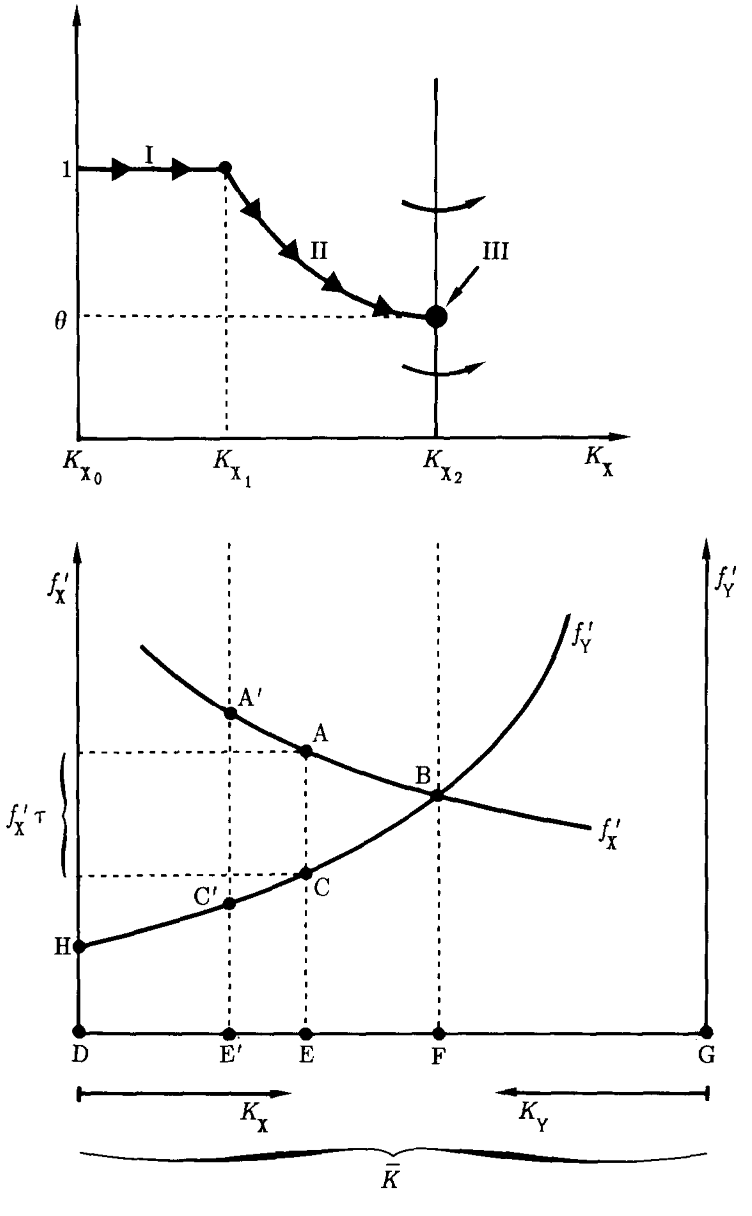

Fig. 2. The vanishing Harberger triangle.

believed to result from the dividend tax is characterized by $K_X = DE$ and $K_Y = EG$, for this allocation implies that the marginal product of capital in the taxed sector exceeds that in the untaxed sector by an amount sufficient to compensate for the tax discrimination. Obviously, output of the untaxed sector exceeds its optimum level by the amount $CBFE$, but this is overcompensated by a comparative output loss of size $ABFE$ in the taxed sector. The comparative net output loss of both sectors together is measured by the Harberger triangle ABC, and this triangle persists for as long as the dividend tax is levied.

The model set up in the previous section is not compatible with this result. To see this, only a few steps are necessary.

3.2. The Harberger problem in dynamic perspective

Notice first that the decision problem of a firm in the non-taxed sector can be seen as a special case of that model where $\theta = 1$; obviously this firm will invest up to the point where $f'_Y(K_Y) = r$. Because of (19), the equilibrium level of the market rate of interest is given by

$$r = f'_Y(\bar{K} - K_X) \tag{20}$$

as in the Harberger model.

