2002

Tax Incidence

Don Fullerton, *University of Texas at Austin*
Gilbert E. Metcalf

Chapter 26

TAX INCIDENCE *

DON FULLERTON

Department of Economics, University of Texas, Austin; and NBER

GILBERT E. METCALF

Department of Economics, Tufts University; and NBER

Contents

* We are grateful for helpful suggestions from Alan Auerbach, Tom Barthold, Leora Friedberg, Michael Keen, and Rob Williams. This paper is part of the NBER's research program in Public Economics. Any opinions expressed in this paper are those of the authors and not those of the National Bureau of Economic Research.

Handbook of Public Economics, Volume 4, Edited by A.J. Auerbach and M. Feldstein

Abstract

This chapter reviews the concepts, methods, and results of studies that analyze the incidence of taxes. The purpose of such studies is to determine how the burden of a particular tax is allocated among consumers through higher product prices, workers through a lower wage rate, or other factors of production through lower rates of return to those factors. The methods might involve simple partial equilibrium models, analytical general equilibrium models, or computable general equilibrium models.

In a partial equilibrium model, the burden of a tax is shown to depend on the elasticity of supply relative to the elasticity of demand. Partial equilibrium models also are used to consider cases with imperfect competition.

In a two-sector general equilibrium model, a tax might be imposed on either commodity, on either factor of production, or on a factor used in one sector. The original use of this model is to analyze the corporate income tax as a tax on capital used only in one sector, the corporate sector. The model can be used to show when the burden falls only on capital or when the burden is shared with labor. The model also has been applied to the property tax, and results of the model have been used to calculate the overall burden on each income group.

Because the total stock of capital is fixed in that model, however, dynamic models are required to show how a tax on capital affects capital accumulation, future wage rates, and overall burdens. Such models might also provide analytical results or computational results. The most elaborate recent models calculate the lifetime incidence of each group. Finally, the chapter reviews the use of such incidence methods and results in the policy process.

Keywords

economic incidence, statutory incidence, tax shifting, distributional effects, payroll taxes, corporate income taxes, personal taxes, sales and excise taxes, general equilibrium models

JEL classification: H22

Introduction

Tax incidence is the study of who bears the economic burden of a tax. Broadly put, it is the positive analysis of the impact of taxes on the distribution of welfare within a society. It begins with the very basic insight that the person who has the legal obligation to make a tax payment may not be the person whose welfare is reduced by the presence of the tax. The statutory incidence of a tax refers to the distribution of tax payments based on the legal obligation to remit taxes to the government. Thus, for example, the statutory burden of the payroll tax in the United States is shared equally between employers and employees. Economists, quite rightly, focus on the economic incidence, which measures the changes in economic welfare in society arising from a tax. The standard view of the economic burden of the payroll tax in the United States is that it is borne entirely by employees.

Economic incidence differs from the statutory incidence because of changes in behavior and consequent changes in equilibrium prices. Consumers buy less of a taxed product, so firms produce less and buy fewer inputs – which changes the net price of each input. Thus, the job of the incidence analyst is to determine how those other prices change, and how those changes affect different kinds of individuals.

Incidence analyses abound in the literature, but they can be roughly classified into a few categories. In particular, when these studies analyze distributional effects of taxes across groups, Atkinson and Stiglitz (1980) note that we economists have used five different ways of dividing taxpayers into groups. First, we can focus on the impact of taxes on consumers as opposed to producers. A partial equilibrium diagram can identify both the loss of consumer surplus and the loss of producer surplus resulting from a tax. Second, we can narrow the focus to analyze the impact of a tax specifically on the relative demands for different factors and the returns to those factors (such as capital, labor, or land). The pathbreaking general equilibrium analysis of Harberger (1962) simply ignores the consumer side by assuming that everybody spends their money the same way, and then he derives the burden of a tax on capital as opposed to labor. Third, we can group individuals by some measure of economic well-being, in order to analyze the progressivity of a tax or tax system. Pechman and Okner (1974) is perhaps the classic analysis of the U.S. tax system that groups taxpayers by annual income, while Fullerton and Rogers (1993) group taxpayers by a measure of lifetime resources. Fourth, taxes can be evaluated on the basis of regional incidence. Such an analysis might focus on regional differences within a country [e.g., Bull, Hassett and Metcalf (1994)], or it might focus on international differences. Finally, taxes can have intergenerational effects. For example, insufficient social security taxes could bring about a transfer from future generations to the current generation. These effects can be captured by the generational accounting approach of Auerbach, Gokhale and Kotlikoff (1991), but see Barro (1974) for a dissenting view.

We begin in Section 1 with some definitions and concepts that will be used throughout this chapter. Next, we turn to a review of static analytical models of tax incidence. We begin with a simple partial equilibrium model, and then proceed to general equilibrium models. While many of the principles and lessons from partial equilibrium analysis carry over to general equilibrium analysis, the latter affords a greater richness and insight than do the partial equilibrium models. In addition, we find a number of instances of results that are "surprising", in the sense that the outcome in the general equilibrium model could not occur in a partial equilibrium model. Along the way, we present examples of empirical incidence analyses with estimates of the burden of the U.S. tax system or individual taxes in the U.S. system. All of these analyses assume perfectly competitive markets, and Section 3 provides a discussion of incidence in imperfectly competitive markets.

In Section 4, we turn to dynamic models. Allowing for endogenous capital accumulation adds both an important type of behavioral change and considerable complexity. Dynamic models also allow the researcher to distinguish between "old" and "new" capital, a source of considerable redistribution in the case of tax reforms. Section 5 continues the analysis in a dynamic framework by investigating the incidence of tax systems over the life cycle. If individuals make consumption decisions on the basis of lifetime income [Modigliani and Brumberg (1954)], then annual income analyses of consumption taxes might be biased towards finding regressivity. Fullerton and Rogers (1993) have looked most thoroughly at this question, and interestingly, they find that the bias predicted by others is not nearly as severe as predicted.

Section 6 focuses on the use of distributional analysis in the policy process. Policy economists face an inherent tradeoff between theoretical rigor and the need for rapid, easily-comprehensible distributional analysis. Economists at several government agencies have refined the available techniques for measuring and reporting incidence impacts of taxes. In this section, we describe both the techniques used to analyze taxes and methods of presenting information to policy makers so that they can make informed decisions. Naturally, other economists have criticized many of the techniques used in the policy process, and we review some of those criticisms here.

Finally, we note that incidence analysis can be more broadly applied than we do in this chapter. We ignore incidence analyses of government spending programs [e.g., Musgrave, Case and Leonard (1974) or McClellan and Skinner (1997)]. Such a spending program can also affect relative prices, and so economic incidence again can differ from statutory incidence. The principles and concepts described in this chapter are not limited to tax analysis and can easily be applied to government spending programs as well.

1. Basic machinery of incidence analysis

In this section we sketch out various concepts and definitions that are commonly used in incidence analysis. We also describe and provide some motivation for analytic techniques that we will use frequently in this chapter.

1.1. Definitions and concepts

A number of concepts are used in incidence analyses. In the introduction, we already drew a distinction between *statutory incidence* (the legal payers of the tax) and *economic incidence* (those who lose real income). We now make further distinctions that are useful to sharpen our understanding of the incidence of various taxes.

To begin, economists might say that a commodity tax is *passed forward,* which means that the consumer price rises and consumers of that good bear the burden. The price received by the supplier might be unchanged. On the other hand, if the consumer price is unchanged when a commodity tax is imposed, then the price received by the supplier must fall. In that case, the burden is *passed backward* onto suppliers (or more precisely, onto labor, capital, or other factors in production). Similarly, a tax that is passed forward to consumers has burdens on the "*uses side*" (depending on how people use their income), while a tax that is passed backward has burdens on the "*sources side*" (because labor and capital are sources of income).

All of these terms must be employed with care. A longstanding principle in tax incidence analysis is that real burdens depend on real allocations, not on the price level or choice of numeraire. Thus, even for a tax on a particular commodity, the true incidence does not depend on whether monetary authorities accommodate by allowing an increase in that price (and thus in the overall price level). Only relative prices matter. Because the price level is irrelevant, however, so must be the question about whether the overall burden is on the uses side or the sources side! Instead, what matters is how changes in relative output prices affect different groups (if some spend more than the average share of income on the taxed good), and how changes in relative factor prices affect different groups (if some earn more than the average share of income from the factor employed intensively in the taxed industry).

Thus, the first job for a complete incidence study is to determine effects on all relative prices. A study might legitimately focus just on the uses side if groups have different spending patterns but all have the same sources of income (or if the taxed industry uses the average capital/labor ratio so that reduced production does not affect relative factor prices). Conversely, a study might focus just on the sources side if all groups spend the same fraction of income on the taxed good (and the taxed industry makes intensive use of labor, capital, or other factors). If the tax affects both output prices and factor prices, then a complete study would divide individuals into groups based on some measure of income, obtain data on all sources of income and all uses of income of each group, and use that data to calculate each group's net economic burden from a tax.

Regardless of how the burden is calculated, for each income group, their relative burdens of a tax can be compared using the ratio of the economic burden to income. A tax is said to be *progressive* if this ratio rises with income, *regressive* if it falls with income, and *proportional* if the ratio is constant. A common misconception is that progressivity is defined by rising marginal tax rates. For example, a flat tax or negative income tax can have a constant marginal tax rate and still be progressive. Let the tax liability (T) be the following linear function of income (Y):

$$T = m(Y - A), \tag{1.1}$$

where m is the marginal tax rate, and $A > 0$ is a family allowance. If income falls below A, then T can be negative (the taxpayer receives a payment from the government)[1]. With this tax system, the average tax rate (T/Y) starts at negative infinity, rises to zero at an income level equal to A, and then continues to rise with income (approaching m asymptotically). This tax is progressive, because the average tax rate rises with income, despite the fact that it has a constant marginal tax rate. For a different example, the Medicare portion of the payroll tax on employees has a constant marginal rate of 2.9%, but this tax is regressive because it applies only to wage income (while non-wage income tends to be concentrated in higher income groups)[2].

Care also is required when we define the incidence experiment. In particular, when we want to determine the distributional effects of raising a particular tax, we need to specify what is done with the revenues. While partial equilibrium incidence analyses often ignore the distribution of the proceeds, a more complete analysis takes into account what is done with the tax revenue. Logically, we have three alternatives. First, *absolute* incidence analysis refers to the assumption that the proceeds of the tax under investigation are simply held by government, but then a full analysis would need to consider the effects of the change in government debt. Second, a *balanced-budget* incidence analysis is one that assumes the revenue is spent, but then the distributional effects depend on how the revenue is spent[3]. Third, a *differential* incidence analysis assumes that the revenue is used to reduce some other tax, but then the distributional effects depend on the effects of the tax being reduced. None of these alternatives isolates the effects of the tax being raised! Still, however, one way to neutralize the effects of the use of the revenue is to assume that the government spends it exactly the same way that consumers would have spent it [as in Harberger (1962)]. This balanced-budget incidence analysis is equivalent to a differential analysis that uses the revenue to reduce lump-sum taxes on consumers – but only if the money goes to exactly the same individuals who were bearing the burden, so that they can spend it the same way

[1] The Flat Tax has been proposed in many forms. Perhaps the most well-known variant is due to Hall and Rabushka (1995). Some plans have $T = \max[0, m(Y - A)]$, so taxes are only positive, but $A > 0$ still means that the system is progressive: the average tax rate (T/Y) is zero up to income $Y = A$, and then it starts to rise with Y. Because T can be negative in Equation (1.1), this system is often called a Negative Income Tax.

[2] This statement ignores the benefits arising from the Medicare system, a point we take up below, as well as the employer portion of the tax. However, our statement about the regressivity of the tax is not affected by the fact that employers pay half the tax.

[3] For example, the regressive effects of the social security payroll tax are substantially modified if one includes the effects of using those revenues to provide progressive social security benefits.

they were spending it before the first tax was imposed. Any other use of the revenue with altered spending could itself affect prices.

An advantage of differential incidence analyses with lump-sum tax rebates is that different analyses are additive in the following sense. If one study considers tax proposal A with proceeds used to lower lump-sum taxes by X, and a second study considers tax proposal B with proceeds used to lower lump-sum taxes by X, then the two studies can be combined to analyze the differential incidence of a shift from tax system A to tax system B (or vice versa). Fullerton and Rogers (1997) illustrate how differential tax incidence can modify conventional thinking in the case of a uniform consumption tax. Normally, a uniform consumption tax has the attractive property that no commodity is tax-advantaged[4]. Yet, Fullerton and Rogers note that relative prices still change, and consumers are differentially affected, if the uniform consumption tax is used to replace an existing system that *does* have differential commodity taxes.

Up to now, we have been a bit vague as to the meaning of the *burden* of a tax. A straightforward measure of the burden of a tax is the equivalent (or compensating) variation. The equivalent variation (EV) is the amount of lump-sum income that a person would give up to avoid a particular tax change (such as the imposition of a tax or a complex change to a system of taxes). So long as the taxpayer can take some action to influence the amount of taxes paid (short of tax evasion), the EV will exceed the tax revenue collected from the taxpayer – and the difference is defined as the deadweight loss of the tax. The true economic burden of a tax, therefore, exceeds the revenue loss to the taxpayer unless the tax is lump-sum in nature. Figure 1.1 illustrates. A commodity (X) is provided with perfectly elastic supply, S. The Marshallian demand curve is D^M. Prior to a tax, CF is purchased at a price of 0C. When a tax on X is imposed, the supply curve shifts up to S' (to reflect the cost of production inclusive of the tax). Demand falls to AB and tax revenue of ABDC is collected. The equivalent variation for this tax is the area between the old and new prices to the left of the compensated demand curve (D^C) and equals ABEC. It exceeds the taxes collected by the deadweight loss triangle BDE.

Note the strong informational requirements for this measure of tax burden. The researcher needs to know the utility function (or equivalently the expenditure function) to measure EV[5]. As we shall note below, a number of alternative measures of the burden of a tax are used in practice. A second approach is to measure the change in consumer's surplus. Willig (1976) provides bounds on the income elasticity of demand under which the change in consumer's surplus provides a good approximation of EV. In Figure 1.1, the change in consumer's surplus is ABFC. A third approach is to measure

[4] Note, however, that Ramsey (1928) considerations provide no optimal tax rationale for uniform consumption taxation except in certain circumstances.

[5] Hausman (1981) shows how to recover the utility function and thus to derive the EV from observed Marshallian demand functions. While this insight is important, it simply pushes back the information problem from that of specifying the utility function correctly to that of specifying the demand function correctly.

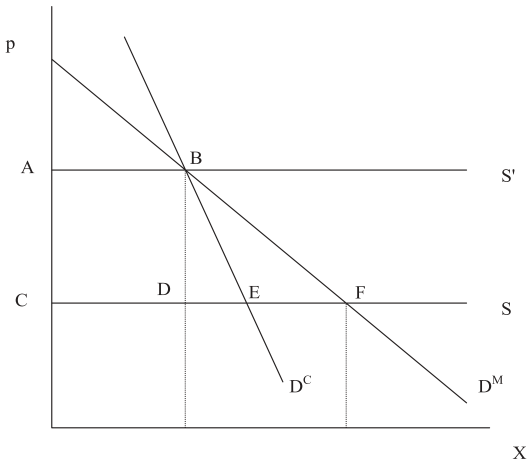

Fig. 1.1. The economic burden of a tax on X.

the tax actually paid (ABDC in Figure 1.1)[6]. This approach ignores the component of the economic burden arising from the deadweight loss. For small taxes, this can provide a good approximation to the true burden of the tax, but for large taxes it can significantly underestimate the true burden. Finally, another common approach is simply to look at the change in net-of-tax prices following tax changes. In Figure 1.1, only the consumer price changes (by AC), and the full burden of the tax is said to be on the consumer.

Before finishing basic concepts and definitions, we have a few other useful terms. A *unit* tax (t) is applied at a particular dollar amount per unit of the good or factor, and so it raises a price from p to $p + t$. An example is a "specific" excise tax. In contrast, an *ad valorem* tax (τ) is some fraction or percentage of the product price, and so it raises a price from p to $p(1 + \tau)$. An example is a local 8% sales tax. Any particular tax law might be worded either way, and it might be analyzed either way so long as the researcher is careful to employ the proper correspondences (such as $\tau = t/p$)[7]. For consistency, we use just *ad valorem* rates below.

Another definitional device useful to incidence analysts is the *unit convention,* which is just a way to define what is one unit of a good. Apples can be priced per pound, per ton, or per bushel, and this choice has no real effect even though the price looks very different. Therefore, we can define a unit as whatever amount costs one dollar (before taxes). Then the initial price is one, and we can focus on tax changes that

[6] The EV is the measure of burden in computational general equilibrium (CGE) models discussed below, while the tax actually paid is used as the measure of burden in studies with incidence assumptions [such as Pechman and Okner (1974)]. For taxes paid by businesses, such studies use specific incidence assumptions to allocate the tax burden among income groups.

[7] The different wording of the tax has been shown to matter in particular models, such as those with imperfect competition. See Section 3 below.

may raise that price or lower it. Similarly, if one person buys a car for \$20 000 while another buys a car for \$10 000, we simply say that the first person has purchased twice as much car. The price they face is the same (\$1 per unit). This convention has the added advantage that a one-cent unit tax is the same as a one percent *ad valorem* tax.

Finally, we must be careful about what is in the denominator of the tax rate. A *tax-exclusive* rate is expressed as a fraction of the price excluding tax, while a *tax-inclusive* rate refers to a fraction of the price including tax. An example of the former is a 50% sales tax that raises the price from \$1 to \$1.50, and an example of the latter is an income tax that takes 33% of all income. These numbers were chosen to make the point that the individual may be indifferent between these two taxes, since government takes one-third of real resources either way. But it matters to the researcher: a 50% sales tax is *not* the same as a 50% income tax! In this chapter, we primarily use a tax-exclusive rate, so the net price is p and the gross price is $p(1+\tau)$. Similarly, if τ is a wage tax, then the net wage is w and the gross wage is $w(1+\tau)$. This latter rate needs to be interpreted carefully since it is not the usual income tax rate.

1.2. Log-linearization

Many recent studies of tax incidence have built large-scale computable general equilibrium models that specify particular functional forms for production and for consumer behavior and then calculate the effects of a large tax change on each product price and on each factor return. Such models are necessary in order to capture much detail with many production sectors, consumer groups that own different factors and buy different goods, and large taxes that have non-marginal effects on prices.

On the other hand, many interesting conceptual questions of tax incidence can be addressed using small models that can be solved analytically. Because we address many such questions in this chapter using analytical "log-linearization" methods attributable to Jones (1965), and because we wish to convey the methods of tax incidence analysis to graduate students in economics, we now explain this method quite fully at the outset. The basic point of this method is to be able to specify a set of general non-linear production functions and consumer behavioral relationships, to convert these equations into a set of simpler linear equations, and then to solve these linear equations in a way that shows quite clearly the effect of a tax change on each price and on each quantity.

To explain why it is called *log*-linearization, consider the wage tax example mentioned above where the net wage is w, the gross wage is $w(1+\tau)$, and the price of consumption is p. Defining W as the real gross wage cost to the firm, we have:

$$W = w(1+\tau)/p. \tag{1.2}$$

To make this nonlinear equation into a linear relationship, take natural logs of both sides,

$$\ln(W) = \ln(w) + \ln(1+\tau) - \ln(p), \tag{1.3}$$

and then differentiate:

$$\mathrm{d}W/W = \mathrm{d}w/w + \mathrm{d}\tau/(1+\tau) - \mathrm{d}p/p. \quad (1.4)$$

Next, use a "hat" to denote a proportional change, so $\hat{W} \equiv \mathrm{d}W/W$ and $\hat{p} \equiv \mathrm{d}p/p$. For convenience, every tax rate is treated a little differently, where $\hat{\tau} \equiv \mathrm{d}\tau/(1+\tau)$. Using these definitions, we have:

$$\hat{W} = \hat{w} + \hat{\tau} - \hat{p}. \quad (1.5)$$

The nonlinear Equation (1.2) might be part of a system of nonlinear equations that is difficult to solve, but this "log-linearization" technique can be applied to every one of those nonlinear equations to produce a system of linear equations like Equation (1.5)[8]. If the system has N equations with N unknowns, then it is easy to solve (using successive substitution or Cramer's Rule). For example, if the goal is to calculate the effects of a tax change, $\hat{\tau}$, then the relevant unknowns might include changes in equilibrium prices ($\hat{W}, \hat{w}, \hat{p}$) and changes in equilibrium quantities such as labor, capital, and output.

Before getting to a general equilibrium system of such equations, however, we provide a complete illustration of the log-linearization technique for a simple partial equilibrium model of just the labor market. Thus, other prices are fixed (so $\hat{p} = 0$, and $\hat{W} = \hat{w} + \hat{\tau}$). Even this simple model yields important and interesting results, however, regarding the difference between statutory and economic incidence. Because workers receive the net wage w, employers bear the statutory burden and face the gross wage cost $w(1+\tau)$. Depending on labor demand and supply behaviors, however, the burden can be shifted through a change in the equilibrium net wage.

To model such behavior, first consider the definition of the elasticity of labor supply (L^S) with respect to the net wage (w):

$$\eta^S \equiv \frac{\mathrm{d}L^S/L^S}{\mathrm{d}w/w}. \quad (1.6)$$

Using the hat notation ($\hat{L}^S = \mathrm{d}L^S/L^S$), the nonlinear relationship in Equation (1.6) can be rewritten as $\eta^S \equiv \hat{L}^S/\hat{w}$, and further re-arrangement provides:

$$\hat{L}^S = \eta^S \hat{w}. \quad (1.7)$$

The point here is that we have taken a definition and turned it into a behavioral equation: if the net wage changes by a certain amount, then Equation (1.7) tells us

[8] Log-linearization is simply a first-order Taylor series approximation around the initial equilibrium. It is completely appropriate for calculating the effects of a small tax change, but sometimes the method has been applied to a large tax change such as the repeal of a tax – as if all of the derivatives were constant.

how labor supply responds[9]. It is one linear equation for our system. Next, if η^D is the elasticity of labor demand (L^D) with respect to the gross wage (W), then similar rearrangement provides

$$\hat{L}^D = \eta^D(\hat{w} + \hat{\tau}). \tag{1.8}$$

In this model, we assume that $\eta^D \leqslant 0$ and $\eta^S \geqslant 0$ are known parameters. In response to an exogenous tax increase ($\hat{\tau} > 0$), behaviors follow Equations (1.7) and (1.8), but reaching a new equilibrium means that the change in labor demand must equal the change in labor supply:

$$\hat{L}^S = \hat{L}^D. \tag{1.9}$$

We now have a system of three linear Equations (1.7, 1.8 and 1.9) in three unknowns ($\hat{L}^S$, $\hat{L}^D$, and $\hat{w}$). We can solve for $\hat{w}$ in terms of exogenous parameters (η^S, η^D, and $\hat{\tau}$) by setting Equation (1.7) equal to Equation (1.8) and re-arranging:

$$\frac{\hat{w}}{\hat{\tau}} = \frac{\eta^D}{\eta^S - \eta^D}. \tag{1.10}$$

The expression in Equation (1.10) lies between 0 and −1, and it shows what fraction of the tax is shifted from employers to workers[10]. Each side of the market tries to avoid the tax by changing behavior: a larger labor supply elasticity ($\eta^S \gg 0$) in Equation (1.10) means a smaller fall in the net wage to workers ($\hat{w}$)[11]. Or, if employers can be more elastic (larger $\eta^D < 0$), Equation (1.10) implies a larger fall in w (and therefore less increase in the gross wage cost of employers). Certain special cases deserve mention: if labor supply is perfectly inelastic ($\eta^S = 0$), or if labor demand is perfectly elastic (η^D infinite), then the right-hand side of Equation (1.10) is −1, and $\hat{w} = -\hat{\tau}$. Then the net wage w falls by the full amount of the tax, with no change in the gross wage cost to employers.

The principle illustrated in Equation (1.10) extends to a tax in any kind of competitive market. For example, a commodity tax burden will be shared by consumers and producers based on the relative elasticities of demand and supply[12].

[9] These elasticity definitions and resulting behavioral equations provide simple examples of log-linearization, but later sections take more care to derive such behaviors from first principles. In Section 2.2, we formally develop the relationship between the labor supply elasticity and primitive preference parameters.

[10] In terms of the measures of "burden" discussed in Section 1.1, this approach uses the price change itself rather than the dollar amount of tax paid or the equivalent variation.

[11] More precisely, η^S must be large relative to $-\eta^D$.

[12] Hines, Hlinko and Lubke (1995) show that when demand and marginal cost curves are linear, both buyers and sellers face the same *percentage* reduction in surplus upon introduction of a commodity tax regardless of demand and supply elasticities. While the burden on consumers may be higher in absolute terms if demand is relatively less elastic than supply, Hines et al. note that the benefits of the market accrue predominantly to consumers (i.e., consumer surplus prior to the tax is greater than producer surplus). The authors interpret this result as support for viewing commodity taxes as flat rate taxes on market surplus, analogous to flat rate income taxes.

We leave as a simple exercise the derivation of the economic incidence of a tax on wage income when the statutory incidence of the tax is on workers rather than on employers [13]. This exercise demonstrates an important principle: in markets with no impediments to market clearing, the economic incidence of a tax depends only on behavior (η^S and η^D) and not on legislative intent (statutory incidence).

We next show some log-linearization techniques that are useful for building a general equilibrium model where supplies and demands are not specified directly, as above, but are instead based on maximizing behavior. Suppose that an output X is produced using both labor L and capital K with constant returns to scale:

$$X = F(K, L). \tag{1.11}$$

This functional form is very general and nonlinear. Differentiate to get:

$$\mathrm{d}X = F_K \mathrm{d}K + F_L \mathrm{d}L, \tag{1.12}$$

where F_K is the marginal product of capital ($\partial F/\partial K$), and F_L is the marginal product of labor ($\partial F/\partial L$). Divide through by X, and we have:

$$\frac{\mathrm{d}X}{X} = \frac{F_K K}{X} \cdot \frac{\mathrm{d}K}{K} + \frac{F_L L}{X} \cdot \frac{\mathrm{d}L}{L}. \tag{1.13}$$

Define θ as the factor share for capital ($rK/p_X X$), where r is the rental price of capital and p_X is the price of X. With perfect competition, where $r = p_X F_K$ and $W = p_X F_L$, the factor share for capital will equal $F_K K/X$ and the factor share for labor will equal $F_L L/X$. And with constant returns to scale, factor shares sum to one, so Equation (1.13) becomes:

$$\hat{X} = \theta \hat{K} + (1 - \theta)\hat{L}. \tag{1.14}$$

While the production function tells us how total labor and capital yield total output, this differential equation tells us how small changes in labor and capital yield changes in output. It is a linear equation in three of the important unknowns ($\hat{X}$, $\hat{K}$, $\hat{L}$).

Finally, for this section, consider the definition of the elasticity of substitution between capital and labor in production (omitting taxes for the moment):

$$\sigma \equiv \frac{\mathrm{d}(K/L)/(K/L)}{\mathrm{d}(w/r)/(w/r)}. \tag{1.15}$$

If we do the differentiation in the numerator, it becomes

$$\frac{L\mathrm{d}K - K\mathrm{d}L}{L^2} \cdot \frac{L}{K} = \frac{\mathrm{d}K}{K} - \frac{\mathrm{d}L}{L} = \hat{K} - \hat{L}. \tag{1.16}$$

[13] This exercise would require redefinition of w as the gross wage and $w(1 - \tau)$ as the net wage.

