



WP/07/142

IMF Working Paper

VAT Attacks!

Michael Keen

IMF Working Paper

Fiscal Affairs Department

VAT Attacks!¹

Prepared by Michael Keen

June 2007

Abstract

This Working Paper should not be reported as representing the views of the IMF.

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

Like the theory of the second best that the 2006 congress marks, the VAT is now fifty years old. Judged by the extent and speed of its spread around the world, and the revenue that it raises, the VAT would seem to have been a remarkable success. Over the last few years, however, it has come under a series of attacks. This paper considers three of the most prominent of these. One is the fear (raised mainly in the United States) that the VAT actually does too good a job of raising tax revenue. The second is the view that the VAT does a bad job of taxing the informal sector—and that tariffs might be a better revenue-raising instrument for many developing countries. The third attack is the most literal, by criminals rather than theorists: in the European Union and elsewhere, sophisticated VAT fraud, targeting its refund provisions, has become a serious concern.

JEL Classification Numbers: H21, H26

Keywords: value added tax; indirect taxation; tax reform

Author's E-Mail Address: mkeen@imf.org

¹ Presidential Address delivered at the 62nd Congress of the International Institute of Public Finance, Cyprus, in August 2006. I am grateful to Ben Lockwood and Stephen Smith for allowing me to draw on our joint work, and to Sijbren Cnossen, Isaias Coelho, Alain Joustin, Victoria Perry, and Emil Sunley for many helpful comments and suggestions. Views and errors remain mine alone, and should not be attributed to the International Monetary Fund.

Contents	Page
I. Introduction	3
II. Second Best Perspectives on the VAT	5
III. Is the VAT A Money Machine?.....	7
IV. Informality and Borders.....	12
V. VAT Fraud.....	15
VI. Conclusion	18
References.....	19

I. INTRODUCTION

The 2006 congress of the IIPF marks 50 years of the theory of the second best. But there is another 50th (or so) anniversary that our profession should take note of. In 1954, France extended a previous tax and gave it a new name: the “value added tax.” Since then, of course, just as second best thinking has come to dominate theoretical public finance, so the VAT has risen to a position of extraordinary practical importance. Since its birth only 50 years ago, it has come to be adopted by more than 130 countries, including not only all OECD members other than the United States, but also many developing countries. And nor is its spread finished, with several more VATs planned for the coming months and years.

So now is an appropriate time to reflect on the VAT, both to take stock of what it has (and has not) accomplished, and to consider the links (or absence thereof) between its rise to practical preeminence and that of the second best approach to public finance. My purpose in this talk is to touch on both of these issues, and in doing so—indeed above all—to stimulate further interest in the VAT, which has been extraordinary neglected in the academic literature.

In terms of stock-taking, some form of mid-life crisis is natural enough when one reaches 50. And the VAT, it would seem, does do something right: the world likes it enough, in any event, to raise about 20 percent of its tax revenue in this way. Over the last few years, however, the VAT has been subject to a range of attacks, from both theorists and practical people.

Some of these really have little to do with the VAT as such. Take, for example, the argument that the VAT is a regressive tax. Most of the discussions and analyses of this issue would remain exactly unchanged if the VAT were replaced by some other form of general consumption tax—such as a retail sales tax—characterized by the same structure of rates on final consumption.² The arguments are generally not rooted, that is, in the distinctive feature that marks out the VAT from other forms of commodity tax: that it is in principle levied on all transactions, with sellers receiving credit or refund for the VAT charged on their inputs against the VAT chargeable on their own sales.³ It could be that some rate structures are less practicable under the VAT than under, say a retail sales tax, and that this implies some difference in their distributional impacts. But if that is the point, then that is where the discussion of whether the tax is intrinsically regressive should be rooted. Typically, however, it is not.

² An interesting exception is the analysis in Jenkins, Jenkins, and Kuo (2006) of the distributional impact of the VAT in the Dominican Republic, which shows, for example, the critical role of the exclusion of small traders from the VAT.

³ Attention is confined throughout to the invoice-credit method, destination-based VAT: that is, to a VAT charged, in principle, on all sales with credit or refund for all VAT paid on purchases, with imports fully taxable and exports charged at a zero rate. With the sole exception of Japan, there appear now to be no national VATs of any other type.

