

Value Added Tax and Excises

*Ian Crawford, Michael Keen, and Stephen Smith**

Ian Crawford is a Reader in Economics at the University of Oxford, a Fellow of New College, and a Research Fellow at the IFS. His main areas of research are applied microeconomics, consumer theory, and index numbers. In the past he has written about taxation of alcohol, road transport, and domestic energy. Current interests include the analysis of habit formation and the tricky business of unobserved heterogeneity in microeconomic data.

Michael Keen is an Assistant Director of the Fiscal Affairs Department of the IMF, where he was until recently head of Tax Policy. Previously Professor of Economics at the University of Essex, he was President of the International Institute of Public Finance from 2003 to 2006, is currently a member of the Board of the National Tax Association, and has served various journals, including as an early editor of *Fiscal Studies*. He has published on a wide range of public finance issues and is a co-author of the book *The Modern VAT*.

Stephen Smith is Professor of Economics and Executive Dean of the Faculty of Social and Historical Sciences at UCL and a Research Fellow at the IFS. His research covers various topics in the economics of tax policy—especially VAT and excises—and environmental economics. Before joining UCL he worked at the Department of Trade and Industry (1977–85) and at the IFS (1985–97), where he was Deputy Director from 1990 to 1997. He has acted as a consultant to HM Treasury, DEFRA, the UK Environment Agency, the OECD, the European Commission, and the IMF.

* We are grateful to Richard Blundell and Stuart Adam for their editorial input and advice, to Mike Brewer and James Browne for their contribution to the reform simulation discussed in the chapter, to the commentators Richard Bird, Sijbren Cnossen, Ian Dickson, Jon Gruber, and David White, and to participants at the Mirrlees Review conferences in London and Cambridge for vigorous and constructive discussion of earlier drafts of the chapter. Views expressed here should not be attributed to the International Monetary Fund, its Executive Board, or its management. Expenditure and Food Survey data are collected by the Office for National Statistics and distributed by the Economic and Social Data Service. Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

EXECUTIVE SUMMARY

The thirty years since the Meade Report (Meade, 1978) have seen a remarkable shift in the balance of taxation in the UK towards the Value Added Tax (VAT), which now accounts for around one sixth of all tax revenue. Reliance on alcohol and tobacco excises, in contrast, has fallen. This chapter considers the strategic design issues that we believe should shape the future course of UK policy on VAT and these excises. Given the strength of EU restrictions on indirect tax policy, this also means, to some degree, considering their future in Europe more widely.

The chapter is in four parts. The first considers theoretical insights into the role and optimal structure of indirect taxes. The second focuses on the VAT, and the third on alcohol and tobacco excises. The fourth considers the increasingly important international issues bearing on the design and implementation of these taxes.

The appropriate balance between direct and indirect taxation—between income taxes and taxes on goods and services—is one of the oldest issues in public finance, but still imperfectly understood. It is clear, however, that the differences between the two types of tax are less sharp than once thought. In particular, the close equivalence between a uniform tax on consumption and a uniform tax on wage and profit income means that—as long as people take price levels properly into account—the two taxes should have broadly equivalent effects on the labour market. So shifting the balance of taxation towards VAT cannot be expected to have a great impact on work incentives or levels of employment. Indeed, the appropriate mix of direct and indirect taxes may be primarily a matter of administration and compliance. Running a broad-based consumption tax in parallel with taxes on income reduces the risk of revenue losses by spreading taxation across a number of sources, each of which is to some degree independently enforced.

Indirect taxes offer scope, not available with an income tax, for taxing different components of consumption at different rates. One group of reasons for taxing some goods more heavily is to deal with pollution (considered in Fullerton, Leicester, and Smith, Chapter 5) and other external costs. The external costs of alcohol and tobacco consumption are frequently seen as the principal economic justification for continuing to levy high excise duties on them, so policy needs to be informed by a clear appreciation of the nature and scale of these external costs. Nevertheless, conventional externalities may not be the only grounds for high taxation of tobacco and alcohol. Recent economic literature argues that people's lack of consistency and self-control

may justify higher taxes than would be warranted if consumption choices were being made by wholly rational, well-informed consumers.

Whether the opportunity to charge different tax rates on different commodities should be exploited more generally depends on the range of other tax instruments available, and the characteristics of people's demands for those commodities. Lower tax rates on items such as childcare costs, travel costs to work and labour-saving food items (such as ready meals), and higher tax rates on items like leisure-time consumption and DIY inputs, could help to offset the disincentives to work created by the tax system. But it is far from clear how much differentiation would be justified on these grounds, or which commodities should be taxed more and less heavily. There are also plenty of practical obstacles to the complex pattern of tax rates that might be required, and these need to be weighed against any potential labour market gains.

One clear implication of this line of analysis, however, is that the case for using preferential rates of VAT to help the less well-off is weak: there are better redistributive instruments available to the UK government than fine-tuning rates of commodity taxation. The chapter shows that ending all current zero and reduced rates (except for housing and exports) while increasing all means-tested benefit and tax credit rates by 15% would leave the poorest 30% of the population better off, on average, and raise £11 billion that could be used to help them further or for some other purpose. The essence of this result is nothing new. Finding the political will to implement such a change needs to begin with a recognition of the fundamental unfairness—and wastefulness—of the existing rate structure.

Still new at the time of the Meade Report, the VAT systems of EU countries are now starting to look outdated by comparison with more recently designed systems, both in terms of the rate structure and in aspects of basic design. Two issues stand out:

First, are the widespread exemptions—for public agencies and health, education, medical services, and financial services, in particular—which break the chain of VAT paid on sales and reclaimed on purchases of inputs, creating inefficiencies of unknown magnitude. Recent international experience—including that of New Zealand—shows that many of these exemptions are unnecessary, and demonstrates the potential for a base-broadening reform of the VAT. The case for considering alternatives to exemption is especially clear in relation to financial services, though comprehensive solutions are difficult. One possibility meriting close attention is cash flow treatment of financial services—treating all cash inflows to providers of such services (including deposits) as reflecting VAT-able sales to customers and treating all cash outflows (including loans made) as purchases of inputs—but zero-rating

financial transactions with registered businesses. While there are many practical issues that require further attention, movement to such a system would not only remove distortions but likely entail a substantial net revenue gain.

The second key structural issue is the VAT treatment of trade between EU member states. We argue that while the goal of a systematic destination principle VAT treatment of internationally traded goods and services remains desirable, the current mechanisms by which this is achieved need reconsideration. In particular, zero-rating of exported goods exposes the VAT system to significant risks of fraud and evasion—exploited by recent high-profile instances of ‘carousel fraud’, as well as by more mundane evasion. We argue that while the extent of these problems should not be over-stated, a reform of the VAT treatment of intra-EU trade, in which goods would be exported bearing VAT, would reduce the vulnerability of the VAT system to frauds and evasion surrounding international transactions.

4.1. INTRODUCTION

The Meade Report was confined to direct taxation: it was initially intended to review the whole tax system, but time constraints led the committee to exclude indirect taxes.¹ One reason for this prioritization, perhaps, was that the UK had then only recently implemented a major indirect tax reform, introducing a value added tax (VAT) to replace the ‘purchase tax’ (a single-stage sales tax) as a requirement of entry into the European Community. It may have seemed that there would be little appetite for further reforms to the UK’s indirect tax system for some time to come. As it turned out, however, developments in indirect taxation, and particularly the rise of the VAT, have been amongst the most marked changes in the UK tax system since the publication of the Meade Report, and one of those most consistent with the central thrust of the Meade Report—its advocacy of expenditure taxation.

Indirect taxes evidently merit closer attention than the Meade Report was able to provide. Moreover, there have been significant developments in the area since, conceptually and practically:

- Theoretical developments since the Meade Report have considerably altered our understanding of the contribution that can and cannot be made by indirect taxes—both VAT and excises—to raising revenue and pursuing distributional and other social objectives.

¹ The only substantive discussion of indirect taxes is in the context of one of the more radical reform options, the ITVAT, under which existing income taxes would be transformed into a tax on expenditure using a VAT-type mechanism.

- Having been in place for some thirty-five years the UK VAT, like that of many EU members, is now beginning to show its age, and a fundamental review, in the light of experience with more modern VATs, is overdue.
- The elimination of internal frontiers in the EU has brought new and challenging issues of administration, enforcement, and tax competition in relation to the VAT and excises, undreamt of thirty years ago and as yet still unresolved.
- Indirect tax policy in the UK is more explicitly constrained by international agreements than is any other area of tax policy (with the sole exception of tariff design). To a large degree, assessing indirect tax policy in the UK requires assessing it in the EU more widely, a key question being whether EU constraints have been a help or a hindrance.
- Recent developments—notably the growth of trade in international services and e-commerce, and high profile VAT fraud—have raised challenges that question basic design features of the VAT, to a degree that causes some to doubt its future.

The focus of this chapter, for the most part, is on the strategic design issues that we expect and/or hope will shape the development of indirect tax policy in the UK, and the EU, in the coming years.

The chapter is in four main parts. The first considers economic theory and empirical evidence on the optimal structure of commodity taxes.² What, in particular, do taxes on goods and services contribute to an efficient tax structure that cannot be achieved equally or more efficiently by other available tax instruments? We then consider in turn the two principal categories of indirect tax in the UK: the VAT, applied to a considerable proportion of spending, and the excise duties levied at high rates on a limited number of goods. We compare the strengths and weaknesses of the VAT with alternative broad-based taxes on goods and services, in particular the single-stage retail sales taxes common in the US, and discuss issues concerning VAT rates, base, the treatment of small firms, financial services, and compliance. Turning to the excises we discuss externality arguments for levying high excise duties on alcohol and tobacco (those on fuel being considered in a separate chapter), and issues concerning the choice of tax base. The fourth main part outlines

² One key issue not addressed here is the incidence of these taxes. This is only partly for brevity, there also being little hard evidence to draw on (especially for broad-based commodity taxes). It is simply assumed that the incidence is fully on consumers, which such evidence as there is suggests is not unreasonable, at least for the longer run. (See, for example, Carbonnier (2007) and the references therein.)

and assesses international issues in indirect taxation. Conclusions are drawn in Section 4.6.

4.2. THE ROLE AND DESIGN OF INDIRECT TAXES

The years since the Meade Report have seen a significant change in the pattern of indirect taxation in the UK. Over the period as a whole, reliance on indirect taxes has increased only moderately, from around 23% of total tax revenue (including National Insurance) in 1978–79 to about 25% in 2007–08.³ There has, however, been a marked change in the way in which this revenue is raised, with a substantial increase in reliance on the VAT (from about 9 to 16% of total tax revenue⁴) and a reduction in that on the excises on alcohol and tobacco (each falling from around 4% of the total to about 1.5%).⁵ These developments have not occurred evenly or even monotonically over this period. The rise of the VAT began with the dramatic first budget of the Thatcher government in 1979, which raised the standard rate of VAT from 8 to 15% (while reducing income tax rates sharply), and continued with the increase in the standard rate of VAT to 17.5%—its level at the time of writing⁶—in 1991.⁷ Since the early 1990s, when it reached more than 18%, the share of the VAT in total tax revenue has actually decreased. Most of the reduced reliance on excises also dates from around then (though receipts from those on alcohol and tobacco have declined throughout the period), so that the overall share of indirect taxes has also fallen since the early 1990s by about 5 percentage points. The shift towards indirect taxation has thus, to a large extent, come and gone, leaving the compositional change—the shift from excises to VAT—as the most striking structural development.⁸

³ Revenue statistics in this paragraph come from HM Treasury: see <http://www.ifs.org.uk/ff/revenue_composition.xls>.

⁴ HMRC reports VAT revenues net of refunds paid to various public bodies but includes those amounts refunded in total tax revenue, so that the implied VAT shares need to be interpreted with caution.

⁵ For the period as a whole, revenue from fuel excises and other indirect taxes (such as the former car tax and more recent climate change levy) have both been broadly unchanged, at something under 5% and 2% respectively.

⁶ This chapter was finalized before the temporary reduction in the standard rate of VAT, until the end of 2009, from 17.5 to 15%.

⁷ A reduced rate of 8%, mainly for domestic energy use but subsequently extended to other items, was introduced in 1994, and reduced to 5% in 1997.

⁸ The OECD *Revenue Statistics*, which include in VAT revenue amounts refunded to public bodies, do not show the same peak in VAT revenues as do the HMRC figures used here. They do, however, also show a peak in reliance on indirect taxes in the early 1990s.

These developments highlight two key questions for tax policy: what is the appropriate balance between direct and indirect taxation, and what is the best structure of indirect taxes?

4.2.1. Equivalences and the balance between direct and indirect taxation

The appropriate mix between direct and indirect taxes is one of the oldest issues in public finance: it was a key issue, for example, in mid-Victorian politics, when the future—indeed survival—of the income tax remained in doubt.⁹ More recent and formal theory has brought relatively few additional insights. The most important, perhaps, is a recognition that, in principle at least, the balance is to some degree arbitrary, there being a close similarity in terms of their impact on individuals' budget constraints—and hence, in the absence of some form of fiscal illusion, on their behaviour—between a uniform tax on consumption and a uniform tax on wage and profit income. This is immediately clear for a consumer who lives only one period and receives income only from these sources: for them, a tax of 20% on all the income they receive is equivalent to a 25% tax on everything they spend.¹⁰ In such a world, the balance between commodity and wage taxation would be immaterial.

For a consumer who lives for several periods, consumption over the lifetime (and any bequest given at the end of it) has to be financed from the initial stock of savings, wage earnings, and receipts of profit and transfer income. The equivalence is then somewhat more subtle: a uniform consumption tax, levied at an unchanging rate over time,¹¹ is equivalent to a proportional tax on wage, transfer, and profit income, also at an unchanging rate, combined with a tax on initial assets and subsidy to bequests at the same rate.

Such equivalences point to the potential fallacy, for example, in arguing that relatively heavy reliance on commodity rather than wage taxation (whether through the personal income tax or social security contributions)

⁹ See, for example, Matthew (1979), who notes that in his famous comparison of direct and indirect taxes to 'two attractive sisters' to both of whom he felt it allowable to pay his addresses, Gladstone carefully did not in fact say he felt obliged to pay them equal attention.

¹⁰ The only reason these numbers differ is that income tax rates are conventionally described in tax-inclusive form (including the tax itself in the base) and VAT rates in tax-exclusive form (excluding it).

¹¹ If its rate is expected to change over time (to increase, say), then a uniform consumption tax affects (reduces) the return on savings and so is in part equivalent to a (positive) tax on capital income.

is 'good for employment': absent other sources and uses of funds, such a shift has no real impact on labour supply. Other distortions may mean that there is some effect on labour market equilibrium: with a binding minimum for the wage before tax and employee's social contributions, for example, a reduction in the employer's social contribution may reduce gross wage costs and so increase employment. But dealing with such distortions is a matter more for labour market than tax policy. Absent such considerations, such a shift has effects only through its impacts on the real value of other sources and uses of funds. Labour market outcomes may be affected, for instance, if such a shift somehow moves the real burden of taxation to those out of the labour market, for example by reducing the real value of unemployment benefit. In practice, however, this can be hard to do: pressures to raise benefit levels in the face of commodity tax increases, for example, can be hard to resist. Most simulations find that the likely employment effects of such tax shifts are relatively minor: see, for example, the recent analysis of such proposals for France in Besson (2007). There is some evidence, however, that relatively heavy reliance on commodity rather than income taxation has been associated with faster growth (Kneller, Bleaney, and Gemmell (1999)).

These equivalences also imply that the choice across alternative tax mixes can be driven largely by considerations of administration and compliance. And these can plausibly point towards the simultaneous deployment of taxes that would be entirely equivalent if their enforcement were costless, as a means of diversifying enforcement risk. It may be optimal, for example, to deploy both a sales tax and a uniform wage tax levied by withholding: withholding is likely to be a good way of capturing a large part of taxpayers' wage income before they receive it, but perhaps relatively ineffective in reaching the self-employed; sales taxation may be a good way to capture a large part of taxpayers' incomes, including that of the hard-to-reach self-employed, when they spend it. This point is stressed by Boadway, Marchand, and Pestieau (1994), who show that when some income escapes tax it may be helpful to deploy a uniform commodity tax even when there would be no other reason to do so. (Such practical considerations also have implications for the form in which indirect taxes should be levied, notably for the relative merits of the VAT and retail sales taxes, to which we return later.)

Beyond these fairly general observations, the appropriate balance between direct and indirect taxation has received relatively little formal attention. Slemrod and Yitzhaki (1996) provide a conceptual framework for determining the appropriate tax mix when enforcement is imperfect, requiring that

the marginal social cost incurred by raising an additional £1 by some tax instrument, defined to include both administration and compliance costs as well as those from the implied distortion of economic activity, be the same for all such instruments. But it remains unclear, in practice, whether the balance currently struck in the UK is in any sense broadly appropriate.

4.2.2. The structure of indirect taxes

There are broadly two aspects of this issue: the way in which commodities (by which we mean both goods and services) should be taxed, and the rates at which they should be taxed.

How should commodities be taxed?

The starting point is the Diamond–Mirrlees (1971) production efficiency theorem: in the absence of externalities and non-competitive behaviour, and in the absence of restrictions on the distorting tax instruments that can be deployed or the ability to levy firm-specific taxes on pure profits,¹² a necessary feature of an optimal tax system—in the strong (Pareto) sense that if this were not the case then one could find a tax reform from which everyone would benefit—is that production decisions are left undistorted. Put simply, business transactions should not be taxed. The intuition is simply that any distortion of production decisions reduces aggregate output, which cannot be wise so long as there is some useful purpose to which that output could be put.

Strictly, these conditions are unlikely to be met in practice. Externalities are the most obvious (and conceptually easiest) source of failure: commodities generating external effects should on this account be taxed at the same rate whether used as intermediate goods or as final consumption (the damage done by CO₂ emissions being the same, for instance, whether fuelling industry or private travel). The other requirements of the Diamond–Mirrlees theorem are also inherently implausible as descriptions of reality, but—with this one exception—the precise consequences of their failure appear to be sufficiently circumstance-specific, and the political risks from allowing special treatment sufficiently troubling, for production efficiency to remain the best guiding principle for practical tax design.

¹² Pareto efficiency from a worldwide perspective may also involve production inefficiency in the allocation of resources across countries if there are constraints on the effective ability to make international lump sum transfers (Keen and Wildasin (2004)). But this generally calls for the use of trade taxes and so does not bear directly on the domestic indirect tax design on which we focus here.

The requirement of production efficiency proves to be a powerful one. As will be seen, it is both a key reason for the use of the VAT (or, equivalently—in principle—a retail sales tax) in preference to taxes that impact intermediate goods transactions, and also has strong implications for the improvement of the VAT both in the UK and in the EU more generally.

What structure for commodity tax rates?

Turning to the second aspect, the key theoretical question—given the equivalences noted above between a uniform proportionate tax on all elements on final consumption and the combination of a proportionate tax on wage income and other items—is that of whether it is desirable to tax some goods or services more heavily than others.¹³

The optimal design of such commodity tax structures has received substantial attention, from Ramsey (1927) on. One key insight—from the results of Besley and Jewitt (1990), Deaton and Stern (1986), and Atkinson and Stiglitz (1976)—is that the case for such rate differentiation is weaker the greater is the government's ability to pursue its distributional objectives by other means, including, but not only, by taxing (or subsidizing) income.¹⁴ The central point here is that differential commodity taxation is a very blunt instrument for the pursuit of equity objectives, with the zero-rating of food and children's clothing in the UK being a classic example.

Take food, for example. It is, indeed, the case that the less well-off spend a higher *proportion* of their income on food than do the better off. But this is not in itself a good reason—even on distributional grounds, leaving the need to raise revenue aside—for subjecting it to a differentially low rate of tax. This is for two reasons.

First, looking only at a snapshot of spending and income patterns in the population at any moment may be misleading given the variability of income

¹³ Here we confine our discussion to linear commodity taxes of the type that characterize the UK VAT and excises: charges that are simply proportional to purchases. It is possible to envisage more complex non-linear tax structures, as in Mirrlees (1978). Commodity taxes might, for example, vary more or less than proportionately with the quantity consumed. While this may be feasible in some cases—some countries levy non-linear charges on domestic electricity use, for instance—the possibilities of resale or splitting purchases commonly rule it out. Alternatively the tax paid might in principle depend on the history of consumption, or such characteristics as age: this could in principle be helpful in dealing with addiction problems that can arise with alcohol and tobacco (indeed regulatory restrictions on under-age drinking, for example, effectively implement an infinite tax rate).

¹⁴ More precisely, these results imply that the preference restrictions under which commodity taxation is unnecessary become weaker the more flexible is the instrument by which wage income can be taxed.

over a lifetime: those with low incomes now may be the young or elderly who will be, or have been, amongst the high income groups at other times. Put differently, a commodity tax looks regressive when assessed relative to current incomes in part because those with high incomes tend to have high savings, and so appear to escape the tax—but they will face it when they come to spend those savings. One way to address these issues is to relate food spending not to income in any period but to total spending, since the latter may be a better reflection of households' perceptions of their own long-run spending ability. Doing so, as Kay and Davis (1985) show for items zero-rated in the UK—and as subsequent studies have shown for a range of taxes on particular commodities¹⁵—tends greatly to dampen the apparent distributional case for tailoring commodity taxation to consumption patterns.

The second reason—perhaps potentially more persuasive to non-economists—is that even if the better off spend a smaller *proportion* of their current income on such items as food than do the less well-off, they are unlikely to spend a smaller *absolute* amount on them. If there were no other way of transferring resources to the poorest, setting a low tax rate on these items might be sensible policy. But it is unlikely to be so when, as in the UK, there are a range of other instruments—not only the income tax, but tax credits and benefits—that could be targeted more directly upon them. Kay and Davis (1985) and Hemming and Kay (1981) provided early illustrations of this point for the UK, the latter showing, for example, that the distributional impact of eliminating zero-rating could be very largely offset by cutting the standard rate of income tax and increasing the tax threshold. We revisit this simple but crucial insight, using more recent data, in Section 4.3.2—and show that it has lost no force over the years.

With sufficiently rich possibilities for income-related payments, the potential case for differential commodity taxation thus rests primarily on efficiency considerations. This is often taken to mean that tax rates should be especially high on commodities with especially low own-price elasticities of demand (meaning that demand for them is relatively insensitive to their price), the intuition behind this 'inverse-elasticity rule' being that in such cases a relatively high tax rate does little to reduce demand, and so causes little distortion of behaviour, while providing a robust revenue source. This prescription can prove dangerously misleading, however. One evident limitation is that it ignores the effect that increasing the tax on one commodity may have on the

¹⁵ See Poterba (1989) on alcohol and tobacco, and Hassett, Mathur, and Metcalf (2007) for a recent analysis of petrol/fuel taxes.

demands for—and hence distortions associated with, and revenue collected from—other commodities. Unless all cross-price elasticities are zero (so that the demand for each commodity is unaffected by the price of any other), it is quite possible that increasing the tax on some good with a low price elasticity, while increasing revenue from that item, may actually reduce total tax revenue and/or lead to more distortion rather than less. And tempting though it may be to ignore cross-price elasticities, doing so evidently becomes inherently less plausible at degrees of disaggregation over commodities sufficiently fine to make tax differentiation meaningful. A tax on beer, for instance, may have little effect on the generality of spending on other commodities but quite powerful effects on the demand for spirits, wine, or tobacco.

There is an even deeper sense, however, in which the inverse-elasticity rule is misleading. For the fundamental determinant of the case for a differentially high tax on some commodity is the way in which taxing it affects market labour supply. As a general principle, commodity taxes should be heaviest—all else equal—on those items that are most complementary with (or least substitutable for) leisure (by which is meant time not spent in the formal labour markets).¹⁶ Thus season tickets to watch football games should be taxed more heavily, for example, than season tickets to commute to work. Intuitively, taxing either particular commodities or income in general unavoidably discourages labour supply—recall the similarity between a uniform commodity tax and a tax on wage income noted earlier—and the only useful purpose that taxing different commodities can serve is in mitigating this distortion by making leisure less attractive than it otherwise would be: which means taxing most heavily those goods whose consumption tends to be associated with that of leisure. Conversely, if all commodities are equally complementary with leisure—a condition referred to as ‘weak separability’—then all should be taxed at the same (proportionate) rate. This result, due to Atkinson and Stiglitz (1976), is a central insight from the modern theory of indirect taxation.¹⁷ The inverse-elasticity rule is simply a special case

¹⁶ The relevant sense of ‘complementarity’ differs according to the capacity of the government to tax wage income: see for instance Atkinson and Stiglitz (1980). The situation in mind here, as seems most relevant in the UK, is that in which there are no restrictions on the shape of the tax schedule applied to such income.

¹⁷ Atkinson and Stiglitz (1976) derive the optimality of uniform—or, equivalently (since such a tax can be absorbed in the wage tax), no—commodity taxation for an atemporal world in which all individuals have the same preferences and an optimal, potentially non-linear wage tax is deployed. Kaplow (2006) and Laroque (2005) show that the conclusion continues to apply even if the wage tax is not optimal, while Boadway and Pestieau (2003) consider its robustness to various forms of heterogeneity. In a genuinely dynamic context, with earnings evolving stochastically and the possibility of conditioning tax payments on all observable actions, Golosov, Kocherlakota, and Tsyvinski (2003) show that weak separability continues to imply the optimality of uniform taxation.

of these more general observations, applicable in the further special case in which wage income is in itself untaxed (and, even more interesting, all individuals are the same): if, for example, an increase in the price of some good had little effect on the demand for it, and none on demand for other commodities, then taxing it more heavily would mean an increase in total expenditure that could only be paid for by working harder—and so that good must in a broad sense be a complement with leisure. But it is the link with labour supply, not the own-price elasticity, that is the fundamental concern.

Further perspectives on the structure of optimal commodity tax structures are provided by recent contributions which have stressed that what is regarded simply as ‘leisure’ in the standard optimal tax framework described above—time not spent in paid work—may also be put to productive use in household production. In such settings, Kleven, Richter, and Sørensen (2000) and Piggott and Whalley (2001) show, a case emerges for relatively low taxation of commodities that are close substitutes for such self-supply—which generally means such services as home improvement and repair—as a means of mitigating the unavoidable discouragement of market labour. If hiring a gardener is cheap, people might work overtime instead of mowing the lawn themselves. Thus Piggott and Whalley (2001), for example, report calculations suggesting that extension of the Canadian VAT to include such services may have been welfare-reducing. In somewhat similar spirit, a series of papers¹⁸ consider tax design when consuming commodities takes time (such as doing the laundry, or watching a DVD). Boadway and Gahvari (2006) show that it may then be optimal to tax less heavily those commodities for which time spent in consumption is pleasurable (being more like leisure in the everyday sense of relaxing) and tax more heavily those for which it is not (being more like work); and amongst the latter, those that require a lot of time to consume should be taxed more heavily than those that do not (again pointing to relatively light taxation of services that can readily be replicated at home). Thus DVDs should be taxed less heavily than ironing boards, which (since ironing is not only dull but time consuming) should be taxed more heavily than dishwashers (dull to fill, but quick to use). Intuitively, such a structure serves the same broad purpose of counteracting tax-induced disincentives to undertaking paid work, while also recognizing that it is better that any given amount of time not spent in paid work be passed doing pleasant

Though by no means of universal applicability, the result has thus survived as a benchmark for thinking on the uniformity issue.

¹⁸ Christiansen (1984) considers the issue briefly, Gahvari and Yang (1993) and Kleven (2004) do so in detail.

rather than unpleasant things (it making no direct difference for tax revenue, in particular, how that time is spent).

These models, it should be noted, are perfectly consistent with the standard framework described above, which simply takes the enjoyment of time not in paid work and the consumption of marketed commodities to be the objects of choice: one can conceive of individuals as having already decided how to use their time not in paid work before applying that framework. What these models do is add detail—relating for instance to the time intensity of consumption—on the relationship between time spent outside paid work and the household's final well-being, and show how that detail is reflected in the structure of commodity demands and labour supply, and hence in optimal tax rates. Quite how informative for policy design a focus on this additional structure will prove remains to be seen: we know relatively little, for instance, on whether time spent consuming particular goods is enjoyable or not (shopping is a pleasure for some, a chore for others), and, as stressed by Gahvari (2007), these approaches do not obviate the need to know price elasticities in order to calculate optimal tax rates. The important point for present purposes, however, is that the principles of optimal tax design set out above continue to apply. In particular, weak separability between market consumption goods and time in paid work remains sufficient—given the atemporal context and other ancillary assumptions—for uniform commodity taxation to be optimal.

Is it then the case that preferences do, in practice, appear to be weakly separable? The only detailed study of this key issue of which we are aware is that of Browning and Meghir (1991), who are able to reject weak separability with great confidence. Revisiting this issue with more recent UK data (Crawford, Keen, and Smith (2008)), we have arrived at the same conclusion. This finding needs to be interpreted with caution. It is possible, for instance, that commodity demands are linked with hours worked in the data not because of properties of tastes but as reflections of intertemporal considerations that are not modelled within the estimation framework used. Nevertheless, there emerges no clear presumption that one can simply tax all commodities at the same rate—and, by the same token, some presumption that well-designed commodity taxes have—in principle—some role to play.

The question then arises as to precisely what form such rate differentiation should take. Table 4.1 reports estimates from Crawford, Keen, and Smith (2008) of the complementarity of commodity demands with leisure for twenty broad groups of spending by UK households. Commodities found to be complements with leisure (in the sense of time not in paid work)—and so candidates for relatively heavy taxation or, more precisely,

Table 4.1. Estimates of commodity demand complementarities with leisure (Crawford, Keen, and Smith (2008))

	Impact on budget percentage share of an additional hour worked (t statistics in brackets)	
Bread and cereals	-0.024	(64.3)
Meat and fish	-0.060	(-49.2)
Dairy products	-0.045	(-66.6)
Tea and coffee	-0.008	(-29.5)
Fruit and vegetables	-0.037	(-52.8)
Other zero-rated foods	-0.020	(-28.1)
Standard-rated foods	-0.027	(-40.0)
Food eaten out	0.054	(38.5)
Beer	0.020	(13.3)
Wine and spirits	0.020	(21.2)
Tobacco	-0.026	(-16.6)
Domestic fuels	-0.049	(-30.6)
Household goods and services	0.064	(24.2)
Adult clothing	0.000	(-0.0)
Childrens' clothing	-0.006	(-8.7)
Petrol and diesel	0.046	(35.9)
Public transport	-0.006	(-6.2)
Leisure goods	0.019	(9.4)
Books and newspapers	-0.001	(-2.0)
Leisure services	0.086	(28.1)

Note: Results from demand system estimates reported by Crawford, Keen, and Smith (2008), based on household micro-data from 22 years of the UK Family Expenditure Survey (1978–99). The table shows the impact of an additional hour worked on the budget (percentage) share of each commodity group in household spending. Thus, for example, an additional hour worked reduces the (average) percentage of households' spending devoted to bread and cereals by 0.024 points. Commodities for which the coefficient is negative are leisure complements, and those for which the coefficient is positive are leisure substitutes. All coefficients except that on adult clothing are significantly different from zero, implying that weak separability is firmly rejected.

whose consumption should be most strongly discouraged by the indirect tax system—include most foodstuffs, domestic fuels, tobacco, children's clothing, and, perhaps surprisingly, public transport. This list includes some commodities which we might also wish to tax more heavily on externality grounds, but here we are considering only the implications of non-separability with labour supply. Complements with work, on the other hand—commodities whose consumption should be less discouraged by taxation—include alcoholic drink, food eaten out of the home, motor fuels, and leisure items (the last of these perhaps reflecting the use of such goods as substitutes for time in producing relaxation, in line with household production considerations discussed above). Again, the list of commodities

includes some associated with pollution and other externalities, which might have implications for tax rates tempering those arising from the issues of preference structure addressed here.

What this implies for the level and even sign of the various tax rates themselves is less clear-cut, however, since that depends not only on the sign but also on the magnitude of demand responses, and, moreover, also depends on patterns of cross-price effects. In Crawford, Keen, and Smith (2008) we demonstrate the significance of this observation; the pattern of relative tax rates that should be applied to different groups of goods differs quite markedly from the pattern of complementarity and substitutability shown by the estimated coefficients on hours worked in the commodity demand equations. For example, while domestic energy appears as a leisure complement—according perhaps with the intuition that those working less will spend more time at home and incur higher costs of heating and lighting—it appears to be a commodity that, on efficiency grounds, should actually be taxed less than others, once the cross-price effects are taken into account. Likewise, the estimated coefficients on the broad categories of ‘leisure goods’ and ‘household goods and services’ are both positive, indicating that both are leisure substitutes, but taking account of cross-price effects suggests that we would want to tax the former more heavily than other goods and services, and the latter at lower rates. This work is still some way from being able to provide a definitive answer to the question of the pattern of optimal commodity tax rates for the UK. Nevertheless, the significant role played by the pattern of cross-price effects does serve to stress that conventional wisdom and crude intuition can be unreliable guides in thinking about indirect tax design.

It is unclear, however, whether the social gain from moving to an optimally differentiated rate structure would be large. Since the effects on commodity demands of hours worked reported by Crawford, Keen, and Smith (2008) are small, there is reason to suppose that—even assuming we knew the optimal structure—it would not be. And that is important because against any such benefits must be set the evident practical costs of implementing differential rate structures. Applying the very large number of distinct rates to which theory might point would require ensuring, for instance, that commodities are not to an unacceptable degree misrepresented as liable to a lower rate than intended. Issues thus arise concerning the number of distinct rates to apply, and which commodities to combine for identical tax treatment.¹⁹ Further and distinct practical difficulties arise under the VAT, since multiple rates

¹⁹ The optimal partitioning of commodities into a fixed number of rate categories is analysed by Gordon (1989).

increase the reporting burden on traders (there being evidence that this effect is substantial)²⁰ and, even if honestly applied in themselves, multiple rates exacerbate control problems by increasing the likelihood that some traders (producing lightly taxed outputs from highly taxed inputs) will be entitled to refunds, an aspect of VAT implementation that all tax administrations have difficulty with (see, for instance, Harrison and Krelove (2005)). This is especially likely to be the case when—as with domestic zero-rating in the UK—the reduced rate is applied largely to final products. And in a wider context the potential effectiveness, in terms of both targeting and ease of implementation, of alternative spending measures would need to be weighed against any case for rate differentiation: if child care is a strong substitute for leisure, for instance, application of a reduced VAT rate may be inferior to public support for provision.

4.3. THE VALUE ADDED TAX

The rise of the VAT has been one of the central developments in UK tax policy since the Meade Report. This chapter reviews key aspects of the design of the VAT and associated implementation issues. As background, Box 4.1 provides a quick primer on the main features of the VAT, and distinctive terminology.

Box 4.1. A VAT Primer

Value added tax (VAT) is levied on the sale of goods and services by registered businesses (those with annual turnover above some threshold level—the choice of which is discussed in Section 4.3.2—or who choose to register voluntarily). It applies to all sales, whether to private consumers or other businesses (in contrast to the retail sales taxes levied in the US, for example, which aim to tax sales to final consumers only).

Under the ‘invoice-credit’ form of the VAT—and all national level VATs are of this form, except that in Japan²¹—registered businesses offset the VAT they

(*cont.*)

²⁰ Cnossen (2003) reports that firms in the UK subject to more than one output VAT rate have more than twice the compliance costs of those subject only to one.

²¹ The VAT in Japan retains elements of being levied on a ‘subtraction basis’: that is, on the book difference between total sales inputs. There also exist a number of subnational VATs levied on other than an invoice-credit basis. The Italian IRAP (*imposta regionale sulle attività produttive*), for instance, is an origin-based (no remission for exports, or taxation of imports) subtraction method VAT, while both Michigan and New Jersey have implemented what are essentially VATs levied on origin additions basis (that is, applied to the sum of wages and profits). Bird (2000) makes a general case for such taxes as a suitable source of subnational revenue, an issue not considered here.

Box 4.1. (cont.)

have been charged on their purchases ('input VAT') against the liability ('output VAT') on their sales, remitting only the net amount due. The result, if this chain of output tax and input credit remains unbroken, is that no net revenue is collected from the taxation of intermediate goods sales (business-to-business or 'B2B' sales), so that the ultimate base of the tax is final consumption (in the sense, more precisely, of sales other than to registered businesses).

For example, consider a simple chain of production consisting of two firms. Firm F makes sales of £30,000 to final consumers and no B2B sales. In the course of production, it uses inputs purchased from Firm Y at a cost of £10,000 plus VAT. Firm Y makes no sales to consumers and uses no taxed inputs; its entire £10,000 output is sold to firm F.

If the sales of both firms are subject to VAT at the (tax-exclusive) rate of 17.5% Firm Y will be liable for £1,750 (= 17.5% of £10,000) in VAT on its sales to F. Firm F will be liable for output VAT of £5,250 (= 17.5% of £30,000) on its sales of £30,000, but can offset against this the £1,750 tax paid on its own purchases, giving a net VAT liability of £3,500. Total VAT remitted by the two firms taken together is £1,750 + £3,500 = £5,250, which is equivalent to 17.5% of the (tax-exclusive) value of the sales made to final consumers.

'Zero-rating' means that the seller charges a VAT rate of zero on its sales but is still entitled to credit for the input VAT paid, so that no VAT remains: the Australian term 'VAT-free' is perhaps more descriptive. So long as some input VAT has been paid, the business will on this account be due a refund.

For example, if the sales of Firm F in the above example are zero-rated while Firm Y's remain standard rated, Firm F would charge no VAT on its sales and would be due a refund of the £1,750 VAT paid on its purchased inputs. Total VAT collected is zero.

The universal practice is to zero-rate exports and fully subject all imports to the VAT. This is as a way—though, as discussed later, not the only one—of ensuring that the VAT applies only to domestic consumption, consistent with the 'destination principle' discussed, and contrasted with the origin principle (of taxation by place of production), in Section 4.5.1. Quite where the place of consumption is, however—and so what 'export' and 'import' mean—is generally fairly clear-cut for goods. But it is much less so for services, so that 'place of supply' rules, determining where tax should be charged, are especially problematic in that context: this issue is discussed in Section 4.5.3.

'Exemption' means that sales are not subject to VAT but, in contrast to zero-rating, the firm does not have the right to reclaim the VAT paid on its inputs: the Australian term 'input-taxed' is more telling. The input VAT thus 'sticks', and the VAT acquires elements of a tax on production rather than consumption.

If Firm F in the example is selling VAT-exempt goods, it would charge no VAT on its sales but would not be able to reclaim the £1,750 VAT paid on the inputs purchased from Firm Y. Firm F's sales would thus indirectly bear some VAT, in the form of the VAT charged earlier on the inputs purchased from Firm Y. Revenue is lower than it would be if F were taxed, by the amount of the tax due on its own value added.

If instead it is Firm Y, selling the intermediate good, that is exempt, there is no effect on total revenue: Firm Y itself then charges no output VAT (and by assumption, pays no input VAT either) so that F simply has no input VAT to credit against its output VAT of £5,250. This is the self-correcting feature of the VAT noted in the text. Revenue would be affected, however, if Firm Y had paid some tax on its inputs: exemption means it would not be able to recover that VAT, and since the output VAT charged by Firm F would be unchanged—or, if anything, increased, as Y increases the price it charges F in order to meet its increased input costs—the total revenue collected would actually increase.