Then, with a similar differentiation of the denominator, we have:

$$\sigma = \frac{\hat{K} - \hat{L}}{\hat{w} - \hat{r}}. \tag{1.17}$$

In fact, many use Equation (1.17) directly as the definition of the elasticity of substitution. A simple rearrangement of the definition turns it into a statement about behavior:

$$\hat{K} - \hat{L} = \sigma(\hat{w} - \hat{r}). \tag{1.18}$$

This procedure converts the complicated nonlinear Equation (1.15) into a linear equation. With the labor tax, where firms react to the gross wage $w(1 + \tau)$, we would have

$$\hat{K} - \hat{L} = \sigma(\hat{w} + \hat{\tau} - \hat{r}). \tag{1.19}$$

For any exogenous tax change (with endogenous change in the wage and interest rate), Equation (1.19) tells us how the firm reacts by changing its use of labor and capital. It is one more linear equation for our system.

While a computational general equilibrium model must specify a particular functional form for production, such as Cobb–Douglas or Constant Elasticity of Substitution (CES), the production function in Equation (1.11) avoids this limitation. It can be any function with constant returns to scale. However, this log-linearization method is valid only for small changes. It does not require a constant factor share θ (as in Cobb–Douglas) or a constant elasticity of substitution σ (as in CES); instead, it only requires that we know the initial observed θ and σ. In the rest of this chapter, we will use this logic to arrive at equations like (1.14) and (1.19) virtually without explanation.

The main purpose of this subsection was to define log-linearization and to provide a few examples. That purpose is completed, and so we are ready to start using this method to derive important incidence results.

2. Static analytical models

We begin our survey by looking at static economic models of tax incidence. Such models are particularly good for analyzing taxes that do not affect saving or investment. Many of the insights that we can glean from these models are more general and carry over to richer, complex models with a full specification of saving, investment, and intertemporal optimization.

2.1. Two-sector general equilibrium model

We first turn to the two-sector general equilibrium model with two factors of production (capital, K, and labor, L). Production of two goods (X and Y) occurs in a constant returns to scale environment:

$$X = F(K_X, L_X), \qquad Y = G(K_Y, L_Y). \tag{2.1}$$

Each factor has a fixed total supply but can freely migrate to either sector (with no unemployment). Thus

$$K_X + K_Y = \overline{K}, \qquad L_X + L_Y = \overline{L}. \tag{2.2}$$

Also, since each factor is fully mobile between sectors, it must earn the same after-tax return in both sectors [14]. Harberger (1962) used this model to consider a tax on capital in one sector. Before considering Harberger's specific experiment, we set up the model more generally to consider a number of taxes. In all cases, we return the tax proceeds lump sum to consumers, all of whom are identical. Because all consumers spend their money the same way, we can focus on incidence effects on the sources side [15]. Income for capital is $r\overline{K}$ (where r is the nominal return to capital), while income for labor is $w\overline{L}$ (where w is the nominal wage rate). Since $\overline{K}$ and $\overline{L}$ are fixed, we can focus on changes in the ratio of r to w to see how the burden of the tax is shared.

We develop the model using equations of change, the log-linearization method of Jones (1965) described above. Totally differentiate the equations in (2.2) to get

$$\lambda_{LX}\hat{L}_X + \lambda_{LY}\hat{L}_Y = 0, \qquad \lambda_{KX}\hat{K}_X + \lambda_{KY}\hat{K}_Y = 0, \tag{2.3}$$

where λ_{LX} is the fraction of labor used in the production of X (the original $L_X/\overline{L}$, before the change). The other λ terms are defined similarly.

Production technology can be represented by the elasticity of substitution between K and L for each good (σ_X and σ_Y):

$$\hat{K}_X - \hat{L}_X = \sigma_X\left(\hat{w} + \hat{\tau}_{LX} - \hat{r} - \hat{\tau}_{KX}\right), \qquad \hat{K}_Y - \hat{L}_Y = \sigma_Y\left(\hat{w} + \hat{\tau}_{LY} - \hat{r} - \hat{\tau}_{KY}\right), \tag{2.4}$$

where $\hat{\tau}_{ij} = \mathrm{d}\tau_{ij}/1 + \tau_{ij}$ is a tax on factor income ($i = L, K$) in the production of good $j (j = X, Y)$.

[14] This model is characterized by the "perfect" assumptions (such as perfect competition, perfect mobility, perfect information, and perfect certainty). Harberger (1962) provided an extremely useful benchmark case that can be solved easily, and he established a research agenda for virtually all of the following incidence literature: what happens with imperfect competition, imperfect mobility, uncertainty, variable factor supplies, unemployment, nonconstant returns to scale, an open economy, some other distortion such as an externality, more than two factors, more than two sectors, or more than one type of consumer?

[15] Harberger assumed homothetic and identical preferences and that government used the revenue to purchase X and Y in the same proportions as do consumers. With either Harberger's assumption or ours, one can ignore uses side effects of the partial factor tax.

Capital is paid the value of its marginal product in competitive markets:

$$p_X F_K = r\,(1 + \tau_{KX})\,, \qquad p_Y G_K = r\,(1 + \tau_{KY})\,, \tag{2.5}$$

just as labor is paid the value of its marginal product in each industry:

$$p_X F_L = w\,(1 + \tau_{LX})\,, \qquad p_Y G_L = w\,(1 + \tau_{LY})\,, \tag{2.6}$$

where p_X is the producer price of X and p_Y the producer price of Y. Given Equations (2.5) and (2.6), and constant returns to scale, the value of output in each industry must equal factor payments:

$$p_X X = w\,(1 + \tau_{LX})\;L_X + r\,(1 + \tau_{KX})\;K_X, \quad p_Y Y = w\,(1 + \tau_{LY})\;L_Y + r\,(1 + \tau_{KY})\;K_Y. \tag{2.7}$$

Totally differentiate the equations in (2.7) and evaluate at $\tau_{ij} = 0$ to obtain:

$$\begin{aligned} \hat{p}_X + \hat{X} &= \theta_{KX}\left(\hat{r} + \hat{\tau}_{KX} + \hat{K}_X\right) + \theta_{LX}\left(\hat{w} + \hat{\tau}_{LX} + \hat{L}_X\right), \\ \hat{p}_Y + \hat{Y} &= \theta_{KY}\left(\hat{r} + \hat{\tau}_{KY} + \hat{K}_Y\right) + \theta_{LY}\left(\hat{w} + \hat{\tau}_{LY} + \hat{L}_Y\right), \end{aligned} \tag{2.8}$$

where the θ's are the factor shares. For example, θ_{KX} is the share of sales revenue in sector X that is paid for capital ($\theta_{KX} \equiv r(1 + \tau_{KX})K_X/(p_X X)$).

In a similar fashion, we can totally differentiate the production functions in Equation (2.1) and use Equations (2.5) and (2.6) to obtain

$$\hat{X} = \theta_{KX}\hat{K}_X + \theta_{LX}\hat{L}_X, \qquad \hat{Y} = \theta_{KY}\hat{K}_Y + \theta_{LY}\hat{L}_Y. \tag{2.9}$$

Note for future reference that the shares of each factor's use add to one,

$$\lambda_{LX} + \lambda_{LY} = 1, \qquad \lambda_{KX} + \lambda_{KY} = 1, \tag{2.10}$$

and that the value shares going to each factor within an industry must add to one:

$$\theta_{KX} + \theta_{LX} = 1, \qquad \theta_{KY} + \theta_{KY} = 1. \tag{2.11}$$

Finally, we can characterize consumer preferences by the elasticity of substitution (in demand) between X and Y (σ_D):[16]

$$\hat{X} - \hat{Y} = -\sigma_D\left(\hat{p}_X + \hat{\tau}_X - \hat{p}_Y - \hat{\tau}_Y\right), \tag{2.12}$$

where the consumer price for X is $p_x(1 + \tau_x)$ and τ_x is an *ad valorem* tax on X. The consumer price for Y is similarly defined.

[16] Consumer behavior is captured by preferences (as represented by the elasticity of substitution between X and Y) and the budget constraint. Equation (2.12) would also hold in a more general model with a labor–leisure choice if leisure is separable and the sub-utility function for X and Y is homothetic. The consumer budget constraint here is unnecessary, as it is implied by Equation (2.7) and the assumption that tax revenues are rebated lump sum to consumers (an example of Walras's Law).

Equations (2.3), (2.4), (2.8), (2.9) and (2.12) are nine equations in the ten unknowns $\hat{X}$, $\hat{Y}$, $\hat{p}_X$, $\hat{p}_Y$, $\hat{w}$, $\hat{r}$, $\hat{L}_X$, $\hat{L}_Y$, $\hat{K}_X$ and $\hat{K}_Y$. Since we focus on real behavior (no money illusion), we must choose a numeraire (fix one of the price changes to zero), giving us nine equations in nine unknowns.

Setting up the system at this level of generality allows us to illustrate a basic equivalency between two tax options. For plan 1, consider an equal tax increase on labor and capital used in the production of X (with no change of tax rates in Y). Define $\hat{\tau}$ as this common increase ($\hat{\tau} \equiv \hat{\tau}_{KX} = \hat{\tau}_{LX}$). Equations (2.3), (2.4), (2.9) and (2.12) are unchanged. Equation (2.8) becomes

$$\begin{aligned} &\hat{p}_X^1 - \hat{\tau} + \hat{X} = \theta_{KX}\left(\hat{r} + \hat{K}_X\right) + \theta_{LX}\left(\hat{w} + \hat{L}_X\right), \\ &\hat{p}_Y + \hat{Y} = \theta_{KY}\left(\hat{r} + \hat{K}_Y\right) + \theta_{LY}\left(\hat{w} + \hat{L}_Y\right), \end{aligned} \tag{2.8'}$$

where $\hat{p}_X^1$ is the change in p_X under this plan. As an alternative, consider plan 2 with an output tax on X defined by $\hat{\tau} \equiv \hat{\tau}_X$ (and $\hat{\tau}_Y = 0$), where this $\hat{\tau}$ is the same size as the one above. In this case, Equations (2.3), (2.4), (2.8) and (2.9) are unchanged while Equation (2.12) becomes

$$\hat{X} - \hat{Y} = -\sigma_D\left(\hat{p}_X^2 + \hat{\tau} - \hat{p}_Y\right). \tag{2.12'}$$

Then it is easy to show that the equilibria under the two tax systems are the same: so long as $\hat{p}_X^1 = \hat{p}_X^2 + \hat{\tau}$, then all other outcomes are identical. Basically, $\hat{p}_X^1$ is the change in the price paid by consumers in plan 1 where p_X must rise to cover the tax on factors, while $\hat{p}_X^2 + \hat{\tau}$ is the price paid by consumers in plan 2 when the tax is on output. This points out a basic tax equivalence: an equal tax on all factors used in the production of a good yields the same incidence effects as a tax on output of that industry. Below, we discuss other tax equivalencies noted by Break (1974) and McLure (1975).

Before analyzing this system further, we pause to note that this very simple model is quite flexible and can be used to analyze a number of different problems. In the next section, we consider a special case of this model.

2.2. Special cases: one-sector model

With suitable modifications, the general model can be recast for various interesting special cases. We consider a one-sector model in some detail, in which one good is produced using labor and capital. We interpret the good Y in the Harberger model as leisure produced by the production function $Y = L_Y$. We can now interpret the labor market constraint in Equation (2.2) as a time constraint where time can be spent providing labor (L_X) or leisure ($L_Y = Y$). The price of leisure is the net wage rate ($p_Y = w$). No capital is used in the production of leisure, and all capital is used to produce X ($K_X = \overline{K}$). Thus, K_X is fixed in the short run (though competition among firms in X means that capital continues to be paid the value of its marginal product). The equations defining the system now become

$$\begin{aligned}
&\lambda_{LX}\hat{L}_X + \lambda_{LY}\hat{Y} = 0,\\
&\hat{L}_X = \sigma_X\left(\hat{r} + \hat{\tau}_{KX} - \hat{w} - \hat{\tau}_{LX}\right),\\
&\hat{p}_X + \hat{X} = \theta_{KX}\left(\hat{r} + \hat{\tau}_{KX}\right) + \theta_{LX}\left(\hat{w} + \hat{\tau}_{LX} + \hat{L}_X\right),\\
&\hat{X} = \theta_{LX}\hat{L}_X,\\
&\hat{X} - \hat{Y} = \sigma_D\left(\hat{w} - \hat{p}_X - \hat{\tau}_X\right),
\end{aligned} \tag{2.13}$$

from which we can solve for $\hat{X}$, $\hat{Y}$, $\hat{p}_X$, $\hat{w}$, $\hat{r}$ and $\hat{L}_X$ (with one numeraire). To begin solving, we can eliminate leisure (Y) from the system and reduce it to market variables only. Solve the first equation of (2.13) for $\hat{Y}$ and substitute into the fifth equation, to get:

$$\begin{aligned}
&\hat{L} = \sigma_X\left(\hat{r} + \hat{\tau}_K - \hat{w} - \hat{\tau}_L\right),\\
&\hat{p} + \hat{X} = \theta_K\left(\hat{r} + \hat{\tau}_K\right) + \theta_L\left(\hat{w} + \hat{\tau}_L + \hat{L}\right),\\
&\hat{X} = \theta_L\hat{L},\\
&\hat{X} + \phi\hat{L} = \sigma_D\left(\hat{w} - \hat{p} - \hat{\tau}_X\right),
\end{aligned} \tag{2.14}$$

where $\phi = \lambda_{LX}/\lambda_{LY}$ is the ratio of labor to leisure. We also drop the subscript X since the system now has only one market good.

The analysis of a tax on capital is very simple. Note that $\hat{r}$ and $\hat{\tau}_K$ always appear together as $\hat{r} + \hat{\tau}_K$ in all equations of (2.14). Therefore, as long as $\hat{r} = -\hat{\tau}_K$ in the first two equations, nothing else is affected. Thus, the tax on capital is borne fully by owners of capital – an unsurprising result since capital is inelastically supplied.

Next consider just a tax on labor. Using Equation (2.14), we can set $\hat{\tau}_K = \hat{\tau}_X = 0$, choose X as numeraire, and solve for $\hat{p}$, $\hat{L}$, $\hat{X}$, $\hat{r}$ and $\hat{w}$ as functions of $\hat{\tau}_L$. Simple manipulation reduces the system to two equations in two unknowns:

$$\left(\frac{\sigma_D}{\phi + \theta_L}\right)\hat{w} = \sigma_X\left(\hat{r} - \hat{w} - \hat{\tau}_L\right), \qquad \theta_K\hat{r} + \theta_L\left(\hat{w} + \hat{\tau}_L\right) = 0. \tag{2.15}$$

Rather than immediately solve for $\hat{w}$ and $\hat{r}$ as functions of $\hat{\tau}_L$, we first rewrite these two equations in terms of labor demand and supply elasticities. From the second equation in (2.15) we have:

$$\hat{r} = -\left(\theta_L/\theta_K\right)\left(\hat{w} + \hat{\tau}_L\right). \tag{2.16}$$

Next, substitute that into the first equation in (2.14) to get

$$\hat{L} = -\frac{\sigma_X}{\theta_K}\left(\hat{w} + \hat{\tau}_L\right) \equiv \eta^D\left(\hat{w} + \hat{\tau}_L\right), \tag{2.17}$$

where η^D is the elasticity of demand for labor with respect to its cost. This equation shows how the general equilibrium model can be used to generate the earlier simple partial equilibrium behavior as a special case.

To derive the elasticity of supply for labor, it is convenient to work with the individual budget constraint. Defining M as non-labor income (i.e., capital income), this budget constraint is

$$pX = wL + M. \tag{2.18}$$

Retaining the output price for the moment, as if we had not yet assigned a numeraire, totally differentiate this constraint to get

$$\hat{p} + \hat{X} = \theta_L \left(\hat{w} + \hat{L}\right) + \theta_K \hat{M}. \tag{2.19}$$

We next combine Equation (2.19) and the fourth equation in (2.14) and rearrange to get an expression for labor supply as a function of prices and income:

$$(\theta_L + \phi)\,\hat{L} = (\sigma_D - \theta_L)(\hat{w} - \hat{p}) - \theta_K \left(\hat{M} - \hat{p}\right). \tag{2.20}$$

Equation (2.20) is a key equation from which we can recover a number of important behavioral parameters. First, note the absence of money illusion. If all prices and nominal incomes change by the same percentage ($\hat{w} = \hat{p} = \hat{M}$), then Equation (2.20) implies no effect on labor supply ($\hat{L} = 0$). Hence, we can operate with or without the numeraire assumption. Second, note that labor supply can be affected by any change in the real wage (w/p) or in real income (M/p). If we hold real non-labor income constant, then the last term in Equation (2.20) is zero, and the labor supply elasticity (η^S) is defined by

$$\hat{L} = \frac{\sigma_D - \theta_L}{\theta_L + \phi}(\hat{w} - \hat{p}) \equiv \eta^S(\hat{w} - \hat{p}). \tag{2.21}$$

This η^S is an uncompensated labor supply elasticity. The first term in its numerator is the substitution effect, while the second term is the income effect. For the incidence analysis below, we assume no initial taxes and that the revenue from the introduction of this labor tax ($\hat{\tau}_L$) is returned to households in a lump-sum fashion. Thus, income effects are not relevant, and we need the *compensated* labor supply elasticity (η^S_C) [17]. From Equation (2.21) it is evident that this elasticity is [18]:

$$\eta^S_C = \frac{\sigma_D}{\theta_L + \phi}. \tag{2.22}$$

[17] Note that income effects can be ignored if one starts at a Pareto-optimum. Otherwise, income compensation won't eliminate the full income effect.

[18] The compensated labor supply elasticity can also be derived from an application of Slutsky's Equation.

Using $\hat{L} = \eta_C^S \hat{w}$ together with $\hat{L} = \eta^D(\hat{w} + \hat{\tau}_L)$ from Equation (2.17) yields:

$$\frac{\hat{w}}{\hat{\tau}_L} = \frac{\eta^D}{\eta_C^S - \eta^D}, \tag{2.23}$$

and substituting this into Equation (2.16) yields:

$$\frac{\hat{r}}{\hat{\tau}_L} = \left(\frac{\theta_L}{\theta_K}\right)\left(\frac{-\eta_C^S}{\eta_C^S - \eta^D}\right). \tag{2.24}$$

These two equations are the general equilibrium solution for the effects of the labor tax $\hat{\tau}_L$ on each factor price, expressed in terms of parameters. Yet note the similarity between Equation (2.23) in the general equilibrium model and Equation (1.10) in the partial equilibrium model. The only difference is that the partial equilibrium model ignores the use of the revenue and therefore employs an uncompensated elasticity, whereas the general equilibrium model assumes return of the revenue and therefore uses a compensated elasticity [19].

Finally, for the one-sector model of this section, we turn to consideration of an *ad valorem* tax on output at rate τ_X. Since the producer price is fixed at $p = 1$ (our numeraire), the consumer price $p(1 + \tau_X)$ will rise. And since the real wage is $w/(1 + \tau_X)$, the change in the real wage is $\hat{w} - \hat{\tau}_X$. Using steps similar to the derivation of Equations (2.23) and (2.24), we find how real factor prices adjust to a change in τ_X:

$$\frac{\hat{w} - \hat{\tau}_X}{\hat{\tau}_X} = \left(\frac{\eta_C^S}{\eta_C^S - \eta^D}\right) - 1, \tag{2.25}$$

and

$$\frac{\hat{r} - \hat{\tau}_X}{\hat{\tau}_X} = \frac{\theta_L}{\theta_K}\left(\frac{-\eta_C^S}{\eta_C^S - \eta^D}\right) - 1. \tag{2.26}$$

Again, we see how relative elasticities matter.

This section illustrates the circumstances under which a partial equilibrium model can be viewed as a special case of a general equilibrium model [20]. Anybody who writes down only the simple Equations (1.7) and (1.8) for demand and supply of labor can

[19] If the tax revenues were used to finance a government project, which employs some labor L or output X, then earlier equations would have to be re-specified. However, if that government project is separable in the individual's utility function, then the result in Equation (2.23) would be identical to Equation (1.10).

[20] In a model with many consumption goods, the same kind of isolation of the labor market is possible by assuming separability between leisure and consumption and homotheticity in the sub-utility function defined over the consumption goods.

Table 2.1
Two sector–two factor model

$$\left(\hat{X}-\hat{Y}\right)=-\sigma_D\left(\hat{p}_X-\hat{p}_Y\right)-\sigma_D\left(\hat{\tau}_X-\hat{\tau}_Y\right)$$

$$\left(\hat{p}_X-\hat{p}_Y\right)=\left(\theta_{LX}-\theta_{LY}\right)\left(\hat{w}-\hat{r}\right)+\left(\hat{\tau}_{KX}-\hat{\tau}_{KY}\right)+\theta_{LX}\left(\hat{\tau}_{LX}-\hat{\tau}_{KX}\right)-\theta_{LY}\left(\hat{\tau}_{LY}-\hat{\tau}_{KY}\right) \quad (2.27)$$

$$\begin{aligned}\left(\lambda_{LX}-\lambda_{KX}\right)\left(\hat{X}-\hat{Y}\right)&=\left(\sigma_X\left(\lambda_{LX}\theta_{KX}+\lambda_{KX}\theta_{LX}\right)+\sigma_Y\left(\lambda_{LY}\theta_{KY}+\lambda_{KY}\theta_{LY}\right)\right)\left(\hat{w}-\hat{r}\right)\\&+\sigma_X\left(\lambda_{LX}\theta_{KX}+\lambda_{KX}\theta_{LX}\right)\left(\hat{\tau}_{LX}-\hat{\tau}_{KX}\right)+\sigma_Y\left(\lambda_{LY}\theta_{KY}+\lambda_{KY}\theta_{LY}\right)\left(\hat{\tau}_{LY}-\hat{\tau}_{KY}\right)\end{aligned}$$

say it is a *general equilibrium* model with one sector that uses two inputs, where utility is defined over leisure and consumption. A similar procedure, left as an exercise, could develop a model of the market for commodity X with an elasticity of demand for X and supply of X, in order to study the effects of a tax on X. A corresponding general equilibrium model could be constructed to include only two goods in utility (X and Y), one factor like labor that is mobile between production of either good, and another factor that is specific to each industry[21]. Then the elasticity of demand for X would depend primarily on the elasticity of substitution in utility, and the elasticity of supply of X would depend primarily on the elasticity of substitution in production.

Overall, this section has shown how results in the literature that uses a one-sector model can be derived directly from the two-sector model of Harberger (1962).

2.3. *Analysis of the two-sector model*

We now return to the original model in Section 2.1 with two sectors and two factors. Incidence on the uses side is based on the change in p_X/p_Y, while incidence on the sources side is based on the change in w/r. We therefore simplify the analysis by reducing the system of nine equations to three, where the unknowns are $(\hat{p}_X-\hat{p}_Y)$, $(\hat{w}-\hat{r})$ and $(\hat{X}-\hat{Y})$. We solve for these unknowns in terms of exogenous parameters (like the θ and λ shares) and exogenous tax changes (the various $\hat{\tau}$'s).

The first equation of our system is Equation (2.12), repeated below as the first equation of (2.27), shown in Table 2.1. To get the second equation of our system, substitute Equation (2.9) into (2.8) and then subtract the second equation in (2.8) from the first one. The result is the second equation of (2.27) in Table 2.1.

[21] If the production function is $X=F(L_X,K_X)$, where L_X is mobile and K_X is fixed, then the industry will supply more of X as its price rises, by bidding more labor away from the other industry.

Table 2.2
Tax equivalencies

τ_{LX}	and	τ_{KX}	$\rightarrow$	τ_X
and		and		and
τ_{LY}	and	τ_{KY}	$\rightarrow$	τ_Y
$\downarrow$		$\downarrow$		$\downarrow$
τ_L	and	τ_K	$\rightarrow$	τ

To get the third equation, first use Equation (2.9) and subtract its second equation from its first equation. Then use Equation (2.4) to get:

$$\hat{X} - \hat{Y} = \hat{L}_X - \hat{L}_Y + (\theta_{KX}\sigma_X - \theta_{KY}\sigma_Y)(\hat{w} - \hat{r}) + \theta_{KX}\sigma_X\left(\hat{\tau}_{LX} - \hat{\tau}_{KX}\right) - \theta_{KY}\sigma_Y\left(\hat{\tau}_{LY} - \hat{\tau}_{KY}\right). \tag{2.28}$$

Then Equations (2.3) and (2.4) can be combined to show that

$$\begin{aligned}\hat{L}_X - \hat{L}_Y = \frac{1}{\lambda_{LX} - \lambda_{KX}} \big(&(\lambda_{KX}\sigma_X + \lambda_{KY}\sigma_Y)(\hat{w} - \hat{r}) + \lambda_{KX}\sigma_X\left(\hat{\tau}_{LX} - \hat{\tau}_{KX}\right) \\ &+ \lambda_{KY}\sigma_Y\left(\hat{\tau}_{LY} - \hat{\tau}_{KY}\right)\big).\end{aligned} \tag{2.29}$$

Substitute Equation (2.29) into (2.28) and simplify to get the third equation in our system (Equation 2.27 of Table 2.1).

The three equations in (2.27) can be solved for the three unknowns ($\hat{p}_X - \hat{p}_Y$, $\hat{w} - \hat{r}$, and $\hat{X} - \hat{Y}$) as functions of the changes in tax rates. Note that the system in Equation (2.27) has not yet assigned a numeraire, and that it includes all possible tax rates. Before solving, we return to the topic of tax equivalencies, and then provide a graphical analysis of a marginal increase in the tax on capital income in sector X [22].

In our initial setup of the two-sector model (at the end of Section 2.1), we noted that a tax on both factors in one industry (with $\hat{\tau}_{LX} = \hat{\tau}_{KX}$) is equivalent to a tax on the output of that industry ($\hat{\tau}_X$). This result appears in the first row of Table 2.2 [23].

Using the equations in this section, we can now explain the first column of Table 2.2, which says that a tax on both industries' use of labor at the same rate is equivalent to a tax on the consumer's labor income. To show this, using our system of three equations, set all $\hat{\tau}_{LX} = \hat{\tau}_{LY}$ and replace those rates with $\hat{\tau}_L$. Then note that the terms $\hat{w} - \hat{r} + \hat{\tau}_L$ appear together throughout the system of three equations, and thus a new equilibrium holds with $\hat{w} = -\hat{\tau}_L$ and with no change in any quantity or in the ratio of the gross wage to the interest rate. The entire burden of this tax falls on labor, because it applies

22 Our graphical analysis is from Atkinson and Stiglitz (1980), but see McLure (1974) for another graphical exposition.

23 See Break (1974), McLure (1974) and Atkinson and Stiglitz (1980).

at the same rate in both sectors, and labor has fixed total supply. In this model, a tax on a factor in both sectors is a lump-sum tax and affects that factor only.

In the bottom row of Table 2.2, either τ_L or τ_K is a lump-sum tax, so the two together is a lump-sum tax on all income, τ. In the final column, either τ_X or τ_Y alone would change production and impact various prices, but τ_X and τ_Y together at the same rate is equivalent to a lump-sum tax on all income, τ, with no effect on any allocations or relative prices. A simple look at the consumer's budget constraint shows that a tax on both goods at the same rate is the same as a tax on both factors at the same rate.

Next we turn to the graphical analysis of our three equation system (in Equation 2.27). Consider the special case of a tax on capital income in sector X, holding all other tax changes to zero. The first equation in (2.27) relates the relative demand for goods (X/Y) to the ratio of prices (p_X/p_Y). In Figure 2.1, this downward sloping demand equation (D) is graphed in the upper right quadrant[24]. The third equation relates the relative *supply* of goods (X/Y) to relative factor prices (w/r). It is drawn as an upward sloping function in the upper left quadrant of Figure 2.1, for the case where X is relatively labor intensive ($\lambda_{LX} > \lambda_{KX}$). In this case, as the production of X rises relative to the production of Y, the demand for labor rises relative to demand for capital (which raises the wage rate relative to capital return, w/r). Finally, the second Equation (2.27) relates output prices to factor prices. Assuming X is more labor intensive in value ($\theta_{LX} > \theta_{LY}$), an increase in w/r increases the price of X relative to the price of Y. This relationship is graphed in the lower right quadrant of Figure 2.1.