A first point then is that if one is to attack the VAT it is important to be clear whether the criticisms are unique to the VAT, or equally applicable to alternatives. With that proviso, there seem to me three lines of recent attack on the VAT which touch sufficiently close to the structural essence of the tax to deserve close attention:

- The first is close to paradoxical: that the VAT is simply too easy a way for a government to raise money. This, most notably, was one of the key reasons why the recent presidential panel on tax reform in the United States found itself hung on the issue of whether or not to adopt a VAT: “Some panelists were...concerned that introducing a VAT would lead to higher total tax collections over time and facilitate the development of a larger federal government—in other words, that the VAT would be a ‘money machine’.”⁴ Of course proponents of the VAT have traditionally used this as the central argument in favor of the tax, claiming that it is a more than usually nondistortionary and practicable way of raising revenue. Underlying both views, however, is a common presumption that the VAT has indeed proved itself a particularly effective form of taxation. But the simple question then arises: has this in fact to be the case?
- The second line of attack is the argument that the VAT functions poorly when, as in most developing countries, there is a large informal sector. Piggott and Whalley (2001) illustrate very elegantly the general point that broadening the base of a consumption tax may be welfare reducing in the presence of informality. This observation is developed into a more thoroughgoing criticism of the VAT by Joseph Stiglitz at the IIPF congress in 2003 (see Stiglitz (2003)). Aspects of the argument are developed further in Emran and Stiglitz (2005),⁵ who derive a series of results on the welfare effects of shifting from tariffs to (their characterization of) the VAT which they believe “raise serious doubts about the wisdom of the indirect tax reform policies pursued by a number of developing countries.”⁶
- The third criticism is that the VAT has proved vulnerable to significant fraud—the attack on the VAT, in this case, is by criminals rather than theorists. The European Commission (2004), for example, reports that losses from fraud have amounted to 10 percent of net VAT receipts in some member states. Most famously, the cost of carousel fraud in particular—a class of schemes exploiting the zero-rating of exports⁷

⁴ President’s Advisory Panel (2006, p. 192.

⁵ See also Munk (2004).

⁶ Emran and Stiglitz (2005), p. 599 (abstract).

⁷ To ensure that the VAT falls on domestic consumption, exports are invariably zero-rated under the VAT: this means that while no tax is payable on output, tax paid on inputs is fully refundable.

and deferral of tax on intra-EU imports—has been put at around €2.1 billion in Germany (roughly 1.5 percent of VAT revenue) and at £1.12–1.9 billion in the United Kingdom (about 1.5–2.5 percent of VAT revenue).⁸ Nor is fraud a problem only for developed countries: one scam in Sri Lanka reportedly cost SL Rs 24 billion, a sizable part of VAT collections.⁹ Such difficulties have led at least one commentator to argue that “the writing is on the wall for the VAT system.”¹⁰

These attacks require serious attention, and later sections will examine them in turn. Before doing so, however, it is useful to consider the broad issue, raised at the outset: the links between the VAT and second best tax analysis. These, as will be seen, turn out to bear directly on the current attacks on the VAT.

II. SECOND BEST PERSPECTIVES ON THE VAT

There is not, to put it mildly, a large academic literature on the VAT: there are, for example, only four papers in the *Journal of Public Economics* with “VAT” in the title.¹¹ And while many more papers speak of a “VAT,” many do not capture any of the distinctive features of the VAT outlined above: all too often, “VAT” is used as synonymous with “uniform consumption tax.” But, as noted earlier, a VAT has a quite distinct structure in terms of its collection throughout the production chain. What, it is natural to ask, might be the appeal of this particular form of tax? This, surely, is a question that the second best approach to tax analysis should by now have helped us to answer.

When one approaches this issue in second best terms, however, there quickly emerges a tension—perhaps an inconsistency—in the argument usually given for the superiority of the VAT over other types of commodity taxation.¹² The VAT is preferable to a turnover tax, the argument goes, because, when it works as it should—with an unbroken chain of VAT charged and credited or refunded at all stages of production—since it avoids distorting the input prices faced by business and so creating the production inefficiencies that, Diamond and Mirrlees (1971) taught us, are unlikely to have any place in a Pareto efficient tax

⁸ ‘A Tax Net Full of Holes,’ *The Economist*, May 13, 2006.

⁹ <http://www.colombopage.com/archive/April18143912SL.html>

¹⁰ Quoted in *Financial Times*, June 20, 2006.

¹¹ There are, in fairness, some signs that this may be changing. Recent contributions on the VAT other than those cited elsewhere include Aizenman and Jinjark (2005), who examine determinants of the effective VAT base, Desai and Hines (2005), who examine the link between the VAT and export performance, and de Paula and Scheinkman (2007), who develop the notion of VAT chains (firms being less likely to be VAT-compliant if their suppliers or customers are noncompliant).

¹² As, for example, in Ebrill and others (2001).

system.¹³ Indeed in this respect the remarkable spread of the VAT can be seen as the greatest triumph, and vindication, of optimal tax theory.

But there is more to the argument for a VAT than that. Why is it any better than a retail sales tax? Well, the argument goes, this is because if, for some reason, final sales escape tax, then revenue will be protected under the VAT by having been collected at earlier stages in the production process.

So: the VAT is preferable to one alternative because it does not ultimately tax business inputs, and is preferable to another because it does.

