Preface

Commodity taxation – by which we mean, broadly speaking, taxes levied on the sale of goods and services – forms a substantial part of the revenue-raising resources of most governments. Across the OECD countries as a whole, taxes on the sale of goods and services accounted for about one third of all tax revenues in 2017, the overwhelming proportion of which came from a large general tax on the sale of goods and services. In nearly all OECD countries, with the notable exception of the United States, this general tax on sales took the form of a value added tax (VAT), a tax which has spread world-wide over the past sixty years.

The design and structure of commodity taxation has for many years been an active field of both theoretical and empirical research in public economics. In this Element we provide an overview of the major economic issues highlighted by this research, focussing mainly on theoretical issues, and on policy in OECD countries. Section 1 defines the scope of the subject: what, precisely, do we mean by commodity taxation? In Section 2 we sketch the outlines of the economic literature on the optimal structure of commodity taxes. When would uniform taxation of all goods and services be optimal, in the sense of minimizing the excess burden of tax revenue raising, or achieving the socially optimal balance between efficiency and equity? Section 3 discusses the incidence of indirect taxes: who ultimately bears the burden of these taxes? Section 4 discusses the design and properties of the VAT, and Section 5 discusses the particular issues involved in the application of VAT or other sales taxes to financial services. Section 6 discusses international aspects of commodity taxation, including the effects of border tax adjustments on traded goods, the effects of cross-border shopping, and tax exporting. Section 7 looks at the use of commodity taxes to achieve behavioural change, such as in environmental and health policy. Section 8 discusses a number of other issues in the recent literature, and Section 9 highlights some promising directions for future research.

1 Introduction and Overview

Commodity taxes have a long history. Taxes on the production, movement, or sale of certain commodities have for many years been a significant source of state revenues in many countries. In a system without modern accounting conventions, and with limited administrative capacity, the production or movement of commodities provided an observable basis for levying taxes, with a rough-and-ready relationship to citizens' ability to pay. Only in the last century or so have taxes on income overtaken taxes on commodities and on real property as the most powerful revenue instruments available to government.

The long history of taxing commodities left its imprint on the tax systems in many countries well into the twentieth century, in the form of patterns of commodity taxation which reflected the accretion over time of excises and other taxes on particular commodities. Even today, traces of the past can be found, in the form of special taxes or tax exemptions for particular commodities, with no apparent justification other than policy inertia. However, the second half of the twentieth century saw quite dramatic changes in commodity taxation, in particular with the development and spread of broad-based taxes such as VAT that replaced many older commodity taxes with a more considered and coherent system and structure.

The contribution that economics can make to rational and efficient commodity tax policy is the underlying theme of this short Element.

Before turning to the various aspects of commodity taxation that economic research has analysed, we begin by drawing some boundaries to the scope of our analysis, by looking at the definition of our subject matter, and the types of taxation with which we will be concerned. We have had to be selective, and have left out some interesting topics, simply because there is not space to do them justice.

By commodity taxes we generally mean taxes on the sale of commodities, in other words, the sale of goods and services.[1] These can include sales to consumers – retail sales – and also sales of intermediate goods between businesses. The literature also refers to commodity taxes as consumption taxes, where the perception is that the taxes in question are taxes on consumption goods rather than intermediate inputs to production or investment goods. We shall use these more or less close synonyms interchangeably, but will be more exact when necessary.

Commodity taxes can take many different forms. Some are general, in the sense that they are imposed on most commodities, even though in practice hardly any commodity tax is universal. Key examples are value added tax (VAT)[2] and general sales taxes levied on a wide range of goods with limited differentiation of rates. Other commodity taxes – commonly called excises or excise duties – are targeted at specific goods such as fuel and alcohol. They can take the form either of *ad valorem* taxes charged according to the value of a

[1] In common with general practice in economics, we will often use the word 'goods' to encompass all commodities, including services, although there will be places where we may need to distinguish between services and (physical) goods.

[2] Value added tax (VAT) – called Goods and Services Tax (GST) in some countries – takes the form of a sales tax initially levied on all sales, including those to other businesses as well as to final consumers; however, the ultimate effect of VAT is broadly, as we will see, to act as a tax on retail sales alone.

transaction or *specific* taxes, levied per unit, or based on a physical measure such as weight or volume.