4.3.1. The nature and strengths of the VAT

The VAT has now been adopted by more than 130 countries, including all members of the OECD other than the US. While each country had its own reason for adoption, the main reason in the UK was simple: it was (and is) a precondition for entry into the EU. And the central reason for the EU's insistence on the VAT is to provide a transparent means of ensuring that exports are relieved of indirect taxation (or subsidy) and imports brought into tax on an even footing.²² A well-functioning VAT—with unbroken chain of crediting and refund—is also consistent with the theoretical preference, described above, for taxing commodities (other than those generating production externalities) only on their final sale to consumers, so as to preserve production efficiency. But, as we see shortly, there are other ways of achieving the same effect. What then is the particular merit of the VAT?

This question is of more than historical interest. Somewhat ironically, the zero-rating of exports under the VAT that has proved so effective in facilitating trade between member states by removing indirect tax as commodities pass between them has also proved highly problematic when the later stage of economic integration—the removal of fiscal frontiers—was reached. And these difficulties have called into question fundamental design features of

²² Södersten (1999) provides an interesting historical account of the adoption of the VAT in Europe.

(and even the wisdom of retaining) the VAT, as discussed in Sections 4.3.3 and 4.5.2.

One way of pursuing production efficiency and the taxation only of domestic consumption is by exploiting the equivalence between the VAT and other combinations of instruments. For instance, a destination-based VAT, levied on the difference between the domestic sales S of every firm and its purchases of material inputs (the latter comprising investment spending I and other material purchases P) is equivalent to a tax on the firm's wage bill W combined with another on cash flow earnings $S - W - I - P$. (More precisely, the final part of this equivalence is with a destination-based, R-form cash flow tax: see Auerbach, Devereux, and Simpson in Chapter 9.)²³ Thus a VAT would be unnecessary if there were no constraints on the ability to tax labour income or destination-based cash flow profit. Conversely, such a cash flow tax might not be needed if a reasonably functioning VAT and wage taxation were available. It may nevertheless be optimal to deploy all instruments, including both a VAT and a cash flow profit tax, if—in similar spirit to the analysis of Boadway, Marchand, and Pestieau (1994) mentioned above—evasion possibilities differ between them (Gordon and Nielson (1997)).

The second way of pursuing these same two objectives would be deploying a retail sales tax (RST), which levies tax only at the point of final sale. (In the first example of Box 4.1, for instance, the same final effect could be achieved by taxing, at the same rate of 17.5%, only the sales of Firm Y.)

Since a VAT and RST are economically equivalent when both function perfectly, the choice between them must rest on the differential challenges for administration and compliance that they imply, and in the opportunities and incentives for evasion that they create. And here the key difference²⁴ lies in the 'fractional' nature of the VAT, with tax in principle collected at each stage of production, compared with the single-stage nature of the RST.²⁵ Put simply, if the final sale of some commodity escapes tax, no revenue is collected

²³ Or to an origin-based cash flow—in turn equivalent in present value, if the tax rate is unchanging, to a tax on pure profits combined with a levy on the initial capital stock (hence the notion of consumption taxation as bearing on 'old' capital)—and a subsidy to inflows from abroad. The flat tax of Hall and Rabushka (1983) is equivalent to an origin-based VAT, collected by the subtraction method, supplemented by the application of a single marginal rate to labour income (with a non-refundable tax credit).

²⁴ There are also likely to be differences in the number of firms that need to be subject to the tax, but it is not clear a priori whether this will be greater or less under the VAT: it will be greater in that all types of businesses, not just retailers, are brought into the tax, but less to the extent that the fractional nature of the VAT means that many small firms can be excluded from the system, by means of a registration threshold, without major revenue losses (a point explored further).

²⁵ This does not mean that revenue accrues to the government sooner under the VAT, since—if the system is working as intended—the same tax that is collected from the seller of some item used as an intermediate input is also (and quite possibly simultaneously) credited to the buyer.

under an RST; under the VAT, on the other hand, all that is lost is the tax on value added at that final stage, since tax (in principle) will have been collected on the final seller's purchases. This means, of course, that in such circumstances the VAT does indeed tax intermediate transactions, which runs foul of the presumption for production efficiency created by the Diamond–Mirrlees theorem discussed above. Recall, however, that this result presumes that all final sales can be taxed. When they cannot, then, as noted by Newbery (1986), it is generally desirable to tax inputs as a surrogate for the missing output tax. And it is here that the VAT may do a particularly good job, since the unrelieved input tax is likely to arise precisely where, for some reason, output tax is not charged.

This distinctive fractional nature of the VAT is a feature stressed by its advocates, and a prime consideration in arguing, along the pragmatic lines above, for substantial reliance on the VAT within the tax mix. But while influential in practice, the strength of this case for the VAT has received little analytical or empirical attention. Nor has the possibility of exploiting it still further. Keen (2008) notes, for example, that a uniform VAT taxes informal sector purchases from VAT-compliant firms at the same rate as formal sector final sales, when one might in principle want to tax the former more heavily—which could be done by the use of some creditable withholding tax or other supplement on sales likely to be to informal sector operators. Such taxes are indeed quite common in emerging market and developing countries, but less so in developed. That is surprising, in that they are more likely to be able to implement the crediting needed to prevent such taxes becoming an additional burden on legitimate traders, but presumably reflects a lesser concern with informality issues. Indeed, the trend in the EU (though not in the UK, as noted later) has been to set low output tax rates on items potentially subject to informality, not high input rates. This aspect of indirect tax policy, in any event, is pursued no further here.

Building on the fractional nature of the VAT, some suggest that—in its invoice-credit form—the VAT is ‘self-enforcing’ in the sense that each trader has an incentive to ensure that their suppliers have themselves properly paid output VAT, in order that they themselves can then claim an appropriate credit.²⁶ There is an element of truth in this. Certainly businesses registered

²⁶ At the opposite extreme, it is also sometimes claimed that the ability to cross-check invoices—verifying that every credit claim is matched by some payment of output tax—can make the VAT especially abuse-proof. While there is again an element of truth in this—invoices do indeed provide a useful trail for VAT auditors—this too can be overstated. Even with the developments in information technology in recent years, complete cross-checking of invoices remains, at least for the present, effectively impossible.

for VAT can gain nothing (beyond perhaps some cash flow advantage) by purchasing inputs on an untaxed basis, since they are in any event able to claim credit or refund for any tax so paid. Indeed, there is a strict advantage in purchasing from VAT-registered businesses, since unregistered businesses will be unable to reclaim the VAT they themselves have been charged on their inputs, and so may charge a higher output price. Thus traders selling to other businesses may indeed wish to register to charge the VAT even if their annual turnover is below the threshold at which VAT registration is mandatory, and arrangements for such voluntary registration are a key part of any well-designed VAT. There is also a sense in which the VAT is self-correcting: if for some reason output VAT is not charged at some intermediate stage of production, proper payment by a trader later in the chain will replace that missed tax (since they then charge the proper output tax themselves, with their input tax credit reduced by precisely the amount of the output VAT that their suppliers should have charged but did not).

But the strength of these intrinsic features of the VAT should not be over-stated. It remains the case that those selling to private individuals and businesses not registered for VAT have similar incentives to sell without tax as under an RST, although muted to the extent that they have paid VAT on their own purchases. That is, while the VAT contains incentives for the formation of 'good' chains of compliant traders, it also contains incentives for 'bad' chains of non-compliance: de Paula and Scheinkman (2006) argue that the latter has been an important feature in Brazil, for example. Moreover, while registered traders have an incentive to ensure that their suppliers provide them with an invoice that the authorities will accept as establishing their right to refund or credit of input VAT, they have no incentive—unless specific requirements are imposed—to ensure that this tax has actually been remitted to the tax authorities. Furthermore, the credit and refund mechanism of the VAT creates its own opportunities for fraud, as we also discuss later.

Whether the VAT has proved to be a particularly effective form of taxation is, or should be, ultimately to some degree an empirical matter. But the question has received little attention. Such evidence as there is on the performance of the VAT (in Keen and Lockwood (2006, 2008)) suggests that it has indeed proved an effective form of tax: countries with a VAT, especially higher income countries, tend to have higher tax ratios—modestly, but significantly so—than those without, which is as one would expect if the VAT has had the effect of reducing the marginal distortionary and other costs of mobilizing tax revenue.

4.3.2. Key design features of the UK VAT

The design of a VAT requires a wide range of decisions as to its rates and coverage. This section reviews the central choices made in the UK. Those choices are constrained by a variety of EU rules on the common VAT: some key features of these, mostly as they relate to rate structure, are summarized in Box 4.2, and others will be set out and discussed below.

Box 4.2. EU rules on the value added tax

The common VAT rules, with the 1977 Sixth VAT Directive at their core and recently consolidated in directive 2006/112/EC, establish broad commonality in definitions on such core matters as taxable person, taxable event, and place of supply. Beyond this, particular rules apply to permissible rates and exemptions.

Rates

With the adoption of the internal market in 1992, and fear that the removal of internal fiscal controls would put downward pressure on VAT rates, member states agreed not to set their standard VAT rate lower than 15% (a provision recently extended to 2010). Member states may set no more than two reduced rates: at no less than 5% on a positive listing of ('Annex III') items and, on condition that competition is not distorted—in effect restricting the application to final sales—on use of electricity, natural gas, and district heating.

As transitional provisions—until the adoption of a 'definitive' regime for the taxation of intra-Community trade, for which no date is set—member states are allowed to retain a variety of otherwise prohibited measures that were in place at the start of 1991: 'super-reduced' (including zero) rates applied for social reasons—by which means the domestic zero-rating in the UK survives—rates of less than 5% on Annex III items, and a reduced rate of no less than 12% (the 'parking rate') on Non-Annex III items. Many of the new EU members, of course, did not have a VAT in place at this date, and so are unable to benefit from this provision: they are required to be fully compliant by 2010 at the latest: Malta, for example, must by then remove its zero-rating of foodstuffs and pharmaceuticals, though there is no similarly unconditional obligation on the UK.

Member states may also apply (until the end of 2010) a reduced rate to no more than three specified labour-intensive services (such as hairdressing, domestic service, and the renovation and repair of private dwellings).

(cont.)

Box 4.2. (*cont.*)**Exemptions**

EU rules mandate several exemptions, including for medical care, education, social welfare and cultural activities, financial services, and the letting of immovable property (with member states allowed, in the last two cases to provide taxpayers an option to be taxed—which the UK currently does for letting but not for financial services). Supplies by public bodies are also exempt (strictly, are outside the scope of the VAT) so long as this does not distort competition; a few member states provide for the refund of input VAT to such bodies in some cases, as the UK does automatically for local authorities, the government of Northern Ireland, and some other named bodies, and on a discretionary basis for others.

By way of background, Table 4.2 provides comparative information on key design characteristics for all OECD countries (other than the US, which of course does not have a VAT).

Rate structure

At 17.5%, the standard rate is around the OECD average and comfortably above the EU minimum of 15%. Indeed, that minimum is currently binding for only one member state, Luxembourg. It is theoretically conceivable that a lower rate in Luxembourg would have induced the UK to set a lower rate (perhaps because other and larger countries might have been induced to do so), but this seems unlikely: certainly the most direct form of interaction between national indirect tax systems—smuggling and cross-border shopping—seems to be fairly limited in relation to the VAT.²⁷ Nor, by the same token, is there any evidence that the UK is constrained from further increasing the standard rate in a way that further increasing the EU minimum would ease.

More striking than the level of the standard rate is that the UK VAT is marked by very substantial rate differentiation: domestic zero-rating is extensive (including notably most foodstuffs, children's clothing, and residential construction), and a reduced rate of 5% is also applied to domestic power and energy, and a range of other items (such as contraceptives, certain

²⁷ The main exception to this appears to be that between Germany and Denmark, with a differential in the standard rate that was 9 percentage points until the 2007 increase in Germany. Gordon and Nielson (1997) estimate that this caused the former to lose only around 0.8% of its VAT revenue from cross-border shopping, though this was before the removal of fiscal frontiers and may have increased since.

Table 4.2. VAT rates, revenues and C-efficiency in the OECD¹

	Standard rate (percent)	Reduced rates ² (percent)	Threshold ³ (in USD)	C-efficiency ⁴ (percent, 2005)
Australia	10	Zero	35,500	57
Austria	20	10, 12	34,400	60
Belgium	21	Zero, 6, 12	6,400	50
Canada	6	Zero	25,000	52
Czech Republic	19	Zero, 5	70,000	59
Denmark	25	Zero	5,800	62
Finland	22	Zero, 8, 17	9,700	61
France	19.6	2.1, 5.5	87,500	51
Germany	19	7	20,000	54
Greece	19	4.5, 9.0	11,500	46
Hungary	20	5	30,800	49
Iceland	24.5	Zero, 7	4,800	62
Ireland	21	Zero, 4.8, 13.5	63,000	68
Italy	20	Zero, 4, 10	8,000	41
Japan	5	–	80,600	72
Korea	10	Zero	None	71
Luxembourg	15	3, 6, 12	11,500	81
Mexico	15	Zero	None	33
Netherlands	19	6	2,200	61
New Zealand	12	Zero	26,300	105
Norway	25	Zero, 8, 14	5,600	58
Poland	22	Zero, 7	20,900	48
Portugal	21	5, 12	11,500	48
Slovak Republic	19	–	86,700	53
Spain	16	4, 7	None	56
Sweden	25	Zero, 6, 12	None	55
Switzerland	7.6	Zero, 2.4, 3.6	44,100	76
Turkey	18	1,8	None ⁵	53
United Kingdom	17.5	Zero, 5	93,600	49
Unweighted Average	17.7			58

¹ As at 1 January 2007 except for C-efficiency.

² 'Zero' indicates zero-rating of some domestic sales.

³ This is the general threshold. Some countries apply a lower threshold to services.

⁴ VAT revenue divided by the product of the standard rate and final consumption expenditure less VAT revenue.

⁵ Small retailers and taxpayers taxed on a lump sum basis or exempt from personal income tax, and farmers, are not required to register. Personal income tax thresholds and conditions apply to VAT.

Source: OECD (2008). Rates applying only in particular regions—such as the reduced rate applied in border regions by Mexico, and to overseas departments by France—are excluded.

energy-saving products, and children's car seats). HMRC put the revenue cost in 2007–08—the additional VAT that would be raised if tax were instead charged at the standard rate (ignoring any behavioural response)—at around £29 billion for domestic zero-rating (around 40% of this being from food, and around 30% from new dwellings) and £3 billion for the reduced rate

(almost all of this being for domestic fuel and power). Exemptions are reckoned to cost another £12 billion (the largest item, accounting for over one-third of this, being financial services).²⁸ This compares to total VAT revenue of around £80 billion. It all means that C-efficiency in the UK—the ratio of VAT revenues to the product of the standard rate and private consumption, which would be 100% for a textbook VAT levied at a uniform rate on all consumption (and comes close to that in New Zealand)²⁹—is very low by OECD standards.³⁰

It has been recognized for more than twenty years that the policy rationale for the zero-rating of food and children's clothing is extremely weak. Theoretical and empirical developments over the last twenty years, as reviewed above, have only confirmed these doubts. Our own empirical results, reported in Table 4.1 above, do not provide any suggestion that differentially low rates on these items—or indeed any others—is warranted, given the availability of fairly flexible earnings-related instruments: most foodstuffs appear to be complements with leisure. The survival of zero-rating of food and children's clothing appears simply to reflect politicians' doubts of their ability to explain why a package involving its removal need not have a regressive impact: as discussed by Alt, Preston, and Sibieta in Chapter 13, the commitment to this zero-rating appears to have become a signal of commitment to at least a moderately pro-poor policy.

The rationale for the reduced rate is also far from clear: to the extent that its original purpose is to mitigate 'fuel poverty', such measures as the additional winter allowance for pensioners provide reasonably well-targeted relief (though a strong case could be made for an element of income-relation). Indeed, there is some perversity in applying the reduced rate to both energy use and purchases of some energy-saving materials. The deeper issue here is the proper design of energy taxes, and similar mechanisms, such as cap-and-trade systems along the lines of the EU Emissions Trading Scheme, to address environmental and other non-revenue concerns such as supply security (see Fullerton, Leicester, and Smith in Chapter 5). With these in place, there would be little case for differential treatment of final energy use; and without them, as at present, it is hard to make a case for rates on final use that are actually lower. Experience with this lower rate, into which an increasing and

²⁸ Table 1.5 at <http://www.hmrc.gov.uk/stats/tax_expenditures/menu.htm>.

²⁹ The use and limitations of C-efficiency measures are discussed in Ebrill et al. (2001) and OECD (2008), which relabels it the 'VAT revenue ratio'.

³⁰ The OECD calculations reflect a figure for VAT revenue substantially above that reported by HMRC: £83 billion for 2005 compared to £73 billion, the difference largely reflecting the treatment of refunding of VAT to public bodies.

diverse number of items have been moved, also illustrates the further general experience that preferential treatment, once granted, tends to spread, and that in its doing so the wider coherence of the tax system suffers. Quite why parents should pay no VAT when they buy clothes for their children but 5% when they buy them a car seat is by no means clear.

Indeed, in a broader sense the extensive rate differentiation still found in the VATs of EU members is coming to look increasingly quaint. Most new VATs adopted in recent years have a single rate: not only in Australia and New Zealand, but also in developing countries, where the case for a single rate is actually weaker (because of a lesser ability to deal with the distributional implications through other instruments). Others, it seems, have learnt lessons from the EU experience that EU members themselves have not.

Experience also shows, however, that moving towards a uniform VAT rate structure is not easy once differentiation has been admitted (though there are some success stories: the Slovak Republic, for instance, unified its two distinct VAT rates in 2004³¹). What will evidently be needed, to overcome distributional concerns, is the packaging of movement to uniformity with other reforms intended to neutralize the equity impact (just as in the Slovak Republic the VAT reform was part of a much wider package, including increased generosity of in-work support).

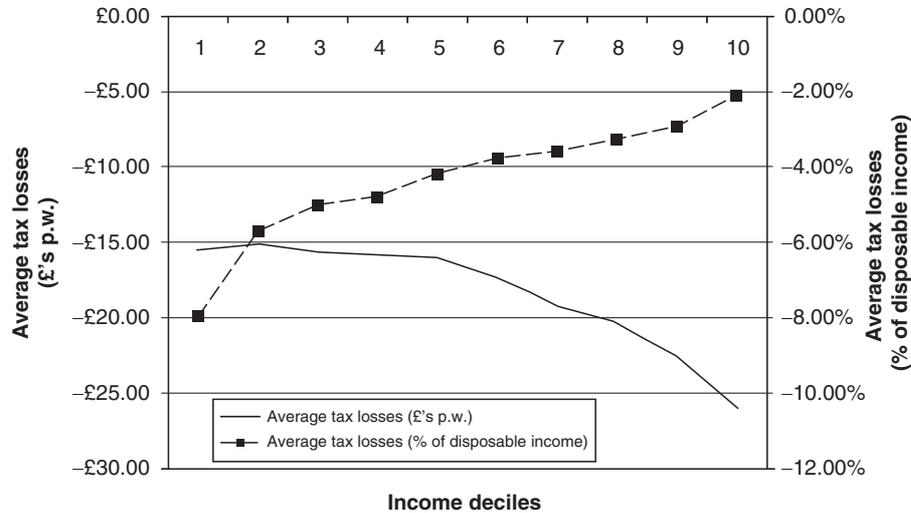
What might such a reform package for the UK look like? By way of illustration, consider applying the current standard rate of 17.5% to all commodities except housing and items currently exempt from VAT³²—eliminating, that is, the current reduced and zero-rate—and combining this with a 15% increase in all income support, income-based jobseeker's allowance and tax credit rates, and in the associated housing benefit and council tax benefit thresholds.³³

Figure 4.1 shows the implied change in VAT payments across the income distribution. As one would expect, the largest absolute financial losses (solid line, left scale) are found amongst richer households simply because they spend more, on average, on reduced- and zero-rated goods than do less well-off households. However, the increased VAT payments are a larger proportion of disposable income (right-hand scale) at lower income levels, since the less

³¹ But, it must be admitted, also reintroduced a reduced rate in 2007.

³² This is in order to respect exemptions currently required under EU rules (though simulations not reported here show that these too could be removed as part of a rate-unifying reform with the same features as that in the text, protecting the poorest and increasing total revenue). The Insurance Premium Tax is kept at its current rate of 5%.

³³ It is assumed that the incidence of the VAT reform is fully on retail prices and that of the compensating measures fully on net incomes.



Note: Income decile groups (1=poorest, 10=richest) are derived by dividing all households into ten equal-sized groups according to disposable income adjusted for household size and composition using the McClements equivalence scale.

Source: IFS calculations using the IFS tax and benefit microsimulation model, TAXBEN, run on updated data from the 2005–06 Expenditure and Food Survey.

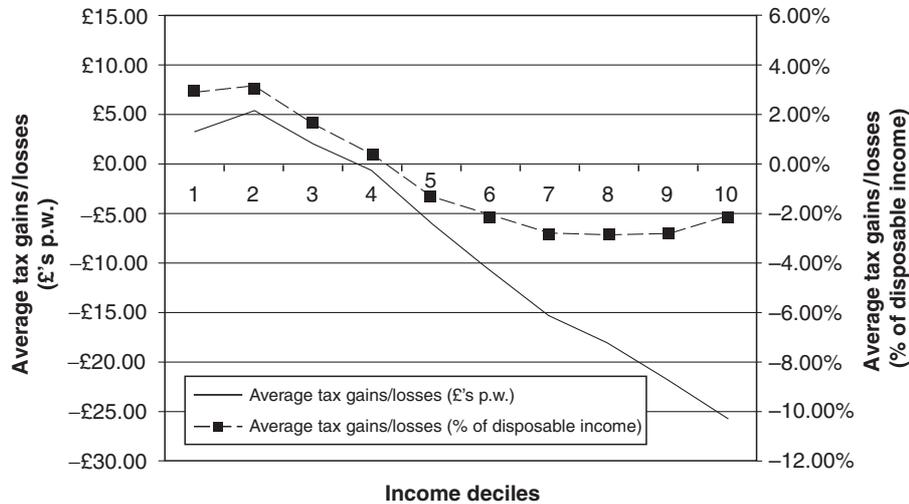
Figure 4.1. Distributional impact of unifying VAT rates at 17.5%

well-off spend a greater proportion of their incomes on food and other items currently taxed preferentially.

When the compensating measures are combined with the VAT reform, however, the net impact across the income distribution is quite different, as shown in Figure 4.2. Households in the lowest income decile group, for example, gain on average about £2.50 per week whereas those at the top lose around £25 per week. It is not quite the case that those on lower incomes always gain more than those on higher—those in the second decile group gain more than those in the first—but the broad distributional impact is clear-cut: the poorest three-tenths gain on average while those on higher incomes tend to lose.

Importantly, the package implies a net increase in tax revenue: the VAT reform raises around £23 billion while the compensating package costs only about £12 billion. If so desired, this £11 billion saving could be used to finance further pro-poor measures.

This package is intended to be essentially illustrative. Behind the aggregate effects shown in the figures are effects that vary across household type. The component of the package relating to child tax credit, for example, means that those with children tend to gain more than those without. There are



Note: Income decile groups (1=poorest, 10=richest) are derived by dividing all households into ten equal-sized groups according to disposable income adjusted for household size and composition using the McClements equivalence scale. Uniform VAT rate excludes housing and non-insurance financial services: see text for details of the reform.

Source: IFS calculations using the IFS tax and benefit microsimulation model, TAXBEN, run on updated data from the 2005–06 Expenditure and Food Survey.

Figure 4.2. Distributional impact of the reform package

also potential impacts on labour supply that are not included in the modelling here but would need to be considered in developing a fully formulated proposal. The key point, however, is that it is not difficult to find ways of compensating the losers from VAT rate unification, arriving at a package of reforms that improve both the progressivity and efficiency (including in terms of administration and compliance) of the overall tax system—and in the process provide revenue that could finance further desirable reforms.³⁴

Two other aspects of the VAT rate structure merit comment. First, the UK has notably not taken up the option under current EU rules (until 2010) to apply reduced rates to specified labour-intensive activities. Such rate reductions can be defended in principle on optimal tax grounds in terms of the arguments cited above for relatively low taxation of services readily self-supplied by consumers. Against this, of course, they carry the administration and compliance burden associated with increasing the extent of rate differentiation. And an assessment by the Commission of the European Communities (2003) is sceptical as to the effectiveness of such provisions in increasing

³⁴ This reform would increase the Retail Prices Index by around 3.5%. Note that this is a one-off jump in the price level, not a permanent change in the rate of inflation.

formal sector employment. It is difficult to make a compelling case that the UK has made a mistake in not taking up this option.

The second striking feature is that the UK is alone among OECD countries in zero-rating construction and sale of residential properties (residential letting is exempt). This creates a potentially significant distortion towards the consumption of housing services, and especially owner-occupation, of a kind that the UK, to its credit, has carefully counteracted by eliminating mortgage interest deductibility. The underlying difficulty, of course, is that the consumption of housing services is not directly observable. The same problem arises to some degree with all durable goods, of course, and the standard if imperfect approach is simply to charge VAT at the first sale: since the price of a durable reflects the present value of the stream of services it is expected to yield, this amounts to prepayment of VAT on those services. The same approach can be adopted for housing, fully taxing the first sale of residential properties (as is already done in the UK for commercial properties).³⁵ Doing so in the UK would be permissible under EU rules, and is already done in Belgium and France, for example. The increase in (VAT-inclusive) house prices this would imply would create difficulties: first-time buyers would suffer, while owners of existing properties would enjoy a windfall gain. The former could be addressed by providing a (perhaps temporary) subsidy to new buyers, as Australia did when introducing its VAT in 2000. The latter would be eliminated (and tax receipts increased) if the tax were instead levied on all house sales after some implementation date (since the same prepayment logic applies to the sale of old but previously untaxed houses).

The political resistance to moving away from the present extensive domestic zero-rating has proved formidable, to such an extent that it is hardly on the policy agenda. Moving instead to exemption might seem a tempting compromise. This is not only untransparent, however, but may not even reduce implementation costs (given the difficulties faced and posed by partially exempt traders). Moreover, as argued next, strengthening the UK VAT requires less exempting, not more. More appealing would be to tax these items at below the standard rate (for some period), perhaps moving them to the present reduced rate. Consideration could also be given to earmarking the proceeds, to pay not only for the compensation but some other popular items (including tax reductions or social security increases). Presenting VAT increases in this way has been considered in France and Japan, for instance,

³⁵ Commercial letting is exempt, but with an option to tax. Thus arrangements for commercial property properly allow for the full operation of the VAT chain.

and recently done with success in Ghana (where resistance to the VAT has historically been particularly high). Against any political advantage of doing so must be weighed, however, the implied inflexibility in the use of funds—the absence of earmarking is one of the wider (if often unremarked) merits of the current UK tax system that should not be taken lightly.

Exemptions

Exemptions under the UK VAT are extensive, reflecting the EU law summarized in Box 4.2 above and including most of what the OECD (2004) calls ‘standard’ exemptions.³⁶

Any exemption is anathema to the logic of the VAT, since it inherently breaks the chain of credit and refund, leading to an element of production taxation. The consequent distortions can take a number of forms. These include an incentive to self-supply (banks providing security services in-house, for example, so as to avoid unrecovered VAT if hiring the services from others), distortions of competition (as financial institutions across the EU face different input costs as a consequence of being charged different rates on their inputs, this effect also cascading into the costs of business using those services; and as exempt public services compete with taxable ones provided by the private sector), and a bias towards imports (since these will have been freed of VAT in the country of export). They also create additional administration and compliance burdens (and opportunities for tax planning) through the need to allocate input VAT between taxable and exempt outputs (credit being available for the former but not the latter) for producers selling both. As Bird and Gendron (2007) note, there is little firm evidence on the quantitative importance of these distortions. The rebating of input VAT to local authorities and similar bodies in the UK may limit those from the exemption of public bodies (the revenue impact of which may also be fairly modest, since full taxation means one part of government paying taxes to another). There is though every sign that these are a serious concern in the financial sector, and certainly its sheer (and increasing) size—Zee (2006) reports that it accounts for 30% of GDP—suggests that the risk of distortion deserves serious attention.

These standard exemptions are to a large degree the hallmark of ‘old’ VATs—such as that laid down for the EU has become. They reflect outdated

³⁶ These include postal services; hospital and medical care; dental care; charitable work; education; non-commercial activities of non-profit making organizations; sporting services; cultural services (except radio and television broadcasting); insurance and reinsurance; letting of immovable property; financial services; betting, lotteries, and gambling; supply of land and buildings; and certain fund-raising events.

presumptions on the arrangement of economic activity—a vision of a large public sector largely insulated from competition with private operators—and, to a large degree, a failure to recognize that the logic of the VAT can be applied to these items. Aujean, Jenkins, and Poddar (1999), for instance, explain how the public sector can be brought fully into the VAT system, and New Zealand in particular now largely does this (even charging VAT on public subsidies used to provide goods and services without charge). There is a very strong case, as argued by Cnossen (2003 and Commentary on this chapter) for eliminating many of these exemptions. These are not matters, however, on which the UK can act unilaterally (or on which we have much to add), so we simply refer the reader to the cited sources, and the general treatments in Ebrill et al. (2001) and Bird and Gendron (2007), for discussion of how the common rules might usefully be restructured.

There is though one aspect of such importance, for reasons noted above, and so central to this wider review of the tax system, as to require some comment. This is the VAT treatment of financial services.³⁷ Those charged for in the form of a fee raise no conceptual difficulty—they can and should be fully taxed, though in practice it seems few do so as a general principle (South Africa being an exception).³⁸ The difficulty arises for those charged for as a margin, between bank's borrowing and lending rates, for example. The choice of exemption in the EU³⁹ to a large degree reflected an inability to see how (continuing with that example) that margin could be broken down into value provided to depositor and borrower, as appeared to be necessary if only that part accruing to any households involved in the transaction is to be taxed. Following Poddar and English (1997), however, it is now understood that such transactions can be properly brought into an invoice-credit VAT by treating all cash inflows to those providing financial services (including principal amounts) as reflecting VAT-able sales to customers, and treating all cash outflows (including loans made) as the purchase of inputs carrying

³⁷ Some have argued that financial services should not be taxed on the grounds that they do not represent consumption since they do not enter the direct utility function. This now seems largely discredited: no one knows what the arguments of final utility are. Boadway and Keen (2003) review the theoretical arguments in this area.

³⁸ One reason for this, as noted by Zee (2006), may be concern that different approaches to the taxation of financial services paid for by fee and those paid for implicitly in a spread is the risk of financial institutions arbitraging between the two. McCann and Edgar (2003) argue that the scope for this is likely to be limited in practice, though the issue remains open.

³⁹ Many other approaches have been adopted: see for instance Bird and Gendron (2007) and Zee (2006). Some—notably using fixed coefficients to simplify the allocation of inputs to taxed and exempt activities (as, in somewhat different ways, do Australia and Singapore)—seem likely to be superior to exemption. Other than outright zero-rating, however (which to large degree gives up on the basic objective of taxing financial services entering final consumption), only the cash flow method about to be described fits within an invoice-credit VAT.

creditable VAT. Intuitively, this ensures that all flows involving registered traders are subject to a tax that is fully credited; the revenue that remains thus derives from transactions involving households.

This approach operates straightforwardly for some financial transactions, not only those charged for as an explicit fee but also term insurance (see for instance Box 8.1 in Ebrill et al. (2001)), and is in such cases already applied in some of the more modern VATs, such as those of Australia, New Zealand, and Singapore. The literature has struggled, however, to find simple ways of administering such a system for other intermediation services. One possibility, suggested by Huizinga (2002) and Poddar (2003), is to zero rate all transactions with registered businesses—as is now done in New Zealand—while providing cash flow treatment of transactions with households, so that responsibility for remittance falls on the financial institution. While this would require financial institutions to distinguish between registered and non-registered customers, that is a distinction which is not only already made in many cases but is also inherent in a wider approach to VAT redesign, the ‘VIVAT’, that we shall argue later has appeal on quite different grounds. The essence of the VIVAT, as it might operate in the EU, is that all sales to registered traders would be taxed at a common EU rate, while other sales are taxed at the final rate specified by each member state. The arrangement just described fits well into such a system: all that is needed is to apply a common intermediate rate of zero on financial services transactions involving unregistered traders and to tax provision to non-registered traders, at the final rate specified by each country, and on a cash flow basis. Box 4.3 illustrates how this would work.

Box 4.3. Cash flow treatment of banking with zero-rating of transactions with businesses

Consider first the case of a bank that deals only with households. In period 1, one household deposits £1,000 and another borrows the same amount. In period 2, the latter repays the loan with interest at 15% and the former withdraws the principal with interest at 5%. The VAT rate is 10%.

The VAT consequences in period 1 are that the bank remits £100 in respect of the deposit (10% of £1,000), this being treated as a taxable sale, but receives a credit of £100 in respect of the funds it loans out, treated as a taxed purchase. In period 2, repayment of the loan creates a VAT liability of £115 (10% of principal and interest of £1,150) while withdrawal of the interest-augmented deposit gives a credit of £105 (10% of £1,050). The only net VAT collected, all in period 2,

(cont.)

Box 4.3. (*cont.*)

is thus £10. This is equal to the product of the VAT rate and the entire value of the bank's spread (£100)—which is as it should be, since in this case the entire value of the intermediation service provided by the bank accrues to final households.

Now suppose that the borrower is a registered trader. With transactions involving the borrower then zero-rated, and so triggering no tax payment, all that remains is the payment of £100 in period 1 and credit of £105 in period 2.⁴⁰ Denoting by R the interest rate available to the government, the net revenue effectively collected, in period 2 equivalent, is 10% of $(R - 0.05) \times £1,000$. In effect, the intermediation services enjoyed by the depositor are valued, and taxed, at the excess of the government's discount rate over the deposit rate.

In terms of equivalences of the kind discussed in Section 4.3.1, such a VIVAT form of cash flow tax would be closely related to the $R + F$ destination-based cash flow tax discussed by Auerbach, Devereux, and Simpson in Chapter 9. It would not quite be equivalent, at firm level, to such a tax (combined with a wage tax), since it would exclude cash flows related to financial services purchased by registered businesses. Summed across firms, however, these flows net to zero, so that such an equivalence would apply at an aggregate level.

There are evidently many challenging practical issues to address before a scheme of this kind could be firmly recommended. But the potential benefits appear significant, not only in terms of easing distortions but also in simple revenue terms. A priori, it is not clear whether such a system (or any form of cash flow treatment) would lead to an increase in VAT revenue (since the unrecovered tax on financial sector inputs under the present exemption system, amplified by further cascading into the taxable value of other commodities, could exceed that due on the value enjoyed by final consumers). In practice, however, it seems likely that it would. For the EU as a whole, Huizinga (2002) estimates, such a scheme would raise an additional

⁴⁰ This simplifies, in that payment of the tax on deposit leaves only £900 available for loan. But this is inessential: taking account of the absence in this case of any tax on the repayment of the loan, it is readily verified that, given the credit on the depositors' withdrawal of £1,050, the bank earns £1,150. Alternatively, arrangements might be put in place (along the lines of the Tax Calculation Account (TCA) of Poddar and Morley English (1997)) to defer (with interest payable) remittance of tax on inflows from households until the credit in respect of withdrawal becomes available. Note too that zero-rating transactions with registrants eliminates any need for TCAs by or on behalf of registered businesses, which has been seen as a major compliance obstacle to cash flow taxation.

€12 billion, while for Germany alone Genser and Winker (1997) estimate a net increase of DM 10 billion (€5 billion).

The threshold

One particular form of exemption is that, de facto, of businesses falling below the VAT threshold (unless, as discussed above, they choose to register voluntarily). At £61,000, the threshold in the UK is the highest in the OECD (Table 4.2). But there is also evidently massive variation in VAT thresholds, some countries having none at all. The natural question is whether the UK has set it too high.

Since any threshold distorts competition between those above and below it, the only rationale for excluding smaller businesses from the tax is to save administration costs to the authorities and compliance costs to the taxpayer. Against this, of course, must be weighed the revenue foregone⁴¹ by excluding those businesses from tax. Box 4.4 sets out a simple framework for thinking systematically about these trade-offs. It shows too that if plausible parameter values (for implementation costs and the social value of tax revenue) are inserted, a threshold at the UK level is not hard to rationalize.

Box 4.4. Setting the VAT threshold

For a benchmark case in which these administration and compliance costs, A and C respectively, are independent of firm size, Keen (2004) show that trading off the implementation costs saved and the revenue foregone by excluding some firms from the VAT implies an optimal VAT threshold of:⁴²

$$Z^* = \frac{\delta A + C}{(\delta - 1)\tau N} \quad (1)$$

where δ denotes the marginal cost of public funds, τ the rate of VAT, and v the ratio of value added to turnover. None of these parameters is known with great certainty. But supposing, for illustrative purposes, that the marginal cost

(cont.)

⁴¹ Although exemptions may in general increase rather than reduce revenue, small traders in the middle of the chain will generally find it advantageous to register voluntarily. The expectation is thus that a threshold will lose revenue.

⁴² The underlying intuition is simple. The social benefit from slightly increasing the threshold from some initial level of Z is the saving in compliance costs C and the administration costs A , with the latter, since it is financed from distorting tax revenue, being weighted by the marginal cost of public funds, all multiplied by the number of taxpayers $f(Z)$ taken out of the VAT. The social cost is the revenue foregone, τN , multiplied by the excess of the marginal cost of public funds over unity (the reduction in revenue being a private gain) and, once more, that number of affected taxpayers. Equating these marginal social costs and benefits to characterize an optimal threshold Z^* gives the result in (1).

Box 4.4. (*cont.*)

of public funds is 1.2, administration and compliance costs £120 and £600 respectively,⁴³ then with a VAT rate of 17.5% and a ratio of value added to sales of 30%, the implied VAT threshold is about £71,000—rather higher than at present, but a similar order of magnitude. While this suggests that the UK value is well within the bounds of the plausibly optimal, it should be noted that the value implied by (1) is quite sensitive to parameter values: changing the marginal cost of public funds to 2.0, for example, reduces the optimal threshold to around £16,000.

The simple rule set out in the box ignores many potentially important considerations. Implementation costs are likely to vary with firm size, for example, and account must also be taken of the inefficiencies created by distorting competition between firms of different size and potentially inducing artificial splitting to remain below, or simply discouraging expanding firm size above, the threshold.⁴⁴ This raises important questions, as Bird (commentary on this chapter) points out, as to whether some simple ‘replacement’ for the VAT should apply to those below the threshold (which may itself affect where that threshold should be).⁴⁵ These considerations significantly complicate the analysis, with the distribution of firm size, for example, playing an important (and theoretically ambiguous) role. Simulations by Keen and Mintz (2004), however, tend to point to thresholds optimally higher than that implied by (1); the somewhat different model of the determinants of firm size in Zee (2005), on the other hand, points in the opposite direction.⁴⁶ Further considerations, awaiting closer analysis—such as the propagation of beneficial VAT chains of the kind described above—may point to lower thresholds. But others, such as the need to control registration as a defence against

⁴³ A recent study of tax compliance costs in the UK by KPMG (2006) implies an average VAT compliance burden per registrant of £562 (combining figures in table 3 and in section 3.3.4). Another for New Zealand (Colmar Brunton (2005)) implies a cost in terms of internal time alone (so neglecting bought-in advice and equipment) of around £660 (though varying fairly substantially, if not entirely systematically, with firm size). It seems unlikely that it is cheaper to comply with the relatively complex UK than with the simpler one in New Zealand, so that £600 seems a reasonable order of magnitude. Earlier estimates by Cnossen (1994) suggest administrative costs of around 20% of compliance costs: hence the £120 figure.