The tension this implies may be more apparent than real, however. When there are constraints on the distorting tax instruments that the government can deploy, production inefficiency, in an appropriate form, generally becomes desirable (Newbery, 1986). Thus it may be that the input taxation that remains under the VAT when some final sales cannot be taxed serves, under those circumstances, a constructive purpose. But, on the other hand, maybe a cascading turnover tax—or a tariff—would then be even better than a VAT? To probe further, some understanding of why the VAT chain might be incomplete is needed.

There are in some cases technical reasons that prevent some transactions being brought into tax: it is not easy, for example, to allocate the value added in the provision of financial services between buyer and seller in such a way as to ensure proper functioning of the crediting mechanism.¹⁴ And costs of administration and compliance may warrant excluding some perfectly honest traders from the VAT. But noncompliance is clearly one key reason for incompleteness of coverage. This indeed is the essence of the informality issue raised above. And noncompliance immediately creates quite distinctive problems for the VAT. Under most taxes, the amount actually payable is bounded below at zero—or at least, for refundable credits such as those under the EITC, is bounded below. For VAT, the provisions for refunding mean that liability is not bounded below. Hence the attraction of VAT fraud for serious and sophisticated criminals.

Some of the most distinctive and challenging second best issues raised by the VAT thus stem from the incompleteness of its coverage, whether due to outright noncompliance or the constraints implied by the challenges of administration and compliance. It is these that are at the root of the various attacks outlined above—including the money machine argument, though in this case it is the absence of these limitations, not their presence, that is seen as the root of the problem. In particular:

¹³ There are, of course, many qualifications to the desirability of production efficiency (including those implied by an inability to tax pure profits, restrictions on the use of distorting tax instruments, and, in an international context, limitations on the ability to make international lump sum transfers).

¹⁴ Zee (2006) provides a thorough account of the difficulties that arise in bringing the financial sector into the VAT, and of alternative approaches.

- Since it is in principle unclear—even when its coverage is complete, but especially when it is not—whether the VAT is indeed superior to other forms of indirect taxation, the question arises: is there any empirical evidence that the VAT has indeed proved a particularly efficient tax? Viewed from another perspective—has it proved a money machine?
- The informality criticism raises a question to which our second best intuition should quickly call us to pay attention: how should VAT be designed when some transactions will escape the tax? In particular: for developing countries, would tariffs (or some other tax) do a better job?
- Do distinct opportunities for fraud arise under the VAT that are so profound as to offset its potential merits, or can these difficulties be kept to manageable proportions?

It is striking, and troubling, that after 50 years of both the VAT and the second best approach to tax design, these questions have hardly been addressed. The rest of this paper tries to make some progress on each of these three issues in turn.

III. IS THE VAT A MONEY MACHINE?¹⁵

The central claim made by advocates of the VAT is that it is a particularly efficient¹⁶ way of raising revenue. This has generally been seen as a potential strength of the tax—including, not least, in developing countries, where strengthening tax revenues is a common element of growth and poverty reduction strategies. But some observers instead see in this supposed efficiency a danger that adoption of the VAT would in itself cause an unwelcome growth of government. This argument is most commonly raised in the United States—the presidential panel cited above being only the most recent expression of a view of the VAT that dates back at least to Brennan and Buchanan (1977). And indeed the emergence of the differential in the size of government across the Atlantic is more or less coincident with the development of the VAT in the latter: between 1965 and 2002 the tax ratio in United States was broadly unchanged, at around 26.5 percent; in the European OECD countries, it rose from only a percentage point or so higher than that in the United States to about 12.5 points more. This has made it tempting to see the rise of the VAT as a cause, rather than a consequence, of the growth of government.

¹⁵ This section draws heavily on Keen and Lockwood (2006a, 2006b).

¹⁶ Proponents of the VAT rarely argue that it has distinct attractions in equity terms (partly for the reason noted in the introduction), typically arguing that it is no worse than alternatives and/or that distributional concerns are best addressed on the spending side, including, not least, by use of the revenue raised by the VAT.

Weak and strong money machines

To dissect these arguments, Keen and Lockwood (2006b) distinguish between two forms of money machine hypotheses:

- *Weak form*: This is the claim that countries with a VAT raise more revenue, all else equal, than do those without.

The issue here—whether adoption of the VAT has indeed been associated with increased revenue mobilization—is clearly central to assessing whether the tax has indeed lived up to the promises made by its proponents. Yet is symptomatic of the lack of academic attention to the VAT that it has received almost no attention.

- *Strong form*: The claim that use of the VAT has in itself been a cause of increased government size.