The decision problem of a taxed firm was formulated in section 2 for the case of an arbitrarily given time path of the market rate of interest r, but the phase diagram of fig. 1 was analyzed assuming a constant rate of interest. An important, but straightforward step required is to generalize the discussion of this diagram to the case where the market rate of interest is endogenously determined by (20).[17]

When r is endogenous, nearly everything that has been said concerning Phase I and Phase III stays valid. In particular, it will still be true that $q = 1$ in Phase I and $q = \theta$ in Phase III. Again, Phase Ib cannot exist. On the one hand, $Q > 0$, $D = 0$, and (20) imply that $\dot{K}_X > 0$ and $\dot{r} = -f''_Y \cdot \dot{K}_X > 0$. On the other hand, (13) and $\dot{r} > 0$ indicate that $\dot{K}_X = \dot{r}/f''_X < 0$, a clear contradiction. The only important addition to the previous analysis is that the steady-state stock of capital, call it now K_{X_2}, that was previously determined by (16), is now implicitly given by $f'_X(K_{X_2}) = f'_Y(\bar{K} - K_{X_2})$. There are minor changes with the possible paths during Phase II, as defined by (14) and (15). Instead of (17), the slope of a path is now given by $dq/dK_X = q[f'_Y(\bar{K} - K_X) - f'_X(K_X)]/f_X(K_X)$. This equation reveals that the paths are more strongly

[17]Note that the time path of r is endogenous to the equilibrium, but not to the firm's planning problem. The firm is assumed to be a price-taker. It is not assumed that it has market power and believes that it can affect the time path of the market rate of interest through its own actions.

curved than in the case of a constant r, but clearly none of the qualitative properties described in fig. 1 and Propositions 1 and 2 is affected. Moreover, the appendix shows that (18) remains a sufficient condition for the initial cost of capital to exceed the Harberger value r/θ, and, in fact, an even weaker condition is provided. Thus, Proposition 3 also stays valid.

Fig. 2 illustrates the implications of intertemporal market equilibrium for the Harberger problem. In the beginning, there is only the untaxed sector or country Y so that the total amount of capital, DG, is invested there. Then, at t_1, the new sector defined by the new marginal product curve f'_X and the dividend tax factor θ appears on the scene. By issuing shares or attracting direct investment this sector will immediately withdraw the amount of capital DE' from the non-taxed sector and bid up the interest rate from HD to $C'E'$. The withdrawal is less than is required by efficiency and it is less than what is implied by the Harberger-type equilibrium condition, $\theta f'_X(K_X) = f'_Y(K_Y)$. There is a welfare loss in terms of an insufficient size of aggregate output and it is even bigger than Harberger thought. However, the welfare loss is only temporary. With the passage of time, the new sector will build up equity capital through profit retentions and claim a growing proportion of the economy's, or the world's, available stock of real assets. This improves the allocation of resources and increases aggregate output. The process comes to a halt when the Harberger triangle has vanished, the taxed sector's stock of capital has increased to DF, and the untaxed sector's stock has fallen to FG. The economy is then in an efficient steady state where its output is maximized given the available stock of capital and where all profits are distributed to the shareholders. The following proposition summarizes this conclusion.

Proposition 4. Initially, when the taxed sector is young and reinvests its profits, there is a Harberger triangle which, under mild conditions, exceeds the one predicted by Harberger's own theory. However, the triangle gradually vanishes with the passage of time. In finite time, the economy reaches a stage of maturity where the taxed sector distributes its earnings and the available stock of capital is being efficiently allocated to the two sectors.

3.3. The correlation between dividend taxes and the distortions they cause

There is a final proposition generated by the model which follows directly from the observation that no dividends, and hence no dividend taxes, are paid during the adjustment to the steady state.

Proposition 5. The dividend tax distorts the intersectoral allocation of resources when it is not *paid and it is neutral when it* is *paid.*

The payment of dividend taxes signals that the firm is in a stage of maturity where the new view of corporate taxation holds and where the dividend tax no longer affects the investment decisions. The burden of the tax is capitalized in share prices, and there is no way for a firm to reduce this burden by changing its investment policy. The very fact that dividend taxes are paid means that a marginal subsidy equal to the dividend tax rate is available for new investment, which is financed with a reduction in dividends. This subsidy exactly compensates for the tax on the returns from this investment and explains why dividend taxes are neutral when they are paid. By way of contrast, corporations that do not pay dividend taxes signal that they are in need of equity capital and have not yet reached the stage of maturity and efficiency. For them, the marginal subsidy implicit in the dividend tax is not immediately available. If they want to invest, they either have to rely on non-subsidized share issues or to wait and reinvest their future profits. The higher the tax rate, the more they are inclined to chose the second alternative and the larger are the distortions.