We now use those two curves together to "derive" the supply curve (S) in the upper right quadrant. First, start with a given output price ratio (point A_1 on the horizontal axis). Through the curve in the lower right quadrant, this output price ratio implies a particular factor price ratio (point A_2). Follow this factor price ratio through the 45° line in the lower left quadrant to the upper left quadrant where the factor price ratio (point A_3) implies a particular output ratio. Together with the original output price ratio at A_1, this output ratio gives us a point on a "supply" schedule (A_4). Then, starting at a different output price ratio, (e.g., B_1), we can find another output ratio and thus sketch out the upward-sloping supply schedule (S). The intersection of this supply schedule with the demand curve (the first equation) indicates equilibrium in Figure 2.1.

Next consider how a capital income tax (τ_{KX}) changes the equilibrium (see Figure 2.2). The point E_0 indicates the pre-tax equilibrium, and E_1 indicates the post-tax equilibrium. In the lower right quadrant, the tax on capital in sector X shifts the output price curve to the right, reflecting a higher price for good X (for any given

[24] The first equation is a *linear* equation in the form $(\hat{X} - \hat{Y}) = a + b(\hat{p}_X - \hat{p}_Y)$, but this linear equation is derived from a *nonlinear* equation in the form $X/Y = A(p_X/p_Y)^b$. Starting with the latter equation, take the natural log of both sides and differentiate to get the former equation. Thus (X/Y) in Figure 2.1 is a nonlinear function of (p_X/p_Y).

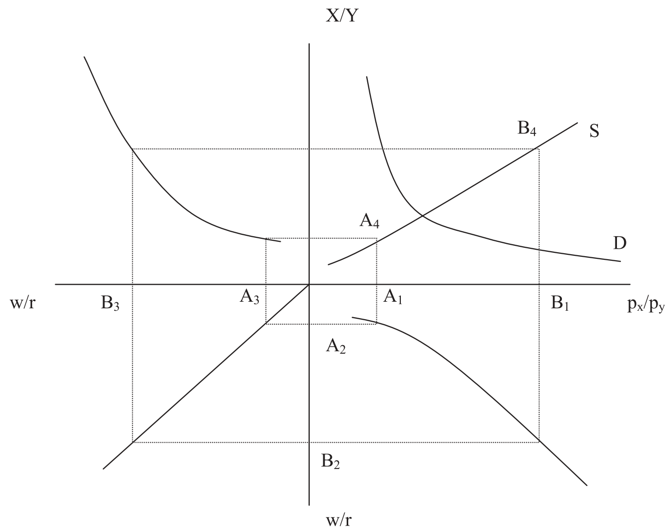

Fig. 2.1. Equilibrium in output and factor markets.

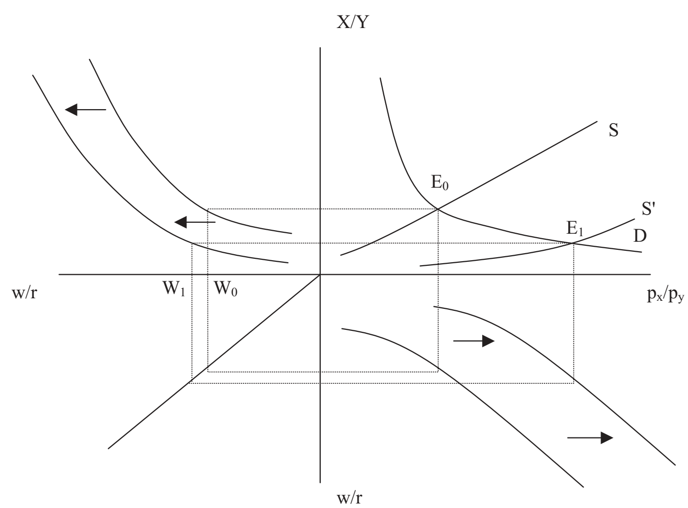

Fig. 2.2. Partial factor tax in general equilibrium.

factor price ratio). Meanwhile, in the upper left quadrant, the tax also shifts the factor demand curve to the left, reflecting a desire to shift out of capital and into labor (for any given output combination). The desired shift from capital to labor raises the wage rate relative to the interest rate (point W_1 relative to W_0).

These new curves can be used to trace out a new "supply" curve in the upper right quadrant that is unambiguously shifted down and to the right, which means an unambiguous increase in p_X/p_Y and an unambiguous decrease in X/Y. The effect on w/r is ambiguous, however. This tax may be borne disproportionately by capital ($\hat{w} - \hat{r} > 0$), by labor ($\hat{w} - \hat{r} < 0$), or in proportion to their income shares ($\hat{w} - \hat{r} = 0$)[25]. We now derive the effect on w/r from a change in τ_{KX} (holding all other tax changes to zero). The three-equation system above can be solved to obtain our version of the famous Harberger (1962) equation:

$$\hat{w} - \hat{r} = \frac{1}{D} \left[\sigma_X a_X - \sigma_D \lambda^* \theta_{KX} \right] \hat{\tau}_{KX}, \tag{2.30}$$

where $a_i = \theta_{Ki}\lambda_{Li} + \theta_{Li}\lambda_{Ki}$, $i = X, Y$, $\lambda^* = \lambda_{LX} - \lambda_{KX}$, $\theta^* = \theta_{LX} - \theta_{LY}$ and $D = \sigma_D \lambda^* \theta^* + \sigma_X a_X + \sigma_Y a_Y$. This denominator is unambiguously positive[26]. In the numerator, the first term in brackets is positive, while the sign of the second term depends on the relative capital intensity of the taxed sector. If X is capital intensive ($\lambda_{LX} < \lambda_{KX}$), then this subtracted term is negative, the whole numerator is positive, and τ_{KX} raises w/r (the burden is disproportionately on capital). This case is clear because the tax applies to capital, but only in the capital-intensive sector! If X is labor intensive, however, then the outcome is ambiguous: the tax is a partial factor tax on capital, but it is imposed on the labor-intensive sector.

The impact of a tax on capital in sector X can be decomposed into two components: a substitution effect and an output effect. The first term in brackets on the right-hand side of Equation (2.30) represents the substitution effect and is unambiguously positive (indicating how the burden of the partial capital tax falls on capital). Its magnitude depends on the degree of factor substitution in the taxed industry (σ_X). The second term reflects the fact that the tax applies only in one industry, so it raises the price of that good and thereby induces a shift in demand from X to Y. As capital and labor are shed by the taxed sector, they must be absorbed by the other sector. If sector X is labor intensive ($\lambda_{LX} > \lambda_{KX}$), the wage rate must fall for sector Y to be willing to hire the excess labor. The magnitude of this output effect depends on the elasticity of substitution in demand (σ_D). In this case (and as drawn in Figure 2.2), the output and substitution effects offset, and it is impossible to say whether w/r will rise or fall in response to a tax on K_X.

[25] With no change in relative factor prices, burdens cannot differ on the sources side: the tax merely raises the price of the taxed good, relative to factor prices. Capital and labor spend the same fraction of their incomes on the taxed good, so the two factors bear burdens in proportion to their shares of national income. Thus, capital's burden can only be larger than labor's burden if r falls relative to w.

[26] To show that D is positive, first note that all parameters in the second and third terms are positive. Then, to show that the first term is positive, we show that $(\lambda_{LX} - \lambda_{KX})$ and $(\theta_{LX} - \theta_{LY})$ must be of the same sign (either both positive or both negative). We have $\theta_{LX} - \theta_{LY}$ equal to $\theta_{LX}\theta_{KY} - \theta_{KX}\theta_{LY}$, which in turn equals $(wr/(p_X X p_Y Y))(L_X K_Y - K_X L_Y)$. If $\theta_{LX} - \theta_{LY} > 0$, then $L_X K_Y - K_X L_Y > 0$, which implies that $\lambda_{LX}\lambda_{KY} - \lambda_{LY}\lambda_{KX} = \lambda_{LX} - \lambda_{KX} > 0$. However, this result is only guaranteed in a model with no other taxes.

The system of three equations in (2.27) includes many possible tax rates to analyze, but the methods are all similar to the methods just employed to analyze τ_{KX}. We just make one last point about one other tax rate, a tax on the sale of X ($\hat{\tau}_X > 0$). Solving for $\hat{w} - \hat{r}$ as a function of $\hat{\tau}_X$ yields

$$\hat{w} - \hat{r} = -\frac{\sigma_D (\lambda_{LX} - \lambda_{KX})}{D} \hat{\tau}_X, \tag{2.31}$$

where D is as defined above. Note that this is precisely the output effect from the partial factor tax in Equation (2.30). This result follows because either $\theta_{KX} \hat{\tau}_{KX}$ or $\hat{\tau}_X$ equals the change in tax revenue as a fraction of the consumer expenditure on X.

Equation (2.30) can be generalized to allow for non-homothetic preferences and public demand for consumption goods that differs from private demand. Vandendorpe and Friedlaender (1976) have carried out this analysis. Their model also allows for pre-existing distortionary taxation. Consider an experiment in which the partial factor tax on capital used in the production of X is increased, with revenues returned lump sum to consumers. Thus, public demands for X and Y are fixed, while private demands (X^P, Y^P) can change. This more-general model now provides a demand-side force affecting the change in w/r. Equation (2.30) becomes

$$\hat{w} - \hat{r} = \frac{1}{\tilde{D}} \left[\sigma_X a_X - \tilde{\sigma}_D \lambda^* \theta_{KX} + \lambda^* \tilde{\eta} B \right] \hat{\tau}_{KX}, \tag{2.32}$$

where $\tilde{\sigma}_D$ is the elasticity of substitution between X and Y in consumption suitably modified to account for government consumption of a fraction of output,

$$\tilde{\eta} = \left(\frac{X^P}{X} \right) \eta_X - \left(\frac{Y^P}{Y} \right) \eta_Y$$

is the difference in income elasticities weighted by the share of private consumption (X^P, Y^P) in total output (X, Y), and B is a measure of the initial excess burden of pre-existing taxes. This B will be negative to the extent that X is initially taxed more heavily than Y.

Relative to the original Harberger Equation (2.30), the third term inside the brackets in Equation (2.32) is an added demand-side effect. Intuitively, we can track this added effect in three steps, in the case where $B < 0$. First, an increase in taxes on capital used in the production of X increases the relative burden on X and adds to the excess burden of the tax system. This burden effect is first order and constitutes a reduction in real output (and hence income) to society. Second, the term $\tilde{\eta}$ translates this real income loss into a relative shift in demands for X and Y. Imagine that preferences were still homothetic, so that income elasticities equal 1, but that the private share of consumption of X is less than the corresponding share for Y. In that case, $\tilde{\eta} < 0$. The loss in income induces a drop in both private demands (X^P and Y^P), but public demands are fixed. In this case, $(X^P/X) < (Y^P/Y)$ means that the drop in total demand for X will be less than the drop in total demand for Y. Factors must shift over from production of Y to X. Third, λ^* translates the change in relative-output demands into changes in relative-factor demands. If production of X is more labor intensive ($\lambda^* > 0$), the shift in production from Y to X will increase the demand for labor. This will drive the wage rate up relative to the interest rate. Note that $\lambda^* \tilde{\eta} B$ is positive, based on our assumptions in this example, so we get the desired positive effect on $\hat{w} - \hat{r}$. Equation (2.32) indicates precisely how Harberger eliminates this demand-side effect. He assumes homotheticity and that public consumption of X and Y are in the same proportions as private consumption, which together ensure that $\tilde{\eta}$ equals zero.

Recognizing the tremendous usefulness of the basic Harberger model, many economists in the following decades developed many other extensions, generalizations and applications. As one example, Mieszkowski (1972) considers the incidence of the local property tax in an extended model with three factors of production (land, as well as labor and capital). As another example, McLure (1970) considers the effects of imperfect factor mobility. These extensions and generalizations are important, but beyond the scope (and page limits) of this chapter. Readers can find thorough reviews of this literature in McLure (1975) and in Shoven and Whalley (1984).

2.4. The corporate income tax

The original paper by Harberger (1962) uses the general equilibrium model to analyze the corporate income tax. To do this in a two-sector model, he must assume that the whole corporate sector produces only one output (X), and that the corporate income tax is effectively a tax on all capital used in that sector (τ_{KX}). We now turn to some of the special cases of his model, to illustrate the impact of a tax on corporate capital. As in Harberger, we can choose the wage as numeraire ($\hat{w} = 0$) and focus on the return to capital to indicate relative factor returns.

First, when do we know that $\hat{r} < 0$ (the burden of the tax falls disproportionately on capital)? From Equation (2.30), a sufficient condition for this outcome is that the corporate sector is capital intensive. However, a different sufficient condition can be found by a rearrangement of the numerator to include $(\sigma_X - \sigma_D)\,\lambda_{LX}\theta_{KX}$ as the only term with ambiguous sign. Then $\sigma_X > \sigma_D$ is a sufficient condition for $\hat{r} < 0$. In other words, this tax on capital in X disproportionately burdens capital if firms in X can shift out of capital more readily than consumers can shift out of X. In fact, higher σ_X always raises the burden on capital; as it approaches infinity, the limit of Equation (2.30) is $-\hat{r} = \hat{\tau}_{KX}$ (the rate of return falls by the full amount of the tax). Because the return falls by the full amount of the tax in *both* sectors, the total burden on capital is *more* than the revenue. The cost of capital is unchanged in X, and lower in Y, so labor gains!

Second, we can ask, under what conditions is the tax burden shared equally between labor and capital ($\hat{r} = 0$)? As σ_Y in the denominator of Equation (2.30) approaches infinity, we can see that $\hat{r}$ approaches zero. A large value of σ_Y just means that the untaxed sector can absorb whatever excess capital is no longer used in the taxed sector. Another way to guarantee that $\hat{r} = 0$, from Equation (2.30), is to have

$$\sigma_D\,(\lambda_{LX} - \lambda_{KX})\,\theta_{KX} = \sigma_X\,(\lambda_{LX}\theta_{KX} + \lambda_{KX}\theta_{LX})\,. \tag{2.33}$$

Necessary conditions are that the corporate sector is labor intensive ($\lambda_{LX} > \lambda_{KX}$) and that consumers can readily substitute ($\sigma_D > \sigma_X$)[27].

Third, when can this partial tax on capital fall disproportionately on labor? The taxed sector must be very labor intensive for the output effect to dominate the substitution effect and not just to offset part of it.

Fourth, when does the entire burden of the tax fall on capital? This special outcome occurs where $\mathrm{d}r(\bar{K}) = -\mathrm{d}\tau_{KX}(rK_X)$, which says that the fall in capital income equals the tax revenue collected. For the initial imposition of the tax, where $\mathrm{d}\tau_{KX} = \hat{\tau}_{KX}$, this equation can be rewritten as

$$\frac{\hat{r}}{\hat{\tau}_{KX}} = -\lambda_{KX}. \tag{2.34}$$

In the special case where $\sigma_X = \sigma_Y = \sigma_D$, substitution of this single σ into Equation (2.30) shows that it is multiplied times everything in the numerator, and everything in the denominator, so it factors out and disappears. Further rearrangement finds that $D = 1$ in the denominator and that the bracketed expression in the numerator equals λ_{KX}. Thus, the case with all the same elasticities of substitution yields the result that capital bears the entire burden of the corporate income tax. A further special case of this special case is the Cobb–Douglas case where all elasticities of substitution are one[28].

The original paper by Harberger (1962) considered plausible parameter values and likely empirical outcomes. First, he finds that the corporate sector is indeed labor intensive. This result itself is sometimes surprising to those who think about the corporate sector's large manufacturing plants, but remember also the number of workers at those plants: labor intensity is relative, and the non-corporate sector includes a lot of agriculture where a single worker can sit atop a large harvester covering many acres of valuable land (which is part of capital in the aggregation with only two factors). The labor intensity of the corporate sector is important because it means that the burden of this tax on capital might be on labor.

Next, Harberger considers alternative values for the key elasticities of substitution (σ_X, σ_Y, and σ_D). He considers some of the 27 possible combinations that can arise when each of those three parameters can take any of three values (0.5, 1.0, and 1.5).

[27] The second condition follows from the fact that $\lambda_{LX}\theta_{KX} + \lambda_{KX}\theta_{LX} = \theta_{KX}(\theta_{LX} - \lambda_{KX}) + \lambda_{KX}$.

[28] In fact, as shown in McLure and Thirsk (1975), the case where all utility and production functions are Cobb–Douglas yields an easy analytical solution for the incidence of a large tax (without using log-linearization techniques that are limited to small changes).

Sometimes capital bears less than the full burden of the corporate income tax (τ_{KX}), and sometimes it bears more than the full burden of the tax, but the main message coming out of his original 1962 paper is that capital is likely to bear approximately the full burden of the corporate income tax, more or less. And capital mobility means that the burden is on all capital, not just corporate capital.

To explain this empirical result, it is important to remember the conceptual result above that capital bears the full burden of the tax anytime the three elasticities are equal ($\sigma_X = \sigma_Y = \sigma_D$). Then, if one of those parameters varies above or below the common value, capital's burden will be somewhat more or less than the full burden of the tax.

Harberger's main focus was the sources side, finding the change in relative factor prices (r/w). He ignored the uses side by assuming that all consumers as well as the government buy X and Y in the same proportions. Although his 1962 paper did not solve for relative goods prices, the same model can also be used to solve for the other unknowns such as $\hat{p}_X$ and $\hat{p}_Y$ (where labor is numeraire). Interestingly, even while capital is bearing the full burden of the tax, τ_{KX} also raises the price of X (thus placing additional burden on those who in fact consume disproportionate amounts of X) and lowers the price of Y (thus providing gains to those who consume more than the average amount of Y). That untaxed industry experiences a fall in their cost of capital, while the wage is fixed at 1.0, so competition among firms in the industry means that the output price must fall. In other words, even though the main effect of this tax is that government confiscates resources *from* the private sector, one of the effects is that some individuals are made better off – anybody who earns most of their income from labor and who spends disproportionately on products of the non-corporate sector.

Many of the empirical studies reviewed below choose to follow the original Harberger (1962) result that all capital income bears approximately the full burden of the corporate income tax, and thus they allocate that tax in proportion to the capital income of each household. However, Harberger assumed (1) a fixed capital stock, (2) a closed economy, (3) no financing decisions, and (4) no uncertainty. We therefore note four challenges to his modeling of corporate tax burdens.

First, in an intertemporal model, the corporate tax might reduce the net rate of return only in the short run, until savings fall enough to reduce the future capital stock and raise the return back up to its long run rate. The smaller capital stock means a lower wage rate, so labor can bear *more* than the full burden of the tax [e.g., Judd (1985a)]. This possibility is discussed in Section 4 below (dynamic models).

Second, in a small open economy with international capital mobility, the corporate tax might just drive capital elsewhere so that domestic savers earn the same net return as before. This drives down the domestic capital stock, and thus the domestic wage rate, so again the burden falls on labor [e.g., Mutti and Grubert (1985)]. Yet Bradford (1978) shows that capital does indeed bear the burden of a local tax on capital, in the aggregate. The tax burden is not on local investors but is spread across all investors worldwide[29].

Third, if investment is financed by debt, then the return is paid as tax-deductible interest. If investment proceeds to the point where the marginal unit just breaks even, with no return above and beyond the interest paid, then no corporate tax applies to the marginal investment. Indeed, as pointed out by Stiglitz (1973), all corporate investment may be financed by debt at the margin. If so, then the corporate tax is a lump-sum tax on infra-marginal investments financed by equity. Then it does not distort the allocation of resources, and it does not affect the return in the non-corporate sector.

Fourth, as pointed out by Gentry and Hubbard (1997), much of the corporate tax applies not just to the risk-free portion of the return to equity-financed investment, but also to a risk premium, to infra-marginal profits, and to lucky windfalls. This has implications for a differential tax incidence analysis of a switch from an income tax to a consumption tax. Such a switch would eliminate the tax only on the first component, and it would continue to tax the other components. Then, since those other components of capital income are concentrated in the top income brackets, they argue that a consumption tax is more progressive than estimated under conventional incidence assumptions. In other words, typical differential incidence studies of a shift from income to consumption taxation err by assuming that the burden of the corporate income tax falls on all capital income, which is disproportionately concentrated in high income brackets, because most of that capital income would still be taxed under a consumption tax. The corporate income tax adds only the burden on risk-free returns, which are not so concentrated in high-income brackets.

2.5. The property tax

Local jurisdictions typically impose a yearly tax on the value of real property – both land and improvements. Alternative "views" of the incidence of this tax have been hotly debated, and general equilibrium analysis has radically changed economists' thinking. First, the property tax has been viewed as an excise tax on housing services that is regressive because housing expenditures are a high proportion of the budgets of low-income families. This "old view" is typically associated with Simon (1943), but it dates back to Edgeworth (1897). Second, the property tax has been viewed as a profits tax on capital income that is progressive because that source of income is a high proportion for high-income families. This view is called the "new view," although it originates with Brown (1924). Perhaps it *is* new relative to Edgeworth (1897)! [30]

[29] See the discussion in Kotlikoff and Summers (1987). In contrast, Gravelle and Smetters (2001) argue that imperfect substitutability of domestic and foreign products can limit or even eliminate the incidence borne by labor, even in an open economy model. They find that the tax is borne by domestic capital, as in the original Harberger model.

[30] The property tax has also been viewed as a tax on site rents that is shifted to landowners. Marshall (1890) provides an early statement of this "classical" view, but Simon (1943) points out that classical

Mieszkowski (1972) reconciles these views in a Harberger general equilibrium modeling framework. If τ_i is the tax rate on property in community i, we can decompose the rate into two components as $\tau_i = \bar{\tau} + \varepsilon_i$ where $\bar{\tau}$ is the average property tax rate over the entire country, and ε_i is the deviation of the local rate from the national average. By construction, the average of ε_i across all communities is zero. Mieszkowski argues that the first component of τ_i can be viewed as a national tax on housing capital at rate $\bar{\tau}$. Using the Harberger framework, he then argues that this tax burdens *all* capital. The second component, Mieszkowski continues, can be viewed as a differential tax that can be positive or negative. This differential tax might be passed forward to consumers of housing or passed backwards to immobile factors (workers or landowners). Mieszkowski concludes that the bulk of this differential tax is passed forward to consumers.

Even in Mieszkowski's model, note that the regressivity of the tax depends on what sort of tax change is contemplated. A uniform nation-wide increase in property tax would impact capital income, which is progressive under the "new" view. In contrast, a single community's increase in property tax would likely raise that town's cost of housing, which is regressive under the "old" view[31].

Next, Hamilton (1976) articulates a third view, called the "benefit" view, that the property tax is neither regressive nor progressive because it is really no tax at all[32]. Building on Tiebout (1956), Hamilton argues that mobile taxpayers would not live in any jurisdiction that charges a tax higher than the value of its local public goods and services – unless property values adjusted to reflect the differential between the value of services received and taxes paid (the "fiscal surplus"). In other words, house prices would rise by the capitalized value of any positive stream of fiscal surpluses or fall by the capitalized value of any negative stream (where taxes exceed services). If the local property tax becomes a voluntary price paid for those local goods and services, then it is no tax at all. Thus, we have the "old" view, the "new" view, and the "no" view of the property tax[33].

Hamilton's focus is on the efficiency impact of property taxes. He argues that the property tax *per se* has no distributional impact because of capitalization. His story is

economists divide the property tax into a portion falling on land rents and a portion falling on improvements.

31 Part of the early "debate" is published in two papers by Musgrave (1974) and Aaron (1974), but they also point out the importance of institutional detail when doing incidence analyses. Musgrave generally supports the old view, and he notes that many rental markets in urban areas are likely to be imperfectly competitive. Thus, some of the insights from Section 3 below may be useful for thinking about property tax incidence. Aaron generally supports the new view. He notes that, even under the old view, the portion of the property tax falling on rental housing may well be progressive since the ratio of market value to rent rises with rent (more expensive houses have relatively low monthly rent).

32 Hamilton (1975) first states this argument, but Hamilton (1976) extends it to heterogeneous communities.

33 Zodrow and Mieszkowski (1983) review this literature, and Zodrow (2001) provides a possible reconciliation of these various views.

not complete yet, as he notes that the value of land is higher when used to construct housing that is below the average value of housing in the community. Because the property tax on such a house would be less than the (uniform) services provided, the fiscal surplus for such a house will be positive, and the landowner can extract those rents when selling the site. This shift in the mix of housing will lead to a shift in the burden of the property tax from owners of below-average-value housing to owners of above-average-value housing. In response, a countervailing political force will limit this shift (zoning or some other form of regulation). The outcome of this political process cannot be predicted in an economic model, and zoning could be so restrictive as to limit the amount of low-value housing to levels that are inefficient (and that lead to a shift of the burden of property tax from high-value homeowners to low-value homeowners). Hamilton concludes that it is impossible to determine the incidence of property taxes until we have a better understanding of the political forces influencing land-use policy.

2.6. Empirical work

Remaining with the property tax for the moment, we note that Oates (1969) first attempts to measure empirically the degree of capitalization of property taxes into property values. This type of measurement turns out to be a complicated statistical exercise, however, and economists continue to disagree about the degree of capitalization. Many economists believe that the benefit view should imply complete capitalization of property taxes (holding public services and other amenities constant). If so, then perhaps an empirical test of capitalization could help us choose between views. Alas, Mieszkowski and Zodrow (1989) point out that property taxes may be capitalized under both the benefit view and the new view. Thus, while capitalization is an important phenomenon in tax incidence theory, it is not useful as an empirical test among views of the property tax.

One interesting study by Carroll and Yinger (1994) looks at property taxes in rental markets rather than homeowner markets. They find that nearly one-half of property tax differentials are passed back to landlords, a result consistent (at least partially) with the new view.

Turning to the corporate income tax, we note the attempt by Krzyzaniak and Musgrave (1963) to estimate the burden econometrically using a time-series regression of the corporate output price on the corporate tax rate and other control variables such as the unemployment rate. They obtain the surprising result that the corporate income tax is "overshifted", meaning that the corporate sector is able to raise prices by *more* than the amount of the tax – and increase their profits. While this overshifting may provide evidence of imperfect competition, as we discuss below, this approach was largely discredited subsequently by considerations of reverse causality. Especially during war years, shortages mean that corporations can raise prices and make profits, which induces Congress to raise the tax rate. We know of no other subsequent attempt to estimate corporate tax incidence econometrically. Thus, while the Harberger model is extremely useful for analysis, the predictions have not exactly been “tested”. Debate continues about the incidence of the corporate income tax as well.

Without resolving any of these debates, another empirical approach can apply the theoretical developments just described to find the implications for a large number of households across the income spectrum [Pechman and Okner (1974), Musgrave, Case and Leonard (1974)]. First, this approach must specify how the burden of each tax is shifted (and can specify more than one outcome, for sensitivity analysis). Then, each scenario is applied to micro-data on households’ sources and uses of income. Pechman and Okner (1974) merge data files for a sample of 72 000 households. They use information on demographic characteristics such as age and family size, and tax return items such as income from dividends, interest, rent, capital gains, and wages and salaries. They classify households into annual income groups using a measure of economic income that includes transfers, the household’s share of corporate retained earnings, and the imputed net rental income from owner-occupied homes. They use tax actually paid as the total burden of each tax to be allocated. Then, for each set of assumptions about the shifting of each tax, they add up the burdens on each household.

Pechman and Okner assume for all cases that the burden of the personal income tax remains with the household, the employee part of the payroll tax remains with the worker, and the burden of sales and excise taxes falls on households according to their consumption patterns. The employer share of the payroll tax is sometimes allocated entirely to workers, and it is sometimes allocated equally between workers and consumers. The property tax is assumed to affect either the return to landowners specifically or all capital owners generally. Finally, for the corporate income tax, they consider several cases with different proportions of the burden on shareholders, capital owners, wage-earners, and consumers. They look only at taxes and ignore the distributional effects of any government spending [34].