To see the significance of and inter-relation between these hypotheses, and how they can be tested, suppose first that policy is made so as to maximize a welfare function of the form:

$$W = U(R) - \left(\frac{1}{2}\right)\theta R^2 \quad (1)$$

where R denotes the level of tax revenue (all used to finance public expenditure, generating welfare according to U) and θ parameterizes the (in)efficiency of the single tax instrument assumed to be at the government's disposal: the more efficient the tax instrument, the lower is θ . The necessary condition on R ,

$$U'(R) - \theta R = 0 \quad , \quad (2)$$

is then easily seen to imply that R is decreasing in θ : access to a more effective tax results in more revenue optimally being raised. This points to another interpretation of the weak form of money machine hypothesis: if the VAT has indeed proved a particularly efficient tax, then one would expect countries with a VAT to raise more aggregate revenue, all else equal, than countries without: that is, higher revenue is indicative of underlying efficiency of the tax system. The empirical strategy this points to is running regressions of the form

$$R_{it} = \alpha V_{it} + \beta'_v V_{it} X_{it} + \beta' X_{it} + \mu_i + \lambda_t + u_{it} \quad (3)$$

where now R_{it} denotes the ratio of tax revenue to GDP in country i at time t , V_{it} is a dummy variable indicating the presence ($V = 1$) or absence of a VAT,¹⁷ X_{it} is a column vector of

¹⁷ Not all VATs are the same, of course, so that use of a dummy of this kind oversimplifies. But data deficiencies, especially outside the OECD, leave little scope for more subtle characterizations.

other variables that may affect tax revenue (the usual suspects including openness, GDP per capita, and dependency ratios), and the remaining terms are country- and time-specific effects and an idiosyncratic error. In effect, equation (3) is a standard “tax effort” equation augmented by inclusion of a VAT dummy, which in turn is allowed to affect tax revenue both directly and in interaction with other variables. The argument that the VAT would be associated with higher tax revenue if its presence significantly increased the effectiveness of the tax system then translates into the hypothesis that $\alpha + \beta'_V X_{it} > 0$.

The first attempt to address what is termed here the weak form of money machine hypothesis was that of Ebrill and others (2001), who estimate equations of the form in (3) for a large cross-section of countries. This, however, suffers from evident problems of heterogeneity bias. Potentially more persuasively, Keen and Lockwood (2006a) estimate a series of equations of the general form in (3) for an unbalanced panel comprising the thirty current OECD member countries for the period 1965–2004 (including, for each, any years prior to OECD membership).¹⁸ The picture that emerges is quite complex, with important roles emerging for both dynamics, through the inclusion of the lagged dependent variable, and interactions between the VAT dummy and country characteristics. On balance, the impression is of a revenue gain associated with the presence of a VAT that is significantly positive—supportive of the weak form money machine hypothesis—but modest in size. For example, evaluating the revenue gain from the presence of a VAT at the mean values of the controls, \bar{X} , for those countries and years in which a VAT is not in place, the predicted revenue gain $\Delta R = \alpha + \beta'_V \bar{X}$ from adoption of the VAT by a typical country-year in the sample without a VAT varies across specifications from about 0.5 to 2.1 percent of the initial tax ratio.

With the weak form of money machine hypothesis broadly confirmed, what of the strong form? How might one identify whether the VAT itself has caused the increase in government size? Now it becomes useful to amend the simple framework above to have two tax instruments, indicated by subscripts A and B , and a parameter λ increasing in the strength of taste for government spending. So policy now seeks to maximize:

$$W = \lambda U(R_A + R_B) - \left(\frac{1}{2}\right)\theta_A (R_A)^2 - \left(\frac{1}{2}\right)\theta_B (R_B)^2 \quad (4)$$

In this case it is easy to see that while an increase in the efficiency of either tax instrument, say A , leads to an increase in total revenue—consistent with the weak form hypothesis—it also leads to a *reduction* in the revenue optimally raised from the other. So to the extent that increased revenue is driven by access to a more efficient tax instrument, one would expect, to some degree, a partially offsetting reduction in revenue from other taxes. It may at first seem surprising that the clue to the VAT having been a cause of increased government size being

¹⁸ In work in progress, Keen and Lockwood (2006a) estimate equations of the same form for an unbalanced panel of 143 countries observed between 1975–2000, recognizing too the endogeneity of V by modeling revenue equation jointly with a probit for adoption of the VAT. The preliminary results there point to similar complexities, with quite substantial heterogeneity in the impact of the VAT.

that government size has increased by less than the amount raised by the VAT. But the reason is straightforward: while access to a more efficient tax instrument naturally leads to more revenue being raised, part of the benefit is also taken in the form of reduced use of more distorting tax instruments. This stands in contrast to a second implication of the framework above: if the growth of government is fueled by increased taste for government, parameterized by an increase in λ , then this will be reflected in increased revenue from *both* instruments, *A* and *B*.

Empirically, these observations suggest testing the strong form of money machine hypothesis by estimating equations of the form:

$$R_{it} = \gamma_V RV_{it} + \sigma V_{it} + \gamma Z_{it} + \mu_i + \varsigma_t + \xi_{it} \quad (5)$$

where RV_{it} denotes revenue from the value added tax (as a share of GDP), Z a vector of additional variables, δ , γ_V , σ and γ parameters to be estimated, and the last three terms are again country- and time-effects and an idiosyncratic error. Under the strong form of money machine hypothesis, $\gamma_V < 1$.