The lesson of the model is that dividend taxes create distortions *before* they are paid. The threat of dividend taxes that will have to be paid in the future makes shareholders reluctant to inject more than a nucleus of equity capital into their firms. However, when these taxes are actually paid, the process of reinvesting profits must have generated enough equity to compensate for this threat and to eliminate the Harberger triangle. In any given period of time there is a negative correlation between the size of the tax burden and the magnitude of the Harberger triangle.

4. Extensions and qualifications

4.1. Productivity shocks

An obvious extension of the model is to think of an initial steady-state equilibrium with two existing sectors that is disturbed by a new and unforeseen invention or economic revolution which shifts sector X's marginal product curve upward. This case can be captured with the formal approach derived above by changing the initial condition of problem (4) from $K_0=0$ to $K_0>0$. This would not affect the time path depicted in figs. 1 and 2 but would simply imply that the economy starts at a later stage on these paths. Consider first a *large* productivity shock. Suppose the new marginal product curve after the invention is the one illustrated in fig. 2 and the old curve intersected the marginal product curve of sector Y to the left of point C'. In this case, sector X issues new shares at the time the invention occurs and jumps immediately to point C'. After this, there is again the finite period of internal growth ending with the stage of maturity, B, where dividends are paid. Once again the dividend tax retards the adjustment process towards an efficient equilibrium, but does not prevent this equilibrium from being reached eventually.

Instead of the initial equilibrium being located to the left of point C' in fig. 2, it might also have been somewhere between C' and B, which is the case of a comparatively *small* productivity shock. Sector X will now not react to the inventions by issuing new shares but will merely stop paying dividends to its shareholders, entering a period of internal growth which eventually leads to the stage of efficiency and maturity.

4.2. Steady-state growth

Another extension is the introduction of steady economic growth where the aggregate supply of capital expands and the firm's marginal product curve shifts upward because of technological progress or the increased supply of other factors. In this environment the three phases of a firm's growth path still show up. However, since there is ongoing investment in the stage of maturity, it is no longer true that all profits are paid out as dividends. Instead, part of the firm's profit is retained to finance this investment and only the remainder is distributed as dividends.[18] Nevertheless, it remains true that dividend payments signal efficiency, and the absence of dividend payments signal distortions.

4.3. Escapes from the equity trap

A third extension would be to relax the constraint that dividends are the only way of distributing cash to shareholders. This constraint is certainly not realistic since most countries allow for a tax-exempt return of original capital. If such a return were permitted to replace or precede dividend payments, the firm's optimal growth path would be strongly affected. However, the typical provision – one that definitely applies in the United States, for example – is that a return of capital cannot occur before current profits and all accumulated reserves have been paid out. Formally, the possibility of returning the original capital *after* distributing the reserves implies that the constraint $Q \geqq 0$ of problem (4) is removed for $K \leqq K_1$ and maintained for $K > K_1$, where K_1 is the original capital (see figs. 1 and 2). Since this means that a flow constraint is removed for values of the state variable K where it is not binding ($q \geqq 1$), and retained where it is ($q < 1$), the

[18]It is not possible that net investment exceeds the return to capital permanently, since this would imply that the rate of growth of the firm's capital stock exceeds the market rate of interest and would hence violate the transversality condition of the firm's planning problem. Apart from that, the case where the economy's net investment exceeds its profit has little empirical relevance in OECD countries.

firm's optimal growth path and all of the conclusions based on it would remain unaffected.