One of the oldest forms of commodity taxes are trade tariffs. Historically these were a major source of public revenue. However, due to the liberalization of international trade over the years, the contribution of export and import tariffs to public revenue has diminished and is now very small in developed countries. They raise distinctive economic issues, which go beyond the scope of this Element, and we shall not consider tariffs in our discussion.

The economic and policy literature on taxation frequently makes use of a distinction between 'direct' taxes, such as the personal income tax, and 'indirect' taxes, such as VAT and other taxes on the sale of goods and services. Defining the precise meaning of 'indirect taxation', or setting the precise boundary between direct and indirect taxes so that all taxes fall in one or other category, is not wholly straightforward. The most common approach to systematic definition is one which distinguishes between taxes that are levied on identifiable persons, and which potentially can be differentiated according to their individual characteristics (such as total income, family size, etc.), and those which are levied on transactions (such as the sale of goods and services) and which impose an 'anonymous' tax burden, without reference to the characteristics or circumstances of the individuals affected.

Typically, with commodity taxes such as VAT, sellers of goods are required to account for the total value of their sales of taxable goods (possibly in various categories subject to different tax rates). A tax is calculated on this amount, and the seller must remit this tax to the government. The perception that this is an indirect tax reflects a view that the sellers would charge the tax as part of the price and hence the buyers are taxed in an indirect way. From this perspective the sellers are not the ultimate taxpayers, but merely operate as tax collectors for the government. To what extent sellers actually pass such taxes through to customers is a key economic question, to which we shall return.

In an important sense, most of our categories of taxation concern organizational form, rather than economic substance. A general sales tax could, in principle, be levied on all consumer spending, and if levied at a uniform rate on all goods and services (and if fully passed-on in prices) the total tax on a consumer's spending would be equivalent to applying the tax rate to the consumer's total spending. The same outcome could be achieved by levying an 'expenditure tax' on the individual's annual aggregate individual spending (computed as the difference between individual income and net saving). This (direct) expenditure tax was advocated, for example, by the Meade Report (1978) as a replacement for income tax, but such a tax has not been fully implemented anywhere. If the direct expenditure tax took account of no

circumstances of the individual other than aggregate spending, its economic substance may be identical to the effect of a sales tax levied at a uniform rate on all categories of spending. However, the direct expenditure tax could instead be levied as a non-linear function of total spending, and also offers the possibility of taking account of various characteristics of the taxpayer that may be relevant in ensuring an equitable distribution of the overall tax burden between taxpayers. By contrast, the burden of a sales tax can only reflect taxpayer circumstances very indirectly, by applying different rates of tax to goods that are more or less important in the spending of different types of household. Whether this is a serious drawback of indirect taxation is an issue that has been extensively discussed in the economics literature, but the contrast in principle between what can be achieved through direct and indirect taxation should be clear.

Another 'boundary' issue that is worth noting is the relationship between commodity taxes and various payments that governments may levy under the names of fees, tolls, charges, or duties to charge people for using certain goods or public services. Although the fees may be intended to charge people for the costs they impose on public services (such as publicly supplied energy, waste collection, transport, telecommunications services, etc.), they may be set at a level that raises net revenue for the government, in which case they are in effect commodity taxes without being formally defined as such. While we do not explicitly discuss public sector pricing here, we note the close similarities with the economic theory of commodity taxation. This is reflected in the term Ramsey–Boiteux pricing which integrates Ramsey's (1927) theory of optimal indirect taxation and Boiteux's (1956) optimal public sector pricing rule – the rule according to which prices ought to deviate from marginal costs as derived from second-best theory.

2 The Optimal Structure of Commodity Taxes

One of the enduring, seminal, papers in public economics lies at the heart of the theoretical literature on indirect taxation. A remarkable 1927 paper, written by Frank Ramsey, sought to give an answer to the question of what the pattern of tax rates across different goods and services should be, if the government's objective is to raise a given tax revenue at the least economic cost. The answer set out by Ramsey – loosely speaking, that tax rates should be set in inverse proportion to the elasticities of demand and supply – has been repeatedly revisited by public economists, leading to a series of papers which probe the precise conditions under which Ramsey's results hold.