Running regressions of the form in (5) for the same OECD sample as above,¹⁹ Keen and Lockwood (2006a) again find a somewhat complex story, including possible nonlinearities in the degree to which increases in VAT are associated with changes in revenue from other sources. The weight of the evidence, however, suggests that increases in VAT revenue are indeed associated, all else equal, with reduced revenue from other sources. Taking, for example, the ‘average’ OECD country, which now collects about 7.2 percent of its GDP in VAT revenue, the associated long run increase in total revenue is about 2.4 percent of GDP: that is, around two-thirds of the revenue raised by the VAT is offset by reduced revenue from other taxes. This support for the strong form of the money machine hypothesis is indeed consistent with the relatively soft support for the weak form: while the presence of a VAT is associated with higher total revenue, this impact is muted by reductions in receipts from other taxes that to some degree offsets that from the VAT itself.

Isn’t a money machine a good thing?

The underlying view of the world of those who fear rather than welcome the efficiency of the VAT, of course, is one in which government, if unchecked, tends to become too big. This then leads to the argument that one way of constraining government size—in the absence of more direct checks—is to constrain the government to use inefficient tax instruments. The point dates back to the analysis of Brennan and Buchanan (1977) who indeed cite exactly the VAT as a leading instance.²⁰ Their framework is of a world in which citizens believe that policy in a post-constitutional phase will be made by Leviathans who, diverting some fixed

¹⁹ Keen and Lockwood (2006a) also run Granger causality tests, which prove essentially inconclusive.

²⁰ Brennan and Buchanan (1977), p. 272.

proportion of tax revenue to their own use, will simply seek to maximize tax revenue within the limits of the taxing powers granted to them in the constitution. Having in mind some desired level of tax revenue—reflecting that part of it from which they will benefit—citizens will then rationally support constitutional restrictions which limit to that level the maximized amount of revenue the leviathan can subsequently achieve.

The conclusion that welfare may be reduced by the use of efficient tax instruments is open to a number of objections, even if one accepts the view of policymakers as self-seeking. Fiscal rules of various forms, for example, may directly constrain overall revenues and spending, while allowing them to be financed in the most efficient way. Moreover, a simple and reasonably plausible relaxation of the extreme Brennan-Buchanan assumptions substantially modifies the conclusion. Suppose that policymakers in the post-constitutional phase do not simply seek to maximize the amount of spending, C , devoted to their own use, but also attach at least some positive weight to the well-being of the citizenry. More precisely, suppose they seek to maximize some function $\Omega(C, W)$ defined over not only C but also the citizenry's welfare W , with the latter reflecting the public expenditure $R - C$ from which they benefit. So (returning for simplicity to the case in which there is only one tax instrument), private welfare is now:

$$W = V(R - C) - \left(\frac{1}{2}\right)\theta R^2 . \quad (6)$$

It is easy to see that an increase in the efficiency of available instruments shifts outwards the possibility frontier in (C, W) -space from which the policymaker chooses. So long as W is normal in the policymaker's preferences, the income effect of this policy innovation will thus lead them to choose a higher value not only of C but also of W : on this account, the citizenry benefits. There is, however, also a substitution effect at work, and this points in the opposite direction: with a more efficient tax instrument available, raising the revenue to finance a marginal increase in C now requires, if anything, less of a reduction in W , in effect making C cheaper. The strength of this effect, in turn, can be shown to depend on the curvature of $V(\cdot)$: if $V'' = 0$, the substitution effect vanishes, since the cost in terms of private utility of diverting revenue towards the use of the leviathan remains unchanged. In this case, citizens' welfare is unambiguously increased by access to a more efficient tax instrument, even though policymakers divert some tax revenue to their own purposes.

More subtle political economy models also lead to ambiguous conclusions. Besley and Smart (2007), for example, construct a model in which policymakers do not have the mixed motive of $\Omega(C, W)$ above but instead differ in type, some caring only about C and others only about W , and in which voters try to elect only the latter. In this case, access to a more efficient tax instrument tends to be welfare-worsening to the extent that it enables an incumbent leviathan to extract more rent but welfare-improving to the extent that this in itself makes it more likely that such incumbents will reveal their type and hence be ejected, and replaced by a benevolent policymaker, at the next election. Only if access to the more efficient instrument shifts the equilibrium from separating to pooling—so that the electoral discipline effect is

removed—can welfare fall, with the likelihood of this then depending on such quantities as the discount rate and the proportion of politicians who are bad.