Cash payments to shareholders that would undermine both Harberger's result and those derived in this paper are profit-financed share repurchases and acquisitions. Share purchases by corporations that are financed out of past and present profits largely avoid the double taxation of dividends and indicate a loophole in the classical and partial imputation systems of capital income taxation.[19] Most countries have effectively closed this loophole by declaring share repurchases illegal. This is particularly true for a subsidiary's share repurchases from its parent.[20] However, with publicly-owned U.S. corporations the situation is ambiguous. On the one hand, Section 302 of the Internal Revenue Code prohibits firms from repurchasing shares in lieu of dividend payments. On the other hand, share purchases by corporations have recently increased dramatically, constituting a large fraction of corporate cash distributions. A very extensive record of this phenomenon is provided by Shoven (1986).[21] He showed that, since 1984, the volume of corporate share purchases, predominantly acquisitions, exceeded ordinary dividend payments, yet he also found that the share purchases did not simply replace the dividends but reflected a leverage phenomenon. Quite remarkably, the time path of dividends was unaffected by the rising repurchase volume and debt rather than profits seemed to have been the source of the additional cash that shareholders received. Shoven's findings do not suggest that the recent increase in corporate share purchases was a use of profits from marginal investment projects that would eliminate distortions of Harberger type, including those examined in this paper.

5. Conclusion

This paper sits between the two stools of the 'old' and the 'new' view on dividend taxation.

It confirms the old view that a dividend tax creates efficiency losses. The tax is an obstacle to the foundation of firms and prevents capital from being used in the taxed sector or economy although it could usefully be employed there.

The paper also confirms the new view that the dividend tax is neutral when the economy is mature. The dividend tax does not create permanent

[19]If the share repurchases are a fixed fraction of the firm's profits, they affect the firm's investment decisions in the same way as a reduced dividend tax rate does. The initial stock of capital would be larger and the phase of internal growth shorter, but unless all profits are used for share repurchases, the qualitative predictions of the model studied in this paper would not change. Cf. Sinn (1990a, c).

[20]For example, the Subpart-F rules in the U.S. tax code classify share repurchases by controlled foreign subsidiaries as profit repatriations and require that they be taxed accordingly.

[21]Cf. also Poterba (1987, p. 471), Bagwell and Shoven (1989), and Sinn (1985, ch. 6).

distortions but merely retards the speed with which an efficient allocation is reached. There is a Harberger triangle, but it vanishes with the passage of time.

Yet, the model presented cannot easily be constructed from the existing building blocks offered by 'new' and 'old' models of dividend taxation. The cost-of-capital expressions used and developed in those models offer poor guidelines for predicting the magnitude and time patterns of distortions that this model generates. The reason is the emergence of a phase of purely internal growth where no dividends are paid and no equity is injected into the firm. Since the marginal profit from investment projects that are carried out before and during this phase is not distributed to shareholders, but is retained and reinvested at a rate of return above the market rate of interest, neither the 'new' nor the 'old' cost-of-capital formulae are applicable. As shown in sub-section 2.5, these formulae tend to underestimate the true cost of equity finance.

The three phases in the life of a firm that were shown to result from the dividend tax carry over to a whole sector or a whole economy when the firms are sufficiently homogeneous. This is the case underlying section 3, and to some degree it may well approximate existing economies. Surely the heartlands of western Europe are now basically in a stage of maturity, while eastern Europe prepares for Phase I.

In general, however, more account should be taken of the possibility of firm heterogeneity than was done here. Even in mature economies, there are always new people who tinker at the workbench in the garage and then found corporations by issuing new shares. Moreover, mature firms are often set back to Phase II or even Phase I because inventions or changed market conditions create more investment opportunities than current profits can finance. Even though the model does not allow for firms with different degrees of maturity, it has immediate implications for this case.

One such implication is the confirmation of the widespread belief that dividend taxation will have a negative impact on the entrants' starting stocks of equity and the incumbents' share issues after inventions. The tax is a continuing impediment to corporate investment, and there is a Harberger triangle all the time. Perhaps this is the explanation for the negative correlation between the level of investment and the statutory dividend tax rate that was mentioned in the Introduction.