The question Ramsey posed was beguilingly simple, albeit that the mathematics required to prove Ramsey's answer was far from straightforward. (The

subsequent development of the theory of duality has greatly simplified the proof.) In Ramsey's problem, the government is constrained to use taxes on commodities to meet its given revenue requirement, and is taxing a population of identical individuals. Ramsey's conclusion – frequently interpreted as an 'inverse elasticity' rule – was explored in more detail by Corlett and Hague (1953), who demonstrated that the efficient pattern of tax rates involves higher taxes on goods that are more complementary with leisure – a result that has the appealing intuition that the indirect tax rates can be interpreted as compensating for the absence of a direct tax on leisure. Nevertheless, both Ramsey and Corlett and Hague derive their results in a context where the taxpaying population is homogeneous. In this context, it is unclear why indirect taxes are needed at all: a poll tax could raise the required revenue without incurring the costs of consumption distortions.

Subsequent literature has explored Ramsey's question in less-restrictive contexts: where individuals differ, so government objectives in revenue raising may need to take account of both efficiency and equity objectives; and where governments have a wider range of available tax instruments, including taxes on income as well as on spending. The most influential contribution has been Atkinson and Stiglitz (1976) who showed that when governments have the ability to levy sophisticated non-linear taxes on income, there are circumstances in which it will be optimal to levy taxes at a uniform rate on all goods and services – specifically where labour supply and commodity demands are separable. We shall single out for further discussion a number of aspects of the optimal structure of indirect taxes in the following sections. More issues are surveyed in Crawford et al. (2010), Boadway (2012), and Nygård and Revesz (2016).

2.1 A Closer Look at Corlett and Hague

While the inverse elasticity rule derived from Ramsey and the Corlett–Hague rule are both widely known, the literature does not often spell out their full implications and the relationship between them. We therefore believe that elaborating certain aspects of these tax rules is worthwhile.

The underlying model assumes that there is a single representative consumer or, equivalently, a population of identical consumers. There are two taxed market commodities and (untaxed) leisure. These goods are indexed 1, 2, and 0, respectively. For an agent, choosing leisure is obviously equivalent to choosing labour supply. Producer prices are fixed and equal to marginal (and average) cost. There is a fixed wage rate set equal to 1. Denote by t_i the tax on commodity i and let q_i be the consumer price. Let σ_{ij} denote the compensated elasticity of demand for

good i with respect to the price of commodity j. It is trivial that taxing any market commodity will distort the trade-off between the taxed commodities and leisure, which is equivalent to distorting labour supply. Where the market commodities are taxed at different relative rates, there is also a distortion of the trade-off between them, which we shall refer to as distortion of the consumption bundle.

The optimal tax structure is the t_1, t_2 pair that maximizes utility, subject to a pre-set tax revenue requirement. In formal terms the objective is to maximize the indirect utility function $V(q_1, q_2)$ subject to $t_1 x_1(q_1, q_2) + t_2 x_2(q_1, q_2) = R_0$ for a given R_0, where $x_1(.)$ and $x_2(.)$ are demand functions. Optimum taxation can then be characterized by the following, known as the Corlett–Hague rule:

$$\frac{t_1/q_1}{t_2/q_2} = \frac{-\sigma_{11} - \sigma_{22} - \sigma_{10}}{-\sigma_{11} - \sigma_{22} - \sigma_{20}}.$$

We immediately see that commodity 1 or commodity 2 is taxed at a higher percentage rate according as $-\sigma_{10} > -\sigma_{20}$ $(\sigma_{20} > \sigma_{10})$ or $-\sigma_{10} < -\sigma_{20}$ $(\sigma_{10} > \sigma_{20})$. This means that a higher percentage tax rate is imposed on the commodity that is more complementary with leisure. Obviously, there is no differentiation in terms of relative tax rates if $\sigma_{10} = \sigma_{20}$. The interpretation is that the distortion of labour supply can be mitigated by taxing a commodity that is complementary with leisure. In the absence of a tax levied directly on leisure, this is an indirect way to tax it.