The possibility that access to a more efficient tax instrument will ultimately reduce citizens' welfare thus cannot be ignored. And the highly stylized political economy models that the literature provides given little direct guidance as to whether or when this is likely to be the case in practice. They do suggest, however, that adverse welfare effects cannot be taken for granted unless one takes views of policymaking that are likely to be extreme even for quite repressive regimes.

IV. INFORMALITY AND BORDERS²¹

The importance of the challenges that informality poses for both tax design and administration are well-known. It is, of course, extremely hard to measure with any great accuracy the size of the informal sector—by which will be meant here simply the set of businesses and individuals that are less than fully tax compliant. Widely cited estimates nevertheless confirm the widespread presumption that it is often substantial, especially in the developing countries that are particularly in mind in this section: Schneider (2002), for example, puts the informal economy in developing countries at, on average, 41 percent of gross national income. The existence of such a large part of the economy not fully in the tax system represents a potentially severe distortion both directly in the misallocation of resources between formal and informal activities and indirectly through the loss of revenue that it entails. It is likely, moreover, to induce inequities, and to reinforce governance problems.

The charge leveled against the VAT is that it is essentially a tax on the formal sector, and so is ill-suited for circumstances in which informality is a significant concern. A natural response is to ask: what would be better? And here interest has turned to the potential role of import tariffs. These, the argument goes, are easy to administer and, in particular, will be borne by the informal sector in so far as they import some of their inputs. Thus Emran and Stiglitz (2005), in particular, establish a series of results establishing circumstances in which a revenue neutral shift from import duties to a “VAT” reduces welfare.

This comparison between the VAT and tariffs, it should be stressed, is of more than conceptual importance. Many developing countries still rely heavily on the revenue raised by tariffs: in Sub-Saharan Africa, for instance, trade taxes still account, on average, for about one-quarter of all government revenues, and in the developing countries of Asia and the Pacific they account for around 15 percent. As trade liberalization proceeds, the question arises as to how such countries might replace lost trade tax revenue from domestic sources. And here the conventional policy advice attaches a key role to the VAT, along with excises. On these issues, however, the emerging empirical evidence is troubling. Looking back at past episodes of lost trade tax revenue, Baunsgaard and Keen (2005) find that while

²¹ This section draws on Keen (2006).

middle-income countries have generally recovered lost trade tax revenue from other sources, many low-income countries have not. Especially important for present purposes, they also find that the degree of revenue recovery is not significantly greater in countries that have a VAT than in those that do not. Part of the reason for that may be that the simple 0/1 dummy used in these regressions does not capture the importance of key design features of the VAT—the number of rates, the level of the threshold, and so on. But might it also be that the presence of a large informal sector in these countries renders the VAT structurally inappropriate for replacing lost trade tax revenue in this way?

VAT, tariffs, and withholding taxes

In thinking through these issues, it is important to be clear on the nature and range of the tax instruments available to developing countries. And a key point here is simply that the VAT is levied on imports, including by those not registered for the VAT. Those who are registered for VAT will, of course, be able to claim a credit or refund against the tax due on their own sales. Those not registered, however, will not be able to recover the VAT charged on their imports. For them, the VAT on imports functions exactly like a tariff; and indeed is administered and collected in almost exactly the same way. Nor is this point merely one of principle: in practice, most developing countries raise more than half of their gross VAT collections from imports.²² This point is recognized, it should be noted, by Emran and Stiglitz (2005).²³ In their formal analysis, nevertheless, the ‘VAT’ is in effect defined as that part of the “real” VAT which is not collected at the border. As can be imagined, given the revenue importance of VAT on imports, few practitioners would recommend the adoption of a VAT if, for some reason, it could not be collected at the border.

It is thus important to recognize that the VAT in practice serves in part as a surrogate tariff, with the attractive feature that it bears only on the purchases of informal operators. This does not quite solve all our problems, however. For at the same time, the VAT also functions as a tax on final sales by producers in the formal sector. And in principle, one might want to charge these two taxes—on formal final sales, and on informal imports—at different rates. Absent a kind of rate differentiation that is both unusual and likely to be hard to implement, however, the VAT cannot do this: it taxes these two things at the same rate.

This observation in turn points to there being potential value in having available another tax instrument, explicitly targeted to informal sector imports. And indeed many developing countries do deploy such an instrument, though its conceptual and practical importance appears to have been entirely neglected in the theoretical literature (and indeed in much of the more practically oriented literature too). That instrument is withholding against income tax liability in the form of a charge on imports. Many developing countries impose taxes of this sort, in many cases at fairly high rates of 4 percent or more, sometimes only on traders unable to produce a tax identification number and sometimes more generally. In principle at

²² See for instance Table 4.3 of Ebrill and others (2001) and Table 1 of Keen (2006).