⁴⁴ KPMG (2006, annex C p. 4) report that some firms interviewed indicated a deliberate decision to keep their size below the VAT threshold.

⁴⁵ For discussion of this, and more generally of the potential importance of the VAT threshold in anchoring the tax treatment of SMEs, see International Tax Dialogue (2007).

⁴⁶ Both Keen (2004) and Zee (2005) have firms differing in underlying productivity, the former have them producing a homogeneous product whereas the latter has them producing Dixit–Stiglitz substitutes. Dharmapala, Slemrod, and Wilson (2007) explore the same issue in a model with endogenous entry and exit; the likely quantitative implications for the optimal threshold, however, remain to be analysed.

carousel and other fraud (as discussed below) point to higher. There is good reason to suppose that the relatively high threshold should be counted as a strength of the UK VAT.

4.3.3. VAT enforcement and compliance issues

Like all taxes, the VAT is subject to evasion. For example, traders large enough to be liable to register may fail to do so, they may under-report sales, or, where different commodities are subject to tax at different rates, they may misclassify sales into the category subject to a lower rate of tax. As discussed above, the fractional nature of the VAT in some respects reduces its exposure compared with other systems of sales taxation.

In other respects, however, VAT offers distinctive opportunities for evasion and fraud, especially through abuse of the credit and refund mechanism. Revenue may be lost through exaggerated claims for credit for VAT paid on inputs. Moreover, the opportunity exists for outright fraud through the construction of business activities with the sole purpose of defrauding the exchequer, because some categories of business can be entitled to net refunds of VAT from the revenue authorities. These include firms selling predominantly zero-rated goods that are due refunds of input VAT. The extensive domestic zero-rating means that refunds are particularly extensive in the UK, at around 40% of gross VAT receipts (Harrison and Krelove (2005)). Little is known about the extent to which domestic zero-rating gives rise to fraud problems. The zero-rating of exports, however, has clearly become a significant source of difficulty.

Particular attention has come to be paid to ‘Missing trader intra-community’ (MTIC) frauds, which abuse the refunding of VAT to exporters by means of a series of contrived transactions. Figure 4.3 provides a simple example of a ‘carousel fraud’, the best-known example.

The two key features of the VAT that this exploits are the zero-rating of exports and the system of ‘deferred payment’ for VAT on imports from other EU member states, adopted in the EU since the removal of fiscal frontiers in 1992.⁴⁷ Under deferred payment, VAT on imports from one member state into another is levied not at the border but at the time of the importer’s next periodic VAT return. As a result, there may be a considerable time lag between the date at which the importing firm (Company B in the example) brings the goods into the UK and the time at which the VAT authorities seek payment of the VAT due. In the meantime, the goods are sold on, via complicit—or

⁴⁷ There are indications that the general level of VAT revenue losses rose by about one-third by the mid-1990s compared with pre-1992 levels (see table 2.1 of HM Customs and Excise, *Measuring Indirect Tax Losses*, 2002, <<http://www.hm-treasury.gov.uk/media/389/E5/admeas02-297kb.pdf>>).

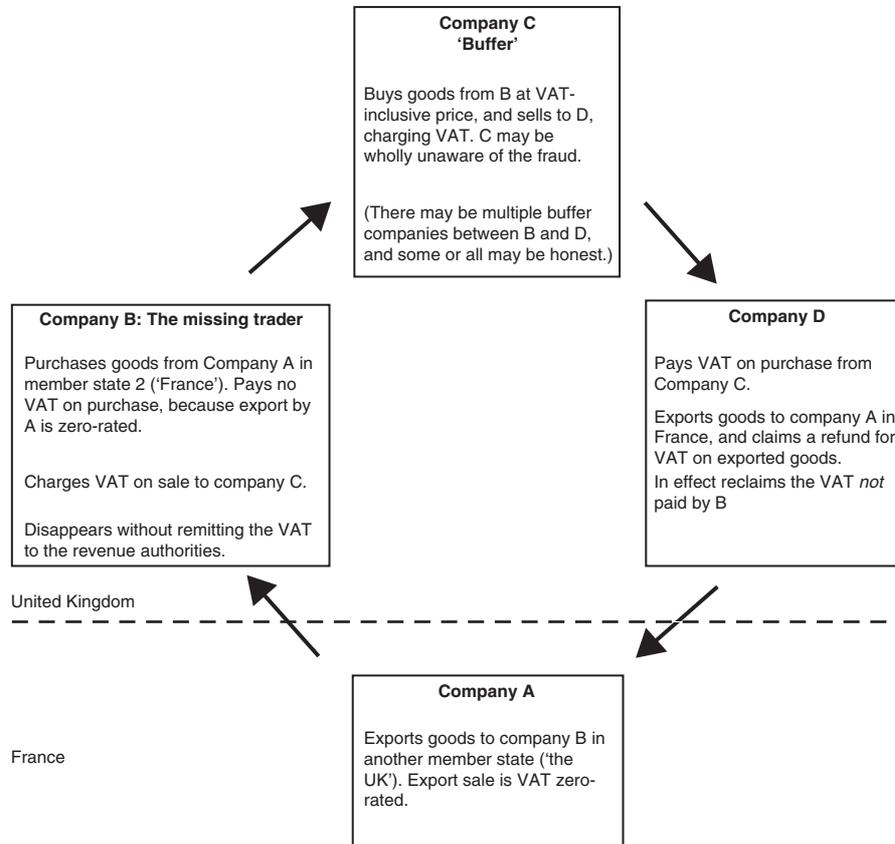


Figure 4.3. Carousel fraud: A simple example

perhaps unwitting—'buffer' companies in the UK, to Company D, which exports the goods, claiming a refund of the VAT paid when it purchased the goods from Company C. In the basic carousel illustrated, the exported goods are then re-imported by Company B—or more likely a new firm, B having gone missing—and so on, following a cycle in which VAT refunds are claimed repeatedly whenever the goods are exported, while the corresponding import VAT liability accumulates but is never paid. This example is extremely simple. In practice, the basic structure of the fraud may be obscured by the use of buffer companies that may or may not be complicit. Moreover, innovation has been a constant feature of these forms of fraud, as those perpetrating them seek to stay one step ahead of the authorities' ability to detect fraudulent transactions.

Key design features of the VAT system influence the extent to which it is exposed to systematic missing trader frauds. These include the ease with which the system allows intending missing traders (such as Company B in

Figure 4.3) to register for VAT, and the relative timing of VAT payments and receipts. *Ex post* audit and investigation, while important, is unlikely to forestall considerable loss of revenue, because the essence of the fraud is that money is made quickly, in the time gap before the missing trader is required to remit the VAT it has supposedly charged on its sales. Once the money has disappeared into the complex web of transactions, tracing and recovering unjustified VAT refunds becomes time-consuming and costly.

Other than more vigorous investigation, two broad approaches may be taken to designing-out the opportunities for fraud based on current treatment of trade between EU member states. One is essentially administrative, in the sense that it retains the zero-rating of intra-community supplies. The other, more fundamental to the structure of the tax itself, removes export zero-rating altogether.

Administrative measures that could be taken while maintaining the current VAT design include, for example: tighter checks on firms seeking to register for VAT (for example, with an on-site visit) and requiring guarantees in dubious cases; slowing down the payment of VAT refunds relative to the collection of VAT due (although this can impose severe cash flow burdens on legitimate businesses); adopting or strengthening joint and several liability rules by which traders can be held responsible for fraud elsewhere in the chain that they might reasonably have been expected to be aware of; and establishing better and quicker information exchange between national tax authorities (so that the country of import can become promptly aware that exports to it that have been reported in another member state have not shown up in its own VAT system). However, while measures of this sort may reduce the risk of VAT fraud, they clearly have undesirable side-effects. More bureaucratic VAT registration procedures and slower payment of VAT refunds might harm legitimate businesses as well as discouraging fraud, for example, and these effects may outweigh the enforcement gains. The authorities have a difficult balance to strike, between ensuring that VAT administration does not impose excessive burdens on business in general and ensuring that it is not unduly exposed to fraud. Some level of VAT evasion has to be tolerated in the wider business interest.

More radical measures within the context of a system that preserves zero-rating of intra-community exports include:⁴⁸

- The use of ‘reverse charging’, by which liability in a business-to-business (B2B) transaction is placed on the buyer rather than the seller. This

⁴⁸ There are others not discussed here, such as the ‘prepayment VAT’ (PVAT) of Poddar and Hutton (2001), under which goods would not be shipped without adequate guarantee of payment of import VAT. Ainsworth (2007) attempts an exhaustive listing.

would close the carousel fraud in Figure 4.3 by making the VAT due on the sale by B (the missing trader) the responsibility of the buyer, C. In turn, the tax due on the sale from C to D would be the responsibility of D. The zero-rating of the subsequent export sale would then offset D's tax liability on its purchases from C, reducing the tax payment by D but not requiring outright refunds. The opportunity to make fraudulent gains by claiming refunds of tax that have not in fact been paid would thereby be eliminated. The UK recently received EU approval to apply reverse charging for mobile phones and computer chips, both having proved popular instruments for carousel fraud. More radically still, Austria and Germany have proposed allowing reverse charging for all B2B transactions above a certain size (€10,000 in the case of the Austrian proposal and €5,000 in the German proposal). The proposals differ in terms of the scale of the reporting obligations placed on firms and their customers: the German proposal would require both parties to a B2B transaction to report it to the tax authorities, and electronic cross-checking of this information, while the Austrian proposal would place fewer reporting burdens on firms.

The difficulty of reverse charging limited to certain products—as proposed by the UK—is that MTIC frauds may simply move on to other goods, not covered by reverse charging. There would also be new enforcement issues, at the 'boundary' between commodities subject to reverse charging and those subject to 'normal' VAT. Universal reverse charging, as proposed by Austria and Germany, avoids these difficulties. But it also, in effect, turns the VAT into something closely akin to a single-stage retail sales tax, with tax payments suspended until goods are sold to final consumers (albeit with the possibility of cumbersome reporting procedures for B2B transactions). The danger of this is obvious: it undermines the fractional nature of the VAT and instead collects all VAT revenue at the final sale, so exposing the system to substantially greater risks of revenue loss through unreported sales to final consumers. In effect, the VAT would be converted into an RST. Extensive reverse charging might help to stem losses from MTIC frauds, but would also expose the VAT to other risks of revenue loss through more mundane forms of evasion.

- 'Reverse withholding' schemes would tackle VAT frauds in a broadly similar way to reverse charging, by requiring the purchaser in a B2B transaction to make a direct payment to the authorities of part or all of the VAT due on its purchase. The difference is that the seller would remain liable for output VAT, receiving a credit for the amount withheld by the purchaser. Depending on the proportion of the VAT that the

purchaser is required to withhold, this would diminish or even eliminate the scope to generate revenues through fraudulent refund claims, since exporting firms will themselves have paid part or all of the VAT on their purchases that they subsequently reclaim on export. The principal drawback of reverse withholding (which is quite common in Latin America but untried in Europe) would be its administrative complexity (which arises because of the need to ensure that the seller is given credit for withholding only when this has actually taken place) and the increased likelihood that traders will be due a refund (because of the reduced output tax they pay).

- Adoption of a system of 'VAT accounts', under which traders would be required to open a distinct bank account into which they would transfer the amount of VAT charged to their customers. VAT refunds would only be paid if the authorities were able to verify that the corresponding VAT payment had been made. This has been proposed by Sinn, Gebauer, and Parsche (2004) as a solution to the problem of VAT fraud, and a system of this sort has been tried in Bulgaria. The key feature is that it requires the VAT payment to be made *earlier* than in the present system, so that when refunds are paid, they can be checked against past payments made. Apart from this matter of timing, however, it does not fundamentally alter the situation. It is not clear that cross-checking refund claims against past payments to a bank account would be any easier, or more reliable, than checking that past payments have been remitted to the revenue authorities themselves. The scheme also eliminates the cash flow benefit that firms enjoy (providing some offset to their compliance costs) by retaining VAT collected until the next periodic payment becomes due. Reflecting these difficulties, Bulgaria has decided to remove its VAT account system.
- The compulsory use of a third party to guarantee VAT payments, either in general or for particular sectors, as set out by Ainsworth (2006). In the example set out in Figure 4.3, Company B, the future missing trader, would be required to obtain a guarantee that its VAT payments would be made. The principal difficulty with this is the cost involved; it is far from clear that banks or other potential guarantors would be any better placed than the revenue authorities to prevent firms disappearing with outstanding VAT liabilities, and the premium required to cover this risk would place substantial burdens on honest firms operating in the sectors most subject to VAT fraud.

These various administrative solutions all have weaknesses, either in creating other opportunities for fraud and/or in increasing taxpayers' compliance

costs. A more durable solution to the problem of missing trader fraud requires a fundamental redesign of the VAT treatment of international transactions. The opportunity to claim fraudulent VAT refunds arises principally because of the break in the VAT chain that occurs as a result of the zero-rating of exports; and the break occurs at an especially vulnerable point in the chain, where control passes from one national tax administration to another. Export zero-rating requires substantial amounts of VAT receipts to be paid back as refunds, and a system that requires refunds on such a large scale creates opportunities for correspondingly large-scale revenue fraud. Ending VAT zero-rating for trade between EU member states would sharply reduce the scale of refunds and eliminate some of the most tempting opportunities for missing trader frauds. We discuss ways in which this could be done in Section 4.5.2.

4.4. EXCISE TAXES ON ALCOHOL AND TOBACCO

The UK, in common with all EU countries, applies excise taxes,⁴⁹ as well as VAT,⁵⁰ to three principal groups of goods: motor fuels, tobacco products, and alcoholic drinks. The scale of these taxes varies widely across the EU (Table 4.3). In the case of all three, UK excise rates are at the top end of the range, and their revenue contribution is significant: some £40 billion in 2006–07 (8.2% of total revenues), of which more than half (some £24 billion) was from motor fuel taxes, with £8 billion each from tobacco and alcoholic drinks. As noted earlier, however, the period since the Meade Report has seen a substantial reduction in the importance of the excises as a source of revenue. Intellectually, too, they have—until recent years—been out of fashion. What role, if any, remains for excises, and on what basis should they be levied?

⁴⁹ By an ‘excise’ we mean a single-stage sales tax applied to a limited group of commodities. The UK excises on alcoholic drinks and tobacco products discussed here are levied at the wholesale stage, with production and movements of these goods before the excise is levied subject to close monitoring and control by the revenue authorities. The UK also levies single-stage sales taxes on a number of other goods and services, including gambling, insurance premiums, and airline tickets, which for practical or historical reasons are not covered by VAT, and also levies some environmental taxes which are basically single-stage excises on certain goods and services including landfill use and quarried aggregates such as gravel. EU rules govern the structure and, to some extent, the rates of the three principal excises—on alcohol, tobacco, and mineral oils—and also discourage proliferation of additional new excises. Nevertheless, some EU members, notably Denmark, have historically levied excises on a much wider range of goods and services, and continue to do so.

⁵⁰ VAT is levied on the price of the product, including the excise. This sequencing makes no difference when excises are ad valorem, but when they are specific it preserves the (potential) role of the VAT as a generalized tax on consumption in the sense that a uniform change in the VAT rate applied to all commodities then leaves their relative consumer prices unchanged.

Table 4.3. Excise duty rates in the UK and the EU, July 2007 (in euros)

Member state	Cigarette excise per pack of 20	Specific excise as % of total tax	Beer per litre, at 12.5 degrees Plato	Still wine per 75 cl bottle	Spirits per 70 cl, 40%
Belgium	2.29	11	0.21	0.47	4.91
Bulgaria	0.70	7	0.10	0	1.57
Czech Rep.	1.16	42	0.11	0	2.62
Denmark	2.29	54	0.34	0.82	5.63
Germany	2.81	46	0.10	0	3.65
Estonia	0.70	39	0.19	0.66	2.72
Greece	1.61	5	0.14	0	3.05
Spain	1.53	9	0.11	0	2.32
France	3.20	8	0.13	0.03	4.06
Ireland	4.28	55	0.99	2.73	10.99
Italy	1.99	5	0.29	0	2.24
Cyprus	1.69	20	0.24	0	1.70
Latvia	0.52	43	0.09	0.43	2.53
Lithuania	0.55	54	0.10	0.43	2.60
Luxembourg	1.76	15	0.10	0	2.91
Hungary	1.00	45	0.25	0	2.40
Malta	2.20	12	0.09	0	6.52
Netherlands	2.28	50	0.25	0.59	4.21
Austria	2.00	21	0.25	0	2.80
Poland	0.91	35	0.22	0.34	3.21
Portugal	1.86	49	0.17	0	2.68
Romania	0.83	39	0.09	0	2.10
Slovenia	1.28	20	0.34	0	1.95
Slovakia	0.91	52	0.17	0	2.12
Finland	2.40	10	0.97	2.12	7.91
Sweden	2.45	18	0.79	2.37	15.04
UK	4.97	52	1.01	2.63	8.08

Source: Authors' calculations from data in *European Commission Excise Duty Tables* ref 1.025 (July 2007), downloaded from <http://ec.europa.eu/taxation_customs/index_en.htm#>.

This section focuses on alcoholic drinks and cigarettes: the taxation of motor fuels is discussed in Fullerton, Leicester, and Smith, Chapter 5 on environmental taxation.

The case for higher taxation of alcohol and tobacco does not immediately entail the use of completely separate additional taxes for these commodities. Why not, instead, implement such differential taxes by differential VAT rates? There are three main reasons. First, excises are a natural way of correcting for damages that arise from the use of the commodity in production as well as in final consumption; by contrast, higher VAT on production inputs would

do little to alter the behaviour of VAT-registered producers. Second, in so far as there is a strong case for taxing some commodity on a specific rather than ad valorem basis, this can be implemented more easily with a separate excise. Third, very high tax rates create particular enforcement risks that call for physical controls beyond the book-based enforcement on which the VAT is based.

4.4.1. Revenue-raising efficiency in the taxation of tobacco and alcohol

Leaving externality arguments aside for the moment, are the characteristics of demand for these commodities such as to call for above-average taxation on standard optimal tax grounds? The question is often posed in terms of the ‘inverse-elasticity rule’: the presumption that commodities with low own-price elasticities of demand provide a relatively efficient revenue source. But, as discussed in Section 4.2.2, the key demand characteristic relevant to revenue-raising efficiency is not the own-price elasticity of demand but rather the degree of complementarity/substitutability with leisure. Is there any reason to believe that cigarettes and/or alcoholic drinks are particularly (relatively) complementary with leisure (in the sense, more precisely, that, conditional on a given level of total expenditure, those who supply more market labour will smoke or drink less)?

It is unclear, a priori, whether cigarette consumption is likely to be a complement with leisure or a substitute. For some people smoking may be a leisure-time pursuit; for others, it may help with the stresses and social interactions involved in working; for some it may be both. The overall relationship between cigarette consumption and labour supply cannot be resolved on the basis of a priori reasoning. Systematic empirical evidence is sparse. Parry (2003) notes that, while some studies have suggested a slightly negative income elasticity of demand for cigarettes, it is difficult to infer much about complementarity or substitutability with labour supply from this; he suggests that workplace smoking restrictions might be a reason to believe that smoking and working may not be complementary. A similar conclusion emerges from the results from Crawford, Keen, and Smith (2008) reported in Table 4.1 above, which suggest that labour supply and tobacco spending are modest substitutes, with an effect that is statistically significant, but relatively small.

Overall, the evidence is far from overwhelming, and its reliability is greatly compromised by the sharp socio-economic differences in smoking patterns

which have emerged in most countries in recent years. In the current state of knowledge it is unlikely that cigarette consumption is sufficiently complementary with leisure to warrant high taxation on these grounds alone, though as Parry (2003) finds, different values for the relative substitution between cigarettes and leisure, within a relatively moderate range, can make a considerable difference to the estimated optimal tax on cigarettes.

The relationship between alcohol and the labour/leisure choice is also likely to be complex. More leisure means more time to consume alcohol and also reduces the risk that the consequences of over-consumption (hangovers and so on) will negatively affect work performance. On the other hand, some level of alcohol consumption may be complementary with time spent working (and hence a leisure substitute), either because drinking helps workers unwind after a stressful day or because of social drinking and networking with colleagues. The balance between these two effects may vary with individual alcohol consumption. At low consumption levels, alcohol may be complementary with work, but at higher levels it may be complementary with leisure. The results from Crawford, Keen, and Smith (2008) reported earlier for two categories of alcoholic drinks, 'beer' and 'wines and spirits', show a positive and significant relationship between their budget shares and hours worked, suggesting that both are, on average, leisure substitutes rather than complements. This result should, however, be viewed with some caution, as it is well known that alcohol expenditure is under-reported in UK household budget surveys, particularly because of under-representation of high consumers. If the relationship between alcohol consumption and labour supply is very different among heavier drinkers than the rest of the population, then the estimates will be biased by the omission of heavy consumers. Nevertheless, and as with cigarettes, the evidence does not create a compelling case for taxing alcohol more heavily than other goods on grounds of revenue-raising efficiency alone.

4.4.2. Consumption externalities and commodity taxation

The excises on alcohol and tobacco might alternatively be justified in terms of their role in discouraging consumption that has consequences beyond the immediate pleasure of the consumer themselves.

Even if we put aside any paternalist inclinations and regard individual consumers as best placed to judge the costs and benefits of consumption to themselves, externality-correcting taxes may be needed to ensure that costs imposed on others are properly considered. The case for using taxes

to discourage *socially* harmful activities was set out by Pigou (1920) and has been developed subsequently in the context, primarily, of environmental policy. In an otherwise first-best context—with prices undistorted by taxation, imperfections of competition or other factors—the tax should be set equal to the marginal external cost of each unit consumed. In more realistic contexts, this conceptually simple policy prescription becomes nuanced. When taxes also serve a revenue-raising purpose, for instance, Sandmo (1976) and Pirtillä and Tuomala (1997)—dealing with the cases in which direct taxes may and may not be imposed—show that optimal commodity tax rules take an additive form in which a Ramsey-type component is supplemented by one related to the marginal external damage, but potentially somewhat lower to reflect any intensification of the distortions from pre-existing taxes. Understanding the nature and extent of external damage is thus key to proper tax design.

The external costs of tobacco and alcohol consumption include direct externalities experienced by other individuals, such as the harm that drunken drivers cause to others, and collectively borne resource costs such as the cost of publicly funded medical treatment for smoking- and alcohol-related conditions. In both cases, individual consumption decisions will not reflect the full social costs of consumption, because the individual consumer does not bear the marginal costs of medical treatment or of direct externalities that follow from their smoking or alcohol consumption decisions. Excise taxes may provide a way of confronting the consumer with these costs.

Frequently, the discussion of whether alcohol consumption and smoking are socially harmful also includes discussion of the consequences for the individual consumer's own health, employment prospects, accident risks, and so on. Such costs to the individual consumer are not, however, externalities. Arguments that they might warrant higher taxation to discourage consumption involve an element of paternalism, perhaps reflecting a concern that individuals may be poorly informed about some of the consequences for themselves of consumption. Viscusi (1995), for example, discusses whether individuals accurately perceive the health risks of cigarette smoking. However, unless there are grounds to believe that consumers are poorly informed about some of the consequences of consumption, paternalist arguments tend to meet with scepticism among economists.

Nevertheless, with tobacco and alcohol, the case for discouraging consumption, in the individual's own interest, should not be dismissed too lightly. Some of the individual costs of smoking and drinking arise as a result of the addictive nature of consumption, and this adds a further dimension to the problem. Current consumption may increase the risk of future addiction.

A well-informed and rational consumer would presumably be less willing to start smoking than if there were no risk of future addiction (Becker and Murphy (1988); Chaloupka (1991)). However, it is unrealistic to assume that all individual users of tobacco and alcohol fully grasp the addictive risks associated with each unit of consumption (Orphanides and Zervos (1995)). In addition (except in an empty and tautological sense), it is unrealistic to assume that all consumption decisions—particularly of alcohol—are based on a rational calculus of the costs and benefits of each unit consumed.

These departures from standard norms of rational behaviour have been emphasized by some authors—notably Gruber (2001) and Gruber and Köszegi (2001, 2004). What is central to these arguments, which are reviewed at greater length in Gruber's commentary on this chapter, is not simply the addictive properties of cigarettes and alcohol, but rather that individuals may show a basic time inconsistency in their behaviour: wanting to quit in the future, for example, but when the time comes preferring not to. They may then welcome devices that help to commit their future behaviour, with high current prices, for example, inducing a reduction in present consumption that will make it easier to amend their future behaviour (Gruber and Mullainathan (2005)). Various ways have been suggested of thinking about this kind of behaviour: Gruber refers to this as an 'internality', while Viscusi (1995) discusses the concept of an 'externality to one's future self'. This can lead, as Gruber points out, to fairly high levels of cigarette taxes—though not ones far out of line with those currently levied in the UK. Quite how persuasive these ideas are remains as yet unclear. One implication, for example, is that smokers would vote for a cigarette tax increase.

To the extent that such behaviour may undermine the case for setting tobacco and alcohol taxes based on the level of traditional externalities, the key consideration in public decision-making becomes the effectiveness of taxes in modifying the behaviour of those most likely to make ill-informed choices and face commitment difficulties. For this reason, the impact of taxes on the smoking and drinking behaviour of children and young adults becomes a major focus. The balance of evidence suggests that higher taxes are more effective in restraining consumption by this group than for the population in general (Lewit et al. (1981); Lewit and Coate (1982); Chaloupka and Wechsler (1997)).

External costs and tobacco taxes

What level of cigarette taxation would be required to reflect the marginal external costs (as traditionally defined) of smoking? These costs fall into

three broad categories. The first consists of direct externalities experienced by other individuals, including the annoyance and adverse health effects from passive smoking. These people may include work colleagues and wholly unrelated individuals (other patrons of a bar or restaurant, for instance), but a high proportion will be friends and family members. The second comprises collectively borne costs, such as those of publicly funded medical treatment for smoking-related conditions. The third category of externalities arise through the tax system. Smoking may have consequences for the individual consumer's income and expenditure, especially through a higher rate of sickness absence. While these effects would be wholly internal in a 'first-best' world without distortionary taxes, the presence of income taxes and spending taxes gives rise to what is, in effect, a tax revenue externality. Given the efficiency cost of raising public revenues through distortionary taxation, induced changes in revenues have a clear social significance.

There is an extensive literature on the social costs of smoking. Much of this is not directly relevant to the estimation of corrective taxes, however, because an insufficiently clear distinction is drawn between costs experienced and internalized by the individual consumer and external costs experienced by others.

The most widely quoted US estimates of tobacco consumption externalities, appropriately defined, are those of Manning et al. (1989). These distinguish between the gross costs of smoking (higher costs of medical treatment, and so on, as a result of conditions caused by smoking), and the net costs, which offset against the gross costs a range of cost savings (mainly public expenditure effects) arising from the premature death of smokers. Gross costs were estimated at 42 cents per pack, discounting future costs at 5%, while net costs were 15 cents per pack, on the same basis. Adding amounts to reflect the costs of passive smoking and smoking-related fires raises the net costs to 38 cents per pack—very close to the 37 cents per pack average tax on cigarettes in the US at the time.

These estimates were subsequently updated by Viscusi (1995), with adjustments to reflect the lower tar levels of cigarettes smoked by more recent generations of smokers, and to reflect the lag profile with which this change in tar content feeds through into health effects after consumption. Table 4.4 summarizes some of the key components in the estimates presented by Viscusi, again with total external costs expressed in terms of the externality per pack. The effects of Viscusi's tar and lag adjustments in the example shown are relatively modest, but such adjustments have been controversial in the light of some evidence suggesting that smokers may smoke lower-tar cigarettes more intensively (Evans and Farrelly (1998)).

Table 4.4. Expected external costs of cigarettes for the United States, in cents per pack

	External cost, raw data	External cost, tar adjustment based on 20-year moving average
Medical care before 65	0.326	0.302
Medical care after 65	0.172	0.153
Total medical care	0.498	0.455
Sick leave	0.012	0.011
Group life insurance	0.126	0.114
Nursing home care	-0.221	-0.197
Retirement pension	-1.099	-1.000
Fires	0.016	0.016
Taxes on earnings	0.351	0.326
Total net external costs	-0.317	-0.274

Source: Viscusi (1995); based on 3% discount rate.

The picture in Table 4.4 may be partial. No estimates are included for the direct externalities of passive smoking, in terms of either annoyance, or the substantial items that could arise from the health effects of passive smoking, especially on children and other family members. Evidence on these effects has been accumulating rapidly in recent years, and is reviewed in Adams et al. (1999). Viscusi (1995) argues that passive smoking effects are insufficiently precisely determined to permit point estimates, but shows a range of possible values, leading to a bottom line which varies from a net saving to a larger net cost, among which he suggests the median estimate of zero net external costs may be regarded as a plausible upper bound.

Taking account of public expenditure offsets arising from the premature death of smokers has been controversial in the public debate. It nevertheless seems wholly correct. If these items were private costs (for example, if individuals were responsible for paying actuarially differentiated private insurance premiums to cover future costs arising from their consumption decisions), then a rational consumer would experience the additional health costs and reduced future social care costs in the same way, and there is no reason to believe that they would treat the two insurance premiums differently. Since the purpose of externality taxation is to confront the individual decision-maker with the external costs of their decision, on the same basis as if they were private costs, consistency would require that both additional costs and cost reductions be treated in the same way.

One reason that public expenditure offsets arising from premature mortality have been controversial is their size. On Viscusi's unadjusted figures, additional medical care costs for smokers are approximately 50 cents per pack,

while savings on nursing home care and retirement pensions are equivalent to about \$1.30 per pack. But there is even controversy about the sign of the health care cost component. In a study for the Netherlands, Barendregt et al. (1997) found that the long-term effect of smoking cessation was to increase health care costs. In the short term, they estimated health care costs would fall, as ex-smokers cease to suffer the illnesses caused by smoking, but this 'conventional' effect on the costs of treating smoking-related diseases would then be offset by longer-term costs, as ex-smokers survive long enough to suffer the very costly medical conditions of old age. By contrast, Rasmussen et al. (2004) use a similar approach, but find that the overall lifetime effects of smoking cessation on health care costs are beneficial.

The externality case for cigarette taxes requires clear identification and measurement of the external costs of smoking. Nevertheless, attempts to draw a clear boundary between internalized and external costs are not straightforward, and a number of areas of conceptual and practical difficulty can be identified.

- *Effects on family members.* One controversial area is the treatment of costs borne by family members. Family members of a smoker may experience considerable costs, including ill-health, and pain and distress as a result of the illness and premature death of the smoker. Many of the consequences of smoking during pregnancy, including severe damage to the lifetime health of the child, also come under this heading. There are, in addition, further important costs for the healthcare system in treating these conditions. Whether, and to what extent, costs experienced by other family members should count as externalities has been controversial. As Viscusi (1995) notes, it turns, in principle, on whether the welfare of other members enters into the utility function of the smoker, and, in other areas of policy, it is often assumed that family members are assumed to care for each other's welfare to the extent that the welfare of the household can be considered as a single entity. In the case of smoking, however, this seems an extreme position, and at least some of the harm inflicted on family members and unborn children would seem best treated as externalities.
- *Healthcare Finance and External Costs.* How far are the costs of medical treatment and healthcare internalized when tobacco consumption decisions are made, and how far should they be counted as an externality? In the case of publicly funded, tax-financed systems such as the UK National Health Service, the answer is straightforward: the treatment costs of illnesses resulting from individual consumption decisions are collectively

financed. The same may largely be true of other collectively financed systems, such as employer-financed healthcare. By contrast, a system of private insurance capable of appropriate differentiation of insurance premiums might be able to internalize a large part of the healthcare costs associated with smoking.

- *Effects on wages and productivity.* A large part of the total social cost of smoking estimated by Manning et al. (1989) and Viscusi (1995) consists of foregone tax revenues on earnings, due to the shorter and less-productive working lives of smokers. This seems fairly uncontroversial. Whether more of the income or output loss from sickness absence and other smoking-related productivity effects should be counted as an externality depends primarily on how far the effects of smoking on worker productivity are reflected in wages. If workers are paid their actual marginal product, then there is no externality: smokers who are less productive receive correspondingly lower wages. However, employers may not always be able to differentiate the wages paid to smokers and non-smokers: Employment protection and sick pay legislation may, for example, compel an employer to continue making payments to a worker who has fallen sick as a result of smoking. In this case, the lower productivity of smokers generates externalities of two forms. First, the employer paying a smoker wages that exceed the worker's marginal product will experience a real income externality. Secondly, since the overall marginal product has fallen and the employer is unable to differentiate the wages paid to different workers, both smokers and non-smokers will receive lower wages as a result of the lower productivity of smokers.

External costs and alcohol taxes

The external costs of alcohol consumption, like those of smoking, include (i) direct externalities experienced by other individuals, including the victims of accidents, property damage, and violence caused by other people's drinking, (ii) collectively borne costs, such as the cost of public order policing and publicly funded medical treatment for alcohol-related conditions, and (iii) tax revenue externalities. The most widely quoted US estimates of alcohol consumption externalities, appropriately defined, are those of Manning et al. (1989), who estimate that net external costs are equivalent in value to about 35% of the producer price of alcohol. A very large proportion of these are from alcohol-related traffic fatalities. Parry (2003) notes that there are fewer such accidents in the UK, so that marginal external costs are consequently also likely to be lower.

Estimates of external costs of alcohol consumption for the UK by Maynard, Godfrey, and Hardman (1994) included various categories of direct externality (such as the damage caused in road traffic accidents) and also the cost of defensive measures (such as policing costs and the costs of research on alcohol problems). The costs of collectively funded medical treatment in the UK National Health Service of alcohol-related illnesses are also included. But the largest items are under the heading 'costs to industry', including the substantial costs of sickness absence and unemployment. Quite how far these should count as external costs depends on what is assumed about the effects of alcohol-induced productivity effects on individual wages, an issue discussed above for smoking. Overall, Maynard et al. (1994) calculate the total externalities associated with alcohol consumption in the UK to be around £2.7 billion, equivalent to some 17% of pre-tax alcohol expenditure.

Translating such estimates into appropriate rates of Pigouvian taxation on alcoholic drinks is complicated by the fact that most are of total costs, not marginal. Unlike cigarette externalities, which may well be almost constant across each unit consumed, the externalities from alcohol consumption largely relate to abusive consumption, and the marginal external cost of consumption may well differ sharply from the average. Excessive consumption, leading to serious drunkenness in individual episodes or alcoholism over time, is the prime source of the externalities, and the external costs associated with moderate consumption may be close to zero. While the results of Maynard et al. (1994) suggest that the average external cost from alcohol consumption in the UK might be of the order of 17% of the pre-tax price of alcohol, the marginal external cost may be quite different—and higher.

Diamond (1973) considers the design of corrective taxation when all individuals causing externalities must be taxed at the same rate, but where the externalities from some are more damaging than those from others. Where there is separability between the externality and consumption, the optimal tax is simply the weighted average of the marginal contributions to the externality, across different individuals, where the weights are given by the sensitivities of demand for the externality-generating good. If the primary reason for differences in the external costs of alcohol arises from differences between individuals, rather than between units of drink consumed by a particular individual, then this result may provide some support for regarding the average external costs as a rough-and-ready indicator of the optimal externality tax. But if the external damage caused by alcohol consumption varies across units consumed by each individual (for example, if the 'last drink' causes all the problems), then the appropriate externality tax would be considerably higher.

Pogue and Sgontz (1989) investigate further the implications for optimal alcohol taxation of the requirement that the alcohol sold to abusive and non-abusive drinkers be taxed at the same rate. Their analysis is directed at quantifying the trade-off between the reduction in welfare of non-abusive drinkers and the social benefits from the reduction in consumption by abusive drinkers. The balance will depend on the size of marginal social costs from abusive consumption, the size of the welfare loss from distorted non-abusive consumption, the relative numbers of abusive and non-abusive consumers, and the price elasticities of the two groups. (There is also the difficult issue of principle of whether alcohol-dependent consumers of alcohol should be treated as deriving utility from their abusive consumption, or not.)

4.4.3. Distributional aspects of alcohol and tobacco excises

As argued earlier, the availability of a sophisticated income tax and direct transfer system effectively undercuts any distributional rationale for differential commodity taxation, and any reason to have particular concern about the distributional characteristics of taxes on most commodities. As a general rule, the distributional incidence of individual components of the tax system is of less concern than the overall distributional incidence of all taxes (or, indeed, all government fiscal interventions) taken together. The fact that a particular regressively distributed component of the tax system may bear disproportionately heavily on poorer households may be unimportant if it is counterbalanced by distributional progressivity in other taxes.

Nevertheless, where, as with tobacco taxes, taxes are distributed extremely unevenly across households at similar income levels, there may be reason to look at the separate distributional incidence of the tax, since any adjustment to other taxes can only offset on average the distributional impact of the excise tax on tobacco. The distributional incidence of taxation on heavy smokers will be more heavily influenced by the distributional characteristics of the tobacco tax; and for non-smokers, the incidence of the tobacco tax will be an irrelevance.

Studies for the US have found that alcohol and tobacco taxes appear substantially less regressive when a longer time frame is used for analysis than when analysed on the basis of current income. Poterba (1989) shows that the regressivity in relation to current (annual) income of alcohol and tobacco excises in the US is diminished sharply if distributional incidence is assessed in relation to household expenditures, a measure arguably more consistent with life-cycle analysis. Lyon and Schwab (1995) use data from

Table 4.5. Income and expenditure shares of tobacco and alcohol, United Kingdom, 2003

Quintile group	Current income quintiles, and tobacco spending as % of total current income	Expenditure quintiles, and tobacco spending as % of total current expenditures
1 = poorest	2.37	1.60
2	2.04	1.95
3	1.53	1.69
4	0.99	1.32
5 = richest	0.56	0.92

Quintile group	Current income quintiles, and spending on alcoholic drinks as % of total current income	Expenditure quintiles, and spending on alcoholic drinks as % of total current expenditures
1 = poorest	6.09	2.53
2	4.28	4.16
3	3.84	4.86
4	3.88	4.94
5 = richest	3.10	4.72

the US Panel Survey of Income Dynamics (PSID) to compare the effects of analysing distribution based on incomes over different periods. Both alcohol and tobacco taxes appear more regressive when analysed using income data for a single year than five-year data, and a further reduction in regressivity is found if the analysis is based on lifetime, rather than five-year, income.

Table 4.5 shows a similar analysis, using more recent data for the UK. Analysed on the basis of current income, tobacco taxes look massively regressive, because household tobacco spending as a percentage of income for the poorest income quintile is some four times that for the richest quintile. On the other hand, analysed on the basis of household spending quintiles, tobacco taxes still look significantly regressive, but much less so than when based on current income. The average tobacco budget share for the lowest-spending quintile is less than double that for the highest-spending quintile. Likewise with spending on alcoholic drinks, a distributional analysis based on current income would suggest that alcohol taxes would be significantly regressive, with spending as a percentage of income by the poorest income quintile twice that of the richest. Analysed on the basis of household spending quintiles, however, alcohol taxes do not look regressive, and indeed appear somewhat progressive across the first four quintiles.