While only the value of $-\sigma_{10}$ relative to $-\sigma_{20}$ determines which tax rate should be higher, the size of $-\sigma_{11} - \sigma_{22}$ also influences the extent to which one would like to differentiate commodity taxes in case $\sigma_{10} \neq \sigma_{20}$. Letting commodity 1 be the more complementary with leisure such that $\frac{t_1/q_1}{t_2/q_2} > 1$, we note that a larger value of $-\sigma_{11} - \sigma_{22}$ implies a smaller percentage increase of the numerator than of the denominator. The effect is to diminish the ratio and move it closer to 1. Accordingly, there is less differentiation. This effect reflects that we should also take into account the distortion of the consumption bundle in addition to the labour supply distortion. This becomes more transparent when noting that $-\sigma_{11} = \sigma_{12} + \sigma_{10}$ and $-\sigma_{22} = \sigma_{21} + \sigma_{20}$, so that for given values of σ_{10} and σ_{20} larger values of $-\sigma_{11}$ and $-\sigma_{22}$ reflect a larger degree of substitution between the taxed commodities. Substituting for $-\sigma_{11}$ and $-\sigma_{22}$, we can write the Corlett–Hague rule as $\frac{t_1/q_1}{t_2/q_2} = \frac{\sigma_{12}+\sigma_{21}+\sigma_{20}}{\sigma_{12}+\sigma_{21}+\sigma_{10}}$. This may be a more instructive way to express the condition. Larger values of σ_{12} and σ_{21} imply that further differentiation of prices will induce stronger substitution between taxed commodities which exacerbates the distortion of the consumption bundle. This is a case for a more modest differentiation of taxes and prices. For sufficiently large

values of σ_{12} and σ_{21}, the ratio will approach 1 and the tax differentiation will vanish. We can conclude that the optimal tax rule expresses a trade-off between the desire to diminish the labour supply distortion and the desire to avoid a large distortion of the consumption bundle.

We note that when $\sigma_{10} = \sigma_{20} = 0$, uniform percentage taxation is always optimal. Since there is no substitution between taxed commodities and leisure, the only potential inefficiency is a distortion of the consumption bundle which can always be avoided by refraining from differentiating tax rates.

Now consider the case where there is no substitution between taxed commodities: $\sigma_{12} = \sigma_{21} = 0$. Hence $\sigma_{11} + \sigma_{10} = \sigma_{22} + \sigma_{20} = 0$, and $\frac{t_1/q_1}{t_2/q_2} = \frac{-\sigma_{22}}{-\sigma_{11}}$, which alternatively can be written as $\frac{t_1}{q_1} = \frac{\alpha}{-\sigma_{11}}$ and $\frac{t_2}{q_2} = \frac{\alpha}{-\sigma_{22}}$ for some $\alpha > 0$ determined by the tax revenue requirement. This tax rule is the familiar 'inverse elasticity rule', frequently highlighted in basic presentations of optimal commodity taxes, under which the tax rate is proportional to the inverse of the own price elasticity. We should note that since $-\sigma_{11} = \sigma_{10}$ and $-\sigma_{22} = \sigma_{20}$, the inverse own price elasticity rule is equivalent to $\frac{t_1/q_1}{t_2/q_2} = \frac{-\sigma_{20}}{-\sigma_{10}}$ or as $\frac{t_1}{q_1} = \frac{\alpha}{\sigma_{10}}$ and $\frac{t_2}{q_2} = \frac{\alpha}{\sigma_{20}}$. These inverse *cross price* elasticity rules highlight the links to the labour market and may be more illuminating than the standard formulations.

We see that, as in the aforementioned general case, which tax rate is set at a higher value depends on the relative complementarity of commodity demand with leisure. What is different in this case is that since there is no substitution between taxed commodities, no effect on the distortion of the consumption bundle appears in the optimal tax rule. A low absolute value of the own price elasticity now simply reflects weak substitutability with leisure and warrants a relatively high tax rate.

It is important to note that the estimates of σ_{11} and σ_{22} required for implementing the taxes prescribed by the Corlett–Hague rule are those describing demand responses in a setting where commodity demand and labour supply are chosen simultaneously. It is not obvious how low or high estimates of own price elasticities under a fixed total consumption expenditure translate to the commodity demand–labour supply model.