For each combination of incidence assumptions, Pechman and Okner calculate the effective tax rate on each household, defined as the total tax burden as a fraction of economic income. Their results indicate that the most progressive set of assumptions do not yield markedly different results than the least progressive set of assumptions. In either case, the overall U.S. tax system is roughly proportional over the middle eight deciles. The effective tax rate is higher, however, at the top and bottom tails of the income distribution. At very low-income levels, any positive consumption implies a positive sales tax burden divided by a small income in the denominator. At the other end of the distribution, the rate is high because of the progressive personal income tax and assumed corporate tax burdens from disproportionate holding of corporate stock.

[34] Thus, when they allocate the burdens of payroll taxes, they ignore the distributional effects of using those revenues to provide social security benefits. This treatment is most troublesome if a marginal increase in benefits is tied to a marginal payment of tax, because then only the difference is really a “burden”.

This finding of rough proportionality has shaped tax policy debates for the past two decades[35]. The general consensus is that the progressive effects of the personal income tax and the corporate income tax are more or less offset by the regressive impacts of payroll taxes, sales taxes, and excise taxes. Musgrave, Case and Leonard (1974) reach similar conclusions. In contrast, however, Browning and Johnson (1979) find that the U.S. tax system as a whole is highly progressive. They assume that sales and excise taxes raise product prices, but government transfers are indexed to provide the same real benefits, thus protecting low-income transfer recipients. These taxes do not fall on consumption generally, but only on consumption out of factor income.

These studies all have three problems. First, they classify households by annual income rather than by income over some longer time period (such as an entire lifetime)[36]. Second, they assume the allocation of a total tax burden equal to tax actually paid, not a burden based on each group's change in consumer welfare (such as the equivalent variation, EV). Third, they use results from different kinds of models to guide their assumptions about the incidence of each tax, but they do not calculate these effects in a single model.

To address the first such problem, Davies, St. Hilaire and Whalley (1984) construct lifetime histories of earnings, transfers, inheritances, savings, consumption, and bequests. Using Canadian survey data, they measure lifetime income and use it to classify households, and then add up each household's lifetime burdens under each set of incidence assumptions. Thus, they extend the approach of Pechman and Okner to a lifetime context. They find that personal income taxes are less progressive in a lifetime context, while sales and excise taxes are less regressive, so the Canadian tax system is as mildly progressive in the lifetime framework as it is in the annual framework[37].

In a different approach to this first problem, Slemrod (1992) notes that a "snapshot" of one year suffers from fluctuations, while a lifetime income perspective requires

[35] The 1966 data used by Pechman and Okner (1974) were updated by Pechman (1985). There, he finds that progressivity fell due to an upward trend in payroll taxes and downward trend in corporate taxes. Browning (1986) indicates that the new data understate transfers and overstate labor income for the poorest groups, and that appropriate adjustments to the data would make the 1985 tax system appear no less progressive than the 1966 system. Pechman (1987) corrects his data and finds virtually no change in progressivity at the low end of the income distribution, but he still finds reduced progressivity at the very top of the income distribution (due to reduced taxes on capital).

[36] An individual at a given percentile of a particular year's annual income distribution may appear at a different place in the lifetime income distribution, both because annual income is volatile and because it tends to rise systematically and then fall with age. Tax incidence across lifetime income groups may also be affected by the shape of the earnings profile: if those with higher lifetime incomes have earlier peaks in their earnings profiles, then they must save more for retirement and bear more burden from taxes on capital.

[37] Poterba (1989) classifies households by current consumption, as a proxy for lifetime income, and he therefore finds that consumption taxes are less regressive than when using annual income to classify households. Lyon and Schwab (1995) use data from the PSID in a life-cycle model, finding that cigarette taxes are just as regressive when using lifetime income rather than annual income as the classifier. They find that alcohol taxes are slightly less regressive.

heroic data assumptions. Slemrod argues that a "time-exposure" of about seven years may be a reasonable compromise. He compares 1967–73 to 1979–85. While annual income inequality has risen substantially over those decades, Slemrod finds less increase in time-exposure income inequality. However, the effect of taxes on inequality is the same in both cases.

To address the last two problems, other researchers have built explicit general equilibrium models that can calculate the effect of all taxes simultaneously on all prices and quantities, from which they can calculate utility-based measures of consumer welfare. For example, Ballard, Fullerton, Shoven and Whalley (1985) specify production functions for 19 industries that use both primary factors and intermediate inputs. Each tax may affect the demand for each factor in each industry. They also specify 12 income groups that receive different shares of income from labor, capital, and indexed government transfers. Assuming utility maximization, they calculate demands for each good by each group that depend on product prices and on after-tax income, while factor supplies depend on net factor returns. The imposition of any tax may then affect prices, and they calculate the EV to measure the burden of each group [38].

A different type of general equilibrium model is built by Auerbach, Kotlikoff and Skinner (1983) and fully described in Auerbach and Kotlikoff (1987). Auerbach and Kotlikoff sacrifice intragenerational heterogeneity to concentrate on intergenerational redistribution. Their model has only one sector but allows for 55 overlapping generations with life-cycle savings decisions. Instead of calculating the incidence of a tax across 12 income groups, they calculate the incidence across age groups. In particular, they find that the switch from an income tax to a wage tax would reduce the burden on the elderly, while the switch to a consumption tax would substantially raise tax burdens on the elderly.

Auerbach and Kotlikoff provide the first computational model of lifetime tax incidence for different age groups, but cannot calculate progressivity across different income groups. Later efforts proceed to calculate lifetime tax incidence for different income groups at each age [Fullerton and Rogers (1993), Altig, Auerbach, Kotlikoff, Smetters and Walliser (2001)]. All of these computational general equilibrium models can calculate the incidence of each tax using explicit production functions and utility-

[38] This type of CGE model captures many behaviors and employs utility-based measures of welfare rather than accounting measures, but it does not capture some other behaviors and effects on utility. Mulligan and Philipson (2000) have a unique "reverse" view of the effect of some redistributive tax policies and other programs. For an example, consider a hypothetical tax credit for health insurance. Under the usual "accounting" approach to tax incidence, this tax credit would seem to be progressive since it provides a flat dollar benefit that is a higher fraction of a poor family's income. They point out that this "merit" good is provided because the rich want for the poor to purchase more health care. If the rich have positive "willingness to pay" for the government to induce the poor to buy more health care, then the program makes the rich better off. It also contrains the choices of the poor more than of the rich. Thus, under this reverse view, such a program is regressive rather than progressive.

based measures of welfare (such as EV), but computational feasibility requires some aggregation across households – such as considering only 12 income groups.

In contrast, the approach based on Pechman and Okner (1974) must assume the incidence of each tax without utility or production functions, but can employ detailed micro data on many thousands of households. This detailed approach also allows calculations of incidence across dimensions other than income (by region, race, gender, or other demographic characteristics). For these reasons, several recent efforts also build upon the original approach of Pechman and Okner. For example, Kasten, Sammartino and Toder (1994) combine data from the Labor Department's Consumer Expenditure Survey, the Commerce Department's Census Bureau, and the Treasury Department's tax returns. Instead of trying to construct a "full" measure of economic income, however, they classify households by a measure of realized cash income. They calculate federal income taxes and payroll taxes for each household, and they assign corporate taxes and federal excise tax burdens according to assumptions about their incidence (but they omit all state and local taxes on income, sales, and property). Despite major changes in federal tax policy between 1980 and 1993, they find virtually no change in the overall level of taxation or in the distribution of burdens, except a slight decline in the effective tax rate for those in the top one percent of the income distribution.

As another example, Gale, Houser and Scholz (1996) use data from the Survey of Income and Program Participation (SIPP) and classify households by "expanded" income that includes some imputations (e.g., employer-provided health insurance) but not others (e.g., imputed rental income from owner-occupied homes). They consider federal and state income taxes, corporate taxes, and payroll taxes, but not federal excise taxes, state sales taxes, or local property taxes. They do consider transfer income. Like prior authors, they find that the current tax system is progressive.

While the three studies mentioned above appear quite similar, it is important to note that they differ in subtle but important ways that can affect the incidence results obtained: each such study makes its own choices about where to get the data, whether to use individual taxpayers or families, which set of taxes to put in the numerator of the effective tax rate (ETR) calculation, and what definition of income to use to classify taxpayers (and to put in the denominator of the ETR calculation). Even once the ETR is calculated at each income level, these studies could choose from among many measures of progressivity [39].

We now turn to empirical tests of these incidence assumptions. First, for the payroll tax, virtually all applied incidence studies assume that both the employee share and the employer share are borne by the employee (through a fall in the net wage by the full

[39] Kiefer (1984) reviews indices of progressivity. For example, the Pechman and Okner (1974) index is calculated as the Gini coefficient after taxes minus the Gini coefficient before taxes, all divided by the latter $((\text{Gini}_{AT} - \text{Gini}_{BT})/\text{Gini}_{BT})$. Other measures such as the Suits Index [Suits (1977)] are based on the tax concentration curve.

amount of payroll tax). This assumption has been tested and confirmed repeatedly, going back to Brittain (1971) who used a 1958 cross-section of 13 industries in 64 nations and found full burdens on labor. Gruber (1997) reviews other more recent empirical studies that use both cross-section and time-series data, consistently finding full burdens on labor. Gruber (1997) himself uses data from a survey of manufacturing plants in Chile over the 1979–86 period to estimate the effects of dramatic 1981 cuts in that country's payroll tax, and finds that "the reduced costs of payroll taxation to firms appear to have been fully passed on to workers in the form of higher wages . . . " (p. S99)[40].

Second, for sales and excise taxes, the standard assumption is that burdens fall on the consumers of taxed products (through higher prices). For example, Fullerton (1996) and Metcalf (1999) employ a model with constant returns to scale and perfect competition, such that the long-run supply curve is flat, and any product tax logically must be passed on to purchasers. They then use input–output evidence on each industry's purchases of taxed products to calculate the increase in the cost of production of each industry – and thus the increase in each equilibrium output price. Finally, data on consumer expenditures can be used to indicate which consumers pay those higher prices[41].

This assumption, too, has been tested, but results are mixed. If the flat supply curve in the above analysis is replaced by an upward-sloping supply curve, then the burden of an excise tax might be shared in any proportions between consumers and producers, such that product price rises by *less* than the tax. In contrast, several studies reviewed by Poterba (1996) find "overshifting", such that the product price rises by *more* than the tax. In his own analysis, however, Poterba uses city-specific clothing price indices for 14 cities during 1925–39 (finding less-than-complete forward shifting) and eight cities during 1947–77 (finding mild, if any, overshifting). On the other hand, Besley and Rosen (1999) point out that overshifting is perfectly consistent with several models of imperfect competition (as discussed more in the next section). They find substantial overshifting for more than half of the 12 goods they study in 155 cities. This result would make excise taxes even more regressive than conventionally thought.

Finally, for the personal income tax, applied studies have consistently assumed that economic incidence is the same as statutory incidence – on the taxpayer – even though this assumption has never been tested.

In summary, few of the standard assumptions about tax incidence have been tested and confirmed (e.g., payroll tax). Most others have never been reliably tested (the

[40] In a survey of all labor economists at top-40 U.S. institutions, Fuchs, Krueger and Poterba (1998) find that the median belief about the payroll tax is that 20% of the burden is borne by employers.

[41] Metcalf (1999) uses the methodology of Caspersen and Metcalf (1994) to compute a measure of lifetime income for each household, and thus can calculate the incidence of these excise taxes across lifetime income groups or across annual income groups. He finds that excise taxes on fuels are regressive when measured annually by themselves, but can be slightly progressive when measured on a lifetime basis if the revenue is used to reduce payroll and personal income taxes in a progressive fashion.

personal income tax, corporate income tax, and local property tax). The standard assumption about the corporate income tax that the burden falls 100% on capital remains the standard assumption even though it is commonly believed to be false (because of international capital mobility and endogenous saving)[42]. The standard assumption about sales and excise taxes is that the burden is shifted 100% to consumers, and this assumption has been tested several times. Some of these studies cannot reject 100% shifting to consumers, while others find significantly less than 100% shifting, and still others find significantly more than 100% shifting.

Many general equilibrium simulation studies "calculate" the incidence of each tax based on carefully-articulated theories, and many data-intensive studies use these results to "assume" the incidence of each tax. But competing theories are rarely tested, and so econometric estimation remains fertile ground for new research.

3. Imperfect competition

In this section, we consider the effects of taxation in imperfectly competitive markets. The analysis, for the most part, is partial equilibrium in nature, and we consider both *ad valorem* and specific taxes on output[43]. Imperfectly competitive markets can appear in a wide variety of forms, and the tax analyst faces the difficult task of determining which model is appropriate in each application (see Tirole (1988) for an excellent discussion of different models). Broadly speaking, we can first classify models on the basis of whether they consider homogeneous or heterogeneous products. Models with different firms producing identical products include the Bertrand oligopoly and the Cournot–Nash oligopoly model. Those with heterogeneous goods include the monopolistic competition models [e.g., Dixit and Stiglitz (1977) and Spence (1976)], location models [e.g., Hotelling (1929) and Salop (1979)], and models of vertical differentiation [e.g., Gabszewicz and Thisse (1979) and Shaked and Sutton (1982)]. Whether products are homogeneous or heterogeneous, we will find that the impact of taxes on prices works through both direct and indirect channels (with the indirect channels differing across models).

3.1. Oligopolies

Let us first turn to the case of Bertrand oligopoly with identical firms and a constant returns to scale production function. Bertrand competition is a Nash equilibrium

[42] Fuchs, Krueger and Poterba (1998) surveyed public finance economists at top-40 U.S. institutions and found that the median belief about the corporate income tax is that 40% of the burden is borne by capital.

[43] Unlike perfect competition, the incidence impact of equal revenue *ad valorem* and specific taxes differs in imperfectly competitive markets. See Delipalla and Keen (1992) for a comparison of these two taxes.

concept in which firms compete in prices. The price equilibrium is quite simple: firms compete by lowering prices until all firms set price equal to their common marginal cost. No firms earn economic profits, leaving no incentive for entry or exit. The effects of a unit tax on output in such a model is straightforward. Since the producer price cannot fall below marginal cost, the entire tax is passed forward to consumers. More generally, even with a positive aggregate supply elasticity, the Bertrand model and perfect competition produce the same equilibrium outcome.

We next turn to the Cournot–Nash oligopoly model in which identical firms compete by choosing levels of output conditional on their expectations of their competitors' output levels. We proceed in two steps: first by fixing the number of firms in the market at N and then by allowing free entry. To simplify matters, we will assume firms are identical and that the equilibrium is symmetric[44].

Consider firm i in the market. Its profit function is given by

$$\pi_i(q_i) = (1-\tau_v)\,p(q_i + Q_{-i})\,q_i - c(q_i) - \tau_s q_i, \tag{3.1}$$

where q_i is the output of the ith firm, Q_{-i} is the output of all other firms in the market, and $p(Q)$ is the inverse demand function for market demand Q. The cost function is $c(q_i)$, and τ_v and τ_s are *ad valorem* and specific taxes on q with statutory incidence on the firm.

The first order condition for the ith firm is given by

$$(1-\tau_v)\,p' q_i + (1-\tau_v)\,p - c' - \tau_s = 0, \tag{3.2}$$

where a prime indicates a first derivative. Second order conditions are

$$(1-\tau_v)\,p'' q_i + 2(1-\tau_v)\,p' - c'' < 0, \tag{3.3}$$

or

$$\frac{\tilde{p}'}{N}\,(\eta + N + Nk) < 0, \tag{3.4}$$

where $\tilde{p} \equiv (1-\tau_v)\,p$ is the producer price, $\eta = Qp''/p'$ is the elasticity of the slope of the inverse demand function and $k = 1 - \frac{c''}{\tilde{p}'}$ measures the relative slopes of the demand and marginal cost curves. Since $p' < 0$, the second order conditions require $\eta + N + Nk > 0$.

[44] The Cournot–Nash assumption is that firm i optimizes assuming that other firms do not adjust output in response. An alternative approach is to apply a conjectural variation assumption. Let $\lambda = \mathrm{d}Q/\mathrm{d}q_i - 1$ be the conjectured response in output of all other firms as firm i increases output by 1 unit. The Cournot–Nash assumption is equivalent to assuming that λ equals 0. Papers that employ the conjectural variations approach include Katz and Rosen (1985), Seade (1985) and Stern (1987). They also consider tax incidence in a Cournot model with a fixed number of firms.

In a symmetric equilibrium, we need only solve for p and q using the two equations:

$$p = p(Nq), \tag{3.5}$$

and

$$(1-\tau_v)\,p'(Nq)\,q + (1-\tau_v)\,p(Nq) - c'(q) = \tau_s. \tag{3.6}$$

Differentiating Equation (3.6) with respect to τ_s and rewriting, we get:

$$\frac{\mathrm{d}q}{\mathrm{d}\tau_s} = \frac{1}{\tilde{p}'(\eta + N + k)}. \tag{3.7}$$

It follows directly that

$$\frac{\mathrm{d}Q}{\mathrm{d}\tau_s} = \frac{N}{\tilde{p}'(\eta + N + k)}, \tag{3.8}$$

and

$$\frac{\mathrm{d}\tilde{p}}{\mathrm{d}\tau_s} = (1-\tau_v)\,p'\left(\frac{\mathrm{d}Q}{\mathrm{d}\tau_s}\right) = \frac{N}{\eta + N + k}. \tag{3.9}$$

If second-order conditions hold, then $\eta + N + Nk > 0$. And, if $\eta + N + k > 0$, then output falls and the tax is (to some extent) passed forward to consumers. The degree of forward shifting of the unit tax on output depends on the elasticity of the slope of the inverse demand function (η), the number of firms (N), and the relative slopes of the marginal cost and inverse demand functions (k).

Overshifting occurs when the producer price rises by more than the excise tax. As we showed in an earlier section, this outcome is impossible in perfectly competitive markets. Once imperfectly competitive markets are allowed, overshifting becomes a possibility and can be guaranteed in some model specifications. Overshifting can occur because of the existence of market power and strategic behavior among firms. Firms recognize that forward shifting of the tax will decrease demand for their product. Thus, under some circumstances, they will wish to raise the price more than the increase in tax to compensate for the revenue loss from decreased demand[45].

By definition, overshifting occurs if the derivative in Equation (3.9) is greater than 1, which means that $\eta + k < 0$. If costs are linear in output, then $c'' = 0$ and $k = 1$, so a necessary and sufficient condition for overshifting ($\mathrm{d}\tilde{p}/\mathrm{d}\tau_s > 1$) is that $\eta < -1$. Consider a constant elasticity demand function with demand elasticity $\varepsilon < 0$. In that

[45] Note that overshifting does not imply an increase in profits for the firm. In fact, if demand is Cobb–Douglas, profits are unaffected by a marginal increase in a specific tax despite the existence of overshifting.

case, $\eta = (1-\varepsilon)/\varepsilon < -1$ for all $\varepsilon < 0$. Overshifting will *always* occur, and it increases as demand becomes less elastic (as η increases in absolute value).

Producer prices rise with an increase in an *ad valorem* tax as follows:

$$\frac{\mathrm{d}\tilde{p}}{\mathrm{d}\tau_v} = \frac{Np(1+1/\varepsilon)}{\eta + N + k}, \tag{3.10}$$

where $\varepsilon < 0$ is the price elasticity of demand. Overshifting of an *ad valorem* tax occurs when the percentage change in the producer price exceeds 100%, and it occurs in this model when $-N < \eta + k < N/\varepsilon$.

Having analyzed tax incidence in the fixed-N Cournot oligopoly, analysis of monopoly markets is straightforward (simply set $N = 1$)[46]. Assuming no pre-existing *ad valorem* tax, a monopolist can shift more than 100% of an excise tax (τ_s) when $1/(\eta + 1 + k) > 1$, or $-1 < \eta + k < 0$. With linear costs, overshifting occurs when $-2 < \eta < -1$. Overshifting cannot occur in the simple case of linear demand and linear costs (because $\eta = 0$ and $k = 1$). From Equation (3.9), $\mathrm{d}\tilde{p}/\mathrm{d}\tau_s$ equals 1/2 in the linear demand/cost case. On the contrary, if demand is of the constant elasticity type, and costs are linear, then overshifting will always occur in the monopoly model. Thus, the two models most typically assumed (constant slope or constant elasticity) each impose a particular incidence pattern in the monopoly model with constant marginal costs of production [see, for example, Musgrave (1959)].

Returning to the general oligopoly model with fixed number of firms, note that N does not affect the overshifting condition for excise taxes but does affect the degree of overshifting. Again, we consider the case with no pre-existing *ad valorem* tax. Assume that $-N < \eta + k < 0$ (so that $\mathrm{d}\tilde{p}/\mathrm{d}\tau_s > 1$). Then $\mathrm{d}^2\tilde{p}/\mathrm{d}\tau_s\mathrm{d}N < 0$. In other words, for given values of η and k, overshifting is maximized for a monopolist and disappears as N approaches infinity.

Now allow for free entry in the Cournot model. In addition to Equations (3.5) and (3.6), we need a third equation to pin down the equilibrium number of firms. Firms will enter until the marginal firm earns zero profits. With identical firms, the zero profit condition becomes

$$(1-\tau_v)\,p(Nq)\,q - c(q) - \tau_s q = 0. \tag{3.11}$$

Equations (3.5), (3.6) and (3.11) determine p, q, and N. We now limit our discussion to changes in excise taxes and assume τ_v equals zero. Thus, $\tilde{p}$ equals p, and

$$\frac{\mathrm{d}p}{\mathrm{d}\tau_s} = \frac{N(k+1)}{\eta + N + Nk}. \tag{3.12}$$

With a linear cost function ($k = 1$), $\mathrm{d}p/\mathrm{d}\tau_s > 1$ iff $\eta < 0$. We now have a wider class of aggregate demands for which overshifting will occur [see Besley (1989) for a fuller

[46] See Bishop (1968) for an early treatment of *ad valorem* and unit taxes under monopoly.

analysis of this point]. The indirect effect of the tax on industry structure contributes to overshifting (where structure here means the number of firms). To see this, note that

$$\frac{\mathrm{d}N}{\mathrm{d}\tau_s} = \frac{N}{p'q}\left(\frac{\eta + k + 1}{\eta + N + Nk}\right), \tag{3.13}$$

and $\mathrm{d}N/\mathrm{d}\tau_s < 0$ if $\eta + k + 1 > 0$. With a positive fixed cost and constant marginal cost, then the equilibrium number of firms will fall in the range of η between −2 and zero. This decrease in firms tends to drive up prices, and the effect is that overshifting occurs for η between −1 and 0 in the variable-N case but not in the fixed-N case. Note that this overshifting does not lead to increased economic rents for producers: in the free-entry model, profits are always zero, so the effect of the unit tax is to drive up costs of production and to induce exit if aggregate demand is sufficiently elastic.

More generally,

$$\frac{\mathrm{d}p}{\mathrm{d}\tau_s} - \left.\frac{\mathrm{d}p}{\mathrm{d}\tau_s}\right|_N = \left(\frac{p - c - t}{\eta + N + k}\right)\left(-\frac{\mathrm{d}N}{\mathrm{d}\tau_s}\right). \tag{3.14}$$

Entry and exit affect the degree of forward shifting through changes in the equilibrium number of firms. Assuming $\eta + N + k > 0$ (consumer prices rise with a unit tax in the fixed-N case), then consumer prices rise more as the equilibrium number of firms falls so long as some market power is in effect ($p - c - t > 0$). This indirect price effect arises because the decrease in the equilibrium number of firms yields increased market power for the remaining firms. Interestingly, if we start at an efficient equilibrium with no market power, then taxes have no indirect effect on prices. The result that part of the incidence impact of a tax occurs through changes in the equilibrium number of firms is a result that will occur in a number of models of imperfectly competitive firms, as we shall see later.

Delipalla and Keen (1992) show that in both the Cournot–Nash and free-entry oligopoly models, *ad valorem* taxes are less likely to lead to overshifting than unit taxes. Venables (1986) notes that *ad valorem* taxes dampen the impact of output changes on prices and thus make the market act more like a perfectly competitive market. Applying that insight here, *ad valorem* taxes will have an impact more like taxes in perfectly competitive markets and so should lead to less overshifting than unit taxes.

Support for overshifting in imperfectly competitive markets appears in a number of empirical studies. Karp and Perloff (1989) econometrically estimate the conjectural variations parameter ($\lambda = \mathrm{d}Q/\mathrm{d}q_i - 1$) in the Japanese market for televisions. They find evidence for imperfectly competitive markets and, based on that conclusion, derive the incidence of a domestic luxury tax on televisions. They find more than 100% forward shifting. Their conclusions depend heavily on the structural assumptions imposed in their model. Harris (1987) analyzed the 1983 increase in the U.S. federal excise tax on cigarettes from \$.08 to \$.16 per pack[47]. He finds that the \$.08 tax increase led to a consumer price rise of \$.16 per pack. As mentioned in the previous section, Besley and Rosen (1999) investigate the impact of changes in state and local sales taxes on product prices for a highly disaggregated set of commodities[48]. They employ quarterly data for 12 goods in 155 cities over a nine-year period (1982–1990), about 4200 observations per commodity. They find overshifting for a number of commodities, including bread, shampoo, soda, and underwear. They cite evidence by Anderson (1990) for market power in many local grocery markets, and estimated markups that are 2.355 times price for retail trade, from Hall (1988).

Poterba (1996), in contrast, finds no evidence for overshifting of sales taxes. The major difference between the Besley and Rosen study and the Poterba study is the level of disaggregation; it is possible that any overshifting in the latter study is obscured by changes in composition of the items in the bundles studied[49]. Doyle (1997) also finds evidence of overshifting in the new car market, where a one-dollar increase in tax is associated with a price increase ranging from \$2.19 for luxury cars to \$2.97 for trucks.

3.2. Differentiated products

The oligopoly models discussed above suffer from the restrictive assumptions that goods are identical and that no distinction can be made between different brands. In some markets (e.g., agricultural commodity markets), this may be a reasonable assumption. In most other markets, however, producers go to great length to differentiate their products. Product differentiation creates some monopoly power, and the results in the fixed-N oligopoly model indicate that the ability to pass taxes forward depends importantly on the number of competitors in the market. In this section, we consider several models of differentiated products and examine the relationship between product competition and tax incidence.

We begin with Dixit and Stiglitz (1977) and their model of monopolistic competition. This is a somewhat special model in that each product competes with all other products, and the main thrust of the model is to illustrate the benefits of product variety. It is useful to begin with this model, however, as it highlights the importance of product differentiation – a feature left out of the homogeneous-good oligopoly model. Consider the following simplified Dixit–Stiglitz model of product variety based on Krugman (1980). Consumers are identical and maximize a utility function

$$U(x_1, x_2, \ldots, x_N) = \frac{1}{\theta} \sum_{i=1}^{N} x_i^{\theta}, 0 < \theta < 1. \tag{3.15}$$

[47] The increase was first temporary but was made permanent in 1986. See Harris (1987) for details.

[48] They consider such items as a three-pound can of Crisco, a dozen large Grade-A eggs, a 200-count box of Kleenex facial tissues, and (naturally) the board game Monopoly.

[49] The studies also differ by cities and time periods examined and econometric specifications employed.

Consumption goods enter utility symmetrically but are not perfect substitutes (unless θ equals 1). Individuals maximize utility subject to the budget constraint that (exogenous) income (M) equals expenditures:

$$\sum_{i=1}^{N} p_i x_i = M, \tag{3.16}$$

where p_i is the consumer price of the ith good. From the first-order conditions, we can derive the demand functions:

$$x_i = \lambda^{-\varepsilon} p_i^{-\varepsilon}, \varepsilon = \frac{1}{1-\theta} > 1, \tag{3.17}$$

where λ is the private marginal utility of income. If N is large, we can assume that the pricing decisions of an individual firm will have negligible effect on λ and demand can be written as

$$x_i = A p_i^{-\varepsilon}, \tag{3.18}$$

and is of the constant elasticity variety.