4.4.4. Specific versus ad valorem taxes

Unlike VAT, which simply taxes commodities in proportion to their selling price at one of a limited number of percentage tax rates, the separate excises for alcohol and tobacco offer—in principle at least—a wider menu of possibilities for the precise specification of the tax base. These are constrained by EU rules. For cigarettes, member states must levy an excise comprising both ad valorem (price-related) and specific (per 1,000 cigarettes) components, and the specific element should be between 5 and 55% of the total tax (including VAT) levied on the ‘Most Popular Price Category’ of cigarette in the member state concerned. In 2008–09, the UK tobacco excise (in conformity with EU rules) includes both ad valorem and specific components: cigarettes bear a specific excise of £112.07 per 1,000 cigarettes, and an ad valorem excise of 22% of the retail selling price. The UK alcohol excises, by contrast, are wholly specific taxes, based—as in all EU member states—on product volume and/or alcohol content.

Issues arising in the choice between specific and ad valorem taxes are surveyed in Keen (1998). The latter have a distinctive ‘multiplier effect’, in the sense that actions which change the producer price by some given amount will have a larger effect on the price charged to the consumer. As a result, ad valorem taxes will tend to discourage costly improvements in product quality and to promote more vigorous price competition between producers. Specific taxes, on the other hand, will tend to have an upgrading effect on product quality. The overall implications for consumer welfare and tax revenue of the choice between specific and ad valorem taxation will vary, depending on the structure of preferences and on market structure.⁵¹ Ad valorem taxation of goods will also tend to be more attractive where producers exert a degree of monopoly power and where there is little product differentiation. Specific taxes would be preferred where there are reasons to want to maintain product quality or where taxation is partly intended to affect an externality that is broadly related to the quantity of the product rather than to its value. Other considerations include the potential distributional advantage of ad valorem taxation in retaining relatively low price variants (though again this is unlikely to be a powerful concern, given the range of distribution-sensitive instruments available in the UK), the perhaps greater ease of administering specific

⁵¹ Delipalla and Keen (2006) show that, in competitive circumstances, the mix should be such as to have minimal effects on product quality, in the sense that, at an optimum, uniform intensification of both taxes would leave quality unchanged. Recognition that ad valorem taxation leads to a lower consumer price than does specific, for a given level of revenue, in circumstances of imperfect competition, dates back to Musgrave and Suits (1953), and is explored further in Delipalla and Keen (1992).

taxes (which require physical checks rather than valuation), and the differing responsiveness of revenues to variations in the underlying producer price (being more stable under specific taxation, for example, when the demand elasticity is low—but not a concern if tax rates can be adjusted frequently relative to these underlying price movements).

For tobacco products, the importance of potential differences in product quality—the wide variation in prices across bands suggests that product differentiation in this market is not simply horizontal—point to substantial reliance on specific taxation. The most plausible argument for ad valorem taxation is likely to lie in market dominance concerns, but the objective of maintaining a low price for the final product does not sit easily with the corrective arguments for tobacco taxation. As for the form of the specific element of the tax base, EU rules require this to be based on the number of cigarettes, in contrast to earlier UK excises based on weight of tobacco (a shift which led to considerable changes in the pattern of competition within the industry). It is sometimes suggested that there might be merit in targeting more precisely the active or harmful content within cigarettes (taxing in proportion to nicotine or tar content), but the case for this is somewhat undermined by evidence that smokers can compensate for changes in strength by smoking weaker cigarettes more intensively (Evans and Farrelly (1998)).

In the case of the alcohol excises, one particular issue frequently raised (and where there is scope for UK reform, despite the current EU constraints on the definition of the excise tax base) is the role of alcohol content in defining the tax payable, both within a single category of drink (beer, wine, or spirits) and across the three principal categories of drink.

Spirits are currently taxed much more heavily per unit of alcohol than beer and wine. More uniform taxation of alcohol content across the three principal categories of drink has been advocated persuasively by Crooks (1989) among others. It would avoid distortion between competing drinks. It would also seem a natural way to structure the Pigouvian taxation of alcohol, in that it would appear to target the externality tax precisely to the underlying source of the externalities. Nevertheless, the issue is not straightforward, because different drinks may differ in their propensity to generate external costs per unit of alcohol content. Arguably, spirits offer the greatest potential to get very drunk very quickly, which may exacerbate some of the externality problems. If uniform taxation of alcohol content would make low-cost industrially produced spirits the cheapest form of alcohol, the shift of abusers to spirits consumption might then not be a matter of social indifference.

4.5. INTERNATIONAL ISSUES

International considerations in the setting of indirect taxes have come to be far more pressing than they were at the time of Meade (as in other areas of tax policy too), with the increased cross-border mobility of commodities, fostered within the EU by the elimination of internal fiscal frontiers, posing new and potentially profound challenges for both the VAT and excises. These concern the fundamental basis on which commodities are taxed, the difficulties created by zero-rating exports under the VAT, the treatment of border-crossing services—an issue that has required little attention until recent years—and interactions in tax-setting.

4.5.1. Origin or destination taxation?

The international norm is for commodities to be taxed where they are consumed: the destination principle. As seen above, however, implementing this—particularly with the zero-rating of exports under the VAT—has become problematic. One broad solution that might suggest itself, and is also consistent with a long-standing EU objective of allowing traders to treat sales anywhere within the union in the same way, is to shift to taxation in the place of production: an origin principle. This has brought back to the fore a long-running debate on the comparison between destination and origin taxation.

It is important first to be clearer on terminology than much of the policy discussion has been. With consumption and production occurring in different jurisdictions, there are a wide variety of ways in which commodities entering international trade might be taxed, differing in which jurisdiction (including, potentially, a supranational one) sets the applicable tax base and rate(s), which collects the revenue, and which ultimately benefits from the revenue (Messere (1994)). The current international rule, WTO-approved, is that the country in which consumption occurs does all three of these things. But such a dichotomy cannot do full justice to the range of possibilities. The point is of some importance, since, as will be seen, some of the schemes proposed to overcome present difficulties disassociate the three aspects, having, for instance, one country collecting tax but transferring the proceeds to another. For definiteness, we here take the destination (respectively, origin) principle to refer to a situation in which the applicable tax rate is set by the country of consumption (production), though as will be seen, even this needs further elaboration for the VAT.

Implementing the destination principle requires some mechanism for border tax adjustment, taking commodities out of tax in the exporting country and bringing them into tax in the importing country. Conceptually straightforward for the excises, this is achieved under the VAT by zero-rating exports and bringing imports fully into tax.⁵² This means that purchasers, whether final consumers or registered businesses, have no tax reason to prefer domestic products over imports or vice versa. This procedure typically uses border controls to monitor exports and imports, but as noted in Section 4.3.3 this is implemented in the EU in the course of traders' normal periodic returns.⁵³ These arrangements are explicitly regarded as transitional, the ultimate objective being movement to 'origin taxation' which (an example of the potential for linguistic confusion in this area) in this context simply seems to refer to the collection of tax by the country of export.

Under the origin principle, no border tax adjustment is needed for excises. Matters are not so straightforward, however, for the VAT. While an origin-based VAT (in the general sense above) could take a number of forms, the most natural is that in which the value added in the production of any commodity is taxed at the rate of the country in which it is added. This requires that VAT be charged on exports at the rate of the exporting country, but credit then given in the importing country not for the export VAT actually paid but the hypothetical amount that would have been paid had VAT been charged at the rate of the country of import.⁵⁴ This is referred to as the 'stage of processing method' by Grossman (1980) and the 'notional credit method' by Genser et al. (1995).

These then are quite different methods of taxing commodities entering international trade. The destination principle is well established as the international norm. Nevertheless, the transitional nature of current arrangements within the EU, and more generally the difficulties that federal countries experience in operating destination-based invoice-credit VATs levied by

⁵² Unregistered businesses and cross-border shopping final consumers are currently taxed on an origin basis in that sense (for the former, within limits of reasonable personal use). Mail order firms and other distance sellers are required to register for VAT in countries to which they sell once their sales exceed threshold levels.

⁵³ A second-order distortion this creates is the possible cash flow advantage which may arise as a result of the tax treatment of intra-Community transactions. In the present system this favours importing goods from a supplier in another member state over purchases from a domestic supplier, although the effect is small, and depends on the timing of VAT payments and recovery. Vanistendael (1995) gives greater weight to this issue.

⁵⁴ Suppose for example that a good with producer price €100 is exported from a country with a VAT rate of 10% to another where the VAT rate is 25%: the tax-inclusive price to the importer will then be €110 and the available credit $(€110/1.25)0.25 = €22$. This ensures that the part of the value that is added in the country of export is taxed at 10% and sets the stage for taxing further value added beyond that at 25%.

lower-level governments (which in turn arise from the same problems of zero rating exports in the absence of border controls as now faced in the EU) suggest that, as international integration deepens, so the case for applying VAT on a destination basis—or at least the way it is currently implemented—requires reconsideration.

Conditions for equivalence

At first sight, origin taxation seems to run counter to the whole thrust of EU policy to eliminate impediments and distortions to intra-EU trade: it would mean that a British firm purchasing goods from other member states would no longer treat the VAT rates of other member states as a matter of indifference. In comparing possible suppliers in Denmark and Germany, for example, it would need to take into account the respective VAT rates there (currently 25% and 17.5% respectively), since these would affect the input tax credit available. It might seem that such a system would involve massive, and costly, tax-induced distortions of business purchasing decisions, unless severe restrictions were placed on member states' powers to determine their own VAT rates.

It is certainly likely that the system would be *perceived* as one in which tax differences between member states are much more likely to distort business purchases than under the 'level playing field' provided by the destination principle. Cnossen and Shoup (1987), in reviewing the relative merits of origin and destination principles for VAT, see this as a key obstacle to adoption of an origin-basis VAT and it is hard to disagree. Nevertheless, they and many others have noted that there are circumstances in which origin and destination principles are equivalent in the sense that they lead to exactly the same patterns of trade and economic welfare. Despite appearances, the playing field is then no more level under one principle than under the other.

The most general statement of such results is in Lockwood, de Meza, and Myles (1994a), but the gist is easily stated. Start with the simple case of a world lasting for a single period and comprising two countries, each of which taxes all commodities at a uniform proportionate rate (which may differ between the two countries). It is then easily seen that the shift from destination to origin principles need have no effect on the real allocation of resources, with no need even for compensating international transfers. All that is required to restore the initial allocation of resources is either a devaluation by, or a reduction in the general price level of, the country with the higher tax rate: this will ensure, for instance, that the change of basis has no effect on the foreign currency price of its exports. There are then neither substitution effects

through changes in relative prices nor income effects through tax revenues. The intuition is straightforward. A uniform tax levied on a destination basis is a tax on the value of aggregate consumption, while a uniform tax levied on an origin basis is a tax on the value of aggregate production; and with balanced trade, these are the same thing.

Recent work has shown the equivalence between destination and origin taxation to hold in rather more general circumstances too. It continues to hold, for example, under imperfect competition, and in intertemporal trade models which have the feature that trade must be balanced in present value.⁵⁵ Striking as these equivalence results are, however, their practical applicability is likely to be limited:

- *Problems with partial application.* A ‘restricted origin’ regime—in which the origin principle applies to trade between some subset of countries (such as the EU member states) while trade with the rest of the world remains subject to the destination principle—raises two sets of difficulties (Shibata (1967)). First, it opens up the possibility of trade deflection: traders operating in a high tax member state, for example, would have an incentive to export to a low tax member via the rest of the world (so incurring the lower tax rate of the country of final consumption)⁵⁶ rather than directly (incurring the high rate of their own jurisdiction). Second, since trade flows between the countries adopting the origin system need not be balanced, a change in tax basis generally redistributes revenue between them (with net exporters on trade within the group being the winners); this latter is less of a concern in pure efficiency terms, since it could in principle be undone by a simple lump sum transfer between countries. As Lockwood, de Meza, and Myles (1994b) observe, both problems can be avoided if member states adopt the origin principle not only with regard to each other but also with regard to the rest of the world: a ‘non-reciprocal restricted origin system’. This provides a clear analytical solution, but it is equally clear that it would be hard to persuade practitioners and policy-makers of the neutrality of a scheme in which exports from the EU are taxed twice (once in the EU, once in the country of destination) while imports into it are not taxed at all.
- *Intergenerational redistribution.* A change of tax basis can have complex and potentially powerful intergenerational effects (Bovenberg (1994)). A shift from destination to origin taxation, for example—somewhat akin to a shift from taxing consumption to taxing wages, as discussed in

⁵⁵ Genser, Haufler, and Sorensen (1995).

⁵⁶ The story is complicated, but not fundamentally overturned, by the possible existence of tariffs.

Section 4.2—would tend to benefit the relatively wealthy and old (financing high consumption levels from past earnings) at the expense of the relatively young and poor (saving a high proportion of their factor incomes), with potentially adverse effects on capital accumulation.

- *Non-uniform taxation.* Unless all commodities are taxed at the same proportionate rate in each country—or a somewhat weaker similarity of tax structures holds⁵⁷—adjustment of the exchange rate or general level of internal prices alone cannot offset the effects of a change in tax basis. In practice, of course, indirect tax systems are far from uniform, reflecting not only multiple rates of VAT but also the excises. There appears to be little evidence, however, on whether the non-equivalence that this implies is in practice likely to be quantitatively substantial.
- *Altered incentives in tax-setting.* Still more fundamentally, the equivalence results presume that tax rates remain unchanged with a shift from one basis to another. National incentives in setting those tax rates, however, are likely to be quite different in the two cases, a point returned to shortly.

When equivalence fails, which is better?

Since origin and destination bases are thus not in general equivalent, the question arises: which is to be preferred? Three groups of issues are relevant.

A first set concerns that comparison if taxes are set cooperatively. Attention then focuses on the quite different arbitrage conditions they imply. Under the destination principle, residents in any country pay the same tax on both imports and domestically produced items, and are thus led to equate producer prices across countries. With perfect competition, this in turn implies that relative marginal costs prices are equated across countries: which means an efficient pattern of production. Under the origin principle in contrast, consumer prices will tend to be equalized across countries, leading to exchange efficiency (that is, an efficient allocation of consumption across countries). The choice thus resolves to one between production and exchange efficiency. It might then seem that, under conditions outlined there, the Diamond–Mirrlees (1971) theorem discussed in Section 4.2 applies, so that production efficiency has priority. As Keen and Wildasin (2004) point out, however, the Diamond–Mirrlees theorem is not directly applicable in international settings when distinct governments face distinct revenue constraints.

⁵⁷ When tax rates vary across commodities, devaluation or general price movements can neutralize the real effects of a change in basis if for every commodity the tax factor (unity plus the ad valorem tax rate) in one country is the same multiple of the tax factor in the other. Uniform taxation within each country is sufficient for this, but not necessary.

In the absence of international transfers, Pareto-efficient tax design may then require deploying taxes on trade or other instruments that drive a wedge between producer prices in different countries. Given the ability to transfer resources across members of the EU, however—effectively consolidating, to some degree, their governments' budget constraints—their results suggest a presumption for production efficiency within the union.

Even with cooperative taxation, however, any presumption for destination taxation becomes less clear-cut under conditions of imperfect competition. The destination principle no longer implies production efficiency, because equalizing after-tax prices across producers does not necessarily equalize marginal costs (Keen and Lahiri (1998)); in addition, taxing intermediate transactions could in principle be desirable (in the absence of more directly targeted measures) to offset monopolistic output distortions.

Second, important distinctions arise if taxes are set non-cooperatively. The incentives for strategic tax-setting are fundamentally dissimilar between the two: towards base-stealing and (pushing in the opposite direction) tax exportation under the origin principle; towards taxation of importables (to induce favourable movements in the terms of trade and to some degree capturing profits otherwise accruing to foreign producers) under the destination principle. Not surprising, the two regimes cannot in general be unambiguously ranked (Lockwood (1993); Keen and Lahiri (1998)).

Third, there is a major practical difference between the two: under an origin-based VAT, producers would have an incentive—not present under the destination principle—to transfer price value-added into jurisdictions characterized by relatively low effective rates on value-added (Genser and Schulze (1997)). Given the considerable difficulties faced in monitoring transfer pricing devices driven by corporate taxation, this is a potentially major drawback.

Implications

Is there then a case for shifting from the destination to the origin principle? One advantage is that it would resolve the present asymmetry between the tax treatment of cross-frontier purchases of goods by individuals and by businesses. Under the current VAT system, differences in tax rates between countries are sustainable only to the extent that there are significant transport costs or other impediments to individual purchasing. If the pressure on the current system from individual cross-border shopping intensifies greatly (as anticipated by Sinn (1990)), movement to the origin principle could be the best (or only) option available. It is also the case that recent results have placed the origin principle in a less-unfavourable light than before. But it remains

hard to make a strong case for clear gains from moving to its wholesale adoption, at least for the foreseeable future. Doing so would throw away the potentially useful ability to tailor national tax systems to national preferences, and risks inducing unwelcome production inefficiencies. It would fundamentally change strategic incentives in tax-setting, with unclear consequences. And it would open up potentially powerful possibilities for transfer pricing. Abandoning the destination principle in favour of a full-blown origin system seems unlikely to offer sufficient gains to offset these various costs and risks.

4.5.2. Alternative VAT mechanisms for intra-EU trade

There are therefore good reasons to retain the destination principle as the underlying basis for the treatment of international trade. In relation to the VAT, however, there remain a range of important issues about the practical mechanism, export zero-rating, which currently implements it. For it is important to recognize that although both the zero-rating of exports and the adherence to the destination principle are effectively universal, the former is not a necessary condition for the latter. Indeed, export zero-rating has two significant drawbacks. One, discussed in Section 4.3.3, is that it creates problems for VAT enforcement. It breaks the chain of VAT revenue cumulation whenever the chain of production and distribution crosses national boundaries, a point of particular enforcement vulnerability. Goods supposedly exported, and therefore zero-rated, may be diverted to the domestic market bearing no tax. And the refund of VAT to exporters coupled with the deferred payment of VAT on imports gives scope for profitable criminal exploitation through large-scale 'carousel frauds' and similar schemes.

The second drawback, highlighted in the European Commission's 1987 proposals for an alternative VAT mechanism to accompany the elimination of intra-EU border formalities at the end of 1992, is the sharp difference in VAT procedures applied to domestic sales and exports when exports are zero-rated. As a consequence, businesses' VAT compliance costs are increased by the need to treat differently domestic sales and exports to other EU member states. Views differ on the importance of removing such differential treatment, and achieving what Keen and Smith (1996) call 'compliance symmetry': it is not achieved, for example, within the US, given different treatment of inter- and intra-state sales. Nevertheless, it has played a significant role in the EU debate, where removing impediments to intra-union trade, not least by smaller and medium-sized enterprises (most of them likely to be within the VAT system, even with a threshold as high as in the UK).

Much of the '1992' programme of measures to complete the internal market of the Community was motivated by a concern that border formalities could increase the costs to a firm of doing business in other member states. Indeed, there was a concern that on occasion member states may have employed frontier bureaucracy as a form of trade protection against products from other member states. In order to remove the opportunities for such non-tariff barriers to arise, the 1992 programme abolished internal fiscal frontier formalities. In addition, in its original proposals, the Commission sought (but failed) to put in place after 1992 a VAT mechanism for cross-frontier transactions which enabled registered traders to treat sales to all member states, including their own, identically so as to ease closer integration within the internal market. In practice, however, the measures adopted, which continue to apply, treat trade within and between member states very differently; and the new procedures then introduced to prevent VAT evasion on international transactions in the absence of frontier controls may well have magnified tax compliance costs on export transactions (Verwaal and Cnossen (2002)).

Several VAT mechanisms have been proposed that would achieve economic outcomes conforming to the destination principle but without zero-rating exports:

Exporter rating

Export rating, advocated by Cnossen (1983) and adopted as the initial proposal of the Commission (1987) for the post-1992 world, means taxing intra-EU exports at the rate of the country from which the goods were exported, with credit then being available in the country of import and some clearing house mechanism put in place to ensure that revenue ultimately accrues, as at present, to the country in which consumption takes place. Exports and domestic sales would thus be taxed identically, which may be an advantage, since it reduces the danger that firms may be deterred from exporting by the need to deal with an unfamiliar export tax regime.

The central problems with this scheme relate to the operation of clearing. Under the initial proposal this was to have been done on the basis of individual transactions, necessitating complicated (and permanent) arrangements for measuring the required clearing flows. This would suffer, moreover, from the fundamental flaw of undermining the incentive for member states to check the validity of claims for VAT credit on imported goods: the cost of giving this credit would be underwritten by the Clearing House, and there would be little incentive for member states to spend resources in reducing

fraudulent claims (Lee, Pearson, and Smith (1988)). An alternative would be a one-off compensation settlement (perhaps involving agreed annual revenue flows), reflecting the scale of the anticipated revenue changes. But while this would restore the incentive for member states to detect fraudulent claims for VAT credit on imports, it would introduce a new problem, giving rise to undesirable incentives for member states to raise their VAT rates so as to increase revenues from the taxation of exports. Since the importing member states would be required to give credit for whatever rate of tax was applied to exports, there would be no competitive restraint on this; the only limit would be the willingness of domestic customers to accept the higher VAT rates that would also apply to domestic sales.⁵⁸

Uniform rating

Under the uniform rating scheme, intra-EU exports would continue to be zero-rated for the national VAT in the country of export but would then be subject to a uniform VAT rate, determined by the Community. This fixes, to the extent of that uniform VAT rate, the break in the VAT chain on intra-union exports, though compliance symmetry is not achieved, since exports and domestic sales are subject to different tax treatments. Such uniform rating can in turn be achieved in a number of ways.

CVAT

In the CVAT regime advocated by Varsano (2000) for Brazil (which faces similar problems subnationally) and McLure (1999, 2000) more generally, this special tax rate for exports is operated as a separate tax, by a single tax authority operating across all the jurisdictions concerned. Since the additional tax on intra-EU exports (the ‘compensating VAT’) would be collected by the same authority that then gives credit for the input tax paid on imported goods, and since these amounts in principle cancel out—except that revenue would be raised to the extent of imports by other than registered traders—this system avoids any revenue redistribution between member states without any need for clearing between them. But it does so by establishing a parallel VAT operation and bureaucracy (for the taxation of intra-EU trade) that generates little net revenue—which seems likely to be a hard political sell.

⁵⁸ While Article 90 of the Treaty of Rome proscribes the use of domestic taxes to provide indirect protection, it is not clear that the use of domestic taxation as a tax exporting device is similarly prohibited (just as WTO rules do not prohibit export taxes). Experience with Article 90 suggests, in any event, that it is effective only against the most egregious uses of domestic taxes to such trade-related ends.

Quite how large the administrative and compliance costs would be in this parallel VAT administration is unclear, and it may be able to draw heavily on VAT information already gathered in national systems. But unlike exporter rating there is a clear danger of asymmetry in VAT compliance costs between exports and domestic sales.

One of the two options recently proposed by the Commission for dealing with VAT fraud—the other (with many hesitations) being reverse charging, as discussed in Section 4.3.3—applies essentially the CVAT logic, in applying a uniform common VAT, at 15%, on trade between member states (Commission (2008)). Implementation, however, would remain wholly with national tax agencies, with proceeds from export VAT being reallocated to the country of import. This has the substantial appeal (in contrast to reverse charging) of strengthening the VAT chain. But by opting against separate administration of the CVAT—no doubt for understandable political reasons—it effectively dissipates the key advantage of the CVAT in avoiding any need for revenue reallocation. And it does nothing to simplify VAT compliance or render it more symmetric.

Dual VAT

Some similar properties are exhibited by the dual VAT regime proposed by Bird and Gendron (1998), who draw on the Canadian experience from the combination of a federal VAT and the provincial VAT levied in Quebec. This envisages the continuation of national VAT systems, including the zero-rating of intra-union exports, but run in parallel with a new Community VAT, set at a uniform rate EU-wide, and operated without regard to national boundaries. This Community VAT generates revenues for the Community, and at the same time ensures that intra-EU exports, although zero-rated by the member state VATs, do move between member states bearing at least some tax. An EU-wide VAT has some appeal as a source of own resources for the Union (being more transparent, for instance, than the current use of a notional VAT base to determine, in part, national contributions). The political prospects do not currently appear bright, however, and the revenue need would readily be met by a VAT rate so low—in the order of 2–3% would suffice to finance all the Unions' activities—that the impact on fraud is likely to be limited. Compliance symmetry would also not be achieved.

VIVAT

The main feature of the VIVAT, proposed by Keen and Smith (1996, 2000), is that a uniform Community-wide rate of VAT would be applied to

transactions between all VAT-registered traders in the EU (perhaps with a rate of zero applied to financial services as part of the cash flow approach described in Section 4.3.2), while member states would retain the power to determine the rate of VAT on sales by traders to final consumers. This would achieve compliance symmetry, fix the break in the VAT chain between member states (with a firmness that depends, of course, on the level of the intermediate rate), and avoid some of the ‘incentive’ defects of the other main schemes, in terms of both enforcement and rate-setting incentives, noted above. And it does so without altering in any way the current ability of member states to choose the rate applied to final sales, and hence—since it is that, not the rate applied to intermediate transactions, which determines it—the total revenue finally raised (at least in so far as the crediting mechanism works properly). There is thus, in that sense, no real loss of national sovereignty. While some form of clearing would be needed to ensure that all revenue continues to accrue to countries of final consumption, the VIVAT lends itself more readily than does exporter-rating to some form of macro-related settlement—so ensuring adequate incentive, for instance, to control claims for credit of intermediate tax on imports—since the commonality of the intermediate rate removes scope for game-playing.

A further attraction of VIVAT is that it could ease some of the difficulties created by the current VAT exemptions, discussed in Section 4.3.2. The application of different national rates currently distorts competition (reflecting the cascading effects of differing degrees of unrecovered VAT implicit in prices paid for financial services, for instance) and requires complex (almost unworkable) rules for hospitals, universities, local governments, and others to declare their purchases in other member states so as to negate the incentive to buy abroad purely for tax reasons by ensuring that the appropriate VAT adjustments are made (by the revenue authorities of the importing country). A VIVAT would enable these distortions to be removed simply by enabling those producing exempt items to purchase at the intermediate rate, removing both the competitive distortion and the need to adjust tax on purchases from other member states. This would not remove the distortions that the widespread exemptions in the EU create, so the case for reforming them out of the system, stressed in Section 4.3.2, would remain. But it would ease them.

The principal disadvantage of VIVAT, beyond the need to arrange revenue transfers between member states, is that it requires an ‘end user’ distinction to be made—and enforced—between the sales which a business makes to other VAT-registered businesses and those sales it makes to final consumers: between, that is, B2B and B2C transactions. These would be taxed differently

under VIVAT, and there would be additional compliance costs to businesses and extra administration costs for the tax authorities in accounting separately for the two categories and in handling difficult borderline cases. Assuming that the VAT rate on intermediate sales was never higher than the rate on final sales (the uniform Community-wide rate on intermediate sales might be 15%, say, with rates on final sales ranging as at present from 15% to 25%), the central risk is of sales to final consumers being disguised as sales to registered traders. This would of course rely heavily on the use of VAT registration numbers, as at present, to identify VAT status. But it would be possible to apply the rules rather more stringently than at present, without serious damage to the firms concerned. If a firm failed to substantiate a claim to be allowed to apply the intermediate rate, it would have to apply the final goods rate, which in some countries would be very little higher. (It might also be possible to credit input VAT claims at that rate where it could be shown that the final consumer rate had been wrongly applied to an intermediate goods transaction.) And it may even be that the opportunity to charge an inappropriately low rate would in some cases reduce VAT evasion on final sales: instead of entirely concealing transactions and remitting no VAT, the better bet for those so inclined could be to offer to charge and remit the intermediate rate.

More fundamentally, however, the central distinction needed to operate a VIVAT, between B2B and B2C transactions, is one that is in any event likely to become increasingly important in the operation of the VAT, and is indeed at the core of recent initiatives in connection with the VAT treatment of international services.

4.5.3. VAT and internationally traded services

International trade in services has grown much more rapidly than trade in goods in recent decades, placing increasing strain on some aspects of current consumption tax legislation and procedures, in the EU and elsewhere. The EU's VAT rules and procedures were primarily developed to handle goods trade, for which the notion of a trade transaction can be defined clearly in terms of the physical movement of the taxed commodities. For services, no such physical movement exists, making it much less straightforward to define transactions to which the tax provisions relating to international trade should apply. The main concern is that some transactions might end up untaxed or taxed twice, and that loopholes may be extensively exploited in the organization of international services businesses. Similar anxieties arise

about the continuing viability of consumption tax systems in the face of the rapid growth of e-commerce. This has undermined national regulation and taxation in certain areas (such as gambling services), and turned some readily taxed goods (records and video-tapes, for instance) into hard-to-tax quasi-services (digital downloads).

Many of the problems in the VAT treatment of traded services can be traced to the ‘absence of internationally agreed principles to determine where taxation should take place’ (OECD (2004)): in other words, to the lack of a systematic and uniformly applied definition of the ‘place of supply’. Current EU rules provide ad hoc treatment for particular categories of service. Some are taxed on the basis of the place of the supplier, others on the basis of the place of consumption; further complexity arises through the use of proxies (such as the place of establishment, or the place of performance) to define the place of either production or consumption of certain services.

The consequences of alternative approaches to the taxation of international services vary considerably across different types of transaction:

B2B transactions between VAT-registered traders

Which side of the transaction applies VAT is in this case unimportant in terms of economic efficiency because the tax rate has no ultimate impact: any tax paid is recovered later through the natural mechanism of the VAT (assuming, as current EU rules seek to ensure, that recovery operates effectively across any borders that the transaction spans), and there is no advantage (apart from second-order cash flow considerations) to purchasing from a lower-taxed source. Having the seller charge tax—by deeming the place of supply to be that in which they are located—has the merit of securing revenue by maintaining a chain of VAT payments, but requires some mechanism for reallocating revenue if it is to accrue to the country in which final consumption occurs. Instead having the buyer reverse charge themselves tax on such purchases—deeming the place of supply to be where they are located—avoids any need for revenue reallocation, but at the cost of breaking the VAT chain.

Sales to exempt traders

Since VAT paid on inputs then ‘sticks’, the VAT treatment (in particular, the tax rate at which supply is taxed) matters for efficiency. The problems here mirror similar problems in the taxation of domestic sales of services to exempt firms, and again point to the importance of scaling back exemptions as discussed in Section 4.3.2. These difficulties are exacerbated to the extent

that domestic and cross-border sales of services might be taxed differently: for example, if the former is subject to normal VAT, and the latter zero-rated.

B2C sales

Distortions arise if some final consumers face different tax rates for similar products. For goods, these problems arise primarily through cross-border shopping, because the country of purchase may differ from the country of consumption, or of residence of the final consumer. Distance selling (mail order) raises similar issues of distortion, mitigated by the special regime applied to such trades. For services, distortions would arise through the equivalent of cross-border shopping or distance selling: in other words, the possibility that a consumer might be able to obtain equivalent services from suppliers located in different member states, and these might, under some definitions of the place of supply, be taxed differently.

Key issues

The key efficiency issues in defining the Place of Supply for services thus primarily concern cross-border sales of services to final consumers and entities not registered for VAT, and are broadly analogous to the issues of cross-border shopping or distance selling for goods.

As with cross-border shopping for goods, the practical significance of the problem, in terms of the risk of significant economic distortion, will vary considerably between products. At one extreme there are services which are effectively different if they are performed in different locations. Examples include passenger transport (a tram journey in Brussels is no use to the Manchester commuter trying to get to work), service performed on fixed property (by a plumber, for example), and various entertainment services (a cinema ticket in Stockholm is not a close substitute for a ticket for the identical film in Rome). At the other extreme there are services (including most e-services) where the location of the supplier is irrelevant to the performance of the service, and where a private customer could, in principle, purchase the service from a supplier in another member state without any penalty in terms of cost, inconvenience, or quality as compared with the equivalent domestic supply.

For the first group of services, the location of the customer coincides with the location of the supplier when the service is performed, and defining the place of supply as either would have an equivalent effect on economic efficiency. Problems of distortion in purchasers' decisions only arise if the

place of supply is defined in terms of the location of the supplier, and if this can differ from the place where the service is performed. For the second group of services the definition of the place of supply is an issue of much greater significance for economic efficiency. For these services, distortion in sales to final consumers and B2B sales to exempt traders can be avoided only by taxing where the customer is located.

There is no simple way, however, of implementing this outcome.

- If the supplier remains liable for the tax, a small-volume EU supplier with customers spread thinly across all member states would have to deal with twenty-seven separate tax authorities. A potential solution to this would be some form of ‘one-stop shop’, allowing a supplier to deal with a single agency in respect of all EU-wide sales. There are various ways in which this could be organized, but the key idea is that the agency would charge VAT at the rate appropriate to each country, and remit the tax collected, and information about the transaction, to the VAT authorities in all member states to which sales were made.
- In some specific cases, such as digital downloads, the supplier may have no idea of the location of the customer, and so be unable to apply tax on the basis of the customer’s location. This suggests that some EU-wide regime (a common rate of tax on all exports of digital services, or even on all sales of digital services within the EU) would be the most appropriate solution.
- If the supplier is located outside the EU, enforcement poses evident difficulties.

And while the alternative of making the customer liable for the tax on purchases of services is in principle a complete solution to the problem of ensuring economic efficiency, there are considerable practical difficulties. For B2B transactions, such reverse charging breaks the VAT chain. And for final consumers, it is evidently unworkable—the experience of the use tax in the US, which in principle requires taxpayers to declare out of state purchases but in practice is almost entirely ignored, illustrates the point.

The approach recently adopted by the EU (with effect from 2010), which follows OECD guidelines, is to tax B2B service purchases in the place where the customer is located, but B2C purchases where the supplier is located (at the rate applicable to the country of the consumer).⁵⁹ The former opts for

⁵⁹ Council Directive 2008/8/EC of 12 February 2008 amending Directive 2006/112/EC as regards place of supply of services. Application to telecoms is from 2015, and a range of special treatments apply.

avoiding revenue allocation. This is clearly a more coherent approach than at present, albeit one with risks: of ensuring that the break in the VAT chain for B2B trade is not exploited, and of identifying the proper rate at which to charge final consumers and implementing a transactions-based reallocation of revenues. Interestingly, it is built on precisely the distinction between sales to registered and to non-registered traders that is at the heart of the VIVAT, but without securing the full advantages that it offers in maintaining the VAT chain and ensuring compliance symmetry. A more thoroughgoing application of the VIVAT logic would not avoid all the difficulties in this area—such as, in some cases, identifying the location of final consumers—but could provide a more coherent framework for addressing these as other issues.

4.5.4. Excise tax competition and cross-border shopping

Tax-induced cross-border shopping and smuggling pose significant policy challenges for the EU. Not only do they cause a direct revenue loss—or perhaps gain to low tax countries—but they are likely to lead to tax rates being set lower than they otherwise would be.

The extent and impact of cross-border shopping

The problem of revenue losses from legal cross-border shopping by individuals has come into prominence as a result of the abolition of border controls between EU member states at the end of 1992. Before this, most member states applied restrictive travellers' allowances on personal imports of tax-paid goods from other EU countries, as well as from the rest of the world, and this kept legal cross-border shopping within tightly constrained bounds. As a result of the abolition of border controls, individuals can purchase goods in another member state and bring them home without restriction or fiscal adjustment, so long as the goods are for their personal use and not for resale.

The scale of revenue losses from both legal and illegal cross-border movements of alcohol is controversial. HM Customs and Excise estimated the UK revenue loss from *legitimate* cross-Channel shopping for alcohol in 1998 at £285 million, about 5% of total UK alcohol duty revenues. Some part of this revenue loss would have arisen as a result of duty-free purchases, rather than tax-paid purchases in other member states, though it is a reasonable guess that a significant part of the alcohol purchased duty-free would have been

purchased outside the UK in the absence of the opportunity to make duty-free purchases (Christiansen and Smith (2004)).⁶⁰ HM Customs and Excise estimated that *illegal* cross-Channel smuggling of alcohol in 1998 involved a revenue loss to the UK of some £230 million, around 4% of total alcohol excise revenue.

For individual member states in this position, alcohol tax policy needs to take account of the potential revenue losses through legal and illegal cross-border activities. Crawford and Tanner (1995) and Crawford, Smith, and Tanner (1999) consider whether the revenue losses through cross-border shopping induced by duty differentials are sufficiently large that UK revenues could be increased by cutting the rates of UK excise duties. They observe that the post-1992 abolition of border controls acts so as to increase the price elasticity of demand for UK-bought alcohol: as the price is increased, some consumers reduce their consumption of alcohol altogether (the normal effect, in the absence of cross-border shopping opportunities), but in addition others may now switch to buying abroad. The higher elasticity for UK-bought alcohol might suggest that UK tax rates on alcohol should be reduced. Indeed, if the increase in elasticity is sufficiently large, it is possible that the existing rates of duty might exceed the revenue-maximizing duty rates.

Using data from the UK Family Expenditure Survey for the years spanning completion of the Single Market, Crawford, Smith, and Tanner (1999) find, however, no evidence of a significant change in elasticities. Whether a reduction in duty rates on the various categories of alcoholic drinks would increase or reduce UK tax revenues depends on the own- and cross-price elasticities of demand for the various categories of alcoholic drinks. Crawford et al. find that the UK tax rates on beer and wine are still lower than revenue-maximizing tax rates, meaning that a cut in duty on these drinks would reduce total revenues (even though it would repatriate a certain amount of cross-border shopping). On the other hand, they cannot reject the hypothesis that the current UK tax rate on spirits is at around the revenue-maximizing rate. This suggests that the duty on spirits may be closer to the level at which a cut might reduce cross-border shopping by enough to compensate for the revenue lost on each unit.

⁶⁰ In 1999 the EU countries abolished duty-free sales on intra-EU travel. Christiansen and Smith (2004) note that the global welfare gains from abolition of duty-free may be asymmetrically distributed between countries, and that high-tax countries may have relatively less to gain from its abolition than lower-tax countries; in particular it may deny them an opportunity to appropriate rents from sales to foreign travellers that would not otherwise be made. Christiansen and Smith (2008) discuss the implications of duty-free sales for the optimal structure of commodity taxes.

Policy problems and responses

The question for policy is what, if anything, should be done about the implications of such potential tax-induced cross-border purchases, legal and otherwise. In assessing this, it is important to remember, the extent of the problem cannot be inferred simply from the extent of observed cross-border shopping. Observing zero cross-border shopping could mean that there is simply no inclination to exploit tax differentials, as would be the case with literally non-tradable goods. But cross-border shopping would also be zero in equilibrium if intense tax competition drives all countries to charge the same excise tax rates, with all standing to benefit from setting a collectively higher rate.

In response to EU proposals to require member states to set VAT rates within specified bands, Lee, Pearson, and Smith (1988) argued that fiscal externalities arising from cross-border shopping between member states could warrant some level of EU tax coordination. However, since these negative fiscal externalities primarily arise where member states cut their rates below those of neighbours, this would indicate an EU-wide floor to excise duty (and VAT) levels, but no corresponding need to place an upper limit on member states' tax rates.⁶¹

The subsequent formal literature on tax policy in the presence of cross-border shopping has explored in more detail the potential impact of cross-border shopping (including for this purpose smuggling of tax-paid goods) and the tax competition it can lead to, and the appropriate policy responses. Using a stylized two-country model in which governments set their tax rates non-cooperatively, each concerned only to maximize its own tax revenue, Kanbur and Keen (1993) show that⁶² imposing a minimum tax rate (a tax floor) leads both countries to gain relative to the non-cooperative outcome: even the low tax country benefits despite being forced to raise its tax rate, because this enhances the ability of the other country to increase its rate. Simply imposing a common rate, on the other hand, must cause the low tax country to lose, unless the common rate is set at a sufficiently high level.

⁶¹ A ceiling on tax rates could be justified where countries have a degree of monopoly power, which would otherwise permit tax exporting.