²³ See their footnote 14.

least, these taxes are creditable for those paying income tax; for those in the informal sector escaping the income tax, however, they remain a final tax, again functioning as a de facto tariff. Such withholding taxes, it should be noted, appear to be in principle WTO-consistent, so long as they have no bite for fully compliant taxpayers. In practice, it is far from clear how far formal sector operators do indeed receive appropriate credit (or refund): if they do not, the withholding in effect becomes simply a de facto tariff. The prior question, however, is that of what use, if any, such taxes might have if they were properly implemented.

Optimal tax structures

Consider then the problem faced by a developing country that can levy three types of tax: a VAT, levied at a single rate on both formal sector sales and informal sector imports; a tariff; and an additional tax on informal sector imports. How should these be deployed?

Keen (2006) addresses these issues in a simple model in which formal and informal operators compete in the production of a nontradeable good, using a single imported intermediate. There are constant returns in the formal sector, with the size of the informal sector then determined by decreasing returns.²⁴ There are also assumed, for simplicity, to be fixed coefficients in the informal sector.

A first and crucial lesson that emerges is that if the VAT and withholding tax are optimally deployed, then the tariff should be set to zero. Intuitively, with these other instruments targeted independently to the taxation of final sales and of informal inputs, there remains no useful purpose for tariffs: all they do is distort production in the formal sector.

The characterization of the optimal VAT and withholding tax rates proves somewhat involved. Matters are clearest if informal sector profits have the same social weight as tax revenue—a useful simplifying assumption, rather than an especially plausible one. In this case, the optimal VAT is characterized by the usual Ramsey rule, without direct reference to the existence of an informal sector. And the withholding tax, as one might expect, is higher the lower is the share of the taxed import in informal sector production costs.

One special case deserves some emphasis: if the informal sector imports all its inputs, then the VAT alone is fully optimal. Intuitively, even though some final sales escape the VAT, one can do no better, in this case, than to deploy only the VAT, set at the rate implied by the Ramsey rule. The interest of this result is that it establishes circumstances in which a ‘real-world’ VAT (meaning one that becomes an input tax for unregistered traders) is all that is needed: neither tariffs nor withholding on imports need to be deployed. Of course, the circumstances required for this result are very special. Indeed that, in a sense, is the point. Outside this very restricted case, there is a purposive role for the withholding tax, although not for tariffs.

²⁴ Broadly the same model is also used by Piggott and Whalley (2001) and Stiglitz (2003).

These results suggest a key role in developing country tax policy not only for the feature of the VAT that it acts as an import tariff on the informal sector, but also for the import withholding taxes that the literature has neglected. This also suggests that these taxes may have a role to play in dealing with the revenue replacement issue raised above. Indeed, while the conventional prescription of matching tariff cuts by point-for-point increases in domestic taxation²⁵ does not retain its attractive properties in the presence of an informal sector, it is straightforward to show that an alternative strategy—combining an indirect tax increase with higher withholding on imports—does.

One further point should be noted. This is that while the results reported here focus on the way in which the VAT taxes the informal sector on its imports, the VAT also serves essentially the same purpose on *all* inputs that informal operators purchase from formal. This calls attention too to a wide range of other withholding taxes deployed in many developing countries, such as withholding of VAT on their purchases by public and/or large enterprises. Thus questions also arise as to how these are best designed and deployed. The practice, it should again be stressed, may well be quite different from the theory above: there is very little evidence on the extent to which compliant taxpayers are, in fact, able to take credit or refund for taxes withheld against them, but much suspicion that it is often very limited. While this should naturally temper policy advice, the key point for tax design remains: the borders that matter in thinking about the taxation of the informal sector are not simply the literal borders of the country itself; the borders that are crossed by transactions between formal and informal sectors may be at least as important in designing appropriate tax structures.

V. VAT FRAUD²⁶

Like any tax, the VAT is open to fraud and evasion. Many of the ways in which it can be abused are the same as for other taxes, though modified by the particular structure of the VAT: as under a sales tax, for instance, retailers can try to escape tax by underdeclaring their sales, though their incentive to do so may be less under the VAT if, in order to make the underdeclaration of sales credible, they feel it necessary also to understate their purchases and hence the input credit due. But distinct possibilities for fraud do arise under the VAT, generally reflecting the opportunities provided by the need to pay systematic refunds, notably (though not only) to exporters. This has proved an attractive target for sophisticated fraudsters, including organized crime.

²⁵ This argument, together with some extensions and qualifications, is developed in Keen and Ligthart (2002). For infinitesimal reforms, see also Anderson (1999) and Hatzipanayotou, Michael, and Miller (1994).

²⁶ This section draws on Keen and Smith (2006).

Carousel fraud—what can be done?