Firms maximize profits, and we assume that costs are linear of the form $cx_i + F$, where c is marginal cost and F is fixed cost. Letting $\tilde{p} \equiv (1 - \tau_v)p$ be the producer price with an *ad valorem* tax (τ_v), the firm's pricing rule is given by the standard monopolist's pricing rule:

$$\tilde{p} = \left(\frac{\varepsilon}{\varepsilon - 1}\right)(c + \tau_s). \tag{3.19}$$

For either an excise or *ad valorem* tax applied to a particular industry only, we can differentiate Equation (3.19). Thus,

$$\frac{\mathrm{d}\tilde{p}}{\mathrm{d}\tau_s} = \frac{\varepsilon}{\varepsilon - 1} > 1, \tag{3.20}$$

and

$$\frac{\mathrm{d}\tilde{p}}{\mathrm{d}\tau_v} = 0. \tag{3.21}$$

The insights from monopoly model in the last section carry forward: an excise tax is more than 100% forward shifted (constant elasticity and linear cost result), while an *ad valorem* tax has no impact on the producer price but is entirely shifted forward to consumers.

A disadvantage of the Dixit–Stiglitz model is that all products are treated as equal competitors with other products. A quick look at any number of markets indicates that this assumption is untenable. We next turn to a model of spatial competition where firms locate themselves in product space to capture maximal customers in a simultaneous entry game. We use the Salop (1979) circle model as developed to analyze *ad valorem* and excise taxes by Kay and Keen (1983)[50]. The virtue of the circle model is that it explicitly allows for modeling of the number of firms in equilibrium [unlike the linear model of Hotelling (1929)]. Following Salop, we assume N identical firms simultaneously deciding whether to enter a market where consumers are located uniformly around the circle, and where each consumer wishes to purchase 1 unit of the product. Firms that enter locate equidistantly around a unit circle. Thus, in equilibrium, each firm will face demand of $1/N$ (assuming the market is covered). Each individual will purchase at most one unit of the good, and each prefers to purchase the good of quality or location x that is as close as possible to their most-preferred quality (x^*). Specifically, the consumer's cost of the good is the purchase price (p_i) plus a "transport cost" that is assumed to be a constant h times the distance from their location $|x - x^*|$. Utility for a consumer who purchases a unit of x obtains utility equal to

$$U = \bar{s} - p - h\,|x - x^*|\,, \tag{3.22}$$

where $\bar{s}$ is an arbitrary constant sufficiently large to ensure $U > 0$.

Consider a consumer located at $\hat{x}$, between 0 and $1/N$ from firm i. That consumer will be indifferent between purchasing from firm i and firm $i + 1$ if

$$p_i + h\hat{x} = p + h(1/N - \hat{x}), \tag{3.23}$$

where p_i is the price charged by the ith firm, and p is the price charged by other firms. For that price p_i (making the consumer at $\hat{x}$ indifferent), demand for the ith firm's good, $D(p_i, p)$, will be equal to $2\hat{x}$. Solve Equation (3.23) for $\hat{x}$ and double it, to get:

$$D(p_i, p) = \frac{p - p_i}{h} + \frac{1}{N}. \tag{3.24}$$

The firm maximizes profits by choosing price. It faces an *ad valorem* tax rate τ_v and a unit tax rate τ_s. Profits are given by

$$\pi_i = ((1 - \tau_v)\,p_i - c - \tau_s)\left(\frac{p - p_i}{h} + \frac{1}{N}\right) - F, \tag{3.25}$$

[50] Anderson, de Palma and Kreider (2001) provide a more general analysis that incorporates the Kay and Keen model as a special case. These authors stress the similarity of results under a Bertrand–Nash environment with differentiated products to the Cournot–Nash setting with homogeneous products analyzed by Delipalla and Keen (1992). Metcalf and Norman (2001) extend the Kay and Keen model to allow for price discrimination and costly re-anchoring of product types in response to entry.

where c is marginal cost and F is fixed cost. Take the derivative, set it equal to zero, and set p_i equal to p (assuming identical firms have equal price in equilibrium), to yield:

$$p_i = \frac{h}{N} + \frac{c + \tau_s}{1 - \tau_v}. \tag{3.26}$$

We can rewrite this in terms of the producer price ($\tilde{p}$):

$$\tilde{p} = \frac{(1 - \tau_v)\,h}{N} + c + \tau_s. \tag{3.27}$$

Thus, unit taxes are fully passed forward in the sense that the producer price rises by the full amount of the tax[51]. Strictly speaking, this statement is only true if the equilibrium number of firms is unaffected by changes in the excise tax.

We need a second equation to pin down the equilibrium number of firms. A zero-profit condition for the marginal firm does this. In equilibrium, each firm covers $1/N$ of the market. Plug this supply into the profit function and set profits equal to zero, to get:

$$((1 - \tau_v)\,p_i - c - \tau_s)\left(\frac{1}{N}\right) = F. \tag{3.28}$$

Substituting Equation (3.26) into (3.28) and solving for N, we get:

$$N = \sqrt{\frac{(1 - \tau_v)\,h}{F}}. \tag{3.29}$$

While a change in the excise tax does not affect the equilibrium number of firms, a change in the *ad valorem* tax does.

Ad valorem tax incidence can be decomposed into two components: a direct effect and an indirect effect through the change in the equilibrium number of firms. Fixing N,

$$\left.\frac{\partial \tilde{p}}{\partial \tau_v}\right|_N \equiv \left.\frac{\partial (1 - \tau_v)\,p}{\partial \tau_v}\right|_N = -\frac{h}{N}. \tag{3.30}$$

The complete incidence is given by

$$\frac{\partial \tilde{p}}{\partial \tau_v} = -\frac{1}{2}\sqrt{\frac{Fh}{1 - \tau_v}} = -\frac{1}{2}\frac{h}{N}, \tag{3.31}$$

exactly half the incidence in Equation (3.30) where N is fixed. In other words, firm exit cuts the burden on producers in half (and raises the burden on consumers)[52].

[51] From Equation (3.26), the consumer price rises by more than the unit tax in the presence of an *ad valorem* tax. The increase in price by the firm to cover the unit tax must also cover an increase in *ad valorem* tax collections. It is not the case, however, that the unit tax is more than 100% passed forward.

[52] Firms exit because an increase in *ad valorem* taxation is equivalent (from the firm's point of view) to an increase in fixed cost relative to revenue. See Kay and Keen (1983) for details.

Some of the theories described in this section have also been incorporated into computable general equilibrium models with imperfect competition. For example, Harris (1984) builds an open-economy trade model of Canada with 29 different industries, of which 20 are potentially noncompetitive. He specifies a fixed cost for each plant within an industry, free entry, and two alternative models (with and without product differentiation). He finds that "the estimated welfare gains from trade liberalization are substantial in the industrial organization model and on the order four times larger than the gains estimated from the competitive model" (p. 1031). In terms of incidence, internationally-mobile capital in his model means that capital-owners are unaffected, but his Table 2 (p. 1028) reveals that the gain in labor productivity from trade liberalization can be four to six times higher in the imperfectly competitive models.

Once we allow for heterogeneous products, we see new avenues for taxes to affect equilibrium prices. Consider a duopoly model with heterogeneous goods in which firms compete over price, and product quality is endogenous. Cremer and Thisse (1994) present a model of vertical product differentiation and show that a uniform *ad valorem* tax applied to both firms *reduces* the consumer price in equilibrium. Part of the price decrease arises from a decrease in quality and hence reduction in marginal (and average) production costs. But the authors note that the price decrease exceeds the cost reduction. A reduction in quality differences sharpens price competition and reduces monopoly power of firms.

A general point can be made here. With differentiated products, taxes can affect prices over additional avenues, whether through the degree of product variety as in the Kay and Keen model or through the distribution of product quality as in the Cremer and Thisse model. Non-price competition can substantially affect the degree to which output taxes are passed forward to consumers and can lead to counterintuitive results, as in the Cremer and Thisse model [53].

4. Dynamic models and incidence

Models with intertemporal optimization allow for endogenous saving and investment. The essential engine of long-run incidence in these models is the impact of taxes on capital–labor ratios (and thus factor prices). We shall also see, however, that short-run inelastic capital supply plays an important role through asset price revaluations in response to tax policy. Anticipations also become important.

Beginning in the 1960s, research on factor taxation in a dynamic setting used neoclassical growth models either with exogenously-specified savings functions or with

[53] In the Cremer and Thisse model, the impact of *ad valorem* taxes on market power has obvious welfare implications. They show that a small increase in a uniform *ad valorem* tax from a no-tax equilibrium is always welfare improving. See Auerbach and Hines (2002) in Volume 3 of this Handbook for further discussion.

overlapping generations (OLG models). In a two-period setting, OLG models have been extensively discussed by Kotlikoff and Summers (1987) and Kotlikoff (2002), and we refer the reader there for more detail. Here, we briefly discuss capital income taxation in a growing economy using a model due to Feldstein (1974). We then turn to perfect-foresight models in which savings behavior follows explicitly from consumer preferences. This provides a link between the savings function and the pure rate of time preference that is lacking in the previous literature. Finally, we turn to asset-pricing models and transition dynamics.

4.1. Taxation in a growing economy

Static models of tax incidence cannot easily capture the impact of changes in the capital-labor ratio on factor prices. Consider a simple linearly-homogeneous production function $y = f(k)$, where output per worker (y) is a function of the capital–labor ratio (k). With competitive pricing, each factor price will be a function of k:

$$r(k) = f'(k), \tag{4.1}$$

$$w(k) = f(k) - kf'(k), \tag{4.2}$$

where r is the rental rate of capital and w the wage rate. As k grows, the rental rate decreases and the wage rate increases. If net capital income is taxed at rate τ, and r is the net rental rate, then the marginal product of capital is equal in equilibrium to $(1 + \tau)r$. Feldstein (1974) develops a model to analyze the long-run incidence of a capital income tax and concludes that much (if not all) of the burden of the tax is shifted to workers in the form of lower wages resulting from a decline in the capital–labor ratio. He notes that a change in the tax on capital income per person (rk) has two components:

$$\frac{\mathrm{d}(rk)}{\mathrm{d}\tau} = k\frac{\mathrm{d}r}{\mathrm{d}\tau} + r\frac{\mathrm{d}k}{\mathrm{d}\tau}. \tag{4.3}$$

He argues that the second term should not be viewed as a burden of the tax, but rather as a shift in the timing of consumption. Thus, Feldstein measures the long-run burden of a new capital income tax as the ratio of the loss to capitalists ($-k\mathrm{d}r$) to the new tax revenue ($rk\mathrm{d}\tau$); the burden on owners of capital from an increase in tax is the ratio of ($-k\mathrm{d}r$) to the loss in real income ($-(k\mathrm{d}r + \mathrm{d}w)$).

The conclusions from the model are particularly stark in a two-class world in which all savings is from capital income only. Assuming that the savings rate s is a function of the net rate of return ($s = s(r)$), then saving per person equals $s(r)\,rk$. In the long-run steady state, the capital stock must grow at the rate of growth of the population (n), and equilibrium in capital markets requires

$$s(r)\,rk = nk. \tag{4.4}$$

The net rate of return (r) is a function of the growth rate of the population (n) only, and is unaffected by a change in the capital income tax rate. Thus, capital owners bear none of the burden of the tax in the steady state. Even if the savings rate out of labor income is positive, much of the burden of the capital tax can be shifted to labor[54].

Once saving is endogenous, other "standard" results can also be reversed. For example, because land is inelastically supplied, many presume that a tax on land is borne by the landowner. In a model where land serves not only as a factor of production but also as an asset, however, Feldstein (1977) shows that a tax on land rent then induces investors to increase holdings of other assets in their portfolios. The resulting increase in reproducible, physical capital can then lead to an increase in the wage rate and a decrease in the return to physical capital. Hence, part of the tax on land rent is shifted to capital, with wage rates rising in response to the greater capital–labor ratio.

Boadway (1979) points out that focusing on the steady state provides an incomplete picture of the impact of a capital income tax. He takes Feldstein's (1974) model and parameter assumptions and carries out simulations of a marginal increase in capital income taxation that finances a reduction in labor income taxation. In steady state, labor is made worse off by the shift, with wage rates falling over 7% in the long run[55]. But Boadway shows that the wage rate first rises before falling, and in fact is higher for 65 years in his simulation[56]. A complete picture of the burden would have to discount and add up the workers' gains and losses over time.

One simple way to measure the burden shift would be to compute the present discounted value of the change in wage income assuming some given discount rate. We note four problems with this approach. First, the discount rate is exogenous rather than being linked to consumer preferences. Second, it would be preferable to have some dynamic measure of compensating or equivalent variation for the tax shift. Third, the savings rate $s(r)$ does not follow from consumer preferences. Fourth, it depends only on current information with no anticipations. For example, an announcement today of a temporary surtax on capital income for ten years that would begin five years from now should have an impact on capital accumulation over the next five years. The models of Feldstein, Grieson, and Boadway cannot capture this effect. We turn next to a model based on Judd (1985a) that addresses all four of these concerns.

4.2. Taxation in a perfect foresight model

The essential departure in the model of Judd (1985a) is the assumption of perfect foresight by an infinitely-lived individual. Perfect foresight is an extreme assumption and perhaps should be viewed as one end of a continuum; it has the attractive quality

[54] Feldstein presents an example with Cobb–Douglas production. With equal savings rates for labor and capital, he calculates that 1/3 of the tax is shifted to labor. With a savings rate for capital twice that for labor, half the tax is shifted.

[55] Grieson (1975) also shows that a shift from wage to capital income taxation can make workers worse off in the long run through a decrease in the steady-state capital-labor ratio.

[56] He also reports results where the wage rate rises for over 75 years.

of allowing individuals to look forward and thus to make decisions today on the basis of beliefs about the world in the future.

Consider a very simple world with only two people: a capitalist and a worker, each of whom lives forever[57]. The capitalist earns income only from the rental of capital, while the worker earns income only from labor supply (fixed at one unit). Workers do not save, and the only purpose of taxation is to redistribute income from capitalists to workers[58]. If τ is the tax rate on capital income, we can consider policy experiments of the form $\mathrm{d}\tau = \varepsilon h(t)$ where ε is small and $h(t)$ is used to represent the timing of the policy under consideration. For example, $h(t) = 1$ for $t \geqslant 0$ would be an immediate permanent increase in capital income taxation, while $h(t) = 1$ for $t \geqslant T$ would be a permanent increase beginning at some date T in the future (but announced at time 0). Finally, a temporary tax increase could be modeled by $h(t) = 1$ for $0 \leqslant t \leqslant T$, and $h(t) = 0$ for $t > T$.

Output is produced according to a concave production function $f(k)$ which gives output per worker in terms of capital per worker. The produced good is taken as the numeraire good and can be used for consumption or investment. In equilibrium, factor prices are given by

$$r_t = f'(k_t), \tag{4.5}$$

$$w_t = f(k_t) - k_t f'(k_t), \tag{4.6}$$

where r_t is the rental rate for capital and w_t is the wage paid to the worker.

Whereas neoclassical growth models [e.g., Feldstein (1974), Grieson (1975), Boadway (1979) and Bernheim (1981)] do not directly link savings behavior to key utility parameters (in particular, the pure rate of time preference), Judd models savings behavior directly from the intertemporal optimization problem of capitalists[59]. Specifically, the capitalist maximizes an additively-separable utility function of the isoelastic form:

$$U^k = \int_0^\infty \mathrm{e}^{-\rho t} \frac{(c_t^k)^{1-\beta}}{1-\beta} \mathrm{d}t, \tag{4.7}$$

by choosing a time path of consumption (c_i^k) and capital (k_t) subject to the constraint

$$c_t^k + \dot{k}_t = (1-\tau)\, r k_t, \tag{4.8}$$

and some given level of the capital stock at time zero (k_0). The pure rate of time preference (ρ) is fixed (and the same both for the capitalist and the worker). A dot

[57] The infinitely-lived consumer assumption can be justified in terms of the dynastic model of Barro (1974).

[58] These assumptions are all innocuous. See Judd (1987) for discussion of endogenous labor supply and other generalizations.

[59] To avoid confusion about who is a worker as opposed to a capitalist, Judd specifies that the worker does not save anything. Consumption for the worker is simply the wage received plus a transfer from the government, financed by the capital income tax.

over a variable indicates a time derivative. The parameter β is the elasticity of the marginal utility of consumption. We assume that utility is concave in consumption so that $\beta > 0$. Along an optimal path, the capitalist trades off a unit of consumption today against the benefit of increased consumption in the future from investing the unit and receiving a net return in the future:

$$u'(c_t^k) = (c_t^k)^{-\beta} = \int_t^\infty \mathrm{e}^{-\rho(s-t)}(1-\tau)\, ru'(c_t^k)\,\mathrm{d}s = \int_t^\infty \mathrm{e}^{-\rho(s-t)}(1-\tau)\, r(c_t^k)^{-\beta}\mathrm{d}s. \tag{4.9}$$

The optimal time path of consumption for the capitalist is determined by differentiating Equation (4.9) and substituting in Equation (4.5):

$$\dot{c}^k = \frac{-(\rho - (1-\tau) f'(k))\, c^k}{\beta}, \tag{4.10}$$

where we have omitted the time subscripts. Capital accumulation is given by

$$\dot{k} = (1-\tau)\, k f'(k) - c^k. \tag{4.11}$$

Equations (4.10) and (4.11) are the equations of motion for the system.

In the steady state, Equation (4.10) shows that the net return to capital is constant and equal to ρ. This suggests that capital taxes are shifted entirely to workers through adjustments in the capital–labor ratio. While the net return is fixed in the long run, however, it can vary along a transition path to the new steady state, and redistribution can occur along this transition path. For a complete picture, as we shall see, it is important to focus not only on the steady state but on the entire transition path.

We now entertain a change in capital income taxation where a policy of the form $\mathrm{d}\tau = \varepsilon h(t)$ is announced as of the present time ($t = 0$). Thus, the equations of motion become

$$\dot{c}^k = \frac{-(\rho - (1-\tau-\varepsilon h(t)) f'(k))\, c^k}{\beta}, \tag{4.10'}$$

and

$$\dot{k} = (1-\tau-\varepsilon h(t))\, k f'(k) - c^k. \tag{4.11'}$$

Consumption and capital (as well as their time derivatives) are now functions of ε as well as time. Let $c_\varepsilon^k(t) \equiv \frac{\partial c_t^k}{\partial \varepsilon}$, evaluated at $\varepsilon = 0$ (and similarly for other variables). Judd differentiates Equations (4.10′) and (4.11′) with respect to ε, evaluating the derivatives at $\varepsilon = 0$ and at the initial steady-state level of capital. Defining $\mu > 0$ as

$$\mu = \frac{\rho}{2}\left(1 - \frac{\theta_L}{\sigma} + \sqrt{\left(1 - \frac{\theta_L}{\sigma}\right)^2 + \frac{4\theta_L}{\beta\sigma}}\right), \tag{4.12}$$

where θ_L is labor's share of output and σ is the elasticity of substitution between labor and capital in production, Judd shows that the initial shock to consumption of the capitalist equals[60]

$$c_\varepsilon^k(0) = H(\mu)\frac{\rho}{1-\tau}\left(\frac{\rho - \mu\beta}{\rho}\right)\frac{c^k}{\beta}, \tag{4.13}$$

where $H(\mu)$ is the Laplace transform of $h(t)$ evaluated at μ. For any discount rate s, $H(s) = \int_0^\infty e^{-st}h(t)\,dt$ is the present value of the policy function $h(t)$. It is easy to show that $\mu > \rho$ iff $\beta < 1$. Also, Judd shows that $\mu\beta < \rho$ iff $\beta < 1$. Thus, capitalists may immediately increase or decrease their consumption in response to an announced increase in capital income taxation. Increased future capital income taxation has an income effect that works to reduce present consumption. On the other hand, the substitution effect works to shift consumption from the future to the present. If $\beta < 1$, the substitution effect dominates and consumption increases. For $\beta > 1$, the preference for smooth consumption makes the income effect dominant. The role of the policy duration appears in $H(\mu)$, where $H(\mu)$ increases with the duration of the tax increase. Thus, consumption at time zero falls more for a longer duration tax increase (in present value terms) when the income effect dominates ($\beta > 1$). Note that consumption falls now, even if the start of the tax increase is delayed. But the drop in consumption is attenuated as a tax hike of fixed duration is put further off into the future.

To determine the degree to which the tax and transfer scheme benefits workers, we need to know how the consumption path for workers changes in response to an increase in capital income taxation. Consumption for the worker is given by

$$c_t^w = f(k) - kf'(k) + (\tau + \varepsilon h(t))\,kf'(k), \tag{4.14}$$

where the first two terms are wage income and the last term is the transfer financed by capital income taxation. Define $c_\varepsilon^w(t) \equiv \frac{\partial c_t^w}{\partial \varepsilon}$ evaluated at $\varepsilon = 0$, and B_ε^w as the

[60] Judd solves the linear differential equation system by first taking Laplace transforms. See Judd (1985b) for details on this derivation.

discounted increase in lifetime utility measured in time zero consumption arising from the tax increase,

$$B_\varepsilon^w = \frac{\int_0^\infty e^{-\rho t} u'(c_t^w)\, c_\varepsilon^w(t)\, dt}{u'(c_0^w)}. \tag{4.15}$$

Judd shows that

$$B_\varepsilon^w = kf'H(\rho)\left(1 - \frac{\left[1-\beta+\frac{H(\mu)}{H(\rho)}\left(\frac{\mu\beta}{\rho}-1\right)\right]\left[\left(\frac{\tau}{1-\tau}\right)\frac{\sigma}{\theta_L}+1\right]}{1-\beta}\right). \tag{4.16}$$

We can now specify the policy experiments and evaluate the impact on consumers.

4.2.1. Immediate temporary tax increase

A short-lived tax increase put into place at time zero can be modeled as $h(t)$ equaling 1 for small t and 0 otherwise. If dt is the length of the time the temporary tax increase is in place, then $H(\mu) = H(\rho) = dt$ and $H(\mu)/H(\rho)$ is one (approximately). Thus,

$$B_\varepsilon^w = kf'H(\rho)\left(1 + \underbrace{\frac{\beta}{1-\beta}\left(1-\frac{\mu}{\rho}\right)\left(\frac{\tau}{1-\tau}\frac{\sigma}{\theta_L}+1\right)}_{(A)}\right). \tag{4.17}$$

Recall that $\beta < 1$ iff $\mu > \rho$. Thus, the term in parentheses in Equation (4.17) labeled A is negative and workers are better off from this temporary incremental tax hike if this term is less than 1 in absolute value. If the initial capital income tax is sufficiently low, then workers are better off. This follows from the continuity of B_ε^w in τ and the fact that this expression evaluated at $\tau = 0$ is

$$B_\varepsilon^w = kf'H(\rho)\left(1 + \frac{\rho-\mu\beta}{\rho(1-\beta)}\right), \tag{4.17$'$}$$

as well as the fact that $\beta < 1$ iff $\rho > \mu\beta$. For pre-existing τ sufficiently large, B_ε^w will be negative, and so workers do not always benefit from an increase in the capital income tax. Essentially, the worker would like to save some of the large transfer but is precluded from doing so by high transactions costs or other institutional barriers; in that case, the worker would prefer capital income to be left with the capitalist who will invest it (and so make a portion of it available to the worker in the future through future taxes and transfers).

4.2.2. Immediate permanent tax increase

Now consider a permanent tax increase implemented at time zero. Thus, $h(t)$ equals 1 for all t. The function $H(s) = s^{-1}$ and B_ε^w now equals

$$B_\varepsilon^w = \frac{kf'}{\rho}\frac{\beta}{1-\beta}\left(1-\frac{\rho}{\mu\beta}\right)\left(\frac{\tau}{1-\tau}\frac{\sigma}{\theta_L}+1\right). \tag{4.18}$$

Again, B_ε^w is positive for small τ but becomes negative for τ sufficiently large. Equation (4.18) can be contrasted to the measures of burden in Feldstein (1974) and Boadway (1979). While Boadway makes the point that wages may initially increase as a result of redistributive taxation, he does not provide a utility-based measure of the gains from the tax shift. Equation (4.18) is just such a measure.

4.2.3. Announced permanent tax increase

Finally, consider the announcement today of a permanent tax increase to be put into effect at some later time. Thus, $h(t)$ equals 0 for $t < T$ and equals 1 for $t > T$. The ratio $H(\mu)/H(\rho)$ now equals $\frac{\rho}{\mu}e^{-(\mu-\rho)T}$ and goes to zero as T gets large if $\mu > \rho$ (and explodes if $\rho > \mu$). Now the benefit of redistributions to the worker depends critically on the value of β. If $\beta < 1$, then $\mu > \rho$, and B_ε^w is zero if τ equals zero (and negative if $\tau > 0$). Thus, the worker is made worse off from an announced future increase in capital income taxation, starting at a positive level of taxation, even with the proceeds transferred to the worker. The decrease in the capital stock along the path prior to the enactment of the tax increase will reduce wages, which in present value terms are more valuable than any future increase in transfers.

In the case that $\beta > 1$, then $\rho > \mu$, and $H(\mu)/H(\rho)$ dominates in Equation (4.16). The terms including $H(\mu)/H(\rho)$ will be positive (since $\mu\beta > \rho$), and workers benefit from a tax increase. Highly concave utility (high β) implies strong intertemporal smoothing of consumption and slow capital stock adjustment to new tax rates. Thus, future tax increases will not lead to immediate and rapid reductions in the capital stock (which would hurt the worker). While 100% shifting of the tax eventually occurs, the burden shift can occur quite slowly, allowing a period during which labor benefits from the higher tax [61].

4.2.4. The role of anticipations

The last result indicates the importance of anticipation in perfect foresight models. We can make this point more emphatically by considering policy changes designed in such a way that they lead to no change in the consumption of the capitalist at time

[61] This focus on anticipations distinguishes this analysis from that of other neoclassical growth models with workers and capitalists.

zero. Given the desire to smooth intertemporal consumption in the additively-separable utility function, any deviation from a steady-state consumption path at time zero must arise from a surprise in tax policy. Thus, a policy that leads to $c^k_\varepsilon(0) = 0$ is a policy that is perfectly anticipated by capitalists. From Equation (4.13), $c^k_\varepsilon(0) = 0$ equal to zero implies that $H(\mu)(1 - \frac{\mu\beta}{\rho})$ equals zero, and so

$$B^w_\varepsilon = -H(\rho)\,kf'\frac{\tau}{1-\tau}\frac{\sigma}{\theta_L}, \tag{4.19}$$

which is zero if τ is zero and negative otherwise. In other words, workers cannot benefit from a tax policy that is perfectly anticipated by capital owners. It is the surprise at time zero along with an inelastic short-run supply of capital that generates a benefit for workers from a tax and transfer scheme.

The Judd model illustrates a number of key points. First the incidence of a tax in a dynamic model can have strong effects through changes in saving and investment and consequently the capital–labor ratio. Both the perfect-foresight model and the neoclassical growth model make this point clearly. The perfect-foresight model, however, illustrates the importance of anticipations and surprises and suggests the possibility of lump-sum taxes on existing capital at the time of the announcement of a new tax regime ("old" capital) [62]. Because of the importance of anticipations and lump-sum characteristics of some tax policies, we pursue this further by developing a model in which taxes affect welfare through changes in asset prices. This model will make clear the distinction between "old" and "new" capital and the role of anticipations.

4.3. *Incidence and the market value of capital*

We present a simple partial equilibrium model of capital investment that emphasizes the importance of costs of adjustment in changing the capital stock. In the Judd (1985a) model previously described, capital accumulation depended on preferences and, in particular, the concavity of the utility function. Costs of adjusting the capital stock played no role. However, firms can incur significant costs during the process of major investment projects [63]. Summers (1985) presents a simple model to illustrate how corporate tax policy can affect investment as well as the market value of capital in place.