⁶² In the Kanbur–Keen (1993) framework it is the smaller country that sets the lower tax rate in equilibrium (since in setting a lower rate than the other it has more to gain by attracting the large tax base from abroad than it has to lose from reducing revenue from its own domestic tax base). Location may also matter for tax-setting incentives: Ohsawa (1999) shows that with three otherwise identical countries ordered on a line, that in the middle tends to set the lower tax rate as it has more borders to attract tax base across. Keen (2002) reviews the literature on cross-border shopping.

Some argue, however, that unrestricted tax competition may be a useful mechanism to control the tendency of governments to excessive growth. The case for the EU imposing some form of duty coordination—and indeed potentially coordinating in other areas of tax policy too—thus to some degree turns on fundamental differences of philosophy about the nature and value of governments. Those who view governments as choosing tax and spending policies to maximize social welfare may regard tax coordination as a way to ease what would otherwise be an undesirable constraint on their choices, while those who believe that there is a natural tendency towards excessive growth in government spending may view tax coordination in a less positive light. Even in this latter case, however, some degree of coordination may be desirable if—as seems plausible—policy-makers are not simply Leviathans having no selfless concerns and facing no electoral constraints.⁶³

These considerations suggest to us a strong case for the agreement of minimum excise tax rates. And this is indeed the strategy that the EU has followed since the advent of the single market. The key issue is not the wisdom of the approach, but the low level—zero, in some cases—at which these minima are set. Raising them to more appropriate rates has, however, proved difficult, and this is unlikely to change.

4.6. CONCLUSIONS

One of the most remarkable developments in UK tax policy in the thirty years since Meade has been the shift towards indirect taxation, with the VAT—still new at the time of the report—coming to acquire a pivotal role in the overall tax system. Some aspects of this development, and of indirect tax design more generally, remain less than fully understood. In particular, the proper broad balance between direct and indirect taxes turns on relative efficiency and implementation costs about which little is known. And while much has been learnt about the theoretical case for some degree of differentiation in the rates of indirect taxation across commodities, limitations of empirical knowledge and understanding of implementation costs still leave some disconnect between these results and their practical application. What does seem clear, nevertheless, is the potential for building on the inherent strengths of the VAT to improve its effectiveness within the wider UK tax system, and the need for clearer thinking of the role of the traditional excises.

⁶³ On this wider political economy issue, see Edwards and Keen (1996) and Besley and Smart (2007).

4.6.1. Rate differentiation

Theory, much of it having come to prominence since Meade, suggests that when—as in the UK—government has at its disposal a fairly sophisticated range of instruments for redistribution, the potential contribution of commodity taxes to efficient revenue-raising is likely to be limited. Certainly there are grounds to tax at particularly high rates goods and services associated with external costs in either production or consumption, an issue explored at more length in Fullerton, Leicester, and Smith, Chapter 5, on environmental taxes, and relevant also to the case for the excises on alcoholic drinks and tobacco in the UK. But, except where externalities are involved, theory indicates that differential commodity taxation is unnecessary if patterns of commodity demand are independent of labour market status (the ‘weak separability’ condition). In this situation, patterns of spending convey no information that cannot be better exploited by more directly targeted redistribution, and taxes which vary with patterns of spending—the additional sophistication in the tax structure which commodity taxes offer—achieve nothing that cannot be achieved by other tax instruments. Nevertheless, the available empirical evidence firmly rejects weak separability, implying that appropriate rate differentiation could reduce the overall distortionary costs of taxation.

Precisely what kind of differentiation would be helpful, however, is much harder to say. In principle, tax rates should be chosen so as to reduce most the demand for goods that are, in an appropriate sense, complements for leisure (meaning time not spent in market work). In this way, unavoidable distortions against market work can be mitigated. The empirical work reported in Crawford, Keen, and Smith (2008) suggests that on these grounds we might think of taxing the broad category of ‘leisure goods’ more heavily than other goods and services, and domestic fuels and ‘household goods and services’ at lower rates. But the limitations and uncertainties of both the theory and the empirical work are such that, at least as yet, they provide little firm basis for policy prescription. The finding that the departures from weak separability, while significant, are quantitatively small suggests too that the gains from differentiation may not be large, though this requires more thorough investigation. And the administrative and compliance costs of rate differentiation would certainly need to be weighed against any advantage to be had in exploiting features of commodity demands. There is some evidence, and much anecdotal wisdom, that implementing such rate differentiation through the VAT would be quite costly (in terms, for example, of the likely increased need for refunds—to those selling lightly taxed items using more heavily taxed inputs—and to monitor borderline issues). Experience suggests

that implementation of multiple stand-alone excises is likely to be even more problematic. And, not least, departures from uniformity may make it harder to resist simple lobbying for preferential treatment.⁶⁴ Externality issues aside, there is thus a strong case, albeit a largely pragmatic one, for uniformity in commodity taxation, and this is indeed the path that has been taken in most of the more modern VATs that have been introduced in recent years.

Since the VAT is at present marked by significant rate differentiation, the implications of this for the UK are substantial. The present domestic zero-rating and the reduced rate serve little purpose that could not be better achieved by other means. We have shown here that eliminating both would provide enough revenue to reconfigure tax credit and income support systems so as to ensure that most poorer households gained from the reform and still leave an additional £11 billion, all with only a modest one-off impact on the price level. Similar remarks apply to the present zero-rating of residential construction. Though not explicitly analysed above, here too there is evident scope for eliminating distortions by following others in including new sales fully into the VAT while putting in place measures to compensate the losers.

The essence of this has been known for many years, at least in relation to the zero-rating of food and children's clothing. What has been lacking is political will, a first step towards forming which is likely to be better public understanding of the inherent unfairness of the current rate structure. There may even be a case for some appearance of earmarking the proceeds from such a VAT reform.

4.6.2. Exemptions, international trade, and the VAT

New at the time of Meade, the VAT is now starting to look old, and imperfectly adapted to the changed international circumstances in which the UK finds itself. This is true not only of the rate structure, but more deeply too of aspects of its fundamental structure. Two issues stand out, with the EU in each case playing a key role. The proper response, in each case, is not to compromise but rather to build on the underlying logic of the VAT.

First are the widespread exemptions that are enjoined by current EU rules, notably in relation to public agencies and health, education, medical services, and financial services. In breaking the chain of tax and credit, these violate

⁶⁴ As argued in the context of tariff formation by Panagariya and Rodrik (1993).

the logic of the VAT in a way that rate differentiation does not. There is little quantitative evidence on the costliness of the inefficiencies this creates—including the incentive to self-supply and other distortions to production decisions, increased implementation costs for partially exempt traders, the bias towards imports—but good reason to suppose they are not trivial, and greater now (as a consequence of privatization and increased financial sophistication) than they were thirty years ago. And at a technical level it is now much better understood—and illustrated by experience in New Zealand, discussed in the commentary by Dickson and White—that these exemptions are not inescapable.

The case for considering alternatives to exemption is especially clear in relation to financial services, though this is one area in which no country has implemented a full solution. One possibility discussed above is to apply a cash flow treatment to financial services—treating all inflows as VAT-able purchases and outflows as sales—while zero-rating transactions with registered businesses. Movement to such a system would not only remove distortions but likely to enable a substantial net revenue gain. And while much closer examination is clearly required, it could fit neatly within the VIVAT form of VAT that we believe has other attractions too.

Eliminating these exemptions would require, however, a change in EU rules, which in turn requires unanimity among the twenty-seven member states. It is not hard to imagine change in the VAT treatment of the financial sector, in particular, being vetoed. This raises wider questions as to the appropriateness of EU provisions that may impede member states' adopting measures that may improve the operation of their domestic VAT systems without having any obvious adverse spillover effects. The first step, in this as perhaps in other areas, may be to allow some degree of experimentation by member states.

The second key structural issue is the VAT treatment of trade between EU member states. Ironically, while the VAT proved well-suited to the early stages of integration, as a means of limiting hidden protection and export subsidies, it has proved problematic (as seen in Section 4.3.3 above) at the later stage in which internal fiscal controls are removed. It is important not to over-react to much-publicized missing trader frauds exploiting current arrangements: they are but one part of an overall level of VAT abuse that may be broadly comparable to that under the personal income tax. Nevertheless, action is clearly needed to strengthen the integrity of the VAT.

Part of this response must of course be improved administrative cooperation, but the greatest assurance in addressing the underlying structural problem—the break in the VAT chain from the zero-rating of intra-union

exports—is likely to come from a structural solution. For this it is not necessary to move towards a genuine origin principle for commodity taxation, and indeed doing so would be unwise (not least because it would introduce transfer pricing problems into the domain of the VAT). The most promising approach is to strengthen the implementation of destination-based taxation by building on the fractional nature of the VAT, which is its fundamental strength—not to undermine it (as would be done, for example, by further reverse charging, whether for selected commodities or for all B2B transactions). And in the absence of an over-arching EU-wide VAT, the two most appealing ways to do so are by adopting either a CVAT (which maintains domestic zero-rating of exports but imposes an EU-wide VAT on inter-union trade) or a VIVAT (which taxes all B2B transactions, throughout the Union, at a common rate while maintaining national discretion in rates applied to final sales). Each has its own advantages and disadvantages. One criticism of the VIVAT, for instance, is that it requires traders to treat differently sales to registered traders and to final consumers. This, however, is a distinction that seems likely in any event to become increasingly important: it is already made in respect of intra-union sales, and is at the heart of the Commission's recent proposals for the treatment of international services. Either CVAT or VIVAT, in any event, offers scope for significant improvement over the current (and explicitly transitional) arrangements.

4.6.3. Excises on alcohol and tobacco

The UK levies excises on alcohol and tobacco which are substantially higher than in many other EU member states. In terms of their demand characteristics, these goods are not markedly different from many others. The estimates in Crawford, Keen, and Smith (2008) suggest that beer and wine and spirits may be moderately complementary with hours worked, while tobacco products appear to have moderate substitutability with labour time. In neither case, however, do considerations of revenue-raising efficiency alone seem likely to justify taxing these goods so much more heavily than the generality of consumer spending. The externalities associated with consumption of these commodities are, however, appreciable, and justify taxing these goods at higher rates than the general VAT. Nevertheless, policy towards these goods needs to be informed by a clearer appreciation of the nature and size of the externalities associated with their consumption.

In the case of tobacco taxes, the evidence is highly controversial. Given that smokers die early, saving the rest of society a significant burden in

pension and old-age care costs, it is far from clear that the existing level of tobacco taxes can be justified solely on the basis of smoking externalities. Behavioural issues of time inconsistency and lack of self-control in consumption decisions, as discussed in the commentary by Gruber, may be particularly important in this area, and may justify higher taxes than would be warranted if decisions were being made by wholly rational, well-informed consumers.

Using alcohol taxes to reduce the externalities associated with alcohol consumption (and, especially, abusive over-consumption) involves targeting the incentive somewhat imprecisely to the underlying externality, since alcohol externalities are not proportional to alcohol consumption but are largely confined to abusive over-consumption by a subset of all consumers. Externality taxation of alcohol thus involves a compromise between the potential gains from reducing external costs of abusive consumption and the welfare costs of discouraging non-abusive consumption.

The external costs of tobacco and alcohol consumption are likely to be heavily affected both by the institutional arrangements for financing health-care, pensions, and so on, and by the cultural context in which alcohol is consumed. It is therefore unlikely that estimates from the US can be carried over without modification to the European context, or that externalities will be uniform throughout Europe. Because of the likely differences across European countries in external costs of smoking and alcohol consumption, it is unlikely that the optimal tax treatment of tobacco products and alcoholic drinks will be identical in all members of the EU. Imposing greater uniformity on the very diverse pattern of EU excises may thus involve some economic inefficiency. Experience has confirmed, however, that for these items there is significant risk of revenue erosion—and hence pressures towards setting tax rates lower than would be collectively desirable—from the legal cross-border shopping and the various forms of illegal smuggling and tax evasion that are encouraged by significant tax differences. The EU-wide floors to alcohol and tobacco excises have the potential to play an important role, both in preventing the various forms of smuggling and evasion which exploit the massive duty differentials between member states, and also in ensuring the sustainability of these duties at levels which properly reflect the external costs of consumption of these goods. The levels at which these excise floors are currently set, however, put little constraint on tax rate decisions in member states, and so make little practical contribution to preventing inefficient downward pressure on duty rates. Raising them—though clearly desirable—has, however, proved difficult, and will no doubt remain so.

REFERENCES

- Adams, E. K., Melvin, C., Merritt, R., and Worrall, B. (1999), 'The Costs of Environmental Tobacco Smoke (ETS): An International Review. Background paper', prepared for World Health Organization Framework Convention on Tobacco Control (FCTC), International Consultation on Environmental Tobacco Smoke (ETS) and Child Health. Available at: <<http://www.who.int/tobacco/framework/consultation/ets2/en/>>. Accessed 4 September 2007.
- Ainsworth, R. T. (2006), 'Carousel Fraud in the EU: A Digital VAT Solution', *Tax Notes International*, May, 443–8.
- (2007), 'Tackling VAT Fraud: Thirteen Ways Forward', *International Tax Notes*, 1205.
- Atkinson, A. B., and Stiglitz, J. E. (1976), 'The Design of Tax Structure: Direct versus Indirect Taxation', *Journal of Public Economics*, 6, 55–75.
- (1980), *Lectures on Public Economics*, London: McGraw Hill.
- Aujean, M., Jenkins, P., and Poddar, S. (1999), 'A New Approach to Public Sector Bodies', *International VAT Monitor*, 10, 144–9.
- Barendregt, J. J., Bonneux, L., and van der Maas, P. J. (1997), 'The Health Care Costs of Smoking', *New England Journal of Medicine*, 337, 1052–7.
- Becker, G. S., and Murphy, K. M. (1988), 'A Theory of Rational Addiction', *Journal of Political Economy*, 96, 675–700.
- Besley, T., and Jewitt, I. (1990), 'Optimal Uniform Taxation and the Structure of Consumer Preferences', pp. 131–56 in Myles, G. D. (ed.), *Measurement and Modelling in Economics*, Amsterdam: North Holland.
- and Smart, M. (2007), 'Fiscal Restraints and Voter Welfare', *Journal of Public Economics*, 91, 755–73.
- Besson, E. (2007), 'TVA sociale', Ministry of Economy, Finance, and Industry, Secrétariat d'état chargé de la prospective et de l'évaluation des politiques publiques, Report, September 2007 <http://www.premier-ministre.gouv.fr/IMG/pdf/TVA_sociale_rapport_remis_par_Eric_Besson.pdf>.
- Bird, R. (2000), 'Subnational Revenues: Realities and Prospects', Mimeo: World Bank Institute.
- and Gendron, P.-P. (1998), 'Dual VATs and Cross-border Trade: Two Problems, One Solution?', *International Tax and Public Finance*, 5, 429–42.
- Bird, R., and Gendron, P.-P. (2007), *The VAT in Developing and Transitional Countries*, Cambridge: Cambridge University Press.
- Boadway, R., and Gahvari, F. (2006), 'Optimal Taxation with Consumption Time as a Leisure or Labor Substitute', *Journal of Public Economics*, 90, 1851–78.
- and Keen, M. (2003), 'Theoretical Perspectives on the Taxation of Capital Income and Financial Services', 31–80 in Honahan, P. (ed.), *The Taxation of Financial Intermediation*, World Bank and Oxford University Press.

- Boadway, R., Marchand, M., and Pestieau, P. (1994), 'Towards a Theory of the Direct–Indirect Tax Mix', *Journal of Public Economics*, 55, 71–88.
- and Pestieau, P. (2003), 'Indirect Taxation and Redistribution: The Scope of the Atkinson–Stiglitz Theorem', 387–403 in Arnott, R., Greenwald, B., Kanbur, R., and Nalebuff, B. (eds.), *Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz* Cambridge, Mass: MIT Press.
- Bovenberg, L. (1994), 'Destination- and Origin-based Taxation under International Capital Mobility', *International Tax and Public Finance*, 1, 247–73.
- Browning, M., and Meghir, C. (1991), 'The Effects of Male and Female Labour Supply on Commodity Demands', *Econometrica*, 59, 925–51.
- Carbonnier, C. (2007), 'Who Pays Sales Taxes? Evidence from the French VAT Reforms, 1987–1999', *Journal of Public Economics*, 91, 1219–29.
- Chaloupka, F. J. (1991), 'Rational Addictive Behavior and Cigarette Smoking', *Journal of Political Economy*, 99, 722–42.
- and Wechsler, H. (1997), 'Price, Tobacco Control Policies and Smoking Among Young Adults', *Journal of Health Economics*, 16, 359–73.
- Christiansen, V. (1984), 'Which Commodity Taxes Should Supplement the Income Tax?' *Journal of Public Economics*, 24, 195–220.
- and Smith, S. (2004), 'National Policy Interests in the Duty-free Market', *CESifo Economic Studies*, 50, 351–75.
- — (2008), 'Optimal Commodity Taxation with Duty-free Shopping', *International Tax and Public Finance*, 15, 274–96.
- Cnossen, S. (1983), 'Harmonization of Indirect Taxes in the EEC', in McLure, C. E., Jr. (ed), *Tax Assignment in Federal Countries*, 150–71. Centre for Research on Federal Financial Relations. Canberra: ANU Press.
- (1994), 'Administrative and Compliance Costs of the VAT: A Review of the Evidence', *Tax Notes*, 63, 1609–26.
- (2003), 'Is the VAT's Sixth Directive Becoming an Anachronism?', *European Taxation*, 434–42.
- and Shoup, C. S. (1987), 'Coordination of Value-Added Taxes', in Cnossen, S. (ed.), *Tax Coordination in the European Community*, Kluwer: Deventer.
- Colmar Brunton (2005), *Measuring the Tax Compliance Costs of Small and Medium-Sized Businesses—A Benchmark Survey*, Wellington: Colmar Brunton.
- Commission of the European Communities (1987), 'Commission Communication on Completion of the Internal Market: Approximation of Indirect Tax Rates and Harmonization of Indirect Tax Structures', COM (87), 320.
- (2003), 'Evaluation Report on the Experimental Evaluation of a Reduced Rate of VAT to Certain Labor-intensive Services', Working Paper SEC(2003) 622.
- (2008), 'Communication from the Commission to the Council and the European Parliament on Measures to Change the VAT System to Fight Fraud', SEC(2008) 249.
- Crawford, I., Keen, M., and Smith, S. (2008), 'Preference Structures and Optimal Commodity Taxation', Institute for Fiscal Studies, Working Paper.

- Smith, Z., and Tanner, S. (1999), 'Alcohol Taxes, Tax Revenues and the Single European Market', *Fiscal Studies*, 20, 287–304.
- and Tanner, S. (1995), 'Bringing it all back Home: Alcohol Taxation and Cross-border Shopping', *Fiscal Studies*, 16, 94–114.
- Crooks, E. (1989), *Alcohol Consumption and Taxation*, Report Series 34, London: Institute for Fiscal Studies.
- de Paula, Á., and Scheinkman, J. A. (2006), 'The Informal Sector', mimeo, Princeton University.
- Deaton, A., and Stern, N. (1986), 'Optimally Uniform Commodity Taxes, Taste Differences and Lump Sum Grants', *Economics Letters*, 20, 263–6.
- Delipalla, S., and Keen, M. (1992), 'The Comparison Between ad valorem and Specific Taxation under Imperfect Competition', *Journal of Public Economics*, 49, 351–67.
- — (2006), 'Product Quality and the Optimal Structure of Commodity Taxes', *Journal of Public Economic Theory*, 8, 547–54.
- Dharmapala, D., Slemrod, J., and Wilson, J. D. (2007), 'Optimal Tax Remittance with Firm-level Administrative Costs', mimeo, Michigan State University.
- Diamond, P. (1973), 'Consumption Externalities and Imperfect Corrective Pricing', *Bell Journal of Economics*, 4, 526–38.
- and Mirrlees, J. A. (1971), 'Optimal Taxation and Public Production I: Production Efficiency, II Tax Rules', *American Economic Review*, 61, 8–27 and 261–78.
- Ebrill, L. P., Keen, M., Bodin, J.-P., and Summers, V. (2001), *The Modern VAT*, Washington DC: International Monetary Fund.
- Edwards, J., and Keen, M. (1996), 'Tax Competition and Leviathan', *European Economic Review*, 40, 113–34.
- Evans, W. N., and Farrelly, M. (1998), 'The Compensating Behaviour of Smokers: Taxes, Tar and Nicotine', *RAND Journal of Economics*, 29, 578–95.
- Gahvari, F. (2007), 'On Optimal Commodity Taxation when Consumption is Time-consuming', *Journal of Public Economic Theory*, 9, 1–27.
- and Yang, C. C. (1993), 'Optimal Commodity Taxation and Household Consumption Activities', *Public Finance Quarterly*, 21, 479–87.
- Genser, B. (1996), 'A Generalised Equivalence Property of Mixed International VAT Regimes', *Scandinavian Journal of Economics*, 98, 253–62.
- and Haufler, A. (1996), 'Tax Competition, Tax Coordination and Tax Harmonization: The Effects of EMU', in Holzmann, R. (ed), *Maastricht: Monetary Constitution without a Fiscal Constitution*, Baden-Baden: Nomos Verlagsgesellschaft.
- Haufler, A., and Sørensen, P. B. (1995), 'Indirect Taxation in an Integrated Market: Is there a way of Avoiding Trade Distortions without Sacrificing National Autonomy?' *Journal of Economic Integration*, 10, 178–205.
- and Schulze, G. G. (1997), 'Transfer Pricing under an Origin-based VAT System', *Finanzarchiv*, 54, 51–67.

- Genser, B. and Winker, P. (1997), 'Measuring the Fiscal Revenue Loss of VAT Exemption in Commercial Banking', *Finanzarchiv*, 54, 565–85.
- Golosov, M., Kocherlakota, N., and Tsyvinski, A. (2003), 'Optimal Indirect and Capital Taxation', *Review of Economic Studies*, 70, 569–87.
- Gordon, J. (1989), 'Tax Reform via Commodity Grouping', *Journal of Public Economics*, 39, 67–81.
- Gordon, R., and Nielsen, S. B. (1997), 'Tax Evasion in an Open Economy: Value-added Taxes vs Income Tax', *Journal of Public Economics*, 66, 173–97.
- Grossman, G. M. (1980), 'Border Tax Adjustments: Do they Distort Trade?', *Journal of International Economics*, 10, 117–28.
- Gruber, J. (2001), 'Tobacco at the Crossroads: The Past and Future of Smoking Regulation in the United States', *Journal of Economic Perspectives*, 15, 193–212.
- and Köszegi, B. (2001), 'Is Addiction "Rational"? Theory and Evidence', *Quarterly Journal of Economics*, 116, 1261–303.
- (2004), 'Tax Incidence when Individuals are Time-inconsistent: The Case of Cigarette Excise Taxes', *Journal of Public Economics*, 88, 1959–87.
- and Mullainathan, S. (2005), 'Do Cigarette Taxes make Smokers Happy?' *Advances in Economic Analysis and Policy*, 5, 1–43.
- Hall, R. E., and Rabushka, A. (1983), *Low Tax, Simple Tax, Flat Tax*, New York: McGraw-Hill.
- Harrison, G., and Krelove, R. (2005), 'VAT Refunds: A Review of Country Experience', International Monetary Fund Working Paper WP/05/218 <<http://www.imf.org/external/pubs/ft/wp/2005/wp05218.pdf>>.
- Hassett, K., Mathur, A., and Metcalf, G. (2007), 'The Incidence of a U.S. Carbon Tax: A Lifetime and Regional Analysis', NBER Working Paper 13554.
- Hemming, R., and Kay, J. A. (1981), 'The United Kingdom', 75–89 in Aaron, H. J. (ed.), *The Value-Added Tax: Lessons from Europe*, Washington, DC: Brookings Institution.
- Huizinga, H. (2002), 'Financial Services VAT', *Economic Policy*, 17, 499–534.
- International Tax Dialogue (2007), 'The Tax Treatment of Small and Medium Enterprises', available at <<http://www.itdweb.org>>.
- Kanbur, R., and Keen, M. (1993), 'Jeux sans frontières: Tax Competition and Tax Coordination when Countries Differ in Size', *American Economic Review*, 83, 877–92.
- Kaplow, L. (2006), 'On the Undesirability of Commodity Taxation even when Income Taxation is not Optimal', *Journal of Public Economics*, 90, 1235–50.
- Kay, J. A., and Davis, E. (1985), 'Extending the VAT Base', *Fiscal Studies*, 6, 1–16.
- Keen, M. (1998), 'The Balance Between Specific and ad valorem Taxation', *Fiscal Studies*, 19, 1–37.
- (2002), 'Some International Issues in Commodity Taxation', *Swedish Economic Policy Review*, 9, 11–45.

- (2008), 'VAT, Tariffs and Withholding', *Journal of Public Economics*, 92, 1892–906.
- and Lahiri, S. (1998), 'The Comparison Between Destination and Origin Principles under Imperfect Competition', *Journal of International Economics*, 45, 323–50.
- and Lockwood, B. (2006), 'Is the VAT a Money Machine?', *National Tax Journal*, LIX, 905–28.
- (2008), 'The Value Added Tax: Its Causes and Consequences', forthcoming in the *Journal of Development Economics*.
- and Mintz, J. (2004), 'The Optimal Threshold for a Value-Added Tax', *Journal of Public Economics*, 88, 559–76.
- and Smith, S. (1996), 'The Future of Value-Added Tax in the European Union', *Economic Policy*, 23, 375–411 and 419–20.
- (2000), 'Viva VIVAT!' *International Tax and Public Finance*, 7, 741–51.
- (2006), 'VAT Fraud and Evasion: What do we Know and What can be Done?' *National Tax Journal*, 59, 861–87.
- and Wildasin, D. (2004), 'Pareto Efficient International Taxation', *American Economic Review*, 94, 259–75.
- Kleven, H. J. (2004), 'Optimum Taxation and the Allocation of Time', *Journal of Public Economics*, 88, 545–57.
- Wolfram, R., and Sørensen, P. B. (2000), 'Optimal Taxation with Household Production', *Oxford Economic Papers*, 52, 584–94.
- Kneller, R., Bleaney, M. F., and Gemmell, N. (1999), 'Fiscal Policy and Growth: Evidence from OECD countries', *Journal of Public Economics*, 74, 171–90.
- KPMG (2006), *Administrative Burdens: HMRC Measurement Project* (KPMG LLP).
- Laroque, G. (2005), 'Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof', *Economics Letters*, 87, 141–4.
- Lee, C., Pearson, M., and Smith, S. (1988), *Fiscal Harmonisation: An Analysis of the Commission's Proposals*, Report Series 28, London: Institute for Fiscal Studies.
- Lewit, E. M., and Coate, D. (1982), 'The Potential for Using Excise Taxes to Reduce Smoking', *Journal of Health Economics*, 1, 121–45.
- and Grossman, M. (1981), 'The Effects of Government Regulation on Teenage Smoking', *Journal of Law and Economics*, 24, 545–69.
- Lockwood, B. (1993), 'Commodity Tax Competition under Destination and Origin Principles', *Journal of Public Economics*, 52, 141–62.
- de Meza, D., and Myles, G. (1994a), 'When are Destination and Origin Regimes Equivalent?', *International Tax and Public Finance*, 1, 5–24.
- (1994b), 'The Equivalence Between Destination and Non-reciprocal Restricted Origin Regimes', *Scandinavian Journal of Economics*, 96, 311–28.
- Lyon, A. B., and Schwab, R. M. (1995), 'Consumption Taxes in a Life-cycle Framework: Are Sin Taxes Regressive?', *Review of Economics and Statistics*, 77, 389–406.

- McCann, E., and Edgar, T. (2003), 'VAT Treatment of Interest and Financial Services with Competitive Banking and Insurance Sectors', *Tax Notes International*, 26, 791–808.
- McLure, C. E. (1999), 'Protecting Dual VATs from Evasion on Cross-border Trade: An Addendum to Bird and Gendron', mimeo, Hoover Institution.
- (2000), 'Implementing Subnational VATs on Internal Trade: The Compensating VAT (CVAT)', *International Tax and Public Finance*, 7, 723–40.
- Manning, W. G., Keeler, E. B., Newhouse, J. P., Sloss, E., and Wasserman, J. (1989), 'The Taxes of Sin: Do Smokers and Drinkers Pay Their Way?', *Journal of the American Medical Association*, 261, 1604–9.
- Matthew, H. C. G. (1979), 'Disraeli, Gladstone and the Politics of Mid-Victorian Budgets', *The Historical Journal*, 22, 615–43.
- Maynard, A., Godfrey, C., and Hardman, G. (1994), 'Conceptual Issues in Estimating the Social Costs of Alcohol', paper prepared for an international symposium on the economic costs of substance abuse, Banff, Canada, 11–13 May. Available at <<http://www.ccsa.ca/Costs/maynard.htm>>.
- Meade, J. (1978), *The Structure and Reform of Direct Taxation: Report of a Committee chaired by Professor J. E. Meade for the Institute for Fiscal Studies*, London: George Allen & Unwin. <http://www.ifs.org.uk/publications/3433>
- Messere, K. (1994), 'Consumption Tax Rules', *Bulletin for International Fiscal Documentation*, 48, 665–81.
- Mirrlees, J. A. (1976), 'Optimal Tax Theory: A Synthesis', *Journal of Public Economics*, 6, 327–58.
- Musgrave, R. A., and Suits, D. B. (1953), 'Ad valorem and Unit Taxes Compared', *Quarterly Journal of Economics*, 67, 598–604.
- Newbery, D. M. (1986), 'On the Desirability of Input Taxes', *Economics Letters*, 20, 267–70.
- Ohsawa, Y. (1999), 'Cross-border Shopping and Commodity Tax Competition among Governments', *Regional Science and Urban Economics*, 29, 33–51.
- Organisation for Economic Cooperation and Development (2004), *The Application of Consumption Taxes to the Trade in International Services and Intangibles*. Report. Centre for Tax Policy and Administration, Paris: OECD.
- (2007), *Revenue Statistics 1965–2006*, Paris: OECD.
- (2008), *Consumption Tax Trends, 2008 Edition* (Paris: OECD).
- Orphanides, A., and Zervos, D. (1995), 'Rational Addiction with Learning and Regret', *Journal of Political Economy*, 103, 739–58.
- Panagariya, A., and Rodrik, D. (1993), 'Political-economy Arguments for a Uniform Tariff', *International Economic Review*, 34, 685–703.
- Parry, I. W. H. (2003), 'On the Costs of Excise Taxes and Income Taxes in the UK', *International Tax and Public Finance*, 10, 281–304.
- Piggott, J., and Whalley, J. (2001), 'VAT Base Broadening, Self Supply, and the Informal Sector', *American Economic Review*, 91, 1084–94.
- Pigou, A. C. (1920), *The Economics of Welfare*, London: Macmillan.

- Pirtillä, J., and Tuomala, M. (1997), 'Income Tax, Commodity Tax and Environmental Policy', *International Tax and Public Finance*, 4, 379–93.
- Poddar, S. (2003), 'Consumption Taxes: The Role of the Value-Added Tax', 354–80 in Honohan, P. (ed.), *Taxation of Financial Intermediation: Theory and Practice for Emerging Economies*, Washington, DC: World Bank.
- and English, M. (1997), 'Taxation of Financial Services under a Value-Added Tax: Applying the Cash-flow Method', *National Tax Journal*, 50, 89–111.
- and Hutton, E. (2001), 'Zero-rating of Interstate Sales under a Sub-national VAT', *National Tax Association Proceedings: Ninety Fourth Annual Conference*, 200–7.
- Pogue, T. F., and Sgontz, L. G. (1989), 'Taxing to Control Social Costs: The Case of Alcohol', *American Economic Review*, 79, 235–43.
- Poterba, J. M. (1989), 'Lifetime Incidence and the Distributional Burden of Excise Taxes', *American Economic Review*, 79, 325–30.
- Ramsey, F. (1927), 'A Contribution to the Theory of Taxation', *Economic Journal*, 37, 47–61.
- Rasmussen, S. R., Prescott, E., Sørensen, T. I. A., and Søgaard, J. (2004), 'The Total Lifetime Health Cost Savings of Smoking Cessation to Society', *European Journal of Public Health*, 15, 601–6.
- Sandmo, A. (1976), 'Direct versus Indirect Pigovian Taxation', *European Economic Review*, 7, 337–49.
- Shibata, H. (1967), 'The Theory of Economic Unions: A Comparative Analysis of Customs Unions, Free Trade Areas and Tax Unions', in Shoup, C. S. (ed.), *Fiscal Harmonisation in Common Markets*, New York: Columbia University Press.
- Sinn, H.-W. (1990), 'Tax Harmonization and Tax Competition in Europe', *European Economic Review*, 34, 489–504.
- Gebauer, A., and Parsche, R. (2004), 'The Ifo Institute's Model for Reducing VAT Fraud: Payment First, Refund Later', *CESifo Forum*, 2, 30–4.
- Slemrod, J., and Yitzhaki, S. (1996), 'The Cost of Taxation and the Marginal Efficiency Cost of Funds', *IMF Staff Papers*, 43, 172–98.
- Smith, S. (2005), 'Economic Issues in Alcohol Taxation', in Cnossen, S. (ed.), *Theory and Practice of Excise Taxation*, Oxford: Oxford University Press.
- Södersten, J. (1999), 'Why Europe Adopted the Value Added Tax', Mimeo, Uppsala University.
- Vanistendael, F. (1995), 'A Proposal for a Definitive VAT System: Taxation in the Country of Origin at the Rate of the Country of Destination, Without Clearing', *EC Tax Review*, No. 1995/1, 45–53.
- Varsano, R. (2000). 'Sub-national Taxation and Treatment of Interstate Trade in Brazil: Problems and a Proposed Solution', 339–55 in Burki, S. J., and Perry, G. (eds.), *Decentralization and Accountability of the Public Sector*, Proceedings of the Annual Bank Conference on Development in Latin America and the Caribbean, Washington DC: The World Bank.

- Verwaal, E., and Cnossen, S. (2002), 'Europe's New Border Taxes', *Journal of Common Market Studies*, 40, 309–30.
- Viscusi, W. K. (1995), 'Cigarette Taxation and the Social Consequences of Smoking', *Tax Policy and the Economy*, 9, 51–101.
- Zee, H. (2005), 'Simple Analytics of Setting the Optimal VAT Exemption Threshold', *The Economist*, 143, 461–71.
- (2006), 'VAT Treatment of Financial Services: A Primer on Conceptual Issues and Country Practices', *Intertax*, 34, 458–74.

Commentary by Richard M. Bird

Richard Bird is Professor Emeritus of Economics at the University of Toronto, an Associate of the Institute for International Business of the Rotman School of Management, and Senior Fellow of the Institute for Municipal Finance and Governance of the Munk Centre for International Studies. A Fellow of the Royal Society of Canada, he has taught in the US, the Netherlands, and Australia among other countries and has frequently worked for the International Monetary Fund and other international organizations. His main interests are tax policy, tax administration, and fiscal decentralization issues in developing countries.

The value-added tax (VAT) was unquestionably the most successful fiscal innovation of the last half-century. No other significant tax, not even the income tax, spread so rapidly and quickly around the world to the point where VATs currently exist in over 150 countries, including of course all member states of the EU. Why was the VAT so successful? Essentially, because virtually every country—apart from the always exceptional United States and a few small islands and oil-rich countries—needs some form of mass consumption tax to support state activities, and experience has shown that the VAT is not only the least distorting such tax but also the one that can, despite some problems, be administered most effectively in most countries. Of course, not all VATs are the same, and none is perfect (Bird and Gendron (2007)).

In general, as Crawford, Keen, and Smith (hereafter CKS) show, in principle the VAT approach to imposing a consumption tax is perhaps the most economically efficient way in which countries can raise significant tax revenues. The marginal cost of raising funds for public purposes through VAT is generally lower than it would be if other taxes were employed. This economic advantage may cut two ways politically, however. By lowering the cost of taxation, VAT makes it relatively cheaper to expand the level of state activity—an outcome that some may consider good and others bad. On the other hand, given any level of state activity, VAT does less damage to economic incentives than would equally productive alternative taxes. What more could one ask of a tax?

Picking up this theme, CKS suggest that the apparent increase in tax ratios associated with VAT in developed countries such as the UK reflects the reduced marginal cost of public funds. However, other factors are at work here than simply the lesser distortionary costs of VATs. Canada recently reduced its national VAT (the GST, in Canada) rate from 7% to 5%.¹ Most Canadian economists did not like this move, but it was certainly politically popular in large part because Canada's VAT is much more visible than in those found in other countries. In Canada, the GST is not included in product prices but is added on separately at the time of purchase: it is thus highly visible every time anyone buys anything. Everyone is aware of it every day, and, unsurprisingly, no one likes it. In economic terms, VAT in Canada as in most countries has the lowest Marginal Cost of Public Funds (MCPF) of any significant tax (Baylor and Beausejour (2004)). In political terms, however, matters are very different owing to the high visibility of the VAT. Some may argue that, since income taxes carry high political as well as economic costs and consumption taxes are lower on both metrics, a hidden VAT may yield a better revenue portfolio not just in enforcement terms but also in terms of bringing more closely together the marginal economic and political costs of raising public funds. On the other hand, it is hard to believe that it is a good thing for government revenues to grow more than they otherwise would simply because people do not understand how much tax they are paying. Little is known about these deep and murky waters.² Perhaps, however, it is more than time for those concerned with establishing sound tax policy to pay more careful attention to what may perhaps be called the reality of perception in shaping the design of tax instruments.

The scholarly literature on VAT that is so well summarized and reviewed by CKS provides surprisingly little useful guidance to those responsible for VAT policy and administration. Much of this literature concerns matters that have already been decided in practice—for the most part, fairly satisfactorily—and are hence unlikely to be worth reconsidering in practice. Further discussion here of the essential policy irrelevance of such matters as the direct–indirect 'tax mix' or 'tax balance' question, the origin versus destination issue, and the relative merits of VAT and retail sales taxes would thus serve no point. Only after clearing a good deal of such relatively unproductive theoretical underbrush do CKS get to the more salient policy issues with respect to

¹ As Bird, Mintz, and Wilson (2006) show, Canada is unusual in that it also has significant subnational consumption taxes some of which are VATs. On average, the level of consumption taxation even after the federal tax cut is about 13%.

² For a promising initial empirical foray, see Chetty, Looney, and Kroft (2007).

VAT—rates, exemptions, thresholds, enforcement, compliance, fraud, and the international setting.

What do we really know about these issues? Consider the summary picture of VATs in the OECD depicted in Table 4.2 of the chapter. About the only conclusion one can derive from the information in this table is that countries, it seems, can do almost anything they want with respect to rates, thresholds, coverage, and enforcement (C-efficiency combines these two factors), subject only to a mild EU constraint in the case of EU countries. Since practice provides so little guidance, CKS understandably turn once more to theory for answers. As it turns out, however, apart from the issue of the threshold, where what seems to be an unanswerable case is made that the UK should certainly not lower it, there seem to be no very substantial answers with respect to the flesh and bones of any VAT—rates, zero-rating, and exemptions.

In practice no VAT is quite as good as it might be for a number of reasons: less than general coverage, less than optimal rate structure, and less than perfect administration. In the EU, for example, there are three major components of the base where coverage is arguably less than ideal—real property, the PNC (public, non-profit, charitable) sector, and the financial sector. CKS discuss the last of these sectors in some detail and provide an excellent overview of the relevant literature.³ But they do not provide any clear guidance as to just what, if anything, the UK, the EU, or anyone else should do with respect to the treatment of the financial sector—or for that matter the other two troublesome sectors mentioned above. The reader is, perhaps unintentionally, left with the feeling that it is all just too difficult to find an acceptable solution and we should perhaps just give up.