Dixit (1975) provides a generalization to more than two taxed commodities. A different extension would be to consider more untaxed goods. There may be goods that are administratively difficult to tax. An even more compelling case is goods consumed abroad by tourists from the home country under consideration (sight-seeing, foreign hotel accommodation, etc.). Since this consumption takes place outside the jurisdiction of the government of the home country, taxation by the home country is obviously impossible. We discuss further international aspects of taxation in Section 6.

Let us now add a further untaxable good, denoted good 3, to our model. We can then easily derive a Corlett–Hague rule for this case:

$$\frac{t_1/q_1}{t_2/q_2} = \frac{-\sigma_{11} - \sigma_{22} - \sigma_{10} - \sigma_{13}}{-\sigma_{11} - \sigma_{22} - \sigma_{20} - \sigma_{23}}.$$

Now we note that also the cross effect to the additional untaxed good matters. For example, suppose that good 1 is domestic tourism and good 3 is the fairly close substitute tourism taking place abroad. Then σ_{13} is positive and conceivably fairly large, implying that good 1 (tourism at home) should be taxed more leniently than otherwise. The example should just be considered as an illustration. A full analysis of the taxation of tourism should clearly also allow for taxation of goods typically purchased by foreign tourists in the home country.

We now return to the setting with a single untaxed good, interpreted as leisure. Taking a general income tax as point of departure, Meade (1955) argued that tax rates on various goods should depend on how demands for the goods are related to leisure (or labour). Meade concluded that welfare would be enhanced by a marginal change in the tax system, which raises the price of those things that are jointly demanded with leisure, and lowers the price of work-related things. One may think respectively of leisure goods (e.g. sports equipment or concert tickets) and goods consumed while working (e.g. childcare or clothes worn at work). However, it is not trivial which time uses one should consider as work or leisure, respectively. Often the distinction is made between work for pay in the market and leisure, but then one observes that many 'leisure activities' are similar to activities that other people carry out as paid work (house maintenance, gardening, cleaning, etc.). Other 'leisure activities' require unpleasant use of time (pure travel time, time at the dentist's, household chores, etc.), hardly experienced as significantly different from time doing market work. Even if it is a pleasure, using leisure goods takes time. Fishing rods, skis, and books are not of value unless time is made available to use them. The time used for consuming a leisure good must then compete with time used for market work, for consuming other leisure goods, or for pure leisure (rest) without consuming particular goods (listening to the birds or enjoying the sunset). A series of papers have addressed these various aspects of time use in the context of indirect taxation.

Sandmo (1990) and Kleven et al. (2000) drew attention to the fact that several commodities can either be purchased in the market or produced by own efforts in the household and argued that optimal taxation should favour market-produced services which are close substitutes for home-produced services.

Christiansen (1984) introduced the Becker–Lancaster approach to model the enjoyment of leisure, or recreation, as an activity using time and market goods as inputs. (Becker, 1965; Lancaster, 1966). By taxing leisure goods purchased in

the market, recreation is made less attractive compared to work, but the paper highlights a further effect: when taxed, leisure goods may be substituted by leisure time in the 'production' of recreation and, with strong substitutability, less time may actually be devoted to work. It follows that it is not necessarily appropriate to recommend taxation of leisure goods, even if one is willing to make the plausible assumption that they are technical complements with leisure within a Becker–Lancaster framework. Recreation may simply be made more time-intensive and less commodity-intensive. To give an extreme example, one might substitute a long, cheap, cottage holiday for a short, service-intensive, luxury cruise.

Later papers by Gahvari and Yang (1993) and Kleven (2004) took a similar approach in that they considered taxation of market goods used to transform time into utility in the spirit of Becker and Lancaster. These papers carry out a more extensive analysis with an arbitrary number of activities, but rule out substitution between the factors own time and market goods in each activity. The optimal tax rule is a simple inverse factor share rule: the tax rate on any given market good is inversely proportional to its factor share. In a special case, the consumer allocates his entire time endowment to activities, all of which require an input of time and of market goods with fixed input coefficients. By taxing the input of market goods, one can indirectly tax the entire time endowment of the consumer to achieve the first best. Another case is the one where part of the time endowment is allocated to pure, untaxed leisure alongside the allocation of time to activities also requiring material input. The inverse factor share rule is then modified by a Corlett–Hague factor, reducing to an inverse elasticity factor in the absence of cross-price effects as in the simple Ramsey model.