Costs of adjustment are captured in a simple capital–supply relationship. Consider a good that is produced with capital, K, according to the concave production function $F(K)$. Let the price of this good as well as the market price of capital

[62] Auerbach and Kotlikoff (1987) also emphasize the normative possibilities associated with taxing old capital in a lump-sum fashion.

[63] Large-scale urban transportation projects are a good example of investment projects that generate large-scale costs to businesses and residents in the urban area (for example, the Big Dig in Boston).

equal 1[64]. Firms wish to invest when the market value of the firm's capital exceeds its replacement cost at the margin. Investment is costly, however, and so firms adjust their capital stock towards some desired level slowly according to the function

$$\dot{K} = \left(g\left(\frac{V}{K}\right) - \delta\right) K, \tag{4.20}$$

where V is the value of the firm, δ is the rate of depreciation, and a dot indicates a time derivative. The function g has the property that $g(1) = 0$ and $g' > 0$. Defining $q = V/K$, Equation (4.20) is a standard Tobin investment function [Tobin (1969)].

Firms finance investment out of retained earnings, and the opportunity cost of funds for equity-holders equals ρ. Thus, if equity-holders are to receive a return equal to ρ, the value of the firm must evolve over time according to the relation

$$\rho = \frac{D}{V} + \frac{\dot{V}}{V}, \tag{4.21}$$

where D is the dividend paid to equity-holders. Dividends are equal to

$$D = F'(K)\,K - \tau(F'(K) - \delta)\,K - g(q)\,K, \tag{4.22}$$

where τ is the tax rate on income net of economic depreciation. Combining Equation (4.21) and (4.22), the value of the firm evolves as

$$\dot{V} = \rho V - F'(K)\,K + \tau(F'(K) - \delta)\,K + g(q)\,K. \tag{4.23}$$

We can re-express the change in value of the firm in terms of the change in value per dollar of existing capital ($\dot{q}$):[65]

$$\dot{q} = (\rho + \delta - g(q))\; q + g(q) - (1 - \tau)\,F'(K) - \tau\delta. \tag{4.24}$$

Equations (4.20) and (4.24) form the equations of motion for our system in terms of K and q. In the steady state (with $\dot{K} = 0$ and $\dot{q} = 0$), q takes the value q^* such that $g(q^*) = \delta$, and the steady-state capital stock (K^*) is defined by

$$(1 - \tau)(F'(K^*) - \delta) = \rho q^*. \tag{4.25}$$

Net of depreciation and tax, the return on capital must equal ρ, the return available on other investments. We illustrate the movements of K and q through the use of a phase diagram (Figure 4.1). The diagram breaks the $q - K$ space into four regions bounded

[64] We abstract from inflation.

[65] Differentiate $q = V/K$ to get $\dot{q} = \dot{V}/K - q(\dot{K}/K)$. Then substitute Equation (4.20) for $\dot{K}$ and Equation (4.23) for $\dot{V}$.

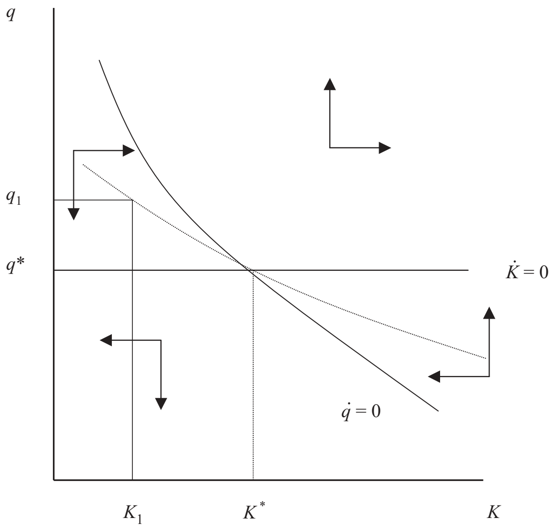

Fig. 4.1. Asset price model phase diagram.

by the $\dot{q} = 0$ and $\dot{K} = 0$ loci. Above the $\dot{K} = 0$ locus, the capital stock grows (depicted in the NE and NW quadrants with a horizontal arrow pointing to the right). Below this locus, the capital stock declines. To the right of the $\dot{q} = 0$ locus, q grows (depicted by the arrows pointing upward) while to the left, q falls. The intersection of these two lines is the steady-state.

The capital stock can only adjust slowly in response to shocks, but q can adjust instantaneously to any level. The dotted line is the saddle-point path moving to the steady-state from either the NW or SE. Consider some catastrophe that reduces the capital stock from K^* to K_1 (an earthquake, say). With perfect foresight, the value of the remaining capital (per unit of K) would immediately jump from q^* to q_1. With q now greater than q^*, investment would exceed depreciation, and the capital stock would slowly return to K^*. With myopic expectations, by contrast, q would jump immediately up further to the $\dot{q} = 0$ locus, as investors do not anticipate the capital loss that follows when new capital comes on line. Such a movement would not be sustainable (in the sense of q moving continuously back to q^*), as movement from the $\dot{q} = 0$ locus would be horizontally to the right, into a region where q and K both increase. This is a region of speculative bubbles, which must collapse at some point (with the price dropping back to the saddle-path).

Along the saddle-path, owners of capital would receive the normal rate of return. While the dividend yield exceeds the required rate of return, the investor incurs a capital loss as new net investment drives down the market price of capital. The only beneficiaries of the destruction of part of the capital stock are the owners of the undestroyed capital who earn a windfall capital gain at time zero.

We first use the model to illustrate a basic point about tax capitalization. Consider an increase in the corporate tax rate (τ). This shifts the $\dot{q} = 0$ locus to the left but

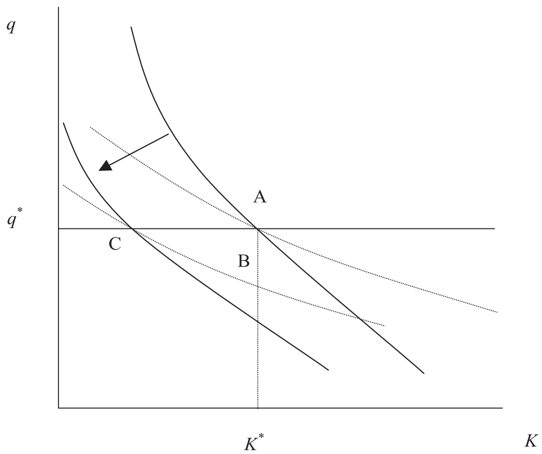

Fig. 4.2. An increase in the corporate tax rate.

leaves the $\dot{K} = 0$ locus unchanged. See Figure 4.2. The result is an immediate drop in the value of capital (a movement from A to B in Figure 4.2). All of the burden of the tax has been capitalized into a price drop at the time of enactment. No future capital owners will be affected, as the return on capital equals ρ along the saddle-path from B to C. Capitalization of taxes into asset prices complicates incidence analysis considerably[66].

The model can also be used to make an important point about the distinction between old and new capital. Old capital is capital in place at the time of a change in tax policy. Consider the enactment of a tax credit for the purchase of new capital. Because of the reduction in taxes, this might ordinarily be viewed as advantageous to all capital owners. To use this model to analyze this policy change, Equation (4.20) must be modified to account for the fact that the price of capital has been reduced from 1 to $1 - s$, where $s < 1$ is the level of the investment tax credit.

$$\dot{K} = \left(g\left(\frac{q}{(1-s)} \right) - \delta \right) K. \tag{4.20$'$}$$

The reduction in taxes increases the funds available to pay out as dividends. Equation (4.24) is accordingly modified:

$$\dot{q} = \left(\rho + \delta - g\left(\frac{q}{1-s} \right) \right) q + (1-s)\, g\left(\frac{q}{1-s} \right) - (1-\tau)\, F'(K) - \tau\delta. \tag{4.21$'$}$$

As s is increased from zero, both the $\dot{q} = 0$ and the $\dot{K} = 0$ loci move. See Figure 4.3. The $\dot{K} = 0$ locus shifts down from q^* to $(1-s)\, q^*$. Simultaneously, the $\dot{q} = 0$ locus

[66] See Aaron (1989) for a discussion of this point along with other issues that complicate the analysis of tax policy.

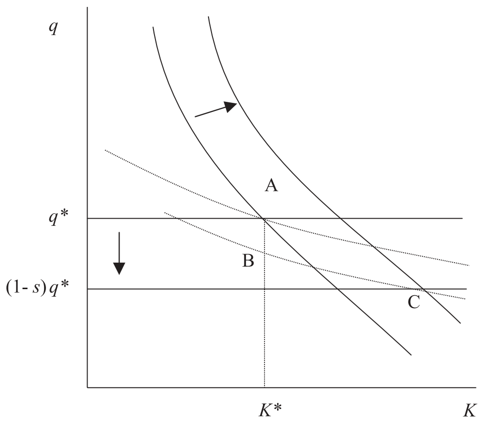

Fig. 4.3. An investment tax credit.

shifts to the right. The immediate impact on q is indeterminate. On the one hand, the rightward shift in the $\dot{q} = 0$ locus operates to create a windfall gain to owners of old capital: any future capital they purchase will be less expensive, and so dividends can be increased. On the other hand, the downward shift in the $\dot{K} = 0$ locus operates to generate a windfall loss: old capital must now compete with new capital that is less expensive. As drawn, the second effect dominates. Prior to the increase in the investment tax credit, the economy is at point A with $q = q^*$. The investment tax credit leads to an immediate drop in q from A to B. Over time, q drops further as the economy moves from B to C. This move does not imply a further loss in value, because the capital loss is exactly offset by an above-normal dividend yield so that investors along the path from B to C receive the normal rate of return (ρ). The tax credit has the desired effect of increasing the capital stock but the unexpected effect of burdening the owners of old capital with a windfall loss at the time of enactment. We leave it as an exercise for the reader to work out the price path for an announcement at time zero of an investment tax credit to be implemented at a given future date.

Dynamic incidence modeling has evolved considerably in the past twenty-five years. With increased computer power, it has become possible to create large-scale computable general equilibrium (CGE) models to evaluate tax policy over the lifetime, as well as to consider questions of capital accumulation and intergenerational redistributions. We turn now to models of lifetime tax incidence analysis, and we consider how these models provide new light on old issues.

5. Lifetime tax incidence

Up to this point, we have focused only indirectly on the relevant time frame for our incidence analysis. To classify households from rich to poor, most of the applied studies reviewed in Section 2.4 use income from one year, but others use income from an

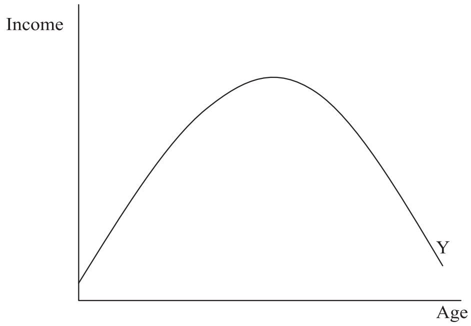

Fig. 5.1. Income over the lifetime.

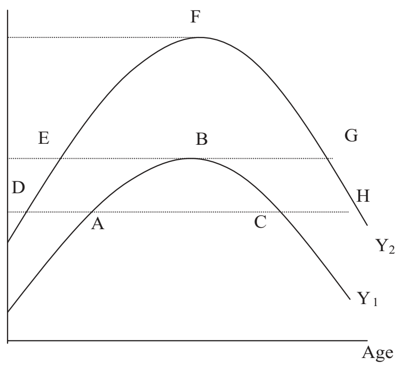

Fig. 5.2. Lifetime income heterogeneity.

entire lifetime. Intermediate choices also are possible, as Slemrod (1992) uses "time exposure" income from a period of seven years.

We now turn to models of lifetime tax incidence and begin with a very simple example to illustrate the importance of the time horizon. Consider a world with *identical* individuals such that one person of each age is alive at any given time. Figure 5.1 illustrates the income profile of each individual throughout life. Income is low at the beginning of life and increases to a peak before decreasing as the individual approaches retirement. Annual income at any given age is measured by the height of the curve, and lifetime income is the area under the curve.

Given our assumptions about identical individuals and the pattern of births and deaths, Figure 5.1 can also be interpreted as the distribution of income in the economy at any given point of time. Young and old have low annual income, while the middle-aged have high annual income. An annual tax incidence analysis using this snapshot of income would give the erroneous impression of considerable income inequality in this economy, despite the fact that everyone is identical. On the basis of the lifetime, the economy has no income inequality at all.

Now let us complicate the economy slightly and allow for two types of people with different lifetime income profiles (see Figure 5.2). Individuals with profile Y_1 earn

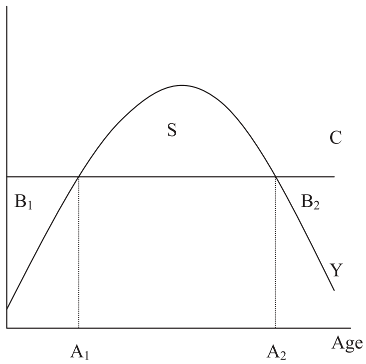

Fig. 5.3. Lifetime income and consumption.

less income at any age than do those with profile Y_2. Now the comparison gets more complicated. An annual income analysis will rank the person at F as the richest person in the economy, followed by the three individuals B, E and G. It then ranks individuals at A, C, D, and H as the poorest. This classification inappropriately groups a lifetime-poor person at the peak of earnings (point B) with lifetime-rich individuals at either the beginning or end of their earnings profiles (E and G).

A lifetime incidence analysis can yield a sharply different conclusion about the progressivity of any given tax as compared to an annual income analysis. Returning to our simple model of identical individuals, one of whom is alive at any given time, consider a consumption-smoothing model as posited by Modigliani and Brumberg (1954). In Figure 5.3, income is hump-shaped as above, and consumption is constant throughout life. At ages below A_1, individuals borrow to finance consumption. Between A_1 and A_2, they repay debt and start to save. In retirement (after A_2), individuals draw down savings to finance consumption. In the absence of bequests, the areas B_1 and B_2 are equal to S (in present value).

An annual incidence analysis of a tax on consumption would compare the average effective tax rate (tax as a percentage of income) across different annual income groups. Consider a flat consumption tax with no exemptions. For the young and the elderly, this tax as a fraction of annual income could be quite high (and possibly exceed 100%). The average tax rate would be lowest for those individuals at the peak of their profile, those whose earnings exceed consumption. Thus, a consumption tax would look highly regressive. On a lifetime income basis, however, the average tax rate (lifetime consumption taxes divided by lifetime income) would simply equal the tax rate on consumption. Then the tax is strictly proportional.

A bit of thought leads to the conclusion that differences in the degree of progressivity between lifetime and annual income analyses will vary depending on the tax under investigation. Continuing with our simple economy, consider a tax equal to a fixed percentage of wage income. On a lifetime basis this tax is proportional, but on an annual basis it will look somewhat regressive since capital income is left out of the tax base. However, the degree of regressivity implied in an annual income analysis will be sharply lower than in the case of our simple consumption tax, because the average tax rate can never exceed the statutory tax rate on wages.

Analyses of lifetime tax incidence have been carried out in a number of fashions. One approach is to build an overlapping generations (OLG) computable general equilibrium (CGE) model of an economy with a representative agent in each cohort [see, for example, Auerbach, Kotlikoff and Skinner (1983) or Auerbach and Kotlikoff (1987)]. Such models are very useful for understanding the intergenerational incidence of government policies [Kotlikoff (2002)]. They are not well suited, however, to studying the intragenerational redistribution brought about by government policies. A second approach is to jettison the CGE analysis of age cohorts, but instead focus on lifetime heterogeneity using incidence assumptions in the style of Pechman and Okner (1974). An example is the Davies, St. Hilaire and Whalley (1984) lifetime model based on Canadian data. A third approach is to combine both intertemporal and intratemporal heterogeneity. Fullerton and Rogers (1993) were one of the first to build a complete CGE model of this type.

Empirical incidence analyses from a lifetime perspective suffer from the lack of data on the entire lifetime income and consumption patterns of households. Thus, any attempt to apply the lifetime approach requires heroic assumptions. In the Davies et al. model, for example, all income streams are exogenous and the consumption path is based on an additive isoelastic utility function. Interest and growth rates are predetermined based on Canadian data, and the model calculates life-cycle consumption, income, tax payments, and government transfers [67].

Other empirical studies use annual data to construct a proxy for lifetime income. Poterba (1989) invokes the Modigliani and Brumberg (1954) consumption-smoothing story to study U.S. federal excise taxes. With perfect life-cycle consumption smoothing, and individuals identical except for lifetime income levels, current consumption is proportional to lifetime income. Thus, Poterba uses current consumption to categorize individuals by lifetime income. For alcohol, fuel, and tobacco taxes, he finds striking differences between annual and lifetime incidence. Metcalf (1994) applied a similar idea to the system of state and local sales taxes in the United States and finds that a case can be made for viewing this system of taxes as progressive, contrary to accepted wisdom. The shift to a lifetime perspective is one important factor blunting the regressivity of state and local sales taxes. In addition, most states exempt a variety of goods with low income elasticities, thereby adding to the progressivity of the system.

Other efforts to carry out lifetime incidence analysis using (primarily) annual data include Lyon and Schwab (1995), Caspersen and Metcalf (1994), Gale, Houser and Scholz (1996) and Feenberg, Mitrusi and Poterba (1997), among others. Caspersen

[67] They find that the incidence of the overall Canadian tax system is mildly progressive under either a lifetime or an annual incidence framework. Personal income taxes look less progressive, while consumption taxes look less regressive under the lifetime incidence framework.

and Metcalf use data from the Panel Study on Income Dynamics (PSID) to estimate age–earnings profiles for individuals based on variables that exist in both the PSID and the Consumer Expenditure Survey (CEX). The PSID has excellent data on income across households and years, so it is a good source for estimating age–earnings profiles that can be used to construct measures of lifetime income. Unfortunately, the PSID has minimal consumption data, which precludes distributional analysis of consumption taxes. The CEX, on the other hand, has excellent consumption data but poor income data. Hence, Caspersen and Metcalf use the PSID to predict age–earnings profiles for households in the CEX. For the introduction of a value added tax (VAT) in the United States, they find that a lifetime incidence analysis sharply reduces regressivity.

In another effort to capture life-cycle considerations, Gale, Houser and Scholz (1996) carry out an analysis in which they restrict their sample to married families with the head between the ages of 40 and 50, arguing that this approach reduces the inappropriate comparisons between people either at the beginning or end of their earnings career with people at the peak of their earnings. They find that this approach does not alter their conclusions about the distributional implications of a shift from income to consumption taxation [68].

5.1. A lifetime utility model

These studies all measure changes in tax liabilities rather than changes in welfare. As we discussed in the introduction, changes in tax liabilities misrepresent the change in welfare for various reasons. An advantage of a general equilibrium model (whether analytical or numerical) is that the researcher can make assumptions about the form of utility and explicitly measure changes in welfare in dollar terms (typically using the equivalent variation). Fullerton and Rogers (1993) construct a lifetime computable general equilibrium model to study the U.S. tax system [69]. We sketch out this model and compare its lifetime results to the classic annual results of Pechman and Okner (1974).

Fullerton and Rogers build a model with consumers of different ages and different lifetime incomes. All have the same lifetime utility function, but differ in labor

[68] Metcalf (1999), however, carries out an incidence analysis of an environmental tax reform using the lifetime methodology of Caspersen and Metcalf (1994) and also using a cohort analysis similar to Gale et al. He finds that the two approaches give very different answers, suggesting that the cohort approach is not a good proxy for a more complete lifetime analysis. One possible reason follows from the permanent income hypothesis [Friedman (1957)]. If people make decisions on the basis of permanent rather than annual income, then any deviations between the two will magnify the perceived regressivity of a consumption tax. Lifetime income approaches are less likely to suffer from this measurement problem.

[69] Other results from this model are presented in Fullerton and Rogers (1991, 1995, 1996, 1997).

productivity (and hence wage rate). Lifetime utility is a nested-CES function with the top-level allocating consumption and labor across time:

$$U = \left[\sum_{t=1}^{T} a_t^{1/\varepsilon_1} x_t^{(\varepsilon_1 - 1)/\varepsilon_1} \right]^{\varepsilon_1/(\varepsilon_1 - 1)}, \tag{5.1}$$

where T is length of life (known with certainty), x_t is the amount of the composite commodity consumed at time t, ε_1 is the intertemporal elasticity of substitution, and a_t is a weighting parameter that reflects the consumer's underlying rate of time preference. Economic life is 60 years, from ages 20 to 79. Lifetime utility is maximized subject to the lifetime budget constraint

$$\sum_{t=1}^{T} x_t \left(\frac{q_t}{(1+r)^{t-1}} \right) = I_{\mathrm{d}}, \tag{5.2}$$

where q_t is the composite price of x_t, r is the net-of-tax rate of return, and I_{d} is the present value of lifetime discretionary income [70]. The composite price, q_t, is implicitly defined by Equation (5.2) and will turn out to be a weighted average of the prices of the components of x_t. A benefit of the nested-CES utility structure is that the demand functions can be solved sequentially beginning at the top nest of the utility function. Defining $\tilde{q}_t = q_t/(1+r)^{t-1}$, then the maximization of Equation (5.1) subject to Equation (5.2) yields standard CES demands in terms of prices $\tilde{q}_t$. In an important simplification, Fullerton and Rogers assume that these prices can be calculated from the current interest rate. These "myopic expectations" mean that each equilibrium period can be calculated before proceeding to the next period, sequentially, whereas perfect foresight would require endogenous calculation of all periods' prices and interest rates simultaneously.

Lifetime income includes bequests received. Rather than model endogenous bequest behavior, Fullerton and Rogers assume that each individual must bequeath the same level bequest at death as received at birth, after adjusting for economic and population growth. Bequests received (and left) as a fraction of income are calibrated to data from Menchik and David (1982).

At the next level of the nest, consumers choose between purchased consumption goods and leisure according to the sub-utility function:

$$x_t = \left[\alpha_t^{1/\varepsilon_2} \bar{c}_t^{(\varepsilon_2 - 1)/\varepsilon_2} + (1-\alpha_t)^{1/\varepsilon_2} \ell_t^{(\varepsilon_2 - 1)/\varepsilon_2} \right]^{\varepsilon_2/(\varepsilon_2 - 1)}, \tag{5.3}$$

where $\bar{c}_t$is a composite commodity consumed at time t, ℓ_t is leisure at time t, ε_2 is the elasticity of substitution between consumption and leisure, and α_t is a weighting

[70] They use a Stone–Geary sub-utility function with minimum required expenditures, so I_d is net of the cost of required expenditures. Only discretionary consumption (in excess of required consumption) is available for lifetime smoothing, so x is defined as discretionary consumption.

parameter. The time endowment is fixed at 4000 hours per year, and the wage rate per effective labor unit is constant, but wage rates can vary across individuals based on individual labor productivity. The individual chooses leisure and labor (L_t) based on maximization of the sub-utility function in Equation (5.3) subject to the budget constraint

$$\bar{p}_t \bar{c}_t + w_t \ell_t = q_t x_t, \tag{5.4}$$

This maximization yields demands for ℓ_t and $\bar{c}_t$. Then composite consumption is modeled as a Stone–Geary function of individual consumption goods (c_{it}):

$$\bar{c}_t = \prod_{i=1}^{N} (c_{it} - b_{it})^{\beta_{it}}. \tag{5.5}$$

The model includes 17 consumer goods ($N = 17$), minimum required consumption (b_{it}), and marginal share parameters (β_{it}). The Stone–Geary function is a parsimonious specification that allows consumption shares to vary across income, and across age groups, as is observed in the data. It also dampens consumption fluctuations, thereby making savings less sensitive to changes in the interest rate[71].

Using the Consumer Expenditure Survey, Fullerton and Rogers estimate 408 parameters: b_{it} and β_{it} for 17 goods for each of 12 different 5-year age brackets. Thus, taxes will affect income groups differentially on the sources side because of different relative factor incomes and on the uses side because of different observed spending shares. And yet the modelers need not assume that the rich are fundamentally different from the poor, in terms of preferences. Here, the fundamental difference between rich and poor is simply their income levels. All 12 groups in the model have the same utility function, with the same 408 parameters, but low-income groups spend much of their money on the minimum required purchases while other groups spend more in proportions given by the marginal expenditure shares.

Next, Fullerton and Rogers convert the vector of 17 consumer goods (C) to a vector of 19 producer goods (Q) using the Leontief transformation $C = ZQ$, where Z is a 17 by 19 transformation matrix. Finally, they distinguish corporate (Q^{c}) and non-corporate (Q^{nc}) output using another sub-utility function

$$Q_j = \left[\gamma_j^{1/\varepsilon_3} \left(Q_j^{\mathrm{c}}\right)^{(\varepsilon_3-1)/\varepsilon_3} + \left(1-\gamma_j\right)^{1/\varepsilon_j} \left(Q_j^{\mathrm{nc}}\right)^{(\varepsilon_3-1)/\varepsilon_3}\right]^{\varepsilon_3/(\varepsilon_3-1)}, \tag{5.6}$$

where ε_3 is the elasticity of substitution, and γ_j is a weighting parameter for industry j. This function explains the co-existence of corporate and non-corporate production within a single industry, and it explains differences in production patterns across

[71] See Starrett (1988) for a discussion of the sensitivity of savings to changes in the interest rate in Stone–Geary and isoelastic utility functions.

industries. Maximization subject to the budget constraint ($p_j^{\mathrm{c}} Q_j^{\mathrm{c}} + p_j^{\mathrm{nc}} Q_j^{\mathrm{nc}} = p_j^Q Q_j$) yields demands for Q_j^{c} and Q_j^{nc} that depend on relative prices – which, in turn, depend on differential taxation of the corporate sector.

Whereas the corporate and non-corporate prices are observable, the various price indices are not. Fullerton and Rogers take the Lagrangian multiplier from this last maximization and invert it, to obtain

$$p_j^Q = \left(\gamma_j (p_j^{\mathrm{c}}) + \left(1 - \gamma_j\right) \left(p_j^{\mathrm{nc}}\right)^{1-\varepsilon_3}\right)^{1/(1-\varepsilon_3)}. \tag{5.7}$$

Knowing these prices, they use the transition matrix Z to recover consumer prices ($p_i = \sum p_j^Q Z_{ji}$). Then, the reciprocal of the Lagrangian multiplier from the maximization of the Stone–Geary utility function is the price of the composite commodity:

$$\bar{p}_t = \prod_{i=1}^{N} \left(\frac{p_i}{\beta_{it}}\right)^{\beta_{it}}, \tag{5.8}$$

and finally,

$$q_t = \left(\alpha_t \bar{p}_t^{1-\varepsilon_2} + (1 - \alpha_t)\, w^{1-\varepsilon_2}\right)^{1/(1-\varepsilon_2)}. \tag{5.9}$$

With an explicit utility function, Fullerton and Rogers can measure the equivalent variation (EV) associated with any change in the tax system. They carry out differential tax incidence experiments where they replace a particular tax with a proportional tax on lifetime labor endowments. If U^0 is lifetime utility under the old tax regime, and U^1 is lifetime utility under the new tax regime, then

$$\mathrm{EV} = \left(U^1 - U^0\right) P^0, \tag{5.10}$$

where P^0 is a price index on the lifetime bundle $\{x_t\}$ calculated at old prices.

Production in each of the 19 industries is based on a similar nested structure. At the top level, value added is combined with intermediate goods from other industries in a Leontief production function. Value added is a CES function of labor (L) and a capital aggregate ($\bar{K}$), where σ_1 is the elasticity of substitution. Aggregate capital is then a CES combination of five capital types, where σ_2 is the elasticity of substitution, to capture differential tax treatment of equipment, structures, land, inventories and intangibles.