Another implication is that the size of the triangle is an increasing function of the economy's growth rate. The higher this rate, the larger are the proportions of immature firms that have recently entered the markets or whose profits are temporarily insufficient to finance the required investment. A high growth rate therefore implies that many firms are in a situation where the dividend tax reduces their investment. A stationary economy, on the other hand, has a comparatively small percentage of firms affected by the dividend tax. This economy may not suffer overly large distortions when a dividend tax is levied.[22]

Measuring the size of the distortions would not be an easy task because parametric cost-of-capital expressions are not readily available. Nevertheless the model at least offers suggestions for a useful empirical approach. First of all, a distinction must be made between firms that pay dividends and those that do not – or that pay only minimal dividends to serve signaling purposes. In the first group, the cost of retained earnings as calculated by models of the Fullerton–King variety is the appropriate cost of capital. However, within the latter group, the cost of capital is higher, depending on the firms' proximity to the phase of new share issues or the phase of dividend payments, respectively. An approximate way of calculating this cost of capital would be to construct a weighted average of the cost of retentions à la Fullerton and King and the true cost of new share issues along the lines of section 2, where the weights are the time span elapsed since the last issues and the time span expected until the next dividend payments.

Alternatively, the firm's capital structure may be used to construct approximate weights. Young firms have a large share of original capital and mature firms a large share of surplus capital. This information could be exploited for calculating the proximity to maturity which turned out to be so important for a firm's cost of capital.

Interestingly enough, this possibility sheds a favorable light on the frequently criticized aspect of the Fullerton–King approach that average rather than marginal financial structures were used as weights in the cost-of-capital formulae, although it does not, of course, rehabilitate the use of the traditional formula for the cost of new share issues. Those who focus on marginal structures and use the Fullerton–King cost-of-capital expressions overlook the high cost of retained earnings for firms that are in Phase II. Since these firms use retained earnings as the only marginal source of finance, it is tempting, but fallacious, to treat them as mature firms. The use of average rather than marginal financial structures, originally motivated by practical considerations very different from those made here, may help avoid this fallacy. The measurement procedure of Fullerton and King therefore receives a new meaning from the phase model developed in this paper.

By way of contrast, Harberger's own measurement method cannot be supported. Harberger and many of his disciples based their welfare estimates on 'effective tax rates' defined as a sector's ratio of total capital income tax liability and total volume of capital income per unit of time. With a classical system of capital income taxation, where the overall tax burden on dividends exceeds that on retained profits, this means that the measured welfare loss will be higher, the higher the proportion of profits paid out as dividends, for

[22] I am grateful to one of the referees for pointing out this growth argument to me.

the higher this proportion, the higher is the measured value of the 'effective' tax rate.

If the spirit of the model presented in this paper is correct, this method of estimating the welfare loss stands the truth on its head. Given the tax law, a high 'effective' tax rate for the corporate sector signals, among other things, that many corporations are mature and pay dividends; and a low tax rate – one that approximates the tax rate of the non-corporate sector – signals that many corporations are in the transitory period of rapid internal growth. A high tax rate therefore signals small, and a low tax rate large, intersectoral distortions or, to put it another way, the true intersectoral distortions are smaller, the larger the distortions Harberger estimated. In a heterogeneous economy with a persistent flow of new inventions and entrants, the Harberger triangle does not vanish. What vanishes is the confidence that focusing on actual tax burdens reveals useful information on the magnitude of this triangle.

Appendix

By studying the functional form of the Phase II paths, this appendix derives a sufficient condition for Phase II to start with a capital stock lower than that implied by $f'(K)\theta = r$, or, equivalently, a sufficient condition for the cost of capital associated with new share issues to exceed r/θ. The proof applies to the general model of section 3 where r is endogenous. The constant value of r assumed in section 2 is a special case of this. It is assumed that $0 < \tau < 1$.

Let $q(K_X)$ be the function that describes the Phase II path in (q, K_X) space (see figs. 1 and 2) and let

$$q^H(K_X) = f'_X(K_X)\theta / f'_Y(\bar{K} - K_X) \tag{A.1}$$

be the value of q implicit in the Harberger approach. Let K^H denote the size of the corporate stock of capital where $q^H = 1$, i.e. the size that would be optimal if the cost of capital were r/θ. Obviously the true cost of new share issues exceeds r/θ if

$$K^H > K_{X_1}. \tag{A.2}$$

Since $q'(K_K) < 0$, this is equivalent to

$$q(K^H) < q^H(K^H). \tag{A.3}$$

Since the steady-state value of K_X, K_{X_2}, is defined by $f'_X(K_{X_2}) = f'_Y(\bar{K} - K_{X_2})$

and since $q(K_{X_2})=\theta$, it follows from (A.1) that $q^H(K_{X_2})=q(K_{X_2})=\theta$. A sufficient condition for (A.3) and hence (A.2) to hold is therefore

$$\frac{q'(K_X)}{q(K_X)}>\frac{q^{H\prime}(K_X)}{q^H(K_X)}, \quad \text{for } K^H \leqq K_X < K_{X_2}. \tag{A.4}$$