Of course, this characterization is not entirely fair. With respect to the financial sector, for example, CKS clearly do think that something should be done, perhaps along Australian lines. However, we are left up in the air as to exactly what should be done and what the costs and consequences of changing the present treatment might be. Similarly, with respect to real property, while an offset to first-time buyers is suggested, again the eager reader is left in the dark with respect to the critical details. Finally, when it comes to the equally troublesome issue of the public and non-profit sectors, neither any discussion of the nature of the problem nor even a hint of a possible solution is to be found. Even after decades of experience around the world with a variety of approaches, there is much more that we do not know than we know about the relevant parameters and behavioural responses determining such matters. These are the sorts of issues that consume the time

³ For a similar overview of the PNC question, see Gendron (2005).

and efforts of VAT policy designers everywhere. The absence of any solid theoretical or empirical framework within which to make such decisions is presumably why CKS do not discuss such issues more fully. Still, it is striking that when they come to the concluding section of the chapter not a word is said about any domestic aspect of VAT.

A more positive approach might have been to note that the EU model in general (and perhaps the UK version in particular) is very much a 'first-generation' VAT. Since a second-generation VAT demonstrating that it is both possible and desirable to have much more comprehensive coverage has been on display in Australia and especially New Zealand for some years, it would have been interesting to consider more systematically the pros and cons (if any) of moving closer to this model in the UK. Such an exercise might also have made clearer the extent if any to which such improvements in the UK VAT are or may be restricted by EU rules. Finally, since New Zealand is not the last word when it comes to VAT, it would also have been interesting if the authors had set out their view of the structure of an 'ideal' (third-generation?) VAT or at least reviewed more systematically how EU rules might need to be altered to enable member states to improve their national VATs.

The analytical discussion in the chapter of some other issues such as the distributional aspects of VAT and the treatment of cross-border transactions seems more directly relevant to VAT policy and administration in the UK and more widely. However, the discussion of these issues also leaves a number of important questions unanswered—and in some cases even unasked.

One such question relates to VAT rate structure. Neither the inconclusive optimal tax literature nor the interesting empirical estimates provided by CKS (which reject, though not strongly, weak separability) provide particularly useful guidance on this point. In the end, the rule implicitly suggested for policy-makers, at least in developed countries like the UK, appears to amount to little more than the old rule of thumb that a uniform rate is likely to be as good, if not better, than any 'scientific' differentiation—let alone the politically driven differentiation that actually prevails in the UK, as in most countries. As CKS demonstrate, there is little case for differential rates on distributional or incidence grounds. Moreover, although CKS do not stress this point, there is considerable experience (though apparently no solid evidence) suggesting that multiple rates increase compliance and administrative costs and perhaps facilitate evasion.

If both theory and practice suggest that rate differentiation (like domestic zero-rating for consumer goods) seems to make little sense, why are these features so pervasive in VATs around the world? The answer may lie deep in the murky waters of perception, electoral psychology, and the political

economy of taxation and hence be well beyond the scope of this chapter. Still, given the existence of such differentiation in the UK VAT, it would have been useful to see, for example, a discussion of whether it makes any difference how 'reduced' (compared to the standard rate) the 'reduced' rate is. For example, if the lower rate is 'too low' it may, like zero rating, generate both excessive refunds and an attractive opportunity for evasion. In principle, presumably the optimal reduced rate for VAT—admitting that, however deplorable it may be, most countries seem to end up with such a rate—should be susceptible to the same kind of analysis as the optimal threshold.

Recently, increasing attention is being paid to some of the important relatively unexplored implications of the way in which VAT is commonly administered through the so-called invoice-credit system. Since input VAT cannot be credited by those not registered as VAT taxpayers, to the extent that taxed inputs—for example, imports—are purchased by 'informal' enterprises, VAT offers a simple and effective way to tax the informal sector. Silver linings seldom come without accompanying clouds, however. Viewed from another perspective, the effective imposition of VAT on enterprises outside the VAT system, like the relatively high compliance costs imposed on small firms within the system, creates a barrier to the growth of the important—and often politically favoured—sector of 'small business'. Unfortunately, although the treatment of small businesses, both within and outside the VAT system, is a matter of increasing interest in many countries, including the UK, the issue is not discussed in this chapter. In response to these conflicting concerns, a growing number of countries in Europe and elsewhere seem to be introducing gross receipts taxes on small business sometimes within the framework of the VAT and sometimes as a separate 'simplified' levy. Such composite transactions taxes deserve closer attention than they appear to have received to date anywhere.⁴ Similarly, the costs and benefits of the various special schemes for small business found in many VATs, including that in the UK, also need closer examination.

Instead of going into such knotty and largely unresolved issues in VAT design, CKS devote a substantial fraction of the chapter to a particular form of VAT fraud currently giving rise to concern in the EU and more generally to the related question of how best to apply VAT to cross-border transactions. Since the UK has much more extensive zero rating than almost any other

⁴ A related issue is how VAT on real property interacts with other taxes that impact the sector in differential ways such as local property taxes, capital gains taxes, inheritance taxes, and, in many countries, special transfer taxes imposed on real property sales. Once again, important policy decisions are being made about an important economic sector with little guidance from either theory or empirical analysis.

developed country it is surprising that the possible linkage between this feature of the UK VAT and the much discussed problem of 'carousel fraud' is not even mentioned. However, the exposition of the evasion game is clear, and the discussion of possible solutions good. CKS find all the solutions currently on offer—such as reverse charging—to be wanting and end by throwing the issue into the more general VIVAT hat discussed next.

On the whole, the chapter is careful not simply to restate the case for the particular alternative mechanism for cross-border trade that some of the authors have argued at length in earlier works (Keen and Smith (2000)). As an EU outsider, however, I still find it difficult to understand the emphasis on 'symmetry' in most EU discussions of these matters, including this chapter. In reality, there must always be some 'asymmetry' in any VAT since any VAT 'includes' some and 'excludes' others involved in transactions. Those within the system are treated differently from those outside. The relevant question in the EU, from the perspective of any member state, is whether those in other member states are 'in' or 'out'. (Those outside the EU are always 'out' of course.) It is true, as CKS note, that 'in' may have importantly different manifestations—whose rate applies, who administers what—but no matter where the lines are drawn with respect to either the basic in/out or the degree of 'in-ness' there is always a line and hence always arbitrage possibilities, and marginal costs and benefits of being on one side or the other. Much of the problem in discussing these matters sensibly in the EU (or anywhere else) is that what is best—and for whom—depends on relationships between various empirical characteristics about which we know little or nothing.

In the end, CKS conclude, perhaps unsurprisingly, that the best solution at least for the UK and the EU is the VIVAT approach. Indeed, it appears that the central idea of this approach, a uniform 'intermediate' rate, is becoming more widely accepted in Europe. Nonetheless there remain many important unanswered questions about the system: the appropriate level of the intermediate rate, the politically significant issue of compensation for 'losing' member states, the increased weight this approach places on the distinction between the characteristics of end-users and the related apparent need for an accessible real-time EU VAT registration base. In tax design as in tax implementation, the devil is always in the details, and there remain many details about which there is a high degree of uncertainty with respect to this issue.

To conclude, CKS provide on the whole an excellent review of the current state of the theoretical and the (considerably less impressive) empirical literature on the VAT. Like the literature on which it draws, however, the chapter does not provide either clear answers to many of the critical questions facing those concerned with designing and implementing VATs in the UK and

around the world or much guidance to where they might look to find such answers. For better or worse, VAT practice is way ahead of VAT theory in the sense that practitioners are daily facing and dealing with problems with little or no theoretical or empirical guidance. Both practitioners and scholars have much to learn from each other with respect to VAT: they need to talk to each other more, both to learn what the relevant questions are and to obtain the information that may in the long run enable either VAT reality to come closer to the ideal VAT of theory or VAT theory to come closer to dealing with the problems confronting VAT in reality. There is much to be done, and this chapter provides an excellent starting point from which to begin doing it.

REFERENCES

- Baylor, M., and Beausejour, L. (2004), 'Taxation and Economic Efficiency: Results from a Canadian CGE Model', Working paper, Department of Finance, Ottawa.
- Bird, R. M., and Gendron, P.-P. (2007), *The VAT in Developing and Transitional Countries*, Cambridge: Cambridge University Press.
- Mintz, J. M., and Wilson, T. A. (2006), 'Coordinating Federal and Provincial Sales Taxes: Lessons from the Canadian Experience', *National Tax Journal*, 49, 889–903.
- Chetty, R., Looney, A., and Kroft, K. (2007), 'Salience and Taxation: Theory and Evidence', NBER Working Paper, National Bureau of Economic Research, Cambridge MA, September.
- Gendron, P.-P., (2005), 'Value-Added Tax Treatment of Public Bodies and Non-Profit Organizations', *Bulletin for International Fiscal Documentation*, 59, 514–26.
- Keen, M., and Smith, S. (2000), 'Viva VIVAT!', *International Tax and Public Finance*, 7, 741–51.

Commentary by Sijbren Cnossen*

Sijbren Cnossen is an Advisor at CPB Netherlands Bureau for Economic Policy Analysis, emeritus Professor of Economics at the University of Maastricht and emeritus Professor of Tax Law at Erasmus University, Rotterdam. He has been a consultant to the IMF, OECD, World Bank, European Commission, US-AID, and Harvard Institute for International Development, in which capacity he advised some thirty countries on the design and implementation of their tax systems. He is a past editor of *De Economist* and *International Tax and Public Finance*, and has published numerous books and articles on the economics of taxation.

1. INTRODUCTION

A large body of research analysing the properties and workings of the value-added tax (VAT) is now available.¹ The best practice that can be distilled from the literature, particularly as it applies to the European Union (EU), strongly indicates that the consumption base of the VAT should be defined as broadly as possible and that all goods and services should be taxed at a uniform rate. This promotes fiscal neutrality and administrative simplicity. On both counts, the VATs of the EU member states leave much to be desired, compared with the new VATs of New Zealand, Canada, Australia, and various other countries. The ‘standard exemptions’ (sic!) of the harmonized EU-VAT² defy the logic and inherent integrity of the VAT, and the differentiated rate structures in most member states are ill-targeted instruments to affect

* The author thanks Richard Bird, Charles McLure, and Ben Terra for their comments on an earlier version of this commentary.

¹ See, for instance, Bird and Gendron (2007) and Ebrill et al. (2001), and the literature cited therein.

² It should be noted that the 2006 VAT Directive on the Common System of Value Added Tax (Council Directive 2006/112/EC) merely recasts the 1977 VAT Sixth Directive (Council Directive 77/388/EEC), incorporating a previous directive and various subsequent amendments. The new date is misleading to the extent it suggests that substantive design changes have been made.

the VAT burden distribution, yet increase administrative complexity and compliance costs.

Another, perhaps more topical, point of discussion concerns the most appropriate treatment of intra-community trade. It is widely agreed that VAT revenue should be allocated among member states in line with consumption, that is, on the basis of the destination principle,³ but there is no consensus on how this principle should be applied in the EU. Currently, intra-community supplies (exports) are zero-rated and intra-community acquisitions (imports) are subject to VAT in the state of importation. This is called the 'transitional regime', because the European Commission believes that it should be replaced by a definitive regime based on exporter rating, that is, the taxation of intra-community exports by exporting member states and the use of the revenue thus collected to finance equivalent tax credits provided in importing states.

Crawford, Keen, and Smith (this chapter) also believe that the EU should change over to exporter rating. In their view (p. 337) zero rating at export breaks 'the chain of VAT revenue cumulation', which invites fraudulent practices. These practices are documented in Keen and Smith (2006), who conclude (p. 39) that 'a fundamental redesign of the VAT treatment of intra-community trade may be required', preferably in the form of a viable integrated VAT (VIVAT) (Keen and Smith (1996)). VIVAT envisages an EU-wide, harmonized single VAT rate on all pre-retail (intermediate) transactions by registered firms within and between member states (with credit for VAT in importing member states), supplemented by a surtax at retail (in fact a retail sales tax (RST)) if governments wish to exploit the VAT base more intensely.

This commentary argues that the current system deserves another hearing. While zero rating and exporter rating are identical in terms of revenue allocation, VIVAT involves substantial additional administrative complexity and may violate tax autonomy (subsidiarity), because it requires some form of central involvement in settling net VAT balances. More importantly, the commentary argues that it is not the break in the VAT-collection chain at intra-community borders that is a matter of concern, but rather the break in the VAT-audit trail, broadly defined as the jurisdictional reach of each VAT administration's ability to control compliance. Accordingly, VAT coordination efforts should focus on improving information exchange between member states and on establishing bilateral and multilateral VAT audit and

³ By contrast, under the origin principle, VAT accrues to the state of production. Accordingly, value added up to the export stage is taxed in the state of production, while imports are not taxed.

investigation units to monitor compliance with VAT obligations regarding intra-community transactions.

Against this background, the commentary is organized as follows. Section 2 dwells briefly on the history and current treatment of intra-community transactions in the EU, essential in understanding and evaluating the current regime against the various reform proposals. Subsequently, Section 3 highlights the proposals that the European Commission has made to replace the transitional regime by some form of exporter rating. Next, Section 4 does the same for the exporter rating proposals made in the tax literature. Whatever proposal is ultimately adopted, the proper treatment of intra-community transactions would be facilitated by various VAT base broadening measures, as argued in Section 5. Section 6 concludes.

In passing, the commentary notes that VAT does not cumulate throughout the production process.⁴ Just as under a retail sales tax (RST), no net VAT is collected within the ring of registered firms, provided, quite plausibly, that the average length of time required for remitting tax and for processing any refunds is the same as the average length of time required for settling accounts receivable and payable.⁵ Net VAT is collected only after the consumer (or unregistered trader) has been invoiced for the full amount of VAT on his purchases. Subsequently, this amount is collected fractionally throughout the production process. In essence, suppliers are made tax collectors on behalf of the government for the consumer's VAT on the rationale that retailers (and other firms) are less likely to default on tax invoiced by their suppliers than on tax payable directly to the tax office (Cnossen (1987)).

2. VAT AND INTRA-COMMUNITY TRANSACTIONS

In discussing the VAT treatment of intra-community transactions (and the full implications of the transitional and various definitive regimes), it is

⁴ Accordingly, in this sense, there is no break in the VAT-collection chain at export. An exporter would still be entitled to a refund for the VAT he has paid to his suppliers if in any VAT return period he would not export the goods he holds in inventory. However, there is a break in the VAT-chargeability chain: no tax is invoiced to foreign buyers.

⁵ To be sure, cash flow benefits (or costs) do arise if a registered firm's collection period (the period during which the tax is collected before being handed over to the tax office) does not coincide with the grace period (the period after the collection period but before the latest day designated for handing over the tax). This will happen under RST if sales are made against cash but the tax is remitted, say, every three months, or if accounts receivable are settled earlier than tax is remitted to the tax office. If the tax payment conditions are similar, the effect also arises under VAT, but part of the tax-induced cash flow benefit may be spread upstream if retail purchases, including VAT, are also made against cash.

useful to make a distinction, generally overlooked in the economics literature, between the cross-border VAT treatment of goods and that of services.

2.1. Cross-border transactions in goods

Prior to 1993 (the year in which intra-community border controls were abolished in the EU), most member states included imported goods in the domestic VAT base by taxing them at their borders under the supervision of customs authorities, while a credit for the VAT was provided when the goods entered the domestic production process. Similarly, the VAT on exports was refunded at the export stage and border controls could be used to check whether the goods had actually left the country. These 'border tax adjustments' (BTAs) ensure that the VAT is levied on the destination principle, while border controls can be viewed as a helpful backstop for enforcement and verification (in lieu of the audit trail).

From 1993 onward, physical controls for VAT at interstate borders have been replaced by accounting controls at the first inland stage of the production process. Registered firms of taxable imported goods have to include these goods (intra-community acquisitions are deemed to be a separate chargeable event) in the return due for the period in which the goods are imported. The VAT is self-assessed, as it were, and a credit is provided at the same time. Accordingly, no net VAT is due and payable, unless the acquisition is made by an exempt firm. This arrangement is called 'deferred payment' (Cnossen (1983), p. 156), because the VAT does not have to be paid upfront at the border.⁶ In 1982, the European Commission proposed the EU-wide introduction of deferred payment (Draft Fourteenth VAT Directive),⁷ but changed its position three years later (European Commission (1985)) in favour of exporter rating. The member states, however, insisted on deferred payment, which had proven its feasibility in Benelux, ever since Belgium, Luxembourg,

⁶ Although correct if looked at from the border VAT point of view, deferred payment is a misnomer when viewed in the domestic context, because it treats imported goods exactly on a par with domestic goods (the right to a tax credit arises at the same time that the VAT on supplies is accounted for). Particularly, it eliminates the previous cash flow disadvantage, i.e. the interest foregone by the importing firm on the VAT paid at import before this VAT was credited (or refunded) against the domestic VAT payable on sales. Some member states used to neutralize this cash flow disadvantage by allowing importing firms a grace period of, say, 6 weeks, before the import VAT had to be paid.

⁷ Without abolishing border controls. Importers of VAT-liable goods still had to hand over VAT documentation to customs authorities at the point of importation without actually being assessed for VAT at that point.

and the Netherlands introduced their VATs in the late 1960s and early 1970s.⁸

Deferred payment places the charge to VAT on the purchaser of imported goods rather than the supplier: the charge to VAT is reversed, as it were. Reverse charging, generally applied, is believed to take care of the 'missing trader', that is, the trader who charges VAT on his supplies but then disappears, while his customer takes credit for VAT that has not been accounted for. The missing trader may be an importer, but can also be a domestic trader. Essentially, contrived insolvency has the same effect; it was practised long before 1992.

Reverse charging, applied throughout the domestic production–distribution process, would nullify the fractional collection nature of the VAT, under which the tax office has a lien, as it were, on suppliers for the tax payable by their customers (ultimately the retailers that collect the VAT from consumers). Austria and Germany have requested permission to apply reverse charging to transactions above €10,000 and €5,000, respectively, but the European Commission (2006) has denied the requests on the ground that reverse charging would increase compliance control problems at the lower end of the production chain where most small businesses can be found.⁹ This commentator believes that comprehensive reverse charging would throw the baby out with the bathwater.

At the time the transitional regime was introduced, customs controls were replaced by a functionally equivalent, if perhaps less certain, VAT information exchange system (VIES) (Council Regulation 1798/2003/EC).¹⁰ VIES requires registered firms to report their intra-community supplies (exports) to registered firms in other member states (indicating their VAT identification numbers preceded by a country code) on a quarterly or monthly basis (listing requirement) to the VAT office. Similarly, the purchaser has to report the total of his intra-community acquisitions. The information is fed into a central data bank and enables the various VAT administrations in the EU to match total intra-community acquisitions per trader against individually reported supplies. VIES imposes differentially higher

⁸ Prior to 1993, deferred payment was also practised for some time in Ireland and the UK where it was called postponed accounting.

⁹ For Germany, see the proposals in Dziadkowski et al. (2002) as well as Gebauer et al. (2007). The European Council (Decision 2007/250/EC), however, has permitted the UK to apply reverse charging to supplies of mobile phones, computer chips, and some other goods.

¹⁰ Furthermore, a statistical data collection system, referred to as the Intrastat system, was set up to collect trade data on goods and (later) services between member states (Council Regulation 638/2004/EC). Recently, the Intrastat obligations have been simplified (COM(2008) 58 final).

compliance costs on interstate traders—a source of discrimination and trade distortion.¹¹

2.2. Cross-border transactions in services

Although services are economically equivalent to goods, their VAT border-crossing treatment differs. Prior to 1993, arrangements for BTAs on goods through border controls could not deal effectively with (non-tangible) services whose location of supply or purchase is difficult to ascertain. Obviously, interstate differences in VAT rates would generate distortions if the liability to tax was determined by the state in which the service was supplied. But if the purchasing firm's state would be the taxing locus, it would be difficult to tax purchases by final consumers.¹²

Article 43 of the 2006 VAT Directive (previously Article 9 of the 1977 Sixth VAT Directive) provides a workable solution to the issue by taxing services, in principle, in the state where they are performed. Highly significant exceptions, however, are made in Article 56 for services rendered by banks, insurance companies, professional firms, advertising agencies, and various other services (nearly all B2B transactions). Upon export, these services are exempt, although the exporter retains the right to a credit for the VAT in respect of any inputs used in performing the services. In the importing state, furthermore, the taxable users of services are liable to VAT on them under Article 196 of the 2006 VAT Directive (previously Article 21 of the 1977 Directive). In essence, prior to 1993, the current transitional regime for goods was already being applied to services.

Initially, there was no need to have separate rules for the acquisition of out-of-state services similar to the rules for goods bought by exempt entities, those sold by mail order firms or for individually imported means of transport—which are all taxable in the destination state. Few high-value services were bought out-of-state in low-VAT countries. This changed dramatically, however, with the advance and privatization of telecommunication, radio, television, and electronic services. Firms providing these services sprung up in Luxembourg (which has one of the lowest VAT rates in the EU) to the chagrin

¹¹ See Verwaal and Cnossen (2002) who point out that most of the differential costs should be attributed to Intrastat obligations (see footnote 10 above), not to VAT obligations.

¹² Note that the much feared wave of cross-border purchases of consumer goods (Sinn (1990)) was already a non-issue for services prior to 1993. In fact, the treatment of border-crossing services, before and after 1993, indicates that the implications of the break-in-the-VAT-chain for goods when border controls were abolished, were not unfamiliar phenomena.

of the member states where the services were consumed and supposed to be taxed.¹³

To remedy this situation and, more generally, to put the VAT treatment of services on a par with the treatment of goods, recently, new rules (which will take effect in 2010) for the place where services are deemed to be rendered have been promulgated with the primary goal of taxing services as much as possible at the place of consumption or use (Council Directive 2008/8/EC of 12 February 2008). In effect, the deferred payment system for goods will be extended to B2B services under an identical reverse charging system. Basically, the exceptions have become the main rule: B2B services will be deemed to be provided where the customer carries on his business. Overriding exceptions are provided for immovable property, cultural services and education, restaurants and catering, transportation of persons, and short-term rentals of vehicles, which are deemed to be provided where the services are actually performed. As is the case with goods, the VAT identification number will play a crucial role in verifying compliance.

The main rule for B2C services, as before, is that they are deemed to be provided at the place where the provider of the services carries on his business. To limit administrative costs, a mini one-stop shop arrangement will be provided for telecommunication, radio, and television services (effective 2015). The tax on these services will be payable in the state where the services are provided. Subsequently, that state distributes the VAT revenue to the member states where the customers are located according to an agreed formula. In fact, this is a mini-form of exporter rating!

2.3. Improving compliance and enforcement symmetry

Basically, the latest directive unifies the VAT treatment of intra-community transactions in goods with similar transactions in services (not, of course, for trade with third countries). Deferred payment/reverse charging (VAT in the destination state) is explicitly applied to goods and B2B services. B2C consumer services are taxed in the state where the services are actually performed, similar to cross-border purchases of goods which are taxed, with minor exceptions, on an origin basis. VIES reporting obligations for goods as well as services potentially extend the verification trail across borders. Compliance symmetry between in-state and out-of-state transactions has been enhanced by requiring in-state suppliers to show their VAT identification number and

¹³ In addition, complications arose regarding third-country providers of services which had to choose a member state of domicile to discharge their VAT obligations.

that of their customers on invoices, as out-of-state suppliers have to do. What remains to be done is the establishment of tax audit units with EU-wide investigative powers to monitor VAT compliance with intra-community transactions. In other words, verification should be complemented by audit control. For all practical purposes, the new arrangements for services seem to transform the transitional regime into the definitive regime, although, as shown below, the debate on some form of exporter rating instead of deferred payment continues.

3. EXPORTER RATING SYSTEMS PROPOSED BY THE EUROPEAN COMMISSION

Very different, perhaps more exciting if less practical, coordination proposals have been made in the tax literature and by the European Commission, which focus mainly on repairing the supposed break in the VAT collection chain. In the early 1980s, Cnossen (1983), argued that the destination principle¹⁴ could be maintained without border controls if intra-community exports would be taxed at the VAT rate of the exporting member state, invoiced to the importer in the importing member state, and credited by him against his VAT liability on sales. To restore the revenue allocation under the destination principle, the importing state would have to reclaim the importer's credit from the exporting state. In essence, only net balances (VAT collections on exports over VAT credits on imports) would have to be settled between member states. This could be done on an EU-wide basis through what Cnossen (1983) called a clearing house system. The idea was adopted by the European Commission (1985) in a paper known as the Cockfield White Paper,¹⁵ and forms the basis of subsequent exporter rating proposals.

¹⁴ Early on, the origin principle was viewed as the only way of doing away with border controls for VAT (Neumark Committee (1963), pp. 145–9). In fact, the doctrine of the 'restricted origin principle' (applied in the EU but not to trade with third countries; Shibata ((1967), pp. 193–4), became a standard tenet in public finance textbooks; see e.g. Due and Friedlaender ((1975), p. 519), Musgrave and Musgrave ((1980), p. 644), and Shoup ((1969), p. 644). For an analysis in the US setting, which reaches the same conclusion, see McLure ((1980), pp. 127–39). Subsequent writings, however, emphasized that origin taxation violated production efficiency—a more important criterion than exchange efficiency (Diamond and Mirrlees (1971))—and that it involved contentious transfer pricing problems (Cnossen and Shoup ((1987), pp. 72–73). After weighing the theoretical and technical arguments, this chapter (Section 4.5.1) concludes that the destination principle is the best lode star for practical VAT design.

¹⁵ Interestingly, as perceived by the European Commission, exporter rating seemed to be in line with the erstwhile pronouncement on origin taxation in Article 4 of the First VAT Directive (which called for the abolition of 'the imposition of tax on importation and the remission of tax on exportation in trade between member states'), since intra-community supplies were taxed in the

The Commission's arguments, however, failed to persuade the governments of the member states which wanted to retain full control over the VAT administration of imports and exports, and, therefore, opted for the deferred payment system, although they agreed that exporter rating should receive a second hearing before 1997. In the tax literature, moreover, Lee et al. (1988) had criticized the Commission's clearing house proposal, pointing out that it had an adverse impact on enforcement incentives, because importing member states might not be inclined to root out fraudulent claims for import VAT credits (after all these would presumably be paid by exporting states), while exporting states would have little incentive to uncover fraudulent failure to charge VAT on exports. Solving this problem would require uncoupling the clearing house flows from taxes actually paid.

Subsequently, the European Commission (1996) made another attempt to persuade the member states of the benefits of exporter rating. The fresh proposal was called home-state taxation, because firms involved in intra-EU trade would have to deal only with the VAT system of the member state in which they were established.¹⁶ Cross-border sales would be taxed in the same fashion as domestic sales, although cross-border movement of goods within the same business would go untaxed. VAT revenues on intra-community transactions would be allocated between member states on the basis of national accounts statistics of aggregate consumption. Complete uniformity in the scope and definition of VAT would also be necessary, as well as close cooperation and EU supervision of VAT administrations.

In an eloquent commentary, Smith (1997) pointed out that the new proposal would put substantial limitations on member states' autonomy to set rates, require an extensive programme of legislative harmonization, cause difficulties in identifying firms entitled to be taxed in a single member state, and a flight of businesses to least-taxed locations or, if rates were the same, to states where VAT evasion would be less tightly controlled. Also revenue allocation rules would undermine incentives to devote adequate resources to VAT collection and enforcement.

state of origin even though revenue would be allocated on the destination principle. Accordingly, to this day the Commission continues to insist on dubbing exporter rating as origin taxation, confusing principle and method for the sake of an outdated point of view.

¹⁶ Since intra-EU exports were zero rated under deferred payment, the European Commission (1996) noted that more than €700 billion worth of goods circulated VAT-free in the internal market, and observed that due to the break in the VAT-collection chain 'some of that amount may well be diverted to the black economy'. In light of the above discussion, it will be noted that all goods and services, not just intra-community traded products, circulate VAT free within the ring of registered firms.

The proposal did not leave the drawing board, but was briefly resurrected in 2004 as the one-stop shop proposal under which exporters would be able to discharge all their obligations with respect to border-crossing transactions at one place only, that is, their place of establishment.¹⁷ This time clearing would not be necessary because exporters would be required to remit the gross VAT collected by them and calculated at the destination-state rate directly to the state of final destination, an idea which had earlier been proposed by Vanistendael (1995).¹⁸ Few details were provided. In the meantime, the Commission seems to believe that the transitional regime will be around for some time to come: in the 2006 VAT Directive, the transitional measures are no longer grouped together, but integrated with related provisions.

4. SHOULD VIVAT BE LAID TO REST?

In the belief that the supposed break in the VAT-collection chain threatens VAT's integrity, Keen and Smith (1996, 2000) have made an imaginative, high-profile proposal for a viable integrated VAT (VIVAT), which would impose a harmonized dual EU-VAT rate, administered by member states, on all pre-retail (intermediate) transactions by registered firms within and between EU member states, supplemented by a surtax at retail (in fact an RST integrated with a member state's VAT) if governments wish to exploit the VAT base more intensely. Clearing would be provided if the VAT collected on exports exceeded the VAT credits provided for imports.

According to Keen and Smith (1996), as well as the authors of this chapter, VIVAT would bolster the destination principle and hence subsidiarity in taxation. The commonality of the EU single rate would lessen the pressure on the clearing system and enforcement. Traders would be able to report exports and imports in aggregate rather than per member state. The uniform single rate would remove the incentive for strategic rate setting, that is, the incentive member states would have under exporter rating with non-harmonized rates to tax exports higher because the VAT would anyway be creditable in importing states.¹⁹ The hassle of the clearing system could be resolved through a

¹⁷ Some details can be gleaned from IP/04/1331 and MEMO/04/249.

¹⁸ The one-stop shop proposal may receive further scrutiny in the run up to the implementation by 2010 of the Services Directive (Council Directive 2006/123/EC), which calls for 'points of single contact' in each member state where traders can discharge all their obligations in other member states.

¹⁹ It is difficult to view this as an advantage of VIVAT, since any attempt at strategic rate setting would be stopped in its tracks by the European Court of Justice on the ground that it would be a gross violation of the non-discrimination principle.

one-off deal, that is, a system of lump sum transfers between member states, obviating the need for future clearing.

VIVAT, however, would not be without its own problems. Uniform exporter rating may appear to repair the break in the VAT-collection chain, but does nothing to solve the break in the VAT-audit trail. Importing member states would still not be able to audit importers' invoices (received from exporters in other member states) for which they have no authority. This would provide a powerful incentive to produce false import invoices, possibly arranged through third countries, showing VAT eligible for credit instead of no VAT as under the current regime. Keen and Smith's (1996) proposal that the excess of collections on exports over imports should be allocated on the basis of export and import listings by businesses of aggregate rather than individual transactions would undermine enforcement efforts, since it would not be possible, as under VIES, to link individual transactions to monitor compliance for audit purposes.

Furthermore, under VIVAT, member states with a greater than average preference for VAT would have to impose an additional RST. In other words, they would have to incur higher administrative and compliance costs than currently and than member states making do with the revenue collected under the VIVAT rate. Registered traders, moreover, would have to make a distinction between sales made to other registered traders (taxable at the VIVAT rate) and sales made to non-registered persons, that is, individuals and exempt entities (taxed at the RST-inclusive rate)—'not a trivial burden', as Keen and Smith (1996, p. 406) admit. The RST-element would have all the drawbacks of a normal RST, noted in the literature (e.g. Cnossen (1987)).

Beyond that, it is difficult to envisage a uniform VIVAT rate in light of the established, if perhaps misguided, preference for greater rate differentials shown by the member states. Also, reduced rates are levied on a product-specific basis. Their application (with revenue consequences) to intermediate transactions would complicate VIVAT and open up other avenues for fraud, particularly if it is not possible to audit import invoices. Substantial refunds might have to be paid out if the VAT rate on a product applied at retail in the importing state would be lower than the VIVAT rate in the exporting state. Politically, VIVAT would further entrench the (high) VAT rate agreement in the EU, making it more difficult to convert the differentiated rate structures into single uniform rates or to reduce the VAT rates in individual member states at some future date.

Finally, a staff working paper issued by the European Commission (SEC/2008/249) notes that member states would become dependent on each

other for, on average, some 10% of total VAT revenues if exporter rating were applied at a uniform rate of 15% (the Netherlands, Germany, Belgium, and Ireland would become large net contributors to the clearing system). The level of mutual trust, therefore, would have to be exceptionally high, particularly in view of the wide variation in VAT fraud levels. Furthermore, mismatches between supply and acquisition listings would arise. In 2006, the excess of reported acquisitions over supplies was €80 billion. At a rate of 15%, the potential amount of VAT involved could be €12 billion. Exporter rating, moreover, would only target missing trader fraud. It would not solve and might even exacerbate VAT losses through the shadow economy, contrived insolvencies, or other domestic fraud.

Other proposals for VAT coordination in the EU have been made by McLure, and Bird and Gendron. McLure (2000), building on a proposal by Varsano (1999) for the Brazilian states, suggests that a separate uniform rate EU-VAT (which he calls ‘compensating value added tax’ or C-VAT) should be introduced, administered by a central agency (or a consortium of states), on all interstate exports matched by a credit for tax on all interstate imports (accordingly, there would be no need for a clearing mechanism). State VATs would be retained along with the deferred payment system for interstate trade. In essence, C-VAT is functionally equivalent to VIES but involves greater administrative complexity.

Bird and Gendron’s (1998) dual VAT (D-VAT), based on the experience in Canada (which has a federal VAT, while Quebec operates a state VAT), envisages a central VAT (next to the state VATs), which would apply to all in-state and interstate sales. Unlike the C-VAT, the D-VAT would raise revenue for the centre. As Keen and Smith (2000) have pointed out both C-VAT and D-VAT currently are not options in the EU if there is to be a central administration. Bird and Gendron ((1998), p. 439), however, do allow for a virtual, functionally equivalent D-VAT in the form of ‘some closely coordinated overarching administrative structure which would, for example, facilitate and ensure information exchanges, development of agreed audit plans, and so on’, in order to give states the capacity to monitor cross-border transactions. It is this idea that is also central to the gist of this commentary.

It is hard to avoid the impression that VIVAT, C-VAT, and D-VAT are heavily predicated on the assumption that a solution has to be found for the VAT treatment of interstate trade in (physical) goods²⁰—so central to the

²⁰ See, for instance, Keen and Smith ((2000), pp. 743–4) who make their case for VIVAT against the background of the abolition of ‘effective border controls’.

EU's 1992 programme—while the practice and experience with (intangible) services is just as or even more important. Bringing services into the equation emphasizes the point that it is the break in the VAT-audit trail that should be the focus of concern. Deferred payment involves fraud, of course, but the answer to fraud, it seems, is audit, investigation, and prosecution, not another system that may be equally susceptible to abuse. Substituting a tried and proven system of deferred payment by some form of exporter rating may turn out to be a costly risk. A simple system of no tax and no valuation of exports would be replaced by a system with a positive tax and perhaps valuation problems. It is quite telling, perhaps, that the Netherlands (which has had deferred payment for nearly forty years!) has the lowest VAT evasion rate (the VAT gap as a percentage of hypothetical revenue) among ten member states for which figures are available (Gebauer and Parsche (2003)), although it is the most open economy of all.

5. VAT BASE BROADENING

One of the contentious issues under deferred payment is the treatment of cross-border acquisitions by exempt entities, which have to self-assess their intra-community acquisitions (which are zero rated in the supplying member state) on a transaction-by-transaction basis. Ad hoc self-assessment with its attendant complexity would not be necessary if the entities were not exempt but registered for VAT purposes like other businesses. More generally, the EU-VAT exemptions violate the logic and functionality of the tax. They distort input choices, harm exports, and complicate administration because the VAT on inputs has to be denied with respect to exempt supplies if performed in combination with taxable transactions. Accordingly, a strong case can be made for repealing most of the 'standard exemptions'.²¹

The case against the exemption of cultural services is particularly strong. Admissions to theatres, concerts, museums, sporting facilities, and the like compete with taxable forms of entertainment, such as travel and reading, and should therefore be taxed. Similarly, public radio and television broadcasts compete with taxable privately financed broadcasts and other forms of communication. This applies also to postal services (which compete with taxable private letter or parcel carriers), newspapers, and periodicals. Since it is difficult to justify these exemptions on externality grounds, withdrawal seems indicated. Various member states, including the UK, are coming around to

²¹ This section draws heavily on Cnossen (2003).

the view that cultural services should be taxed, but taxation is by no means universal in the EU.

Even the exemptions for healthcare and for education services hardly stand up to close scrutiny. Admittedly, the externality arguments are stronger than in the case of cultural services, but if health and education services should be provided below cost, then (an increase in) budget subsidies (or a zero rate) would be the appropriate policy response. The exemptions violate production efficiency because the institutions providing healthcare and education services are induced to perform laundry, cleaning, food preparation, and various other services in-house in order to save the payment of VAT on the labour element of these services, which would be payable had they been acquired from outside, taxable establishments. This hampers the contracting-out of these services (privatization) and thus the efficient functioning of the institutions. Exemption also raises the cost for companies wishing to conduct research through hospitals and universities, because they cannot take implicit credit for the VAT on the inputs used by the exempt institutions. These considerations become more important as the private provision of health care and education grows relative to public provision, as is happening in the EU.

Administratively beyond reach so far are financial transactions, because the intermediation charge, which should be taxed, cannot be separated from the pure interest rate, premium or rate of return which should not be taxed. Perhaps the cash flow approach, pioneered by Poddar and English (see Poddar (2003)), deserves further scrutiny. The EU has closely considered their ingenious idea, but doubts remain about its practicability. There is little doubt about the desirability of taxing public bodies more widely, particularly local and provincial governments. Competition is distorted to the extent government services compete with similar services provided by the private sector. As with hospitals, universities, and financial institutions, taxation would obviate the need for delineation between taxable and exempt government activities as well as for ad hoc self-assessment if taxable goods and services are acquired out-of-state.

Another anomaly is the flat rate schemes for the agricultural sector which exempt farmers from the obligation to register for and pay VAT, but compensate them for the tax borne on inputs. The schemes provide only rough justice and, just like the old turnover tax, can be used to subsidize farmers. Best practice would be to make farmers fully liable for VAT, subject to the small-business exemption. Similar comments can be made about the various small-business schemes, which add greatly to administrative and compliance costs without contributing much to revenue. Experience in new VAT countries indicates that the simplest small-business scheme is a fairly generous

exemption without any strings attached. Indeed, this is also the outcome of the formula discussed in this chapter.

6. CONCLUSIONS

This commentary has argued that exporter rating does not seem to have obvious advantages over deferred payment for the VAT treatment of intra-community transactions and may complicate VAT administration. Instead, the search should be for a workable system that extends the VAT audit trail by setting up cross-border tax audit and investigation units to monitor intra-community transactions. Precedents for this can be found in police and judicial units with cross-border pursuing and investigative powers. If this is done, the current transitional regime can be retained. Furthermore, explicit reverse charging should remain the exception rather than become the rule.

Last but not least, the 2006 VAT Directive, which is based on its 1977 predecessor (long before the wall came down), has not stood the test of time. Efficiency in production and tax collection is not served by the large number of so-called standard exemptions. Admittedly, improvements to the 2006 VAT Directive require the consent of twenty-seven member states, which is hard to come by.²² If the member states are not to be locked into outmoded VATs, perhaps they should be permitted to have better VATs than provided by the 2006 VAT Directive. The time has come to support this chapter's call for allowing more experimentation than has hitherto been the case.