Note that production is constant returns to scale, so firms earn zero profits in a competitive environment. This is a common assumption in many CGE models used to measure tax incidence. Firms solve a simple one-period optimization problem, in contrast to consumers who solve an intertemporal maximization problem. Dynamics are not ignored, however, in that interest rates affect capital accumulation.

The government engages in three activities in this model. First, it makes transfer payments that vary according to age and income. Second, it produces goods and services sold in the market place. In this regard, the government is simply one more producer using capital, labor and intermediate goods for production. Third, government buys goods and services for a public good that enters utility in a separable fashion.

The treatment of taxes in the Fullerton and Rogers model is similar to that of Ballard, Fullerton, Shoven and Whalley (1985). Personal income taxes are specified as a linear function of consumer income, with a constant slope and an intercept that varies across lifetime income categories and age. The slope measures the marginal tax rate, while the intercept captures various deductions and exemptions that vary across consumers. Payroll taxes are treated as *ad valorem* taxes on the use of labor services by industry [72]. Retail sales taxes are treated as *ad valorem* taxes on consumer goods, while excise taxes are *ad valorem* taxes on producer goods. Business tax provisions are incorporated using the cost-of-capital approach of Hall and Jorgenson (1967). This includes corporate taxes at both the federal and state level, property taxes, investment incentives, and depreciation deductions. These tax provisions affect the demand for capital by firms, which affects the interest rate used both in the consumer's problem and in the firm's cost of capital.

Finally, Fullerton and Rogers group households into lifetime income categories through a two-step procedure. Using data from the PSID, they estimate lifetime profiles for wages, taxes and transfers. They estimate wage rate rather than wage income regressions, since labor supply is endogenous in their model. These wage rates vary on the basis of age, education, race and sex. Using the estimated coefficients, they forecast and backcast wages of each individual to create a lifetime wage profile. An initial measure of lifetime income (LI) is then given by the equation

$$\mathrm{LI} = \sum_{t=1}^{60} \frac{4000 \cdot w_t}{(1+r)^{t-1}}, \tag{5.11}$$

where r is a discount rate, and w_t is the actual wage for any year in the sample, or an estimated wage for any other year [73]. In the second step of the procedure, individuals are sorted into 12 groups on the basis of this initial measure of lifetime income [74]. For each group, the log of the wage rate is again regressed on age, age squared, and age cubed. This 2-step procedure allows wage profiles to differ across income groups. Differences in the wage profiles will create differences in savings patterns

[72] No distinction is drawn between the employer and employee share of the payroll tax, under the assumption that statutory incidence does not affect the economic incidence.

[73] The lifetime income measure is adjusted for taxes and transfers. For couples, each individual is given the average income for the two spouses.

[74] They first divide the sample into ten deciles. They then subdivide the top decile into the top 2% and next 8%, and the bottom decile into the bottom 2% and next 8%.

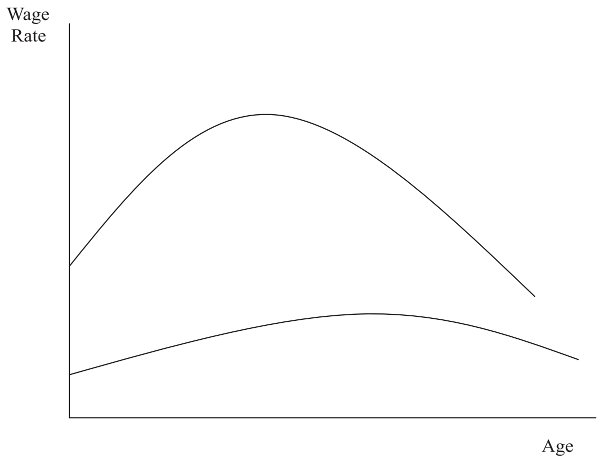

Fig. 5.4. Wage profiles.

across groups, which will play an important role in determining the incidence of capital income taxation.

Figure 5.4 shows estimated wage profiles for different lifetime income groups. Both the curvature of the wage profiles and the location of the peak varies across groups. More sharply curved wage profiles mean individuals must engage in more saving to smooth consumption. An earlier peak also means more savings – for consumption in later years.

Table 5.1 shows the burden of all U.S. taxes in 1984, as measured by the lifetime EV benefit as a percentage of lifetime income of a switch from the existing tax system to a proportional lump-sum labor-endowment tax. Except for the first group (the bottom 2% of the distribution), every income group gains. These benefits are roughly flat from the second through tenth income groups and then rise sharply in the highest two income groups (top ten percent of the population). This pattern of proportionality across the middle of the income distribution with progressivity at the top end matches the findings of Pechman and Okner (1974) and Pechman (1985) in their annual income incidence analyses. Fullerton and Rogers's results differ from Pechman's at the bottom of the income distribution. The former find progressivity at the lowest end, while the latter finds regressivity.

The table shows distributional results in the new steady state. The sum of the 12 groups' gains from shifting to the lump-sum tax is large, measuring 3.5% of their aggregate lifetime income. This large gain comes about, in part, through a substantial tax on endowments of older generations during the transition. In present value terms, the gains are less than half, reflecting the fact that losses accrue to living generations while gains primarily accrue to future generations.

While the degree of progressivity in the U.S. tax system appears similar in either annual or a lifetime incidence analyses, important differences remain for particular

Table 5.1
Lifetime incidence of US tax system in 1984[a]

Lifetime income category	EV as a percentage of lifetime income
1	–0.06
2	3.13
3	1.41
4	2.37
5	3.58
6	1.39
7	3.46
8	2.51
9	2.95
10	3.01
11	5.55
12	11.10
All, in steady state	3.52
PV(EV)/LI	1.29

[a] Source: Fullerton and Rogers (1993, Table 7-15).

taxes. Perhaps the most important difference is that Pechman finds that corporate taxes are progressive because of the sources side of income. Since high-income people disproportionately earn capital income, they are most impacted by a capital income tax. In contrast, Fullerton and Rogers find that the corporate tax does not appreciably affect factor prices (because the statutory corporate rate is largely offset in 1984 by the investment tax credit and accelerated depreciation allowances). Instead, the corporate tax affects relative output prices (because some industries have larger corporate sectors and get more credits and allowances). Thus, it primarily affects individuals on the uses side of income. For the lower part of the distribution, the tax is regressive because the poor tend to spend greater fractions of incomes on goods produced in the corporate sector. At the top end of the distribution, the tax is progressive because of the nature of the replacement tax. The proportional tax on labor endowments does not tax inheritances, and the rich receive larger inheritances, so they benefit from the tax on labor endowment.

Another important finding of the model is that sales and excise taxes continue to be regressive when measured on a lifetime basis – whereas previous work by Poterba (1989) and others hypothesized that consumption taxes would look roughly proportional on a lifetime basis. Fullerton and Rogers note two reasons. First, the utility structure that they employ does not specify a minimum required leisure expenditure. The lifetime poor must spend a greater share of their income on required goods, so

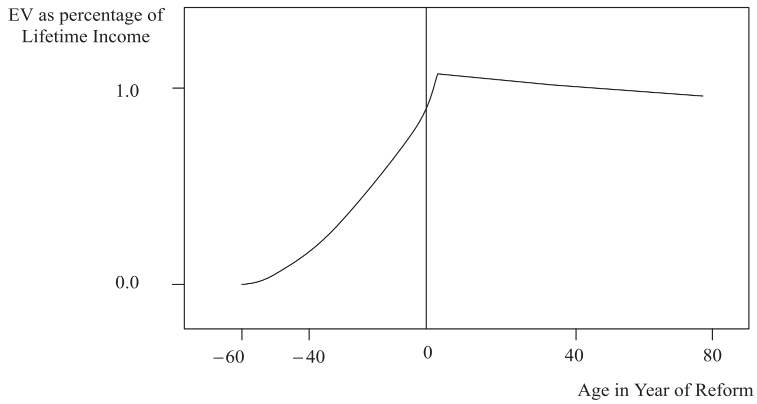

Fig. 5.5. Intergenerational transfers.

they pay more sales tax as a fraction of lifetime income. Thus, some regressivity is built into the model structure. Second, goods with high tax rates tend to be goods with high estimated minimum required purchases (alcohol, tobacco, and gasoline).

Another advantage of lifetime models is the ability to capture intergenerational transfers. Consider sales and excise taxes, for example. Figure 5.5 groups individuals by age of birth, rather than by lifetime income, and it shows the equivalent variation as a percent of lifetime income (for the replacement of sales and excise taxes by a proportional tax on labor endowment). The EV as a fraction of income for the entire population is 0.44%. The figure shows how the EV varies across cohorts. For those born after the tax reform goes into effect (individuals to the right of the vertical line in the middle of the graph), EV is roughly 1% of lifetime income. For those born prior to the reform, EV is substantially lower and approaches zero for the oldest groups. This picture tells a complicated incidence story. Older generations get less of a benefit from the tax shift because the replacement tax is a tax on their time endowment – which translates, for the elderly, into a tax on their leisure time.

Lifetime incidence models can be constructed to focus on both intergenerational and intragenerational redistribution. The Fullerton and Rogers model focuses on both types of redistributions, but assumes myopic expectations about future prices as well as ad hoc bequest behavior. Auerbach and Kotlikoff (1987) construct a dynamic model with a perfect foresight equilibrium[75], but they have a representative agent in each cohort and thus focus only on intergenerational redistribution arising from fiscal policy[76]. Altig, Auerbach, Kotlikoff, Smetters and Walliser (2001) build

[75] An early published version of the model was in Auerbach, Kotlikoff and Skinner (1983). For a brief history of the model's development, see Kotlikoff (2000).

[76] The Auerbach–Kotlikoff model also has only one type of good and makes no distinction between corporate and non-corporate production, thus limiting its ability to provide meaningful incidence results for the existing tax system.

on the Auerbach–Kotlikoff model, but follow Fullerton and Rogers in adding intragenerational heterogeneity. They use the new model to measure the utility gains and losses from different types of fundamental tax reforms. But because their replacement tax is different from the one in Fullerton and Rogers, results from the two models cannot easily be compared.

5.2. Generational accounts

As noted earlier, a complete picture of the incidence of government fiscal policy would take into account transfers as well as taxes [Browning (1985, 1993)]. Auerbach, Gokhale and Kotlikoff (1991) develop "generational accounts" to measure the fiscal impact of government taxes and transfers over each cohort's lifetime. A generational account is simply a measure of a cohort's net tax payments (taxes less transfers) from today until all members of the cohort die. For a cohort born in year k, its account in year t is defined as

$$N_{t,k} = \sum_{s=\upsilon}^{k+D} \frac{T_{s,k} P_{s,k}}{(1+r)^{(s-\upsilon)}}, \tag{5.12}$$

where $T_{s,k}$ is the net tax for cohort k in year s, $P_{s,k}$ is the population weight for cohort k in year s (accounting for mortality and immigration), r is the discount rate, $\upsilon = \max(t,k)$, and D is maximum length of life. For generations already born ($k < t$), the account $N_{t,k}$ is the present value of all future net tax payments discounted back to year t. For future generations ($k > t$), $N_{t,k}$ discounts net tax payments back to year k. For generations alive at time t, net tax payments into the future are based on current law and government projections of changes in tax and transfer programs. For years beyond government projections, taxes and transfers are assumed to grow at the growth rate assumed for the whole economy, thereby keeping net tax payments fixed relative to income. To assess net tax payments for future cohorts, we begin with the government intergenerational budget constraint:

$$\sum_{k=t-D}^{t} N_{t,k} + \frac{\sum_{k=t+1}^{\infty} N_{t,k}}{(1+r)^{k-t}} = \sum_{s=t}^{\infty} \frac{G_s}{(1+r)^{s-t}} - W_t^g. \tag{5.13}$$

Equation (5.13) states that the government budget constraint must be balanced over time. Future net tax payments (left-hand side of Equation 5.13) must equal the present value of future government consumption (G_s) less net government wealth in year t (W_t^g). The first term on the left-hand side is the stream of remaining net tax to be paid by cohorts alive at year t. The second term is the net tax paid by future cohorts. Assuming some path for future government purchases, as well as knowledge of the current net wealth stock, the right-hand side of Equation (5.13) is fixed. The first term on the left-hand side is also known, leaving the second term as a residual. Finally,

Table 5.2
Net tax payments (present value in thousands of $ 1995)[a]

Generation's age in 1995	Male	Female
0	77.4	51.9
20	182.2	115.0
40	171.2	99.0
60	–25.5	–52.0
80	–77.2	–90.2
Future generations	134.6	90.2

[a] Source: Gokhale, Page and Sturrock (1999, Tables 21.1, 21.2).

for these residual net tax payments to be divided across different future cohorts, it is assumed that average per capita tax payments grow at the same rate as productivity growth. Thus, for future generations, net tax liability relative to lifetime income is constant[77]. Table 5.2 gives an example of the calculation of net tax payments, from Gokhale, Page and Sturrock (1999).

Ignoring the newborn for the moment, net tax payments are highest for the young and decline with age. This reflects the fact that the current elderly will pay little in taxes relative to the benefits they receive in future years. Of course, the elderly in 1995 had paid taxes prior to 1995, but the table does not take account of those past taxes. Following Equation (5.12), it focuses only on future net tax liabilities. Women have lower net tax liabilities, reflecting both their smaller tax payments and higher benefit receipts (largely due to social security and mortality differences between men and women). The newborn have a lower net tax liability since their taxes and transfers, for the most part, will not begin for some time into the future and so in present value terms are smaller[78]. For future generations, we see the current fiscal imbalance: taxes will have to be raised on future generations in order to bring the government's budget into balance.

Net tax payments in the tables above cannot be compared for any cohorts other than newborns and future generations, since net tax payments are only computed over a portion of the lives of generations currently alive in 1995. To compare all cohorts both living and not yet born, net tax liabilities can be computed for each cohort over their entire lifetime and discounted back to time zero for each cohort. Similarly, lifetime income can be calculated and discounted back to time zero. Then an average tax liability can be calculated as the ratio of lifetime taxes to lifetime income[79]. Table 5.3

[77] Other assumptions can be made, depending on the experiment under consideration.

[78] Gokhale, Page and Sturrock (1999) use a discount rate of 6%. Adjusting for the fact that newborns enter the work force roughly 20 years in the future, the corresponding net tax payment would be 248.2, which is 36% higher than that of people born in 1975.

[79] This calculation is similar to the methodology of Fullerton and Rogers.

Table 5.3
Lifetime net tax rates [a]

Year of birth	Net tax rate	Gross tax rate	Gross transfer rate
1900	23.9	28.0	4.0
1920	29.6	36.4	6.7
1940	32.5	40.3	7.8
1960	33.3	44.1	10.8
1980	30.8	43.0	12.2
1995	28.6	41.7	13.1
Future generations	49.2	–	–

[a] Source: Gokhale, Page and Sturrock (1999, Table 21.3).

shows lifetime net tax rates for living and future generations, from Gokhale, Page and Sturrock (1999).

For generations born from 1900 to 1960, the increase in net tax rates reflects the growth of government over the first half of the century (see gross tax rates in the middle column). The decline in net tax rates since 1960 reflects longer life expectancies and the rapid increase in medical transfers (see transfers in the last column). The bottom row indicates that the current policy cannot persist. Net tax rates will have to increase from 28.6% (for people born in 1995) to 49.2%, an increase of 72%.

The calculation of these generational accounts is in the spirit of the Pechman and Okner analysis rather than the CGE models of Fullerton and Rogers or Auerbach and Kotlikoff. It takes fiscal policy as given, and it allows neither for behavioral responses nor for changes in factor prices in response to government policies. Fehr and Kotlikoff (1999) compare net tax burdens using both generational accounting and the Auerbach–Kotlikoff CGE model described above. They find that the generational accounts methodology works well for closed economies and for economies with minimal capital adjustment costs.

Generational accounting has been used to look at Social Security and Medicare policy [Auerbach, Gokhale and Kotlikoff (1992)] as well as to compare tax and transfer systems in various countries around the world [Auerbach, Kotlikoff and Leibfritz (1999)].

6. Policy analysis

Applied incidence analysis plays an important role in tax policy making, as the results of government studies help determine the course of actual reform. Most such studies use recent incidence theory, as described above, to allocate the burden of each tax among income groups using much data about the sources and uses of income in each group [as in Pechman and Okner (1974) or Gale, Houser and Scholz (1996)]. This

Table 6.1
Distributional effects of repeal of federal communications excise tax: calendar year 2003[a]

Income category	Change in federal taxes		Effective tax rate	
	Millions ($)	Percent	Present law	Proposal
Less than $10 000	–324	–4.3	9.3%	8.9%
10 000 to 20 000	–621	–2.3	7.4%	7.2%
20 000 to 30 000	–608	–0.9	12.4%	12.3%
30 000 to 40 000	–572	–0.6	16.0%	16.0%
40 000 to 50 000	–490	0.4	17.4%	17.3%
50 000 to 75 000	–920	–0.3	19.9%	19.9%
75 000 to 100 000	–531	–0.2	22.4%	22.3%
100 000 to 200 000	–421	–0.1	25.1%	25.1%
200 000 and over	–371	–0.1	28.6%	28.6%
Total: all taxpayers	–4858	–0.3	21.5%	21.5%

[a] Source: Joint Committee on Taxation (2000).

approach forms the foundation for analyses undertaken by the U.S. Congressional Budget Office (CBO), the Office of Tax Analysis (OTA) of the U.S. Department of the Treasury, and the U.K. Office for National Statistics[80]. We focus here primarily on the incidence analysis by the staff of the Joint Committee on Taxation (JCT) of the U.S. Congress[81].

6.1. *The distributional table*

A key tool used by policy makers in their consideration of changes to the tax system is the distributional table. Table 6.1 presents a distributional table for the repeal of the federal communications excise tax for the calendar year 2003. The first column indicates the income categories over which the tax is distributed. This column has a number of features. First, the unit of observation is the tax-filing unit, so a data point in any of the income categories may be a single taxpayer or a couple filing jointly. Thus, if a married couple each earn $17 000 and file separately, they show up

[80] See Bradford (1995) for a discussion and critique of this type of analysis in the United States. For the United Kingdom, Lakin (2001, p. 35) reports figures that are very similar in nature to those for the USA: “The proportion of gross income paid in direct tax by the top fifth of households is almost double that paid by those in the bottom fifth: 24% compared with 13%. Indirect taxes have the opposite effect to direct taxes taking a higher proportion of income from those with lower incomes”. We cannot know whether the similarity of results is because of similar methodology or because of similar policies.

[81] See Joint Committee on Taxation (1993). Cronin (1999) describes the OTA methodology, while Kasten, Sammartino and Toder (1994) describe work at CBO.

in this table as two data points in the second row of the table. If they file jointly, however, they appear in the fourth row[82]. Second, the annual time frame is used for measuring income. Third, the JCT uses a measure of income called "expanded income". This measure is defined as adjusted gross income (AGI) plus tax-exempt interest, employer contributions for health plans and life insurance, the employer share of payroll taxes, worker's compensation, nontaxable Social Security benefits, the insurance value of Medicare benefits, alternative minimum tax preference items, and excluded income of U.S. citizens living abroad. This measure is an effort to conform more closely to a Haig–Simons definition of income[83]. It is by no means a close proxy for economic income, however, nor is it a close proxy for lifetime income. One advantage of expanded income is its explicit recognition that factor income by itself is inadequate for measuring income, and another advantage is its easy calculation from readily-available data, primarily tax returns. These features help make the measure more readily understandable to policy makers, many of whom have limited economics education[84]. Fourth, the number of tax filing units differs across the income categories. In 1995, for example, the number of tax returns filed in the \$10 000 to \$20 000 AGI category was roughly 20 times the number in the over-\$200 000 AGI category[85]. Fifth, taxpayers are grouped into income categories on the basis of year 2000 income, the first year of analysis in this report. Any changes in income due to either transitory fluctuations or trends do not shift taxpayers across brackets.

The second column of Table 6.1 shows the aggregate change in federal taxes for each income category, while the third column shows the change as a percentage of expanded income. The essential point to understand about this measure is that it is an estimate of the change in tax payments, not the change in tax burden. Figure 6.1 illustrates the distinction for a simple case where supply is perfectly elastic. Consider an existing tax that shifts the supply curve from S_0 to S_1, and an increase that shifts the supply curve from S_1 to S_2. The tax increase will raise revenue by an amount equal to A–F, but the increased tax burden is area A + B. These are quite different sizes, and they may even differ in sign. Depending on the price elasticity of demand, the higher tax rate may increase or decrease tax revenue (area A may be less than area F). However, the increased tax burden given by the area A + B is unambiguously positive[86]. Thus, the use of tax revenue as a proxy for burden can lead to the incorrect

[82] OTA uses the family as the unit of observation, combining tax returns of all members of the family.

[83] OTA uses a measure called Family Economic Income (FEI) that is more comprehensive and therefore closer in spirit to Haig–Simons income. In addition to data from tax returns, FEI requires imputations of certain income sources. See Cronin (1999) for details.

[84] The need for a simple income measure may help explain why imputed rental income for owner-occupied housing is excluded.

[85] U.S. Bureau of the Census (1999, Table 559). Note that these AGI categories do not correspond exactly to the expanded income categories in Table 6.1.

[86] Here we ignore distinctions between the change in consumer surplus and equivalent or compensating variation.

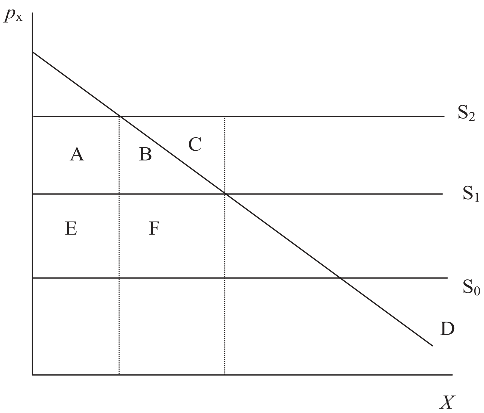

Fig. 6.1. Measuring tax burden for policy analysis.

conclusion that a higher tax rate could reduce tax burden. As discussed below, the Joint Committee on Taxation reported distribution tables based on tax burdens rather than tax revenues for a brief while. OTA reports burden estimates, but only reports area A as the increased burden, ignoring the deadweight loss (area B)[87].

Finally, for each income category, the table reports effective tax rates (the ratio of tax payments to expanded income) under current law and under the proposed policy change. The proposal portrayed in Table 6.1 would be characterized as progressive, since average tax rates fall the most for lower income groups.

This approach is subject to a number of criticisms[88]. In addition to the issues highlighted above, another problem is the failure to take account of asset price changes and implicit taxation. In Section 4.3 above, we made the point that tax capitalization complicates the task of identifying who bears the burden of a tax. Subsequent owners are observed to pay a tax, in distributional tables, but they may not bear any burden if they bought the asset for a reduced amount. Distributional analyses also ignore implicit taxation, which occurs when a tax-favored asset pays a lower rate of return than a comparable non-favored asset. Consider, for example, state and local municipal debt that is exempt from federal tax[89]. If the taxable rate is 8% and the tax-exempt rate is 6%, then the implicit tax on municipal debt is 25%. Distributional tables ignore this implicit tax, despite its equivalence to an explicit 25% tax that is used to pay those who now benefit from the reduced rate on municipal debt[90].

[87] See Cronin (1999) for a discussion of other issues associated with measuring burden.
[88] See, for example, Graetz (1995) and Browning (1995).
[89] State and local debt is often exempt from state taxation also.
[90] Gordon and Slemrod (1983) find that the rich benefit from tax-exempt municipal debt through lowered taxes payments, while the poor benefit from increased expenditures made possible by the lower borrowing rate paid by communities.

6.2. Suggested changes

In 1993, the Joint Committee on Taxation made significant changes in their methodology for distributing the burden of taxes, as described in Joint Committee on Taxation (1993) as well as Barthold and Jack (1995)[91]. Despite the fact that many of the changes were short-lived, they are worth discussing because they illustrate a creative effort to apply economic theory to the policy process. In making the changes, the JCT attempted to adhere to three broad principles: 1) to make calculations on the basis of the economic incidence rather than the statutory burden of a tax, 2) to be consistent in the treatment of taxes expected to have the same economic incidence (regardless of the statutory incidence), and 3) to use a methodology that allows comparisons of unrelated tax proposals.

In addition to the choice of the “expanded income” measure described above, the JCT made two other significant conceptual changes. First, they measured burden from tax changes rather than just distributing tax payments across groups. Above, we noted that using changes in tax revenue as a proxy for changes in burden can lead to the anomalous result that a tax increase is beneficial to the taxpayer (ignoring the use of proceeds from the tax). Like OTA, the JCT did not propose to measure the change in consumer surplus, but rather to use a proxy that could easily be estimated from existing data. Unlike OTA, however, the JCT measured burden by the change in tax revenue that would occur if behavior were fixed. Thus, in Figure 6.1, the JCT’s measure of the burden from a tax increase would be the area $A+B+C$.

Second, the JCT chose to measure the burden of a tax proposal over a five-year window[92]. Prior to that time, the JCT measured burdens within a single year. The second principle noted above was violated in cases where some or all of the burden of a tax fell outside of the one-year window. Shifting to a five-year window does not solve this problem but reduces its impact since less of a tax is likely to fall outside a five-year window (and because the present value of tax changes five years out is lower than the present value one year out). The JCT chose not to go to an infinite window for a number of reasons. Results are sensitive to the choice of discount rate in an infinite-horizon model, and economic forecasting of key variables required for revenue estimation become increasingly unreliable for years further into the future. Furthermore, it is simply not credible to assume that tax policy will remain unchanged into the distant future. Thus, a shorter time horizon was chosen.

The JCT then reports an annuitized measure that accounts for economic growth. To illustrate the idea, we take an example from Joint Committee on Taxation (1993). Assume a discount rate of 10%, and economic growth of 5%, and consider three proposals. First, consider a permanent tax reduction of $100 per year beginning immediately. The JCT assumes that the value of the tax reduction will grow at the

[91] Also, see Barthold, Nunns and Toder (1995) for a comparison of the new JCT methodology and the OTA and CBO methodologies.

[92] The five-year window is similar to the “time-exposure” measure of Slemrod (1992).

Table 6.2
Annuitization of taxes in Joint Committee on Taxation (JCT) methodology

Proposal	Year					Total
	1	2	3	4	5	
Immediate permanent tax reduction of $100/year	100	105	110	116	122	553
Immediate temporary tax reduction of $100	22	23	24	25	26	120
Postponed permanent tax reduction of $100/year	18	19	20	21	22	100

overall rate of economic growth and so will be worth $105 next year and $110, $116, and $122 in subsequent years. The JCT calculates an annuity equivalent for year one that is also assumed to grow at the overall rate of economic growth. In this case, the annuity equivalent is $100 for year one (followed by 105, 110, 116, and 122). Second, consider an immediate tax cut of $100 that lasts only one year, with a present value of simply $100. The five-year annuity equivalent would be $22 in year one (an amount that could grow at five percent per year over the five-year window and be discounted at 10% to yield a present value of $100). For a final example, take a permanent $100 per year tax cut that is postponed for four years, so that the first year of benefits occurs in the last year of the five-year window. The value in the last year is $122, which in present value terms equals $83. The annuity equivalent would be $18 in the first year. Table 6.2 shows the tax reductions that the JCT would report in a five-year window.

The third proposal (with a permanent $100/year tax cut) looks very much like the second proposal (with a $100 tax cut in only one year), because only the first year of the delayed permanent tax cut is counted. A one-time tax reduction in year five would give the same annuity equivalent as is recorded in this third row of Table 6.2. Comparing rows 2 and 3, it is clear that an immediate tax reduction of $100 is worth more than a postponed reduction of $122, a result that follows because the 10% discount rate exceeds the 5% growth rate.