Using (15), the time derivative of q can be written as

$$\dot{q}=q'(K_X)\dot{K}_X=q'(K_X)f_X(K_X).$$

Inserting this into (14) and using (20) to explain r endogenously, one obtains:

$$\frac{q'(K_X)}{q(K_X)}=\frac{f'_Y(\bar{K}-K_X)}{f_X(K_X)}-\frac{f'_X(K_X)}{f_X(K_X)}. \tag{A.5}$$

By way of contrast, it follows from (A.1) that

$$\frac{q^{H\prime}(K_X)}{q^H(K_X)}=\frac{f''_X(K_X)}{f'_X(K_X)}+\frac{f''_Y(\bar{K}-K_X)}{f'_Y(\bar{K}-K_X)}. \tag{A.6}$$

Using (A.5) and (A.6), (A.4) can be transformed to

$$\left(-\frac{f''_X}{f'_X}K_X-\frac{f''_Y}{f'_Y}K_Y\frac{K_X}{K_Y}\right)\frac{f_X}{f'_XK_X}>\frac{f'_X-f'_Y}{f'_X}. \tag{A.7}$$

To interpret condition (A.7), it is useful to hypothesize that the pure profit or rent that the concave functions f_X and f_Y imply are the returns from a hidden second factor of production. Let $\sigma_i(K_i)$, $i=X, Y$, denote the Hicksian substitution elasticity between capital and the hidden factor in section i assuming that the production functions are linearly homogeneous. Moreover, let $\alpha_i(K_i)$ and $\beta_i(K_i)$, $i=X, Y$, denote the partial production elasticities of the hidden factor and of capital in the two sectors, respectively. It is a standard result that (A.7) can then be written as

$$\left(\frac{\alpha_X}{\sigma_X}+\frac{\alpha_Y}{\sigma_Y}\frac{K_X}{K_Y}\right)\frac{1}{\beta_X}>\frac{f'_X-f'_Y}{f'_X}. \tag{A.8}$$

Notice that, by the definition of K^H,

$$\frac{f'_X - f'_Y}{f'_X} \leqq \tau, \quad \text{for } K^H \leqq K_X < K_{X_2}.$$

This implies that it is sufficient for (A.8) and hence for (A.2) to hold if

$$\frac{\alpha_X/\beta_X}{\sigma_X} + \frac{\alpha_Y/\beta_X}{\sigma_Y}\frac{K_X}{K_Y} > \tau > 0.$$

This condition applies to the general case where the rate of interest is endogenously determined by the mechanics of the two-sector model of section 3, which imply that r is a rising function of K_X. However, in the limiting case of a relatively small size of the taxed sector ($K_X/K_Y \to 0$), it also captures the possibility of a constant rate of interest, as assumed in section 2.

References

Auerbach, A.J., 1979, Wealth maximization and the cost of capital, Quarterly Journal of Economics 94, 433–436.

Auerbach, A.J., 1983, Taxation, corporate financial policy and the cost of capital, Journal of Economic Literature 21, 905–940.

Bagwell, L.S. and J. Shoven, 1989, Cash distributions to shareholders, Journal of Economic Perspectives 3, 129–140.

Barlow, E.R. and I.I. Wender, 1955, Foreign investment and taxation (Halliday Lithograph Corp., Englewood Cliffs).

Bradford, D.F., 1980, The economics of tax policy toward savings, in: G.M. von Fürstenberg, ed., The government and capital formation (Ballinger Publ. Co., Cambridge, MA).

Bradford, D.F., 1981, The incidence and allocation effects of a tax on corporate distributions, Journal of Public Economics 15, 1–22.

Edwards, J.S.S. and M.J. Keen, 1984, Wealth maximization and the cost of capital: A comment, Quarterly Journal of Economics 99, 211–214.

Fazzari, S.M., R.G. Hubbard and B.C. Petersen, 1988, Financing constraints and corporate investment, Brookings Papers on Economic Activity, 141–195.