Boadway and Gahvari (2006) brought together many of the aspects of time use mentioned previously. They distinguish between uses of time generating disutility (labour or household work, etc.) and uses of time generating utility either directly as leisure in isolation (pure leisure or rest) or in combination with goods purchased in the market (museum tickets, music, etc.). Consumption of a good requires a fixed amount of time that is either unpleasant and is a perfect labour substitute or is enjoyable and is a perfect substitute for pure leisure.[3] Leaving details and caveats aside, we can roughly summarize the overall insights from the paper as follows: For goods requiring leisure-equivalent time, the optimal tax rules are the standard ones. Goods for which the time spent consuming them is unpleasant and labour-equivalent, should *ceteris paribus* be taxed at a higher rate than those for which the time use is pleasant.

[3] In this respect, the paper is inspired by Gahvari (2007).

The more time-intensive these market goods are in consumption, the greater the need for the tax rates to be higher.

The Corlet–Hague and inverse elasticity rule are derived under the strict assumption that there is a homogenous population. To extend the analysis to models with a heterogeneous population and distributional concerns, we consider a simple setting with two (types of) agents, type 1 and type 2, where the latter is the richer, for instance, due to a higher wage rate. We assume that the policy instruments are two commodity taxes and a lump sum transfer, denoted by a. We denote prices by q_1 and q_2. The social welfare function is the sum of (indirect) utilities: $V^1(q_1, q_2, a) + V^2(q_1, q_2, a)$, and the tax revenue requirement is $t_1\left(x_1^1 + x_1^2\right) + t_2\left(x_2^1 + x_2^2\right) - 2a = R_0$, where superscripts denote types of agent. To maximize welfare, we formulate the Lagrange function:

$$V^1(q_1, q_2, a) + V^2(q_1, q_2, a) + \mu\left(t_1\left(x_1^1 + x_1^2\right) + t_2\left(x_2^1 + x_2^2\right) - 2a - R_0\right).$$

To summarize the trade-offs that characterize the tax optimum, we can derive the following condition which expresses the effects of shifting the tax burden from commodity 2 to commodity 1: $\left[\frac{t_1}{q_1}(\sigma_{11} + \sigma_{22} + \sigma_{20}) - \frac{t_2}{q_2}(\sigma_{22} + \sigma_{11} + \sigma_{10})\right] + (\Lambda^1 - \Lambda^2)\left(\frac{x_2^1}{X_2} - \frac{x_1^1}{X_1}\right) + (m^1 - m^2)\left(\frac{x_2^1}{X_2} - \frac{x_1^1}{X_1}\right) = 0$, where $m^i = t_1\frac{\partial x_1^i}{\partial y^i} + t_2\frac{\partial x_2^i}{\partial y^i}$, to which we shall return later. $\Lambda^i = \frac{\lambda^i}{\mu}$ we interpret as the welfare weight assigned to marginal income accruing to agent i.

The Appendix shows how the first order conditions are derived.

This optimal tax condition comprises three main terms. The first term in brackets consists of the Corlett–Hague terms capturing the labour supply and consumption bundle distortions discussed previously. We note that in the special case of identical agents, the condition reduces to the pure Corlett–Hague rule. The second term captures the distributional effects. Where it is desirable to redistribute income from the richer type 2 to the poorer type 1 (i.e. $\Lambda^1 > \Lambda^2$), shifting taxes from commodity 2 to commodity 1 has a beneficial welfare effect where the poorer agent's consumption share is smaller for commodity 1 than for commodity 2. The reason is that for a marginal tax increase that is fully passed on to the consumer price, the consumption share of an agent reflects the consumer's share of the tax burden.[4] Finally, if we consider the third term, this can be interpreted as an efficiency effect of redistribution. The reason is that the commodity taxes are wedges between the consumers' marginal commodity valuation and the marginal cost. It follows

[4] The burden of a price increase measured as loss of real income is $-\frac{\partial V^i/\partial q_k}{\partial V^i/\partial y^i} = x_k^i$, where x_k^i is agent i's consumption of commodity k. The fraction of the burden is x_k^i/X_k, where $X_k = \sum_i x_k^i$.