Two other issues described in the 1993 JCT publication relate to the treatment of a broad-based consumption tax such as a national retail sales tax. The first issue is whether the general price level rises (to accommodate forward shifting of the tax) or remains unchanged (in which case taxes are shifted backward in the form of lower factor incomes). Real factor prices are the same in either case, and the status of the general price level would appear to have no impact on the measured distribution of the tax burden, but government transfer programs complicate the analysis [Browning and Johnson (1979)]. Some transfers to the poor are stated in nominal dollars, so a consumption tax shifted forward into higher prices will reduce the real purchasing power of these transfers. If the consumption tax is shifted backwards into lower factor prices, however, recipients of these government transfers are not affected[93]. Whether the general price level rises or not depends importantly on monetary policy and cannot be predicted beforehand. But the price level response may have an important impact on the outcome of the analysis, especially as it relates to households with the lowest incomes[94].

The second issue about the consumption tax is when to allocate the tax. We can allocate a consumption tax when consumption occurs, or when the income that finances that consumption is earned. The advantage of the latter approach is that the analysis then conforms to the third principle above, namely, to use techniques that allow analysts to combine proposals. In particular, the JCT says that it facilitates the comparison of consumption taxes to income taxes (the predominant type of tax analyzed by the JCT).

The distinction between allocating consumption taxes when consumption occurs or when the income is earned is only relevant with any saving or dissaving. This, in fact, is the main reason for using lifetime measures of income for consumption tax analysis, as discussed above. Since life-cycle changes in net wealth can be quite large, over periods of more than five years, the JCT measure of the burden of a consumption tax can still be quite different from the burden measured in a lifetime analysis.

Rather than allocating the consumption tax, the JCT converts a broad-based consumption tax into a combined tax on wage income and old capital. To see the equivalence, consider the budget constraint of an individual with k years remaining in life at the time a consumption tax is imposed:

$$W_0 + \sum_{t=0}^{k} \frac{w_t L_t}{(1+r)^t} = \sum_{t=0}^{k} \frac{(1+\tau)\, C_t}{(1+r)^t}, \tag{6.1}$$

where W_0 is the person's net wealth at time 0, $w_t L_t$ is wage income in year t, C_t is consumption, τ is the consumption tax rate, and r is the rate of return available to the individual. The JCT approach works by defining a tax at rate $\tilde{\tau}$ on old capital (W_0) and wage income such that $1 - \tilde{\tau} = 1/(1+\tau)$. Then Equation (6.1) becomes

$$(1-\tilde{\tau})\, W_0 + \sum_{t=0}^{k} \frac{(1-\tilde{\tau})\, w_t L_t}{(1+r)^t} = \sum_{t=0}^{k} \frac{C_t}{(1+r)^t}. \tag{6.2}$$

From the individual's point of view, the consumption tax is equivalent to a tax on wage income plus a capital levy[95].

[93] They would be affected if policy makers reduced transfers in nominal terms, which seems unlikely.

[94] Many transfers in the USA are indexed, including social security, food stamps, and in-kind health care, but other non-indexed transfers are received by the lowest income bracket, as discussed below. Also, the price-level problem and the response of the Federal Reserve to the imposition of a tax is not, in principle, limited to general consumption taxes. Consider an income tax that is assumed to be shifted backwards to labor and capital. The Federal Reserve could increase the monetary supply and allow nominal prices to rise, to keep nominal factor prices from changing (even though real factor prices still fall).

[95] The lump-sum component of a consumption tax with no transition rules is a major source of efficiency gain from a consumption tax relative to a wage tax. See Auerbach and Kotlikoff (1987) for more on

Table 6.3
Distributional impact of a 5% comprehensive consumption tax (as a percentage of pre-tax income)

Income class	(p, C)	(p, Y)	(w, Y)	(w, C)
$0–$10 000	3.70	3.69	2.84	2.85
10 000–20 000	2.66	2.68	2.86	2.83
20 000–30 000	2.90	3.00	3.10	2.99
30 000–40 000	2.92	3.04	3.20	3.07
40 000–50 000	2.94	3.10	3.26	3.10
50 000–75 000	2.77	2.97	3.21	2.99
75 000–100 000	2.63	2.88	3.01	2.74
100 000–200 000	2.50	2.84	2.92	2.57
200 000 and over	1.76	2.78	2.86	1.76

[a] Source: Joint Committee on Taxation (1993, Table 3, p. 55).

These two issues give rise to four possible ways of distributing a consumption tax. Following the JCT's notation, we can distinguish:
(p, C) prices allowed to rise and burden assigned as consumption occurs;
(p, Y) prices allowed to rise and burden assigned as income occurs;
(w, C) factor prices fall and burden assigned as consumption occurs;
(w, Y) factor prices fall and burden assigned as income occurs.
Our Table 6.3, taken from JCT (1993), shows the impact of the four different approaches on the distribution of a comprehensive 5% tax on consumption.

As noted above, whether prices are allowed to rise primarily affects the burden of the tax at the very low end of the income distribution (because some transfers are not indexed). On the other hand, the timing of the tax burden affects the very top of the income distribution (because they undertake most savings). The measured burden of a consumption tax in the highest-income group is roughly one percent of pre-tax income higher when allocated on the basis of income rather than consumption[96].

The first column of Table 6.3 (labeled (p, C)) is the traditional method for distributing consumption taxes, and it makes consumption taxes look sharply regressive. If the (w, Y) method were used to distribute consumption taxes, they would look nearly proportional. Instead, the JCT favors the (p, Y) approach, on the basis of some empirical evidence that the introduction of value added taxes in Europe led to at-least-

this point. As an aside, individuals who have negative net wealth at the time of the imposition of a consumption tax receive a lump-sum subsidy equal to $\tilde{\tau} W_0$. Thus, the consumption tax redistributes from lenders to those in debt (relative to a tax just on wages).

[96] A dollar of saving receives relief from a full dollar of a consumption tax when the tax is allocated as consumption occurs, but it only receives relief equal in value to the annuity that a dollar buys when the consumption tax is allocated as income is earned.

partial forward shifting into higher consumer prices, combined with the JCT's wish to adhere to their third principle of tax comparability.

The JCT used the approach outlined in Joint Committee on Taxation (1993) for a brief while, but it then reverted to an approach that distributes tax payments rather than burdens, on a year-by-year basis instead of using five-year windows. In particular, the analysis in Table 6.1 accords with current JCT policy.

Both Joint Committee on Taxation (1993) and Cronin (1999) illustrate creative efforts to bridge the gap between economic theory and real-world policy analysis. In addition to theoretical rigor, policy analysts need measures that are easily constructed from readily-available data and easily understood both by the public and by policy makers. The retreat at the JCT from the innovations described in Joint Committee on Taxation (1993) is perhaps discouraging, particularly in light of the tentative nature of the steps towards a more-comprehensive lifetime measure of economic burden arising from changes in tax policy. But it should be recognized that much of the policy process occurs in an informal give-and-take between policy makers and staff economists; it may be in this latter environment that incidence theory can be most effective [97].

7. Conclusion

The field of incidence analysis has progressed dramatically in the past twenty years, as new research has yielded fresh insights into the burden of taxes in imperfectly competitive models and in intertemporal models. The increase in computing power and the availability of large-scale data sets have also enriched our understanding of tax incidence. Moreover, the power of recent analytical models and of new data sets is evident in recent attempts by government economists to bring state-of-the-art incidence analysis to policymakers.

Yet, the basic tools of log-linearization in simple two-sector models are just as useful today as they were in Harberger's classic 1962 paper. These techniques are still frequently used in studies of new taxes, externalities, imperfect competition, and other non-tax distortions. Such analytical models can yield important insights that do not follow directly from complicated computable general equilibrium models. In fact, many researchers now combine both approaches within a single paper, as they find it useful to push the analytical results as far as is possible, for intuition, before turning to numerical methods to determine likely magnitudes. Using all of these techniques, the topic of tax incidence will continue to be an area of productive research yielding further insights in the years to come.

References

Aaron, H.J. (1974), "A new view of property tax incidence", American Economic Review 64:212–221.

[97] But see Graetz (1995) for a more pessimistic viewpoint.

Aaron, H.J. (1989), "Politics and the professors revisited", American Economic Review 79:1–15.
Altig, D., A.J. Auerbach, L.J. Kotlikoff, K.A. Smetters and J. Walliser (2001), "Simulating fundamental tax reform in the United States", American Economic Review 91:574–595.
Anderson, K.B. (1990), A Review of Structure-Performance Studies in Grocery Retailing (Bureau of Economics, Federal Trade Commission, Washington, DC).
Anderson, S.P., A. de Palma and B. Kreider (2001), "Tax incidence in differentiated product oligopoly", Journal of Public Economics 81:173–192.
Atkinson, A.B., and J.E. Stiglitz (1980), Lectures on Public Economics (McGraw-Hill, New York).
Auerbach, A.J., and J.R. Hines (2002), "Taxation and economic efficiency", in: A.J. Auerbach and M. Feldstein, eds., Handbook of Public Economics, Vol. 3 (Elsevier, Amsterdam) pp. 1347–1421.
Auerbach, A.J., and L.J. Kotlikoff (1987), Dynamic Fiscal Policy (Cambridge University Press, Cambridge).
Auerbach, A.J., L.J. Kotlikoff and J. Skinner (1983), "The efficiency gains from dynamic tax reform", International Economic Review 24:81–100.
Auerbach, A.J., J. Gokhale and L.J. Kotlikoff (1991), "Generational accounts: a meaningful alternative to deficit accounting", Tax Policy and the Economy 5:55–110.
Auerbach, A.J., J. Gokhale and L.J. Kotlikoff (1992), "Social security and Medicare policy from the perspective of generational accounting", Tax Policy and the Economy 6:129–145.
Auerbach, A.J., L.J. Kotlikoff and W. Leibfritz, eds (1999), Generational Accounting Around the World (University of Chicago Press, Chicago, London).
Ballard, C.L., D. Fullerton, J.B. Shoven and J. Whalley (1985), A General Equilibrium Model for Tax Policy Analysis (University of Chicago Press, Chicago).
Barro, R.J. (1974), "Are government bonds net wealth?", Journal of Political Economy 82:1095–1117.
Barthold, T.A., and W. Jack (1995), "Innovations in the JCT distribution methodology", in: D.F. Bradford, ed., Distributional Analysis of Tax Policy (AEI Press, Washington, DC) pp. 148–163.
Barthold, T.A., J.R. Nunns and E. Toder (1995), "A comparison of distribution methodologies", in: D.F. Bradford, ed., Distributional Analysis of Tax Policy (AEI Press, Washington, DC) pp. 96–110.
Bernheim, B.D. (1981), "A note on dynamic tax incidence", Quarterly Journal of Economics 96: 705–723.
Besley, T. (1989), "Commodity taxation and imperfect competition: a note on the effects of entry", Journal of Public Economics 40:359–367.
Besley, T.J., and H.S. Rosen (1999), "Sales taxes and prices: an empirical analysis", National Tax Journal 52:157–178.
Bishop, R.L. (1968), "The effects of specific and ad valorem taxes", Quarterly Journal of Economics 82:198–218.
Boadway, R. (1979), "Long-run tax incidence: a comparative dynamic approach", Review of Economic Studies 46:505–511.
Bradford, D.F. (1978), "Factor prices may be constant but factor returns are not", Economics Letters 1:199–203.
Bradford, D.F., ed. (1995), Distributional Analysis of Tax Policy (American Enterprise Institute Press, Washington, DC).
Break, G.F. (1974), "The incidence and economic effects of taxation", in: A.S. Blinder, R.M. Solow, G.F. Break, P.O. Steiner and D. Netzer, eds., The Economics of Public Finance (Brookings Institution, Washington, DC).
Brittain, J.A. (1971), "The incidence of social security payroll taxes", American Economic Review 61:110–125.
Brown, H.G. (1924), The Economics of Taxation (Holt, New York).
Browning, E.K. (1985), "Tax incidence, indirect taxes, and transfers", National Tax Journal 38:525–533.
Browning, E.K. (1986), "Pechman's tax incidence study: a note on the data", American Economic Review 76:1214–1218.
Browning, E.K. (1993), "Transfers and tax incidence theory", Public Finance 48:138–141.

Browning, E.K. (1995), "Tax incidence analysis for policy makers", in: D.F. Bradford, ed., Distributional Analysis of Tax Policy (AEI Press, Washington, DC) pp. 164–180.
Browning, E.K., and W.R. Johnson (1979), The Distribution of the Tax Burden (American Enterprise Institute, Washington, DC).
Bull, N., K.A. Hassett and G.E. Metcalf (1994), "Who pays broad-based energy taxes? Computing lifetime and regional incidence", Energy Journal 15:145–164.
Bureau of the Census, U.S. (1999), Statistical Abstract of the United States: 1999 (U.S. Government Printing Office, Washington, DC).
Carroll, R., and J. Yinger (1994), "Is the property tax a benefit tax? The case of rental housing", National Tax Journal 47:295–316.
Caspersen, E., and G.E. Metcalf (1994), "Is a value added tax regressive? Annual versus lifetime incidence measures", National Tax Journal 47:731–746.
Cremer, H., and J.-F. Thisse (1994), "Commodity taxation in a differentiated oligopoly", International Economic Review 35:613–633.
Cronin, J.A. (1999), U.S. Treasury Distributional Analysis Methodology (Office of Tax Analysis, Washington, DC).
Davies, J.B., F. St. Hilaire and J. Whalley (1984), "Some calculations of lifetime tax incidence", American Economic Review 74:633–649.
Delipalla, S., and M. Keen (1992), "The comparison between ad valorem and specific taxation under imperfect competition", Journal of Public Economics 49:351–367.
Dixit, A.K., and J.E. Stiglitz (1977), "Monopolistic competition and optimum product diversity", American Economic Review 67:297–308.
Doyle, M.P. (1997), The Effects of Interest Rates and Taxes on New Car Prices (Board of Governors of the Federal Reserve System, Washington, DC).
Edgeworth, F.Y. (1897), "The pure theory of taxation", The Economic Journal 7:46–70.
Feenberg, D.R., A.W. Mitrusi and J.M. Poterba (1997), "Distributional effects of adopting a national retail sales tax", Tax Policy and the Economy 11:49–90.
Fehr, H., and L.J. Kotlikoff (1999), "Generational accounting in general equilibrium", in: A.J. Auerbach and L.J. Kotlikoff, eds., Generational Accounting around the World (University of Chicago Press, Chicago).
Feldstein, M. (1974), "Incidence of a capital income tax in a growing economy with variable savings rates", Review of Economic Studies 41:505–513.
Feldstein, M. (1977), "The surprising incidence of a tax on pure rent: a new answer to an old question", Journal of Political Economy 85:349–360.
Friedman, M. (1957), A Theory of the Consumption Function (Princeton University Press, Princeton, NJ).
Fuchs, V.R., A.B. Krueger and J.M. Poterba (1998), "Economists' views about parameters, values, and policies: survey results in labor and public economics", Journal of Economic Literature 36: 1387–1425.
Fullerton, D. (1996), "Why have separate environmental taxes?", Tax Policy and the Economy 10:33–70.
Fullerton, D., and D.L. Rogers (1991), "Lifetime versus annual perspectives on tax incidence", National Tax Journal 44:277–287.
Fullerton, D., and D.L. Rogers (1993), Who Bears the Lifetime Tax Burden? (Brookings Institution, Washington, DC).
Fullerton, D., and D.L. Rogers (1995), "Distributional effects on a lifetime basis", in: D.F. Bradford, ed., Distributional Analysis of Tax Policy (AEI Press, Washington, DC) pp. 262–294.
Fullerton, D., and D.L. Rogers (1996), "Lifetime effects of fundamental tax reform", in: H.J. Aaron and W.G. Gale, eds., Economic Effects of Fundamental Tax Reform (Brookings Institution Press, Washington, DC) pp. 321–347.
Fullerton, D., and D.L. Rogers (1997), "Neglected effects on the uses side: even a uniform tax would change relative goods prices", American Economic Review 87:120–125.

Gabszewicz, J.J., and J.-F. Thisse (1979), "Price competition, quality and income disparities", Journal of Economic Theory 20:340–359.

Gale, W., S. Houser and J.K. Scholz (1996), "Distributional effects of fundamental tax reform", in: H. Aaron and W. Gale, eds., Economic Effects of Fundamental Tax Reform (Brookings Institution, Washington, DC) pp. 281–315.

Gentry, W., and R.G. Hubbard (1997), "Distributional implications of introducing a broad-based consumption tax", Tax Policy and the Economy 11:1–47.

Gokhale, J., B.R. Page and J.R. Sturrock (1999), "Generational accounts for the United States: an update", in: A.J. Auerbach, L.J. Kotlikoff and W. Leibfritz, eds., Generational Accounting around the World (University of Chicago Press, Chicago) pp. 489–517.

Gordon, R.H., and J. Slemrod (1983), "A general equilibrium simulation study of subsidies to municipal expenditures", Journal of Finance 38:585–594.

Graetz, M.J. (1995), "Distributional tables, tax legislation, and the illusion of precision", in: D.F. Bradford, ed., Distributional Analysis of Tax Policy (AEI Press, Washington, DC) pp. 15–78.

Gravelle, J.G., and K. Smetters (2001), "Who bears the burden of the corporate tax in the open economy", Working Paper (NBER, Cambridge, MA).

Grieson, R.E. (1975), "The incidence of profit taxes in a neo-classical growth model", Journal of Public Economics 4:75–85.

Gruber, J. (1997), "The incidence of payroll taxation: evidence from Chile", Journal of Labor Economics 15:S72-101.

Hall, R.E. (1988), "The relationship between price and marginal cost in U.S. industry", Journal of Political Economy 96:921–947.

Hall, R.E., and D.W. Jorgenson (1967), "Tax policy and investment behavior", American Economic Review 57:391–414.

Hall, R.E., and A. Rabushka (1995), The Flat Tax (Hoover Institution Press, Stanford).

Hamilton, B.W. (1975), "Zoning and property taxation in a system of local governments", Urban Studies 12:205–211.

Hamilton, B.W. (1976), "Capitalization of intrajurisdictional differences in local tax prices", American Economic Review 66:743–753.

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

Harris, J.E. (1987), "The 1983 increase in the federal cigarette excise tax", Tax Policy and the Economy 1:87–111.

Harris, R. (1984), "Applied general equilibrium analysis of small open economies with scale economies and imperfect competition", American Economic Review 74:1016–1032.

Hausman, J.A. (1981), "Exact consumer's surplus and deadweight loss", American Economic Review 71:662–676.

Hines Jr, J.R., J.C. Hlinko and T.J.F. Lubke (1995), "From each according to his surplus: equi-proportionate sharing of commodity tax burdens", Journal of Public Economics 58:417–428.

Hotelling, H. (1929), "Stability in competition", The Economic Journal 39:41–57.

Joint Committee on Taxation (1993), Methodology and Issues in Measuring Changes in the Distribution of Tax Burdens (U.S. Congress, Washington, DC).

Joint Committee on Taxation (2000), Distributional Effects of the Chairman's Amendment in the Nature of a Substitute to H.R. 3916 (Repeal of the Federal Communications Excise Tax) for Markup by the Committee on Ways and Means on May 17, 2000 (U.S. Congress, Washington, DC).

Jones, R.W. (1965), "The structure of simple general equilibrium models", Journal of Political Economy 73:557–572.

Judd, K.L. (1985a), "Redistributive taxation in a simple perfect foresight model", Journal of Public Economics 28:59–83.

Judd, K.L. (1985b), "Short-run analysis of fiscal policy in a simple perfect foresight model", Journal of Political Economy 93:298–319.

Judd, K.L. (1987), "A dynamic theory of factor taxation", American Economic Review 77:42–48.
Karp, L.S., and J.M. Perloff (1989), "Estimating market structure and tax incidence: the Japanese television market", Journal of Industrial Economics 37:225–239.
Kasten, R., F. Sammartino and E. Toder (1994), "Trends in federal tax progressivity, 1980–93", in: J. Slemrod, ed., Tax Progressivity and Income Inequality (Cambridge University Press, Cambridge) pp. 9–50.
Katz, M., and H.S. Rosen (1985), "Tax analysis in an oligopoly model", Public Finance Quarterly 13:3–19.
Kay, J.A., and M.J. Keen (1983), "How should commodities be taxed?", European Economic Review 23:339–358.
Kiefer, D.W. (1984), "Distributional tax progressivity indexes", National Tax Journal 37:497–513.
Kotlikoff, L.J. (2000), "The A-K Model – its past, present, and future", in: G. Harrison, S.E. Hougaard Jensen, L.H. Pedersen and T.H. Rutherford, eds., Using Dynamic CGE Models for Policy Analysis (North Holland, Amsterdam).
Kotlikoff, L.J. (2002), "Generational policy", in: A.J. Auerbach and M. Feldstein, eds., Handbook of Public Economics, Vol. 4 (Elsevier, Amsterdam) Chapter 27, this volume.
Kotlikoff, L.J., and L.H. Summers (1987), "Tax incidence", in: A.J. Auerbach and M. Feldstein, eds., Handbook of Public Economics, Vol. 2 (Elsevier, Amsterdam) pp. 1043–1092.
Krugman, P. (1980), "Scale economies, product differentiation, and the pattern of trade", American Economic Review 70:950–959.
Krzyzaniak, M., and R.A. Musgrave (1963), The Shifting of the Corporation Income Tax (Johns Hopkins Press, Baltimore).
Lakin, C. (2001), "The effects of taxes and benefits on household income, 1999–2000", Economic Trends 569:35–74.
Lyon, A.B., and R.M. Schwab (1995), "Consumption taxes in a life-cycle framework: are sin taxes regressive?", Review of Economics and Statistics 77:389–406.
Marshall, A. (1890), Principles of Economics, 8th Edition (Macmillan and Co., London).
McClellan, M., and J. Skinner (1997), "The incidence of Medicare", Working Paper (National Bureau of Economic Research, Cambridge, MA).
McLure, C.E. (1970), "Taxation, substitution, and industrial location", Journal of Political Economy 78:112–132.
McLure Jr, C.E. (1974), "A diagrammatic exposition of the Harberger model with one immobile factor", Journal of Political Economy 82:56–82.
McLure Jr, C.E. (1975), "General equilibrium incidence analysis: the Harberger model after ten years", Journal of Public Economics 4:125–161.
McLure Jr, C.E., and W.R. Thirsk (1975), "A simplified exposition of the Harberger model, I: Tax incidence", National Tax Journal 28:1–27.
Menchik, P.L., and M. David (1982), "The incidence of a lifetime consumption tax", National Tax Journal 35:189–203.
Metcalf, G.E. (1994), "The lifetime incidence of state and local taxes: measuring changes during the 1980s", in: J. Slemrod, ed., Tax Progressivity and Income Inequality (Cambridge University Press, Cambridge) pp. 59–88.
Metcalf, G.E. (1999), "A distributional analysis of green tax reforms", National Tax Journal 52:655–681.
Metcalf, G.E., and G. Norman (2001), "Oligopoly deregulation and the taxation of commodities", Working Paper (Tufts University, Medford, MA).
Mieszkowski, P.M. (1972), "The property tax: an excise tax or a profits tax?", Journal of Public Economics 1:73–96.
Mieszkowski, P.M., and G.R. Zodrow (1989), "Taxation and the Tiebout model: the differential effects of head taxes, taxes on land rents and property taxes", Journal of Economic Literature 27(3):1098–1146.
Modigliani, F., and R. Brumberg (1954), "Utility analysis and the consumption function: an interpretation

of cross-section data", in: K.K. Kurihara, ed., Post-Keynesian Economics (Rutgers University Press, New Brunswick) pp. 388–436.

Mulligan, C.B., and T.J. Philipson (2000), "Merit motives and government intervention: Public finance in reverse", Working Paper 7698 (National Bureau of Economic Research, Cambridge, MA).

Musgrave, R.A. (1959), The Theory of Public Finance (McGraw-Hill, New York).

Musgrave, R.A. (1974), "Is a property tax on housing regressive?", American Economic Review 64: 222–229.

Musgrave, R.A., K.E. Case and H. Leonard (1974), "The distribution of fiscal burdens and benefits", Public Finance Quarterly 2:259–311.

Mutti, J., and H. Grubert (1985), "The taxation of capital income in an open economy: the importance of resident–nonresident tax treatment", Journal of Public Economics 27:291–309.

Oates, W.E. (1969), "The effects of property taxes and local public spending on property values: an empirical study of tax capitalization and the Tiebout hypothesis", Journal of Political Economy 77:957–971.

Pechman, J.A. (1985), Who Paid the Taxes: 1966–85? (Brookings Institution Press, Washington, DC).

Pechman, J.A. (1987), "Pechman's tax incidence study: a response", American Economic Review 77: 232–234.

Pechman, J.A., and B.A. Okner (1974), Who Bears the Tax Burden? (Brookings Institution Press, Washington, DC).

Poterba, J. (1989), "Lifetime incidence and the distributional burden of excise taxes", American Economic Review 79:325–330.

Poterba, J. (1996), "Retail price reactions to changes in state and local sales taxes", National Tax Journal 49:165–176.

Ramsey, F.P. (1928), "A contribution to the theory of taxation", The Economic Journal 37:47–61.

Salop, S.C. (1979), "Monopolistic competition with outside goods", Bell Journal of Economics 10: 141–156.

Seade, J.E. (1985), Profitable Cost Increases and the Shifting of Taxation: Equilibrium Responses of Markets in Oligopoly (University of Warwick, Warwick).

Shaked, A., and J. Sutton (1982), "Relaxing price competition through product differentiation", The Review of Economic Studies 49:3–13.

Shoven, J.B., and J. Whalley (1984), "Applied general equilibrium models of taxation and international trade: an introduction and survey", Journal of Economic Literature 22:1007–1051.

Simon, H.A. (1943), "The incidence of a tax on urban real property", Quarterly Journal of Economics 59:398–420.

Slemrod, J.B. (1992), "Taxation and inequality: a time-exposure perspective", Tax Policy and the Economy 6:105–127.

Spence, M. (1976), "Product selection, fixed costs, and monopolistic competition", Review of Economic Studies 43:217–235.

Starrett, D. (1988), "Effects of taxes on savings", in: H.J. Aaron, H. Galper and J.A. Pechman, eds., Uneasy Compromise: Problems of a Hybrid Income-Consumption Tax (Brookings Institution Press, Washington, DC) pp. 237–268.

Stern, N. (1987), "The effects of taxation, price control and government contracts in oligopoly and monopolistic competition", Journal of Public Economics 32:133–158.

Stiglitz, J.E. (1973), "Taxation, corporate financial policy, and the costs of capital", Journal of Public Economics 2:1–34.

Suits, D.B. (1977), "Measurement of tax progressivity", American Economic Review 67:747–752.

Summers, L.H. (1985), "The asset price approach to the analysis of capital income taxation", in: G.R. Feiwil, ed., Issues in Contemporary Macroeconomics and Distribution (Macmillan, London) pp. 429–443.

Tiebout, C.M. (1956), "A pure theory of local expenditures", Journal of Political Economy 64:416–424.

Tirole, J. (1988), The Theory of Industrial Organization (The MIT Press, Cambridge, MA).

Tobin, J. (1969), "A general equilibrium approach to monetary theory", Journal of Money, Credit and Banking 36:227–239.

Vandendorpe, A.L., and A.F. Friedlaender (1976), "Differential incidence in the presence of initial distorting taxes", Journal of Public Economics 6:205–229.

Venables, A.J. (1986), "Production subsidies, import tariffs, and imperfectly competitive trade", in: D. Greenaway and P.K.M. Tharakan, eds., Imperfect Competition and International Trade (Wheatsheaf, Sussex).

Willig, R. (1976), "Consumer's surplus without apology", American Economic Review 66:589–597.

Zodrow, G. (2001), "The property tax as a capital tax: a room with three views", National Tax Journal 54:139–156.

Zodrow, G.R., and P.M. Mieszkowski (1983), "The incidence of the property tax: the benefit view vs. the new view", in: G.R. Zodrow, ed., Local Provision of Public Services: The Tiebout Model after Twenty-Five Years (Academic Press, New York) pp. 109–129.