Fullerton, D. and M. King, 1984, The taxation of income from capital. A comparative study of the United States, The United Kingdom, Sweden and West Germany (University of Chicago Press, Chicago).

Gravelle, J.G. and L.J. Kotlikoff, 1989, The incidence and efficiency costs of corporate taxation when corporate and noncorporate firms produce the same good, Journal of Political Economy 79, 749–780.

Harberger, A.C., 1962, The incidence of the corporation income tax, Journal of Political Economy 70, 215–240.

Harberger, A.C., 1966, Efficiency effects of taxes on income from capital, in: M. Krzyzaniak, ed., Effects of corporation income tax (Wayne State University Press, Detroit).

Hamada, K., 1966, Strategic aspects of taxation on foreign investment income, Quarterly Journal of Economics 80, 361–375.

Hartman, D.G., 1985, Tax policy and foreign direct investment, Journal of Public Economics 26, 107–121.

Hines, J.R., 1989, Credit and deferral as international investment incentives, Unpublished manuscript, Princeton.

Hines, J.R. and R.G. Hubbard, 1989, Coming home to America: Dividend repatriations by U.S. multinationals, National Bureau of Economic Research, Working paper no. 2931.

Howitt, P. and H.-W. Sinn, 1989, Gradual reforms of capital income taxation, American Economic Review 79, 106–124.
Kemp, M.C., 1962, Foreign direct investment and national advantage, Economic Record 38, 56–62.
Kemp, M.C., 1964, The pure theory of international trade (Prentice-Hall, Englewood Cliffs).
King, M.A., 1974a, Taxation and the cost of capital, Review of Economic Studies 41, 21–35.
King, M.A., 1974b, Dividend behavior and the theory of the firm, Economica 41, 25–34.
King, M.A., 1977, Public policy and the corporation (Chapman and Hall, London, Wiley & Sons, New York).
King, M.A., 1989, Economic growth and the life-cycle of firms, European Economic Review 33, 325–334.
MacDougall, G.D.A., 1960, The benefits and costs of private investment from abroad: A theoretical approach, Oxford Bulletin of Economics and Statistics 22, 189–211.
Oberhauser, A., 1963, Finanzpolitik und private Vermögensbildung (Westdeutscher Verlag, Cologne).
Penrose, E.T., 1956, Foreign investment and the growth of the firm, Economic Journal 66, 220–255.
Penrose, E.T., 1959, The theory of the growth of the firm (Basil Blackwell, Oxford).
Poterba, J.M., 1987, Tax policy and corporate saving, Brookings Papers on Economic Activity, 455–515.
Poterba, J.M. and L.H. Summers, 1983, Dividend taxes, corporate investment, and Q, Journal of Public Economics 22, 135–167.
Poterba, J.M. and L.H. Summers, 1985, The economic effects of dividend taxation, in: E. Altman and M. Subrahmanyam, eds., Recent advance in corporate finance (Irwin, Homewood, Ill.).
Shoven, J.B., 1986, New developments in corporate finance and tax avoidance: Some evidence, National Bureau of Economic Research Working paper no. 2091.
Sinn, H.-W., 1984, Die Bedeutung des Accelerated Cost Recovery System für den internationalen Kapitalverkehr, Kyklos 37, 542–576.
Sinn, H.-W., 1985, Kapitaleinkommensbesteuerung, (Mohr, Tübingen); English edition: Capital income taxation and resource allocation (North-Holland, Amsterdam, New York) 1987.
Sinn, H.-W., 1988, The vanishing Harberger triangle, University of Munich discussion paper no. 88-05.
Sinn, H.-W., 1990a, Taxation and the cost of capital: The 'old' view, the 'new' view, and another view, forthcoming in: D. Bradford, ed., Tax policy and the economy (NBER, Cambridge, MA).
Sinn, H.-W., 1990b, Taxation and the birth of foreign subsidiaries, University of Munich discussion paper no. 90-30; forthcoming in: H. Herberg and N.V. Long, eds., Festschrift for Murray Kemp (Michigan University Press).
Sinn, H.-W., 1990c, Share repurchases, the 'new' view, and the cost of capital, Economics Letters, forthcoming.
Stiglitz, J.E., 1973, Taxation, corporate financial policy, and the cost of capital, Journal of Public Economics 2, 1–34.