REFERENCES

- Bird, R. M., and Gendron, P.-P., (1998), 'Dual VATs and Cross-Border Trade: Two Problems, One Solution?', *International Tax and Public Finance*, 3, 429–42.
- (2007), *The VAT in Developing and Transition Countries*, Cambridge, Cambridge University Press.
- Cnossen, S. (1983), 'Harmonization of Indirect Taxes in the EEC', chapter 7 in McLure, C. E. Jr. (ed.), *Tax Assignment in Federal Countries*, 150–68. Canberra: ANU Press.

²² This chapter argues that a common VAT base is helpful in determining the VAT-contribution to the EU's 'own resources', but this contribution is calculated by statistical agencies on the basis of national accounts, divorced from actual VATs. In other words, the actual VAT base does not have to coincide with the agreed VAT-contribution base.

- (1987), 'VAT and RST: A Comparison', *Canadian Tax Journal*, 35, 559–615.
- and Shoup, C.S. (1987), 'Coordination of Value-Added Taxes', chapter 2 in Cnossen, S. (ed.), *Tax Coordination in the European Community*, 59–84, Deventer, the Netherlands: Kluwer Law and Taxation Publishers.
- (2003), 'Is the VAT's Sixth Directive Becoming an Anachronism?', *European Taxation*, 43, 434–42.
- Diamond, P. A., and Mirrlees, J. A. (1971), 'Optimal Taxation and Public Production: I Production Efficiency, II Tax Rules', *American Economic Review*, 61, 8–27.
- Due, J. F. and Friedlaender, A. F. (1975), *Government Finance: Economics of the Public Sector*, 6th edn., Homewood, Ill.: Irwin.
- Dziadkowski, D., Gebauer, A., Lohse, W. C., Nam, C. W., and Parsche, R. (2002), 'Entwicklung des Umsatzsteueraufkommens und finanzielle Auswirkungen neuerer Modelle bei der Umsatzbesteuerung', *Ifo Forschungsberichte* 13, Munich.
- Ebrill, L., Keen, M., Bodin, J-P., and Summers, V. (2001), *The Modern VAT*, Washington DC: International Monetary Fund.
- European Commission (1985), *Completing the Internal Market*, White Paper from the Commission to the European Council ('Cockfield' White Paper), COM(85) 310 final, Brussels.
- (1996), *A Common System of VAT: A Programme for the Single Market*, COM 328(96), final, Brussels.
- (2006), *Communication from the Commission to the Council in Accordance with Article 27(3) of Directive 77/388/EEC*, COM(2006) 404 final, Brussels.
- Gebauer, A., and Parsche, R. (2003), 'Evasion of Value-Added Taxes in Europe: IFO Approach to Estimating the Evasion of Value-Added Taxes on the Basis of National Accounts Data (NAD)', *CESifo DICE Report*, 2, 40–4.
- Nam, C. W., and Parsche, R. (2007), 'Can Reform Models of Value Added Taxation Stop the VAT Evasion and Revenue Shortfalls in the EU?', *Journal of Economic Policy Reform*, 10, 1–13.
- Keen, M., and Smith, S. (1996), 'The Future of Value-Added Tax in the European Union', *Economic Policy*, 11, 373–411 and 419–20.
- — (2000), 'Viva VIVAT!', *International Tax and Public Finance*, 7, 741–51.
- (2006), 'VAT Fraud and Evasion: What Do We Know and What Can Be Done?', *National Tax Journal*, 59, 861–87.
- Lee, C., Pearson, M., and Smith, S. (1988), *Fiscal Harmonisation: An Analysis of the Commission's Proposals*, IFS Report Series 28, London: Institute for Fiscal Studies.
- McLure, C. E., Jr. (1980), 'State and Federal Relations in the Taxation of Value Added', *The Journal of Corporation Law*, 6, 127–39.
- (2000), 'Implementing Sub-National VATs on Internal Trade: The Compensating VAT (CVAT)', *International Tax and Public Finance*, 7, 723–40.
- Musgrave, R. A., and Musgrave, P. B. (1980), *Public Finance in Theory and Practice*, 3rd edition, New York: McGraw-Hill.

- Neumark Committee (1963), Report of the Fiscal and Financial Committee, in *The EEC Reports on Tax Harmonization*, Amsterdam: International Bureau of Fiscal Documentation.
- Poddar, S. (2003), 'Consumption Taxes: The Role of the Value-Added Tax', chapter 12 in Honohan, P. (ed.), *Taxation of Financial Intermediation: Theory and Practice for Emerging Economies*, New York: Published for the World Bank by Oxford University Press, 345–80.
- Shibata, H. (1967), 'The Theory of Economic Unions: A Comparative Analysis of Customs Unions, Free Trade Areas and Tax Unions', in Shoup, C. S. (ed.), *Fiscal Harmonization in Common Markets, vols. 1 and 2*, New York: Columbia University Press.
- Shoup, C. S. (1969), *Public Finance*, Chicago: Aldine Publishing Company.
- Sinn, H-W. (1990), 'Tax Harmonisation and Tax Competition in Europe', *European Economic Review*, 34, 489–504.
- Smith, S. (1997), *The Definitive Regime for VAT*, Commentary 63, London: Institute for Fiscal Studies.
- Vanistendael, F. (1995), 'A Proposal for a Definitive VAT System. Taxation in the Country of Origin at the Rate of the Country of Destination without Clearing', *EC Tax Review* 1, 45–53.
- Varsano, R. (1999), 'Subnational Taxation and Treatment of Interstate Trade in Brazil: Problems and a Proposed Solution', in Burki, S. J., and Perry, G. (eds.), *Proceedings of the Annual Bank Conference on Development in Latin America and the Caribbean*, 339–55. Washington, DC: The World Bank.
- Verwaal, E., and Cnossen, S. (2002), 'Europe's New Border Taxes', *Journal of Common Market Studies*, 40, 309–30.

Commentary by Ian Dickson and David White*

Ian Dickson is a shareholder and director in Burleigh Evatt Consulting, a New Zealand professional services firm. Ian began his career in the New Zealand Treasury, where he was involved in the introduction of the Goods and Services Tax in 1986. Since leaving the Treasury he has been involved in investment banking, corporate finance, and strategy consulting. Recently he has been an adviser contributing to the revenue reform programme in the Kingdom of Tonga. The Kingdom introduced a VAT called 'Consumption Tax' in 2005.

David White has been responsible for the tax research agenda of the Centre for Accounting, Governance and Taxation Research at Victoria University of Wellington, New Zealand, since the Centre was set up in 2001. His research focuses on tax policy and tax policy process design, especially in relation to the Goods and Services Tax and international income taxation. He worked in the New Zealand Treasury tax policy branch 1987–2000. He studied law at Victoria University of Wellington and the London School of Economics and received his doctorate in law from the University of Sydney.

1. INTRODUCTION

The editors of the Mirrlees Review asked us to evaluate the UK VAT and the European Union VAT Directive and compare them to the New Zealand VAT model, which is called GST (Goods and Services Tax). At first sight, this brief might seem strange. After all, the VAT is a European invention that has swept the world, now applying in around 150 countries while GST is merely a more recent, modified European VAT.

Yet, for the whole life of the New Zealand GST, international VAT experts, including from Europe, have made favourable comparisons between the old

* We are grateful to two anonymous reviewers for their helpful comments on the manuscript.

European VAT model and the GST model.¹ In many cases, these comparisons have concluded that the changes that New Zealand made when introducing its VAT in the mid-1980s were improvements on the European model—a case of same wine, new bottle. Countries seeking to introduce VAT have often been advised to consider New Zealand's approach, and many countries have done so.

Surprisingly, there has been little research on GST, particularly of an empirical nature, to test objectively the merits of the innovations and the international experts' advice.² The recent twentieth anniversary of the New Zealand GST, however, did provide an opportunity to begin rectifying that omission and to develop a research agenda.³ A multidisciplinary team of twenty-nine experts from eight countries analysed and discussed the New Zealand GST experience, the influence of the New Zealand model around the world, and current issues for VAT tax systems everywhere. This commentary draws on the work of that team,⁴ our experience as tax policy-makers, and our research.

New Zealand's GST is different from the UK VAT in the important respect that it is applied comprehensively at a single domestic rate. There are no reduced or super-reduced rates, exemptions, or zero rates—other than those necessary to define the appropriate base of the tax.⁵ Food, children's clothing, medical care, education services, publications, energy, and other necessities of life are taxed at 12½% like all other goods and services. This was a deliberate policy choice made at the time of introduction and maintained intact over the intervening years. Applying the most widely used indicator of VAT performance, the C-efficiency ratio,⁶ the New Zealand GST scores twice as highly as the UK VAT.⁷ Indeed, the GST is apparently a huge 43.5 percentage points above the OECD average for C-efficiency.⁸ It is this long-standing, comprehensive application of GST that provides insights for formulation of indirect tax policy for the UK. The commentary will also briefly consider aspects of how the New Zealand GST applies to services and organizations that in Europe are often considered too hard to tax.

¹ See, for example, Cnossen's (2003) recent call for a comprehensive overhaul of the EU VAT Directive.

² Early analysis by New Zealand Institute of Economic Research staff is summarized and discussed in Bollard (1992).

³ White (2007a) especially at 372–4. A shorter version of this article also appeared in (2007b).

⁴ Krever and White (eds.) (2007).

⁵ In the category of 'base-defining' exemptions and zero rates, we include exports, provision of labour services, and investment capital.

⁶ C-efficiency is the ratio of VAT revenue to consumption expenditure, divided by the standard tax rate, expressed as a percentage.

⁷ OECD (2006) at 53.

⁸ If 'departmental GST' is excluded from the calculation (since it results in no net revenue

2. UNITED KINGDOM AND NEW ZEALAND:
COMMONALITIES AND DIFFERENCES

Before analysing the New Zealand GST model, we need to examine the contextual similarities and differences between New Zealand and the UK. The similarities are easiest. Both countries are islands of roughly the same area. New Zealand is a former colony of the UK and is English-speaking. The two countries have a great many cultural similarities and historical ties, including a common legal heritage, political institutions, and experience of geopolitical conflict. Naturally, the development of many areas of government policy in New Zealand has over time followed the lead of policy in the UK—although the development of indirect tax policy in the last twenty years is a notable exception. Britain’s policy choices, such as to enter the Common Market in 1973, have also affected aspects of social, economic, and foreign policy and the sense of national identity in its remote former colony. Both countries have modern information technology systems in both the private and public sectors (including, in the customs and revenue authorities).

New Zealand and the UK also employ broadly similar tax types, with some variation. New Zealand has neither separate capital taxes nor National Insurance contributions (pensions are funded from general revenue on a pay-as-you-go basis). New Zealand relies a little more heavily on its VAT as a source of taxation than the UK while the UK relies a little more heavily on its excise duties on specific goods and services, like motor fuels, alcohol beverages, and tobacco products (Table 1).

Both countries have extensive social welfare systems and both provide income tax credits for families with insufficient market income to provide an acceptable living standard. New Zealand’s tax-funded universal pension is available at the age of 65 years to all who meet residential requirements. The level of payment is reviewed each year and is adjusted to take account of increases in inflation and wages. When wages increase, New Zealand

to government), the New Zealand C-efficiency ratio reduces by about one-third, as shown in the following table.

Years ended 30 June	1999	2000	2001	2002	2003	2004	2005
C-efficiency ratio	105.9	107.6	106.3	88.3	90.9	94.9	93.1
C-efficiency ratio adjusted for departmental GST	78.2	79.4	78.4	65.2	67.1	70.1	68.7

Sources: Statistics New Zealand’s Consolidated Accounts of the Nation; Financial Statements of the Government of New Zealand (Annual).

Table 1. Comparison of VAT and Excises in the United Kingdom and New Zealand

	United Kingdom	New Zealand
Year VAT introduced	1973	1986
Standard VAT rate	17.5	12.5
General threshold (US\$)	93,700	26,846
VAT revenues as % of total taxation (2003)	19.8	26.1
C-efficiency ratio for VAT (2003) ⁹	46.4	96.4
Tax on general consumption as % of GDP ¹⁰	7.0	9.1
Tax on general consumption as % of total taxation	19.5	25.6
Tax on specific goods and services as % of GDP	4.0	2.3
Tax on specific goods and services as % of total taxation	11.2	6.5

Source: OECD (2006), 25–8, 30, 32, 52–3.

Superannuation is adjusted so that it is between 65% and 72.5% of average ordinary time earnings after tax.

The settings of the respective countries, however, are very different. New Zealand is a nation of 4.5 million individuals separated from its nearest neighbour, Australia, by 2,000 kilometres of sea, so that New Zealand has no proximate borders with any other countries. While New Zealand has economic integration agreements with Australia, and a growing number of trade agreements with countries in the Asia–Pacific region that generally have VATs of their own, it is not a member of any multilateral relationship that compares with the UK's membership of the European Union in limiting domestic policy flexibility. Consequently, VAT administration issues that arose for the UK with the elimination of internal frontiers in the European Union, and the potential tax competition, have not been ones for New Zealand.

3. A COMPREHENSIVE BASE AND SINGLE RATE: LESSONS FROM NEW ZEALAND

The Meade Report¹¹ had an immense influence on the development of tax policy thinking in New Zealand in the years following its publication. It

⁹ See note 6 above for the definition of the C-efficiency ratio. Also, in note 8 above, we suggested an adjustment to the C-efficiency ratio for New Zealand in order to exclude departmental GST. Using official New Zealand statistics, we calculate the C-efficiency ratio adjusted for departmental GST in 2003 at 67.1.

¹⁰ These last four tax ratios use 2004 data.

¹¹ Meade (1978).

was much pored over, studied, debated, and quoted at length in officials' reports to New Zealand ministers. It provided guidance about how problems relating to the taxation of individuals and businesses, and wealth, should be addressed. Under Meade's influence, the formulation of tax policy advice became systematic and principles-based.

By the early 1980s, New Zealand was lumbered with an awful wholesale tax that omitted around 67% of the potential indirect tax base, and the country relied significantly upon international trade taxes. Other horrors abounded in the tax, regulation, and subsidization of industry, as well as the conduct of economic management generally. The ad hoc grab bag of taxes could not fund existing expenditure demands, let alone shoulder the burden of a switch towards a greater reliance on indirect taxation that was government policy. Meade could give New Zealand policymakers no direct guidance on how to address this problem, nor on the pressing question of the day, 'which is better, VAT or retail sales tax?' Meade, however, inspired an investigative approach to these problems that eventually resolved these questions.

There are still many questions about the directions for indirect taxation in the twenty-first century. There are two central questions where the evidence from New Zealand can provide insights for the authors' proposed reforms of the UK VAT, which imposes reduced rates on more total consumption expenditure than the VAT in any other EU-15 Member State:¹²

- Are the alleged administration and compliance benefits of a comprehensive single-rate VAT¹³ borne out by the New Zealand experience?
- What is the evidence on the ability to alleviate the distributional impacts of a comprehensive single-rate GST on the poor and lowest paid?

Embodied in these two questions are the essential contrasts between the European-style and New Zealand-style VATs. This section first considers the New Zealand debate on comprehensiveness prior to introduction of GST in 1986, secondly, administration, compliance, and public acceptance issues, and, finally, the evidence on distributional impacts.

¹² Cnossen (2003).

¹³ Cnossen (1994). Cnossen concludes this review by arguing that overall compliance costs and administrative costs can be reduced by, among other things, broadening the VAT base, imposing a single rate, and increasing the threshold for registration.

3.1. The New Zealand debate on GST base comprehensiveness: 1981–86¹⁴

New Zealand policymakers were fortunate in having a dry run at indirect tax reform issues and arguments in 1981–82, several years before the successful campaign to introduce a GST in 1984–86. A task force that reviewed the tax system had proposed that the wholesale sales tax should be converted to a credit-offset basis, the indirect tax base should be broadened and supplies of services should be taxed separately.¹⁵ The then New Zealand Prime Minister adopted much of the task force's advice in 1982 but not their indirect tax proposal.

New Zealand policymakers learnt from the 1981–82 experience that maintaining a comprehensive tax base would be one of the most difficult issues, particularly in relation to food and clothing. They developed arguments for taxing food and other necessities on equity and economic efficiency grounds. They realized, moreover, that they would need to run public relations, consultation, and education campaigns on a scale previously not carried out and fully involving both the public and private sectors.¹⁶ This was a high-risk strategy since it focused on the very things the public would be most sensitive about paying tax on. If that debate were won, the strategy suggested, other people seeking exemptions would find it much harder to get public acceptance of their claim and there was a better chance of enacting a comprehensive indirect tax.

The argument conceded that taxing food was regressive. Evidence available at the time suggested that while the bottom 20% of households allocated between 23 and 29% of their budgets to food, the top two deciles spent between 7 and 10% of weekly expenditure on food. However, the answer to the question 'who spends the most on food?' was that upper-income households spend twice as much as low-income households. Of every \$100 spent on food in New Zealand, the least well off spent \$6.50, whereas the most well off spent \$12. Taxing all food thus made revenue available to redistribute and supplement the income of the poor. This argument, presented by a government with credibility, was broadly accepted and GST went ahead with a comprehensive base and single domestic rate.

¹⁴ This section draws on the historical and policy perspective in Dickson (2007).

¹⁵ McCaw (1982).

¹⁶ See the following chapters in Krever and White (eds.) (2007): Douglas (2007); Green (2007); Todd (2007); Dickson (2007).

3.2. Administration costs, compliance costs, and public acceptance

The first question under this heading is whether the alleged administration benefits of the comprehensive single-rate VAT are borne out by the New Zealand experience. Unfortunately, the administrative costs of operating GST are not separately identified because GST-related work is often not separate from other tax-gathering functions.

It is possible, however, to identify areas of considerable administrative cost saving in the GST system. GST does not require hundreds of pages of classifications of goods, services, and providers and the associated interpretations and rulings. It does not, to take one recent UK example, require officials to distinguish a 'biscuit' from a 'cake', and, if they make a mistake, to consider whether the refund should be for the total amount wrongly charged or whether this would constitute 'unjust enrichment' of the taxpayer.¹⁷ It also does not require officials to identify the public organizations that shall, and shall not, be subject to GST. GST remains quietly agnostic on such questions: the general rules apply.

It is also possible to identify areas of administrative cost and risk in the GST system.¹⁸ First, a comprehensive tax applied in an island country without near neighbours is still a 'cash refund' tax that carries the attendant risks of refunds. Even with the taxation of imports, some missing trader fraud occurs and a particular problem area has been property developers. Secondly, GST requires regular legislative maintenance, in one recent case blocking GST refunds with an estimated revenue cost of up to NZ\$200 million.¹⁹ Thirdly, in contrast to the UK, New Zealand still has separate agencies administering the border and inland collection of GST. The fragmented administration in New Zealand requires agency-to-agency information exchanges to audit the deduction for GST collected at the border and this may be more costly. The New Zealand Customs Service increasingly focuses on border security rather than revenue collection, so differing agency priorities are a risk issue.

¹⁷ The Marks & Spencer teacake case, concerning the misclassification of their chocolate teacakes as standard-rated biscuits rather than zero-rated cakes from 1973 to 1994, is still not settled after 10 years of argument. Marks & Spencer claim repayment of £3.5 million but the Commissioners of Customs & Excise refuse to refund more than 10% of the overpaid VAT.

¹⁸ The consideration of revenue risk draws, in part, on Snell (2007) in Kever and White (eds.) (2007).

¹⁹ The Taxation (Annual Rates, Taxpayer Assessment and Miscellaneous Provisions) Bill 2001 retrospectively blocked some GST refunds affecting inbound tour operators and educational institutes. See officials' letter to the Finance and Expenditure Committee of the New Zealand Parliament dated 31 July 2001, attached as Annex A in R. Oliver, 'Taking a Fixed Tax Position in a Changing World—A Personal Perspective', 19–25, available at <<http://taxpolicy.ird.govt.nz/publications/files/icanz2003.pdf>> (as at 26 February 2008).

While administrative costs of GST may be difficult to ascertain, policy administration costs are demonstrably low. In twenty years, GST has had one major policy review, which produced no recommended structural change. The Inland Revenue Department, responsible for most of the legislative work on GST, has the time of around around $1-1\frac{1}{2}$ professional policy staff dedicated to GST out of a total tax policy complement of 45. The key point is that GST has been a 'low maintenance tax' from a policy administration cost perspective²⁰ and this is generally attributed to the twin policy pillars of comprehensive coverage and a single domestic rate.

The second question is whether the alleged compliance benefits of the comprehensive single-rate VAT are borne out by the New Zealand experience. Here the evidence is less conclusive so this question cannot receive a short answer. There are five main observations that we would like to make. First, as international experts argue, VAT compliance costs are difficult to measure and interpret correctly²¹ and 'international comparisons of administrative and compliance costs should be regarded as tools to raise questions rather than providing immediate answers'.²² Therefore, care is needed in interpreting and drawing conclusions from the various surveys.

Secondly, two major New Zealand surveys in the last fifteen years have attempted to quantify GST compliance costs:

- In 1991–92, a benchmark survey by Sandford and Hasseldine²³ concluded that total compliance costs of the GST were large (7.3% of GST net revenue) and that mean annual compliance costs were regressive (\$1,066 or 1.6% of turnover for firms with \$30,000–\$100,000 turnover; \$3,521 or 0.2% of turnover for firms with \$1 million–\$2 million turnover; \$9,615 or 0.005% of turnover for firms with more than \$50 million turnover). The overall value of the cash flow benefit from the delayed payment of GST reduced these compliance costs figures by around 39%, providing greatest benefit to larger firms. Substantial offsetting managerial benefits from improved accounting practices were reported, mostly by smaller firms. At the time, only 23% of respondents used computers for their GST accounting.

²⁰ This label was used to describe the perspective of another key tax policy advising agency in New Zealand, the Treasury. See Snell (2007).

²¹ Clossen (1994), at 1665.

²² Sandford (2000).

²³ Sandford and Hasseldine (1992). All figures in New Zealand dollars.

- In 2004, a survey of SMEs²⁴ to provide a baseline for evaluating the effectiveness of future policy and administrative initiatives was conducted by an independent research organization for the Inland Revenue.²⁵ On average, respondents reported mean annual compliance costs of \$2,471 for income tax and \$1,553 for GST. GST compliance costs are regressive (\$1,285 for firms with \$40,000–\$99,000 turnover; \$2,646 for firms with more than \$1.3 million turnover). Fifty-three per cent used computers for their GST accounting. The survey revealed that levels of stress in meeting tax compliance requirements and finding the money to pay the tax were higher for GST (3.8%) and provisional income tax (3.7%) than for pay-as-you-earn income tax (3.2%) and fringe benefit tax (3.2%).

The results of both surveys are consistent with what is known about the relatively fixed nature of costs associated with administrative tasks in small businesses. They show that GST compliance costs are very regressive. The 2004 SME study says, for example, that the irreducible cost of GST compliance is around £40 per month.

Thirdly, a difficult issue concerns the marginal cost of GST compliance (that is, the amount that would be saved if the GST did not exist). There are two aspects to this issue. First, compliance costs must be attributed between tax types (in particular, the income tax and the GST). Secondly, compliance costs must be attributed between tax and core accounting costs. We believe that marginal GST compliance costs may not be as high as the New Zealand surveys suggest. A New Zealand case study that required participants to record tax and accounting functions in a weekly diary over 12 months, supplemented by regular interviews, provides an interesting insight. It concluded that the high proportion of tax compliance costs allocated to GST rather than to income tax in earlier New Zealand studies (the main one being Sandford and Hasseldine (1992)) may well be explained by the fact that GST reporting is more frequent (monthly, bimonthly, or six monthly) and that small-business people perceive income tax as the by-product of accounting first prepared for GST returns.²⁶ This suggests it may be fraught to attempt a meaningful attribution of tax compliance costs by tax type. The potential overlap between core external and internal accounting costs and GST compliance costs is even more problematic.²⁷ A strong possibility

²⁴ Firms that employed fifty or more staff and had an annual turnover of more than NZ\$10 million were excluded.

²⁵ Colmar Brunton (2005). All figures in New Zealand dollars.

²⁶ Ritchie (2001), in Evans, Pope, and Hasseldine (2001), at 312.

²⁷ For a discussion of the considerable efforts made to improve the compliance cost estimates in the 2004 survey and some of the outstanding issues, see: Oxley, Turner, and Sullivan (2005); and Oxley and Elwela (2006).

remains that the costs of basic business record keeping are represented as GST compliance, since merchants commonly refer to 'doing their GST' as shorthand for such activities.

Reinforcing this argument is the point that if GST compliance costs were an unreasonable burden on smaller businesses there would have been much more adverse comment than has occurred. Compliance costs with tax, regulatory, health and safety, and statistical obligations have been an on-going complaint from business advocacy groups, but the focus has been on other areas of fiscal encroachment on day-to-day business affairs rather than GST. This speaks to us of a GST system that nevertheless finds acceptance amongst the business community that collects the tax. Compliance with the huge volume of income tax changes, with fringe benefit tax and provisional income tax, where there is an element of uncertainty about establishing liability (as there would be in a VAT system with extensive exemptions), and with requirements that do not mirror ordinary day-to-day business activities (in the way that issuing and receiving invoices do), have attracted more comment and criticism.²⁸

Fourthly, further corroborating evidence on the level of compliance costs of GST may come from the tax advisory (as opposed to the tax compliance) profession. Until recently, the tax teams of all large law and accounting professional firms in New Zealand, with one exception, did not include a specialist GST partner. It was only in 2006 that a small number of other large professional firms considered that giving GST advice justified appointing a specialist GST partner.

Fifthly, a 15-year-old New Zealand and UK compliance cost comparison is available. The two studies had one author in common (Sandford) and used the same methodology. The comparison shows UK VAT compliance costs were lower at low turnovers (US\$0–500,000) and higher at high turnovers (above US\$500,000) than in New Zealand. For example, for turnover under US\$50,000, UK VAT compliance costs as a percentage of taxable turnover were 1.49 compared to 2.06 in New Zealand, for turnover of US\$50,000–100,000, 0.70 compared to 0.91, for turnover of US\$100,000–200,000, 0.50 compared to 0.67, and for turnover of US\$200,000–500,000, 0.44 compared

²⁸ From 2003 onwards, an annual private-sector survey of tax compliance costs for all sizes of enterprise has reported very high tax compliance costs (in 2007, annual total average internal and external compliance costs for all types of tax advice was \$11,592 and the average total cost per full-time equivalent staff was \$402). This survey does not break down tax compliance costs by size of enterprise or tax type. The marked changes in the respondents each year mean that year-on-year comparisons are difficult. Business New Zealand–KPMG Compliance Cost Survey, October 2007, 35–9. See <<http://www.businessnz.org.nz/surveys/504>> (as at 26 February 2008).

to 0.47. On the other hand, for turnover of US\$500,000–1 million UK VAT compliance costs as a percentage of taxable turnover were 0.34 compared to 0.28 in New Zealand and for turnover of US\$1 million–10 million, 0.07 compared to 0.04.²⁹

A common author of these studies suggests the following explanation of the higher New Zealand GST compliance costs for smaller businesses: 'Because compliance cost[s] as a percentage of turnover are negatively correlated with size, the overwhelming reason for the difference will be the much lower registration level for the New Zealand tax (about one-third of the UK threshold of 1986–87) and its wider coverage and the pressure put on small operators in New Zealand to register voluntarily.'³⁰ Another factor may be the different tax return periods most commonly used in the two countries (in most cases, two months for New Zealand and three months for the UK).

We are inclined to put most weight on the threshold, voluntary registration and the tax return period factors. Cnossen would add the registration of non-profits in New Zealand. After preparing the table comparing the 15-year-old UK VAT and the New Zealand GST compliance costs surveys, Cnossen still concludes that broadening the VAT base is one of three ways of reducing overall compliance costs.³¹ We agree with Cnossen and with the conclusion that Sandford earlier reached with his co-author Hasseldine that, 'the wider base of the New Zealand tax . . . could be expected to reduce compliance costs compared with the UK. Only a much more detailed comparison than is currently possible would reveal the areas of significant difference in compliance costs.'³² That work has still not been done.

The final question under this heading concerns public acceptance of the twin pillars policy. Acceptance of GST by the business community, in particular, was aided by extensive consultation and adaptation prior to enactment of legislation. This was the first time that the New Zealand public had been consulted on tax policy and legislation. Moreover, submitters could see the results of expressing their views as the draft legislation contained in the *White Paper* (1985) was completely rewritten before enactment. An extensive public relations programme aimed at educating taxpayers was also undertaken under the auspices of the GST Coordinating Office.

²⁹ Sandford (2000), 132, reporting the comparison and conversion into US dollars made by Cnossen (1994), 1665–7.

³⁰ Sandford (1994), 137. Cf. Sandford and Hasseldine (1992), 112.

³¹ Cnossen (1994).

³² Sandford and Hasseldine (1992), 112–13.

A comprehensive single-rate VAT has found such acceptance in the New Zealand social, political, and economic context that the concept has gone virtually unchallenged for two decades. There have been no serious attempts to challenge the model.³³ In 1987, the year after GST's introduction, the main opposition National Party proposed an 'Extax' that would have introduced exemptions and suspended the credit-offset mechanism. Following their defeat in the general election, the National Party dropped Extax. Later, after the change of government at the 1990 general election, National's Minister of Finance described the introduction of GST as 'a model of tax reform'.³⁴

To what extent does this public acceptance result from unique circumstances in New Zealand not replicable in the UK? One set of circumstances might be the broad national consensus in New Zealand in 1984 that reform was desperately needed. This swept the Fourth Labour government to power and returned the same administration three years later, despite its 'root and branch' reform policies. There is a sense that in a national emergency, such as New Zealand faced in 1984, the public will temporarily accept the 'unacceptable'. By the mid-1990s the emergency had ended, yet acceptance of, indeed popular support for, a comprehensive single-rate VAT system now appears to be as strong as ever. Some other economic reforms of the Fourth Labour government, including the move towards a more comprehensive income tax, may have been partly or wholly reversed in the last twenty years but the twin pillars of the GST model remain standing. This suggests that the public is able to be convinced of the merits of such an approach when the arguments are presented and provides support for the authors' illustrative proposal of a more comprehensive 17.5% VAT applied to all commodities except housing and items currently exempt from VAT for the UK.

3.3. Evidence on distributional impacts

The position of two groups in society is of particular concern when examining the distributional impacts of a proposal to extend VAT to food, clothing, and the other necessities of life:

- pensioners;
- low-paid workers and social security beneficiaries.³⁵

³³ The application of GST to local authority property rates and to tourist expenditure have been continuing issues that successive governments have declined to change.

³⁴ Quoted in Todd (2007).

³⁵ This section draws, in part, on Stephens (2007) especially 77–87.

Distributional impacts for pensioners arising from the introduction of GST were dealt with adequately with a one-off benefit adjustment that reflected the estimated consumer price effect of the reform.³⁶ It should be noted, moreover, that under the universal New Zealand Superannuation scheme (a pay-as-you-go scheme) the pension is explicitly linked to average after-tax wages. Pensioners therefore participate in the higher after-tax incomes of working age individuals and are protected, at least as far as their state-funded pensions are concerned, from the inter-generation equity implications of a direct-to-indirect tax shift. Obviously, no such protection is available in respect of savings and non-indexed private pensions.

Low-paid workers and social security³⁷ beneficiaries present different issues. At GST's introduction there were one-off adjustments to social security benefits as occurred for pensions. The low-paid were recipients of targeted relief through income tax credits. However, the income tax credits were confined to families with dependent children, meaning that low-paid individuals and couples not on a benefit received nothing other than small tax cuts. This decision reflected the results of official studies of the incidence when a 10% GST was substituted for the wholesale sales tax.³⁸ These studies, based on data gathered from household income and expenditure surveys, showed that the vertical incidence of the increase in the indirect tax burden³⁹ would be regressive by income level for each family type, but manageably so within the scope of the available compensatory measures. Three interesting results clearly stood out and influenced the shape of the compensation package:

- The incidence of indirect taxes is regressive, with a much higher average impact on households in low-income deciles⁴⁰ compared to higher-income deciles.⁴¹ For example, the incidence on the second decile was 25% compared to 15% for top income earners.

³⁶ The official estimate of the price impact of GST's introduction was a one-off increase in the general level of prices of 5.5%. Other estimates ranged between 5 and 7%. The actual increase in consumer prices measured in the December quarter 1986 was 8.9% up from 3.3% in the September quarter. The March 1987 quarter recorded a 2.4% increase in consumer prices.

³⁷ Income tested benefits paid to individuals reflecting unemployment, sickness, widowhood, or being a solo parent.

³⁸ New Zealand Planning Council (1990). See also the following unofficial studies: Scott, C. Goss, P. and Davis, H. (1985); Broad and Bacica (1985).

³⁹ The average household burden from indirect taxes rose from 11.4% to 18.4% between 1985–86 (the year before GST's introduction) and 1987–88.

⁴⁰ A decile divides sorted data into ten equal parts, so that each part represents $\frac{1}{10}$ th of the population. The 1st decile cuts off the lowest 10% of data at the 10th percentile and the 9th decile cuts off the lowest 90% of data at the 90th percentile.

⁴¹ As Stephens notes, the incidence for the first decile needs to be treated cautiously as self-employed people comprise a large part of that decile and they can declare income tax losses while spending freely. Stephens (2007), in Keever and White (eds.) (2007), at 79.

- The presence of dependent children within a family type magnified the effect of taxing food and other necessities. This result strongly associated with the number of dependent children in the family unit. In the third decile, for example, it was estimated that the impact of GST would reduce household disposable income for families of 2 adults and 1 child by 6.1% and reduce household disposable income for families of 2 adults and 3 children by 9.5%. The net impact for the top decile for these family types would be just 4.5%.
- The impact on pensioners was less than for a typical household due to their relatively low food consumption.

The identified above-average impacts were addressed through compensatory income supplementation. The compensatory approach is an available and workable alternative to preferring food and other necessities in the VAT. However, it relies on a social bargain between the government and low-paid workers and beneficiaries that the supplementation will keep up with costs. There were periods in the 1990s and the early 2000s when social assistance lagged behind costs.

Stephens⁴² makes an observation that may be important in assessing the ability to translate the New Zealand GST experience to the UK. He argues that a prime reason that mid-1980s New Zealand tax and economic reforms have endured was the attention given to the issue of horizontal equity. Horizontal equity—like treatment of people in similar circumstances—seems to underpin the notion of fair play in the New Zealand context to a greater degree than redistribution.

4. THE HARD-TO-TAX: LESSONS FROM NEW ZEALAND

In this section, we identify aspects of the New Zealand experience that may be relevant in addressing the problem of hard-to-tax services and organizations. We make comments in Table 2 on the following subjects and issues:

- financial services (intermediation and for-fee);
- residential dwellings;
- threshold;
- government services (including health and education);
- overseas travel.

⁴² Ibid. 78.

Table 2. The hard-to-tax: Lessons from New Zealand

Subject	Issue	Lessons from New Zealand
Intermediation financial services	Financial services that are remunerated by a margin or spread are not able to be integrated with the general invoice system of VAT. There is no feasible, conceptually correct, solution. Is it not time to put the search for a 'solution' to intermediation financial services into the category of a quest that has no practical value? The choices are to zero-rate, exempt (without credit), or a hybrid of both.	In an attempt to reduce the cascading effects of exemption for businesses, New Zealand has applied a hybrid system since 2005. To date, the fiscal impact of the new zero-rating for business-to-business financial services and reverse charge rules has been as officials forecast but it is too early to judge whether the new rules have opened up tax planning opportunities. ⁴³
For-fees financial services	New Zealand has successfully taxed for-fee financial services, such as fire and general insurance, giving credit for a notional tax content of claims.	This is a potentially useful expansion of the UK VAT tax base. ⁴⁴
Real property	Taxing supplies of real property causes compliance problems analogous to the missing trader fraud discussed by the authors in Section 4.3.3 at 311–16.	The problem essentially arises from the delay between claiming VAT during the construction phase and the completion tax point. The possible solutions replicate the authors' suggestions in relation to missing traders in Section 4.3.3 at 311–16. ⁴⁵ Fiscal risks can also arise from deferred settlement schemes that take advantage of differing accounting bases (accruals, cash, or hybrid).
Residential dwellings	New Zealand taxes residential construction activities and the first sale of new dwellings while the UK does not. The issue noted by the authors in Section 4.3.2 at 304 appears to be increasing house prices to new buyers (housing affordability) and the windfall gain to existing owners.	The authors correctly make the comment that taxing new construction would give a windfall to existing homeowners. Direct compensation is one approach. However, other factors affecting housing

(cont.)

⁴³ Pallot (2007) especially at 169–70.

⁴⁴ For a discussion of the merits of the New Zealand and Australian regimes for taxing fire and general insurance, see Edgar (2007) at 156–61.

⁴⁵ See also the discussion and solutions suggested in Harley (2007) especially at 234–41. See also Snell (2007) especially at 424 and 429.

Table 2. (cont.)

Subject	Issue	Lessons from New Zealand
	Landlords and owner-occupiers face the same GST treatment. All purchases, insurance, local authority rates, maintenance charges as well as construction costs—the inputs to the provision of shelter—are taxable, but the value added in the provision of shelter is not subject to the tax.	affordability should also be considered before dismissing an application of tax to repairs, improvements, and new dwelling construction. New Zealand made no special arrangements in this regard. Australia did.
Threshold	While it is an elegant exposition, we are not convinced that the Keen and Mintz optimal threshold formula ⁴⁶ is a sufficient basis for the conclusion that a high threshold is justified for the UK. It seems to rely on key assumptions about the variability of administration and compliance costs with firm size, and does not capture the potential distortions of trade from unregistered input-taxed suppliers competing with registered firms, especially where services are being provided with high labour and low goods content.	The New Zealand solution is a low threshold ⁴⁷ combined with a choice of three return periods ⁴⁸ and simplified accounting requirements for micro and small businesses. Complaints about competition from unregistered suppliers are rare. However, the evidence would suggest that a disproportionately higher compliance cost burden on the smallest firms is a negative consequence of the New Zealand approach.
Government services (including health and education services)	The conventional European treatment of government-provided services excludes them from the VAT base. This means that services such as education, medical care, and services to residents by local authorities are outside the VAT base. This is an erosion of the potential VAT base and may exacerbate problems of competition between public and private suppliers. ⁴⁹	New Zealand taxes all government-provided services, including the notional outputs of policy and administrative agencies. Government-provided services, such as healthcare and education, are also taxed, as are such services in the private sector. The cost of this approach is that it may be more

⁴⁶ Keen and Mintz (2004).

⁴⁷ The New Zealand Government is currently seeking submissions on increasing the current GST threshold from NZ\$40,000 to NZ\$50,000. It observes that the New Zealand registration threshold is low by international standards but that about 40% of the taxpayers currently registered for GST are voluntary registrants with an annual turnover of less than NZ\$40,000. The estimated cost of allowing about 24,000 taxpayers to exit the GST base is about NZ\$15 million a year. Cullen and Dunne (2007).

⁴⁸ The standard return period is two months, which is shorter than in many other jurisdictions. Micro businesses may apply to use a six-monthly return period. There is also a one-month return option.

⁴⁹ The case for the VAT in the EU fully taxing all public sector bodies (and non-profit organizations) is well argued in Aujean, Jenkins, and Poddar (1999).

Table 2. (cont.)

Subject	Issue	Lessons from New Zealand
Overseas travel	<p>The New Zealand model goes further and treats the policy and administration services supplied by government departments as taxable activities regardless of whether their funding is appropriated by Parliament or sourced from fees and charges.</p> <p>Tourist expenditure is by convention taxed in the country of destination, but expenditure on travel between countries generally falls outside any countries' tax net. Overseas travel is a significant (and growing) item of household consumption and worthy of attention by tax policymakers.</p>	<p>administratively complicated than exempting agencies, but not more complicated than zero-rating. It also inflates the size of GST receipts and expenditures by corresponding amounts (subject to timing mismatch). The advantage is that it makes agencies indifferent between charging and appropriations, and so preserves a competitive position between the state and private sectors.</p> <p>Very few countries impose VAT on overseas travel. New Zealand does not. Indirect taxation of overseas travel might require an arrangement analogous to postal revenue sharing between nations on either a regional or multilateral basis. Indirect tax reform is worthwhile contemplating.⁵⁰</p>

5. CONCLUDING COMMENTS

The authors have identified a compelling case for reforming the UK VAT supported by empirical evidence on the small potential gains from rate differentiation. Moreover, they have identified an illustrative reform package of a uniform standard rate of 17.5% applied to all non-housing consumption goods and items currently exempt from VAT, which, while not completely

⁵⁰ For a broad consideration of the indirect tax issues and an optimal policy proposal for fuel and ticket taxes, see Keen and Strand (2007).

removing the distortions of the current UK VAT scheme, can be introduced without waiting for the EU ground rules for VAT to be renegotiated.

This commentary has sought to provide insights from the New Zealand GST model for UK indirect tax policy and the authors' proposals on the following three issues:

- Are the alleged administration and compliance benefits of the comprehensive single-rate VAT borne out by the New Zealand GST experience?
- What have been the distributional impacts of a comprehensive single-rate GST on the poor and lowest paid in New Zealand?
- What lessons arise from how New Zealand GST applies to services and organizations that in Europe are often considered too hard to tax?

The broad conclusion from two decades of the New Zealand GST is that the alleged benefits of maintaining a comprehensive base with a single domestic rate are, on balance, borne out by experience. While VAT/GST compliance costs are undoubtedly regressive in relation to turnover, we consider that the New Zealand surveys may overstate marginal GST compliance costs, in part because the regular reporting of VAT/GST on a one-, two-, or six-monthly basis means that GST is the first tax return for which a business must complete accounts each year. The wider base of the GST is likely to have reduced compliance costs compared with the UK VAT but this has not been empirically tested.

We have argued that the public is able to be convinced of the merits of a comprehensive single-rate VAT and that this provides contextual support for the authors' illustrative proposal of a 17.5% VAT applied to all non-housing consumption goods and items currently exempt from VAT for the UK.

To the extent that taxing the basic necessities of life will affect the economic position of the poor, relief can be provided using more effective support tools than GST exemptions. However, it relies on a social bargain between the government and low-paid workers and beneficiaries that the income supplementation will keep up with costs. In New Zealand, the political consensus has been to apply compensating assistance via the income tax and social welfare systems, with an emphasis on supporting working families with children. Critics may point out that, as a result, New Zealand has an army of bureaucrats administering its social welfare and income tax credit systems instead of an army of revenue officers administering the boundaries between rates and exemptions. If the former army is taken as a given in a modern developed country, the absence of the latter is clearly a cost saving to be added to the reduced deadweight costs of the New Zealand style of GST.

REFERENCES

- Aujean, M., Jenkins, P., and Poddar, S. (1999), 'A New Approach to Public Sector Bodies', *International VAT Monitor*, 10, 144–9.
- Bollard, A. E. (1992), 'New Zealand's Experience with Consumption Tax', *Australian Tax Forum*, 473–93.
- Broad, A., and Bacica, I. (1985), *The Incidence of Indirect Taxes, Vol. 2*, Wellington: Institute of Policy Studies.
- Colmar Brunton (2005), *Measuring the Tax Compliance Costs of Small and Medium-Sized Businesses—A Benchmark Survey*, 95–103, Wellington: Colmar Brunton.
- Crossen, S. (1994), 'Administrative and Compliance Costs of the VAT: A Review of the Evidence', *Tax Notes International*, 8, 1649–68.
- (2003), 'Is the VAT's Sixth Directive Becoming an Anachronism?', *European Taxation*, 43, 434–42.
- Cullen, M., and Dunne, P. (2007), *Reducing Tax Compliance Costs for Small and Medium-Sized Enterprises*, 13–14, Wellington: Policy Advice Division of Inland Revenue.
- Dickson, I. (2007), 'The New Zealand GST Policy Choice: An Historical and Policy Perspective', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 45–63, Wellington: Brookers Ltd.
- Douglas, R. (2007), 'The New Zealand GST Policy Choice and its Political Implications', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 3–11, Wellington: Brookers Ltd.
- Edgar, T. (2007), 'The Search for Alternatives to the Exempt Treatment of Financial Service under a Value Added Tax', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 131–61, Wellington: Brookers Ltd.
- Evans, C., Pope, J., and Hasseldine, J. (eds.) (2001), *Taxation Compliance Costs: A Festschrift for Cedric Sandford*, Sydney: Prospect Media.
- Green, R. (2007), 'Consulting the Public in Developing a GST', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 13–25. (2007), Wellington: Brookers Ltd.
- Harley, G. (2007), 'Dilemmas for GST Tax Policy Designers—Land Transactions', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 213–41, Wellington: Brookers Ltd.
- Keen, M., and Mintz, J. (2004), 'The Optimal Threshold for a Value-Added Tax', *Journal of Public Economics*, 88, 559–76.
- and Strand, J. (2007), 'Indirect Taxes on International Aviation', *Fiscal Studies*, 28, 1–41.
- Krever, R., and White, D. I. (eds.) (2007), *GST in Retrospect and Prospect*, Wellington: Brookers Ltd.
- McCaw, P.M. (1982), *Report of the Task Force on Tax Reform*, Wellington: Government Printer.

- Meade, J. (1978), *The Structure and Reform of Direct Taxation: Report of a Committee chaired by Professor J. E. Meade for the Institute for Fiscal Studies*, London: George Allen & Unwin. <http://www.ifs.org.uk/publications/3433>.
- New Zealand Planning Council (1990), *Who Gets What? The Distribution of Income and Wealth in New Zealand*, Wellington: New Zealand Planning Council.
- OECD (2006), *Consumption Tax Trends*, Paris: OECD.
- Oxley, P., and Elwela, D. (2006), 'Tax Compliance Costs of New Zealand Small Businesses, 2004: Designing the Survey for its Policy Purpose', paper presented at the Seventh International Conference on Tax Administration held in Sydney on 20–21 April 2006.
- Turner, S., and Sullivan, C. (2005), 'Tax Compliance Costs of Small Businesses 2004: Improving the Estimates', paper presented at the 2005 IRS Research Conference held in Washington, DC on 7 June 2005, <<http://www.irs.gov/pub/irs-soi/05oxley.pdf>>.
- Pallot, M. (2007), 'GST and Financial Services—Rating Zero Rating', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 163–78, Wellington: Brookers Ltd.
- Ritchie, K. (2001), 'The Tax Compliance Costs of Small Businesses in New Zealand', in Evans, C., Pope, J., and Hasseldine, J. (eds.), *Taxation Compliance Costs: A Festschrift for Cedric Sandford*, 297–31, Sydney: Prospect Media.
- Sandford, C. (2000), *Why Tax Systems Differ: A Comparative Study of the Political Economy of Taxation*, 137, Bath: Fiscal Publications.
- and Hasseldine, J. (1992), *The Compliance Costs of Business Taxes in New Zealand*, 57–79, Wellington: Institute of Policy Studies, Victoria University of Wellington.
- Scott, C., Goss, P., and Davis, H. (1985), *The Incidence of Indirect Taxes, Vol. 1*, Wellington: Institute of Policy Studies.
- Snell, D. (2007), 'GST—Revenue and Business Risk', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 423–30, Wellington: Brookers Ltd.
- Stephens, R. (2007), 'The Economic and Equity Effects of GST in New Zealand', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 65–87, Wellington: Brookers Ltd.
- Todd, J. (2007) 'Implementing GST—Information, Education, Co-ordination', in Krever, R., and White, D. I. (eds.), *GST in Retrospect and Prospect*, 27–43, Wellington: Brookers Ltd.
- White, D. I. (2007a), 'Twenty Years of GST: The Best Path Forward', *New Zealand Journal of Taxation Law and Policy*, 13, 357–80.
- (2007b), *International VAT Monitor*, 18, 343–51.
- White Paper on Goods and Services Tax: Proposals for the Administration of the Goods and Services Tax* (March 1985), Wellington: New Zealand Government.

Commentary by Jonathan Gruber

Jonathan Gruber is Professor of Economics at MIT and Director of the Program on Children at the National Bureau of Economic Research. He is a co-editor of the *Journal of Public Economics* and an Associate Editor of the *Journal of Health Economics*. He has written broadly on public policy, with a focus on health insurance coverage and on retirement, consumption decisions, and labour supply. Dr Gruber was elected to the Institute of Medicine in 2005, and in 2006 he received the American Society of Health Economists' Inaugural Medal for the nation's best health economist aged 40 or under. In 2006 he was appointed to the board of the Massachusetts Insurance Connector, the main implementing body for the state's ambitious health care reform effort.

Public policy makers have long been fascinated with the notion of taxing sinful consumption. The notion that the government could encourage proper behaviour, and, in the process, raise revenues to spend on other beneficial causes, is a powerful one. And developed nations have a long history of specific excise taxes on sin goods like tobacco and alcohol. In the United States, in 1960, 12.8% of government revenues were raised by specific excise taxes. By 2005, however, this had fallen to 3.2% (Gruber (2007)).

While sin taxes may make good politics, however, economists have often claimed that they make bad policy. The traditional economics argument towards sin taxes is straightforward. If sinful consumption levels are chosen by rational and fully informed agents, and those activities cause no harm to others, then the government has no more right to tax these activities than any other activities that are non-externality producing. By revealed preference arguments individuals are pursuing the optimal level of such activities and there is no need for 'correction' through taxation. The only particular role for excise taxation is to correct the externalities that sinful consumption may cause, in addition to their usual role in the optimal system of commodity taxation.

Section 4.4 of the chapter by Crawford, Keen, and Smith provides a very careful discussion of the role of excise taxation on smoking and drinking.

The authors make four important points. First, they integrate these specific commodity taxes into the larger optimal commodity taxation argument developed earlier in the chapter. They conclude, I think rightly, that the evidence is sufficiently unclear to motivate specific excise taxes on these particular goods (relative to the standard VAT rate) solely from elasticity-type arguments.

Second, they raise the well-known equity concerns around such taxes. Since lower income groups spend a much larger share of their income on 'sinful' consumption such as smoking or drinking, then these taxes are typically viewed as more regressive than the general VAT, which would offset any efficiency argument for higher taxes. Third, they review estimates of externalities due to alcohol and tobacco consumption and their implications for optimal taxation, highlighting the difficulties of translating externalities to tax policy when damage is highly non-linear (as is the case with alcohol). Finally, they discuss the same issues in the appropriate structure of excise taxation, particularly the equality of taxation across types of alcohol.

In this commentary, I will take issue with the baseline model that underlies the analysis of the chapter. On the basis of existing evidence and recent developments in behavioural economics, the standard economics model of 'sinful' consumption appears woefully inadequate. When sensible alternatives are applied, traditional conclusions such as those drawn here can be turned on their heads. My discussion will focus on the cases of smoking and alcohol, as discussed in this chapter; I will also include a discussion of perhaps the most important externality of all in the US context, obesity. I will leave aside the issue of proper taxation of motor fuel, which is discussed extensively in recent work by Don Fullerton and co-authors (e.g. Fullerton and Gan (2005)).

1. THE STANDARD MODEL OF SINFUL CONSUMPTION AND IMPLICATIONS FOR GOVERNMENT POLICY

In the standard economics model, there is nothing special about consumption of sinful goods as opposed to standard goods. Individuals optimize a well-defined utility function over a bundle of goods, some of which are sinful and some of which are not. If there are no external effects on others from the sinful consumption, then there is no call for government intervention.

For years, an informal claim was that the addictive nature of some sinful consumption (e.g. cigarette smoking) would justify government intervention even if there were no externalities. This claim was rejected in a pathbreaking article by Becker and Murphy (1988). In the Becker and Murphy model, individuals recognize the addictive nature of choices that they make, but may still make them because the gains from the activity exceed any costs through future addiction. In this 'rational addiction' framework, individuals recognize the full price of addictive consumption goods: both the current monetary price, and the cost in terms of future addiction.

Becker and Murphy model the act of addictive consumption (such as smoking) as the building of an addiction stock. The more cigarettes smoked today, the greater the addiction capital tomorrow. High addiction capital lowers *average* utility but raises the *marginal* utility of smoking. In this way, smoking lowers future utility but also increases the craving for another cigarette. The key feature of any addiction model is on how people deal with this intertemporal problem. In the original Becker–Murphy formulation individuals discounted the future *exponentially*, meaning that they discount k -periods forward by δ^k , where δ is the per-period time discount factor. This assumption has turned out to be central, as will be discussed further below.

Rational addiction subsequently became the standard approach to modeling consumption of goods such as cigarettes. The key normative implication of this approach is that the optimal regulatory role for government related to smoking is solely a function of the interpersonal externalities induced by smoking. Since smoking, like all other consumption decisions, is governed by rational choice, the fact that smokers impose enormous costs on themselves is irrelevant; it is only the costs they impose on others that gives rise to a mandate for government action.

The conclusions of this traditional economics literature create a central policy importance for the measurement of externalities associated with sinful activities. Over the past twenty years, a sizeable literature has been devoted to the measurement of these externalities. This chapter provides a nice review of externality estimates. While there are a number of interesting details, the bottom line is that the estimated externalities from smoking are quite small, while those from drinking are large. This is because most of the damage done by smoking is *internal* (through shorter lives), not external. Indeed, the limited external damage done by smoking (through higher system wide health costs) is largely offset by the savings through reduced pension payments due to earlier death of smokers (the so-called 'death benefit'). On the other hand, most of the damage from drinking is *external*, in the form

primarily of drunk driving and induced criminal activity (the latter being a platform where it is much harder to measure the causal impacts of alcohol *per se*).

It is important to recognize some controversy in these estimates, however. The most important source of controversy is the relevance of external costs imposed on family members, such as through secondhand smoke or alcohol abuse. If smokers are maximizing a family utility function, rather than simply considering individual utility, then these costs should not be considered externalities. That is, if maximizing family utility, the smoker or drinker will take into account any damage done to others in the family in trading off the costs and benefits of smoking. Available evidence, however, suggests that family utility maximization is far from complete. For example, Lundberg, Pollak, and Wales (1997) show that, in contrast to the family utility maximization model (where everyone cares equally about all the family members), shifting the control of household financial resources from husbands to wives significantly increases the expenditures made on behalf of children.

2. PROBLEMS WITH THE STANDARD MODEL

A major problem with the perspective just outlined is that it is not consistent with observed excise tax policy in developed nations. Taxes on tobacco are typically many multiples of the externality estimates, particularly in European countries, while taxes on alcohol are much lower than the externality estimates. One explanation for this inconsistency is a lack of faith in the standard model among policy makers—and, implicitly, among the public as well.

2.1. The irrationality of youth

This lack of faith may reflect two realities about sinful behaviours that are not reflected in the standard model. This first is that engagement in such behaviours often begins at a young age. Of all US adults who smoke, more than 75% begin smoking before their nineteenth birthday (Gruber and Zinman (2001)). One-quarter of the estimated number of illegal drug addicts in the US are under the age of 17 (Pacula et al. (2001)). Despite the illegality of sales to, and consumption of, alcohol by minors in the US, 88% of 15-year-olds report having had a drink, one-quarter report drinking at least weekly, and one-third report having been drunk twice or more in their life. On this

last count, in fact, the US appears to have much less child usage than other nations; the US ranks tenth in developed nations, for example, in the share of 15-year-olds who have been drunk twice or more in their life (Cook and Moore (2001)).

While models such as Becker and Murphy's presume a 'homo economicus' who is making rational and forward-looking decisions, most would admit that such a model does not fit the teenagers who are making potentially addictive decisions.

Indeed, there is some evidence that this monumental decision may not be made in the forward-looking fashion required by rational addiction models. A survey asked high school seniors who smoked a pack a day or more whether they would be smoking in five years and then followed the seniors up five years later. Among those who had said they would be smoking in five years, the smoking rate was 72%—but among those who said they would *not* be smoking in five years, the smoking rate was 74%! This result suggests that teens who smoke may not account for the long-run implications of addiction.

So how do young persons make decisions about whether to smoke, drink, overeat, or use drugs? In fact, we know remarkably little about this question. The state of knowledge is reviewed in the papers in Gruber (2001), and is summarized in the introduction to that volume. There are three key lessons from the analyses that underlie that volume. First, the actions taken by children along a variety of such 'sinful' behaviours have important implications for their behaviour as adults. For example, Gruber and Zinman (2001) document a strong intertemporal correlation between youth smoking and adult smoking; the causal link is established by exploiting variation in the taxation of cigarettes faced by individuals as youths (those who face higher taxes as youths smoke less both as youths *and* as adults). And Cook and Moore (2001) show that individuals who grew up in states and years with higher drinking ages are less likely to engage in binge drinking later in life.

Second, we know that these youth behaviours do respond to economic incentives. Most of the papers in the Gruber (2001) volume show that youth risky behaviours do respond rationally to incentives: youths smoke less, drink less, and use fewer illicit drugs when prices are higher; youths use fewer illicit drugs and drive more safely when the criminal penalties for not doing so are higher; youths stay in school and avoid unprotected sex when the labour market returns to completing education are higher; and so on. Finally, unfortunately, we know that economic incentives and other standard socio-economic controls can only explain a small part of the behaviours of youth over time. There are enormous time trends in smoking, drinking, drug use,

crime, and youth fertility that are not readily explained by our economics models.

As a result, it is extremely unlikely that the complicated mechanics underlying the Becker and Murphy (1988) model can do much to explain the decision of youth to pursue addictive behaviours. This may be a well-earned source of scepticism among policy makers as they seek to apply these conclusions to policy making.

2.2. The inconsistency of adults

The other type of evidence, or at least introspection, that may deter policy makers from applying the standard model is the inconsistency in adult decision making over sinful activities. There is an enormous body of evidence that adults are *time inconsistent* in their behaviour towards such activities: the problem isn't their rational ability to make plans, the problem is that they lack the self-control to carry out those plans.

The term 'rational addiction' obscures the fact that the Becker and Murphy model imposes two assumptions on consumer behaviour. The first is that of forward-looking decision-making, which is a centrepiece of most welfare analysis in economics. Becker and Murphy also assume, however, that individuals *can not only optimize their utility function, but that they can then carry out those optimal plans*. There is much evidence from psychology, however, that contradicts this assumption: individuals are often unable to carry out long-term plans that involve self-control when there are short-term costs to doing so. An excellent example of this is smoking, where there is a short-term cost of quitting (in terms of physical discomfort and perhaps mental distress), but a long-term health benefit. Perhaps as a result, eight in ten US smokers express the desire to quit smoking, but many fewer than that actually do quit. Other examples include retirement savings (short-term cost in terms of foregone consumption today, but long-term benefits in terms of a higher standard of living in retirement), or whether to diet and/or exercise (short-term costs in terms of less food or more work today, but long-term benefits in terms of a longer life). In many arenas, individuals appear unable to control their short-term desires for their own longer-term well-being.

There are three types of evidence for the existence of self-control problems. The first is from laboratory experiments in psychology. In laboratory settings, individuals consistently reveal that they are willing to be patient in the future, but are impatient today, the defining characteristics of self-control problems. A person with self-control problems has the right long-run intentions (he

rationally optimizes his utility function given his budget constraint), but he just can't carry them out. For example, in one experiment, most people preferred a cheque for \$100 they could cash today over a cheque for \$200 they could cash two years from now. Yet the same people prefer a \$200 cheque eight years from now to a \$100 cheque six years from now, even though this is the *same choice*—it's just six years in the future (Ainslie and Haslam (1992)). This is indicative of self-control problems: individuals are willing to be patient in the future, but not today when faced with the same choice.

The second type of evidence for self-control problems is the demand for *commitment devices*. If individuals have self-control problems and are aware of those problems, they will demand some type of device that helps them fight these problems. And the search for such commitment devices is the hallmark of most recommended strategies for quitting smoking: people regularly set up systems to refrain from smoking by betting with others, telling others about the decision, and otherwise making it embarrassing to smoke. These practices help individuals combat their self-control problems by raising the short-run costs of smoking to offset the short-run benefits of smoking. The use of self-control devices is widespread in other arenas as well: individuals set up 'Christmas Clubs' at their banks to make sure they have enough money to buy Christmas presents, and they buy memberships at sports clubs to commit themselves to work out when it would generally be cheaper to just pay each time they go.

The final type of evidence is patterns of data that are inconsistent with the standard model. For example, Angeletos et al. (2001) calibrate a model of the savings decision which shows that asset holding patterns in the US, with many individuals holding savings in illiquid forms but no savings (and often debt at high interest rates) in liquid forms, is inconsistent with the standard model but arises naturally in a model with self-control problems (since illiquid savings provides a commitment device).

Gruber and Mullainathan (2005) provide a test in the context of cigarette smoking. They note that an important distinction between the Becker–Murphy model and models with time inconsistency is the welfare impact of taxation. In the Becker–Murphy model, since an exponential individual makes a time-consistent choice to smoke, a rise in taxes can only *lower* discounted utility today. If it were to raise it, then the rational addict could raise utility by simply reducing smoking by the amount that the tax does, that is, by emulating the tax. So cigarette taxes should reduce the welfare of time-consistent rational addicts. Yet Gruber and Mullainathan find that higher cigarette taxes lead to *higher* levels of reported well-being among smokers. This is inconsistent with the standard model but, as explained below, could

be due to the welfare gains of government provision of a commitment device in the self-control model.

It is equally important to clarify that there is a host of evidence on addictive behaviours that *does not* distinguish time inconsistency. The fact that consumers, even youths, are price sensitive in their consumption decisions of goods such as cigarettes, alcohol, and illegal drugs does not mean that they are time consistent, simply that they are rational (which is true in most alternative models discussed below). Particularly misleading is the sizeable literature growing out of Becker, Grossman, and Murphy (1994) that shows that addictive consumption responds to future prices as well as current prices. This finding has been taken to support the Becker–Murphy model *in toto*, but this is not the case. As highlighted by Gruber and Koszegi (2001), this evidence simply serves to prove that individuals are rational and forward-looking, not that they are time consistent.

3. POLICY IMPLICATIONS OF AN ALTERNATIVE MODEL: THE CASE OF SMOKING

There is by now broad agreement among many economists on the failures of the Becker–Murphy model appropriately to capture sinful behaviours such as smoking. Over the past seven years, several alternatives have been developed. While these alternatives differ, sometimes dramatically, in detail, they all have at their core the same concept: problems of self-control in the face of temptation. In this section I discuss in detail one such model, the quasi-hyperbolic discounting model of Gruber and Koszegi (2001, 2004). Since much of the literature is focused on the case of smoking, I will consider that case here as well; below I will expand the discussion to encompass other ‘sinful’ activities.

3.1. The quasi-hyperbolic discounting model and its implications

A major alternative approach to Becker and Murphy’s (1988) model uses the *quasi-hyperbolic discounting* model developed by Laibson (1997) and O’Donoghue and Rabin (1999). In contrast to the exponential discounting used by Becker and Murphy, in this quasi-hyperbolic formulation, the next period is discounted by $\beta\delta$, the following period by $\beta\delta^2$, and k periods in the future by $\beta\delta^k$, where $\beta < 1$ is an extra discount factor that changes the discounting of this period relative to the entire future. The key feature of

such a hyperbolic model is that individuals will have self-control problems. Specifically, a sophisticated hyperbolic individual (one who knows that he discounts hyperbolically) would like to smoke less in the future than he actually can. The problem arises because he is patient about the future (the relative discount rate between future periods is δ), but impatient about the present (the relative discount rate between today and tomorrow is $\beta\delta < \delta$). This means that when the future arrives he will end up making more impatient choices (i.e. smoke more) than he would like to from today's vantage point.

An alternative formulation of time inconsistency is the *I* case, where individuals do not recognize their own self-control problems (O'Donoghue and Rabin (1999)). One feature that distinguishes time-consistent agents from time-inconsistent agents is an inability to realize desired future levels of smoking. As noted above, unrealized intentions to quit at some future date are a common feature of stated smoker preferences.

Gruber and Koszegi (2001, 2004) augment the Becker–Murphy model by applying such quasi-hyperbolic preferences. This simple change has a radical implication: the discounted utility of a sophisticated hyperbolic consumer can rise if a tax is imposed. The reason is that the tax serves as a self-commitment device, which the private sector cannot perfectly supply.¹ As Gruber and Koszegi (2001) highlight, 'The argument that people act in their best interests so . . . the government should leave them alone, is immediately invalidated in our setting. Therefore . . . a benevolent social planner would want to intervene in this economy.'

That self-control problems could overturn the externalities-only motivation for taxation in the standard model is not very surprising. More striking is the implications of this alternative model for levels of optimal taxation. Gruber and Koszegi (2002) undertake a detailed calibration exercise based on their quasi-hyperbolic formulation to account not only for the externalities but also the *internalities* of smoking—that is, to value the cost to the individual themselves. They do so just for the mortality effects of smoking, ignoring other personal costs such as non-fatal illness and personal discomfort.

For this calibration exercise, they use consensus estimates of the value of a life from Viscusi (1993), and account for the fact that the years lost from smoking are at the end of life and so should be discounted to the present. Their results suggest that the internality from smoking a pack of cigarettes

¹ Gruber and Koszegi rule out perfect private sector commitment devices by assumption. As they discuss, any voluntary commitment device can readily be undercut by a voluntary 'de-commitment' device, and non-binding contracts for long-term commitment cannot be readily enforced. At the same time, they recognize that there are limits on government taxation, such as smuggling. So while they refer to the calibrations discussed below as optimal taxation, they recognize that they really refer to the optimal combination of private commitment and enforceable government policies.

is over \$35 per pack. This is an enormous figure which is of the order of 100 times as large as the externalities associated with smoking. This large figure is driven by the large damage that smoking does to health and the high implied value of life from Viscusi's work.

Gruber and Koszegi (2002) present a range of estimates for the implied optimal tax based on this internality (which would be in addition to any externality-based effects discussed earlier). The optimal tax level depends on a number of parameters in the model, most importantly the shape of the discount rate (determined by the quasi-hyperbolic parameter β) and the level of the discount rate δ . As individuals are more hyperbolic in their decision making, the optimal tax rises; but as they are less patient, the optimal tax falls, since they care less about the damage done at the end of life. The optimal tax is estimated to be about \$1 for individuals who are impatient (10% discount rate) and close to time consistent ($\beta = 0.9$), rising to over \$9 for individuals who are patient (3% discount rate) and far from time consistent ($\beta = 0.6$); the lower level of β is much more consistent with available laboratory and calibration evidence. Thus, this alternative model rationalizes the higher levels of cigarette taxation that we see in developed countries.

This model also has a number of other important implications. First, the optimal level of taxes rises with the stock of addiction; individuals with more past exposure to the addictive behaviour should face higher taxes, since the internalities to them are largest from continued participation in the activity. This implies, for example, that the optimal pattern of taxation of such goods is to have a high tax initially that declines as habits are broken. This is the opposite of the traditional prescription of rising taxes over time, which are usually supported on distributional grounds; but I question such distributional motivations below.

Second, a key underlying feature of the Becker–Murphy model is intertemporal complementarities in smoking decisions. Thus, raising the tax in one period can lower consumption in all periods. This implies that if there are barriers to raising taxes at some stage in life (e.g. on adults), it would motivate higher taxes at a different stage (e.g. on children). Moreover, if one extends this point to space rather than time, it suggests a novel justification for smoking bans in restaurants or at work. Since we cannot effectively regulate smoking in the home, if smoking is complementary across locations, we may want to regulate it more severely in observable locations. That is, since we are undertaxing in one location, we want to overtax in another, so that clean air regulations are not only protecting non-smokers, but potentially the smoker themselves as well.

Third, these conclusions also hold for naive time-inconsistent consumers, but there is an important distinction. In this case the tax corrects not just a self-control problem but a misperception problem as well. Unlike the sophisticated case, therefore, it is possible in the naive case that the appropriate policy prescription is a complete ban rather than just corrective taxation. Thus, if we think that young smokers are ‘irrational’ naive consumers, then a ban on youth smoking with a tax on adults could be the optimal policy prescription.

Finally, contrary to popular perception, taxes on goods such as cigarettes are *not* regressive. Since sinful goods such as cigarettes and alcohol consume a larger share of budgets for lower income households, such taxes are typically viewed as regressive. But the goal of tax incidence analysis is not simply to measure who consumes more of a good, but rather who is ‘hurt’ by taxation of the good. The two are typically viewed as equivalent for incremental taxes by the envelope theorem. But the envelope theorem does not hold for quasi-hyperbolic consumers. In that case, measuring the burden of taxation implies measuring both the burden due to current consumption and the self-control benefits from taxation.

Gruber and Koszegi (2004) develop a ‘self-control adjustment’ to account for this point. This adjustment is rising in the price sensitivity of smoking decisions; the more price sensitive are consumers, the larger is the self-control value of taxation. It is also rising in the degree of time inconsistency, but falling in the discount rate (for reasons discussed above) and falling in the value of life. They undertake a number of calibrations which account most importantly for the fact that lower income groups are much more price sensitive in their smoking decisions than are higher income groups. As a result, for a broad range of parameters, they find that cigarette taxes are *not* very regressive, and may actually be progressive. This contradicts the presumption made towards regressivity in this chapter by Crawford, Keen, and Smith.

4. POLICY IMPLICATIONS FOR TAXING ALCOHOL AND OBESITY

Determining the optimal taxation of cigarettes is hard in one sense, which is the weight to put on externalities. But in other senses it is quite simple, because cigarettes have the nice feature that damage is monotonic and essentially linear: every cigarette is bad for you, and each additional cigarette is

worse by a roughly equal amount (at least as far as science can tell today). This is not true for other important sources of 'sinful' consumption.

4.1. Alcohol

These problems are easily illustrated for alcohol. Much, if not most, alcohol consumption is likely done by those who will exert little external impact on others. Indeed, when Manning et al. (1991) measure the externalities due to alcohol, they focus on 'heavy drinkers' who report having more than two drinks per day (and likely have many more given under-reporting of consumption); consumption by such drinkers is less than half of total alcohol consumption. It is infeasible, however, to tax only heavy drinkers and not others. This non-linearity in external damage may, in fact, explain the low taxation of alcohol relative to its external damage: the typical drinker is not responsible for such damage and so would view as highly unfair taxation to compensate for it.

As noted in this chapter, this issue was initially addressed by Diamond (1973), who discusses externality taxation in a context where all individuals causing externalities must be taxed at the same rate, but where the externalities from some are more damaging than those from others. Diamond shows that if the externality function is linear, then the corrective tax should be a weighted average of the marginal externalities, where the weights are the elasticities of demand for the good. If the externality function is non-linear, however, then the corrective tax is related to the total cost of the externality divided by the number of consumers, which can be much *higher* than the weighted average marginal externality.

A behavioural perspective on this question is taken by O'Donoghue and Rabin (2006). They use a model with self-control problems to investigate the importance of population heterogeneity in driving the optimal tax. They find that even if only a small share of individuals have self-control problems, if externalities are large, the optimal tax on the full population can be large (when revenues are lump sum rebated). Thus, under both the standard model and this alternative, broad-based taxation of alcohol can be justified.

Of course, this may not be enough to sway public opinion! It is therefore worth considering whether more targeted tax policies towards alcohol are merited. One is to move to a purely alcohol-content based tax, rather than the rough approximations (taxes on beer vs. wine vs. spirits) used in most nations. This is a difficult issue, as noted in this chapter, because different alcohol 'delivery devices' have different potential for damage. For example,

spirits have about eight times as much alcohol per unit volume as does beer. Yet individuals who so desired could consume more than six times as much alcohol in a short period of time drinking spirits than they could drinking beer. Technically, the appropriate basis for alcohol taxation might be something like blood alcohol content per unit time of concentrated drinking.

That said, within type of alcohol, there is clear scope for differential taxation by alcoholic content. For example, all US states have only one tax on beer despite alcoholic content that varies from 4% (light beer) to 8% (dark beer). They also typically have just one tax rate on spirits despite alcoholic content that varies from liqueurs like Amaretto (28% alcohol) to grain alcohol (up to 95% alcohol).

A second approach is to consider differential taxation of alcohol in settings where it is differentially harmful. In the US, for example, it seems reasonable to conjecture that much of the damage due to drunk driving involves individuals leaving drinking establishments rather than drinking at home. This might suggest differentially high taxes on alcohol served in establishments. To overcome political opposition from bar owners, the revenue could be redistributed to them through a one-time tax credit to overcome their loss in asset value.

4.2. Obesity

Given the enormous growth in obesity in the US, it is rapidly becoming this nation's most important health problem. In principle, this issue could also be addressed with corrective taxation. But this is an even more challenging case than alcohol, and certainly more challenging than cigarettes, for three reasons.

First, damage from poor eating habits is not only non-linear—it is non-monotonic. Any food is unambiguously good for individuals who would otherwise starve, and it is only past a certain point that we would worry about the damage caused by overeating. Second, there remains considerable uncertainty in the science linking consumption of particular goods to the ultimate externalities and internalities of obesity. Finally, there remains uncertainty about the substitution patterns across inputs that might impact the shape of tax policy. If individuals, for example, substitute from fats to sugars when the former is taxed, is that obviously health-improving?

For all of these reasons, addressing obesity issues through taxes on inputs is very difficult. But the problem of obesity at least raises concerns about a common pattern of consumption taxation world-wide: excluding food from

consumption taxes. For example, in the UK, most foodstuffs for home consumption are exempt from tax; similarly, in the US food for home consumption is often excluded from state sales taxes. These exclusions can be clearly justified on distributional grounds, using a distribution-weighted approach to optimal commodity taxation.

Yet if there are obesity externalities from overeating, this provides a further rationale (along with relative inelasticity) against tax favouring foodstuffs. The UK VAT tries to address the obesity issue by standard-rating some food stuffs, such as ice cream. Yet other obesity-causing foodstuffs, such as cakes or chocolate chip cookies from a bakery, are zero-rated; potato-chips are standard-rated, yet tortilla chips are zero-rated. The difficulty of assigning obesity externalities to particular types of foodstuffs suggests that this is an extremely difficult approach to addressing them.

A better approach might be to follow the approach of a number of US states, which is to include food at the standard sales tax rate, but to provide rebates to low-income families to approximate the burden of food taxation (Johnson and Lav (1998)). Since these rebates would not depend directly on food spending, this system does not provide a marginal subsidy to food consumption, while in theory preserving the redistributive properties of a food exemption. In practice, however, a problem with this approach is that the rebate typically falls below the taxation of food expenditures for low income families, as politicians do not sufficiently inflate the rebate to match the rise in food expenditures (Johnson and Lav (1998)).

Unlike the case with smoking (where you can't tax those with lung cancer) or driving (where you can't sufficiently tax those who kill others through drink driving), there is a feasible output based taxation mechanism here: taxing body fat. If direct scientific evidence shows a strong link between certain measures of body fat and health costs, then the tax system could include body weight taxes. Of course, such a system is more easily proposed than implemented. Proxies for body fat, such as Body Mass Index (BMI), are only indirect; for example, African-Americans have a much higher BMI for any given level of body fat. And gathering data on even such proxies would be daunting.

REFERENCES

- Ainslie, G., and Haslam, N. (1992), 'Hyperbolic Discounting', in Loewenstein, G. and Elster, J. eds, *Choice Over Time*, New York: Russell Sage Foundation.

- Angeletos, G. M., Laibson, D., Tobacman, J., Weinberg, S., and Repetto, A. (2001), 'The Hyperbolic Buffer Stock Model: Calibration, Simulation and Empirical Evaluation', *Journal of Economic Perspectives*, 15, 47–68.
- Becker, G. S., and Murphy, K. M. (1988), 'A Theory of Rational Addiction', *Journal of Political Economy*, 96, 675–700.
- M. G., and Murphy, K. (1994), 'An Empirical Analysis of Cigarette Addiction', *American Economic Review*, 84, 396–418.
- Cook, P., and Moore, M. (2001), 'Environment and Persistence in Youthful Drinking Patterns', in Gruber, J. (ed.), *Risky Behavior Among Youths: An Economic Analysis*, Chicago: University of Chicago Press.
- Diamond, P. (1973), 'Consumption Externalities and Imperfect Corrective Pricing', *The Bell Journal of Economics and Management Science*, 4, 526–38.
- Fullerton, D. and Gan, L. (2005), 'Cost Effective Policies to Reduce Vehicle Emissions', NBER Working Paper No. 11174, March.
- Gruber, J. (2001), *Risky Behavior Among Youths: An Economic Analysis*, Chicago: University of Chicago Press.
- (2007), *Public Finance and Public Policy, 2nd Edn.* New York: Worth Publishers.
- and Koszegi, B. (2001), 'Is Addiction "Rational"? Theory and Evidence', *Quarterly Journal of Economics*, 116, 1261–303.
- (2002), 'A Theory of Government Regulation of Addictive Bads: Optimal Tax Levels and Tax Incidence for Cigarette Excise Taxation', NBER Working Paper No 8777, February.
- (2004), 'Tax Incidence when Individuals are Time Inconsistent: The Case of Cigarette Excise Taxes', *Journal of Public Economics*, 88, August, 1959–88.
- and Mullainathan, S. (2005), 'Do Cigarette Taxes Make Smokers Happier?', *Advances in Economic Analysis and Policy*, 5, art. 4 (2005). Available at <<http://www.bepress.com/bejeap/advances/vol5/iss1/art4>>.
- and Zinman, J. (2001), 'Youth Smoking in the U.S.: Evidence and Implications', in Gruber, J., ed., *Risky Behavior Among Youth: An Economic Analysis*, Chicago: University of Chicago Press, 69–120.
- Johnson, N., and Lav, I. (1998), *Should States Tax Food?*, Washington, DC: Center on Budget and Policy Priorities. Available at <<http://www.cbpp.org/stfdtax98.pdf>>.
- Laibson, D. (1997), 'Golden Eggs and Hyperbolic Discounting', *Quarterly Journal of Economics*, 112, 443–77.
- Lundberg, S. J., Pollak, R. A., and Wales, T. J. (1997), 'Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit', *Journal of Human Resources*, 32, 463–80.
- Manning, W. G., Keeler, E. B., Newhouse, J. P., Sloss, E. M., and Wasserman, J. (1991), *The Costs of Poor Health Habits*, Cambridge, MA: Harvard University Press.

- O'Donoghue, T., and Rabin, M. (1999), 'Doing it Now or Later', *American Economic Review*, **89**, 103–24.
- (2006), 'Optimal Sin Taxes', *Journal of Public Economics*, **90**, 1825–49.
- Pacula, R. L., Grossman, M., Chaloupka, F. J., O'Malley, P. M., Johnston, L. D., and Farrelly, M. C. (2001), 'Marijuana and Youth', in Gruber, J., ed., *Risky Behavior Among Youths: An Economic Analysis*, Chicago: University of Chicago Press.
- Viscusi, W. K. (1993), 'The Value of Risks to Life and Health', *Journal of Economic Literature*, **31**, 1912–46.