While fraudulent refund claims can be generated in many ways, the archetypal form—at the center of the current debate in the EU—is *carousel fraud*, which highlights the depth and potential scope of the difficulties associated with the control of VAT refunds.²⁷

In its simplest form, this works as follows. Firm A, registered for VAT in one member state, imports goods from another. Under the deferred payment arrangements of the EU, it should account for the import VAT due in its next periodic return, crediting this amount against the output tax due on its own sales. Firm A does indeed invoice VAT to its customer, firm B, but fails to remit the VAT payable, and disappears. Firm B then exports the goods, and—on the strength of the invoice issued to it by A—claims a refund for VAT that has not, in fact, been paid. The goods, often high value items easy to transport (such as mobile phones or computer chips) are then brought back into the country once more, and the carousel keeps on turning. In practice, schemes are often (stunningly) more complex than this example, with buffer companies, perhaps perfectly innocent, placed between A and B, and/or more complex routing of goods. But the essence remains the generation of refund claims that do not correspond to VAT actually paid.

A variety of responses to such frauds have been proposed. Many are essentially administrative, in the sense that they retain the zero-rating of exports that creates the possibility of fraudulent VAT refund claims. At one level, these may essentially involve tightening current practices: more extensive controls on registration, for example, and more effective international information exchange (to verify that goods recorded as exported from one member state to another duly attract VAT in the latter). There are also more radical administrative proposals:

- One is to extend the scope of *reverse charging*, by which liability for the VAT is placed on the purchaser rather than the seller. This would block the simple fraud above, for example, since firm B would be liable for the input VAT itself. To this end, a number of EU countries have adopted reverse charging for a number of specific goods. Most radically, Germany and Austria have proposed adopting reverse charging for *all* business-to-business transactions above some modest size. The difficulty with this, however, is that it in effects defers VAT liability until the final sale to a nonregistered person, converting the VAT into a retail sales tax and so losing the distinctive merit of the VAT that revenue is protected by being collected throughout the production chain.
- Another set of proposals, differing somewhat in their detail, is for 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: see for

²⁷ Harrison and Krelove (2005) provide an extensive discussion of problems and experience in the management of VAT refunds.

example Sinn, Gebauer, and Parsche (2004). In the example above, firm B would be denied a refund on the grounds that firm A had not paid the corresponding amount into the required account. The difficulty with this lies in the additional compliance costs imposed on traders, and the interest that traders would forego by paying VAT earlier than under a system of periodic payments. It is for these reasons that Bulgaria, which has operated a system of this kind, recently announced its removal.

None of these administrative solutions addresses the ultimate source of the difficulty: the breaking of the VAT chain as goods pass between countries—which is a particularly bad place to have such a break, moreover, as it is the point at which control passes from one national tax administration to another. Such is the sophistication of the frauds that have emerged in recent years, and so great is their potential payoff, that difficulties seem likely to continue so long as this key structural feature of the VAT remains.

Lasting solutions may thus need to be found in more radical restructuring of the VAT to close this break in the VAT chain. There are several candidates. One is the CVAT of Varsano (2000), under which exports would continue to be zero-rated in the country of export but become liable to a special compensating tax fully creditable to the importer. Another is the VIVAT of Keen and Smith (1996, 2000), under which all business-to-business transactions, including those across member states, would be subject to a common VAT rate. And there are other possibilities. The central point is that the most effective way to attack refund-based fraud is likely to be by pursuing the logic of the VAT more forcefully.

How important is it?

VAT fraud and evasion is a serious concern. Even apart from revenue losses of the order cited in the introduction, it can significantly distort trade statistics²⁸ and so interfere with macroeconomic management. More generally, the dramatic nature of many of these frauds, and their regular appearance in the media, runs the risk of bringing the VAT more widely into disrepute, and weakening compliance with it more generally.

But it is important to keep the scale of losses in perspective. In the United Kingdom, for example, official estimates put the VAT gap—the shortfall of collections from the amount properly due, arising from all sources, as a percentage of the latter—at about 13.5 percent for 2004–05. While there are no comparable estimates for the gap under the personal income tax, estimates for the United States, generally regarded as reasonably accurate, consistently put this at around 14 percent.²⁹ Given the more extensive withholding in the United Kingdom, it thus seems likely that the VAT and personal income tax gaps are of much the same order. While such a rough equality of gaps is not necessarily optimal, and no reason for complacency in the face of what are evidently quite profound difficulties, they do suggest that the time has not yet come to panic.

²⁸ See for instance Ruffles and others (2003).

²⁹ See for example Slemrod (2006).

VI. CONCLUSION

After 50 years of experience with the VAT, and 50 years of progress in better understanding the principles by which tax systems should be designed, there remains much that is not known about the VAT.

Why, given its obvious importance, has the VAT received so little attention? One reason, surely, is the relative lack of interest in the tax in the United States, which drives so much of the research agenda in our profession.³⁰ Conversely, the VAT is especially important to developing countries, indeed for them VAT and trade reform are arguably *the* issues; and tax issues in developing countries continue to be largely absent from the academic public finance research agenda. A second and more prosaic reason may be the relative lack of comparable data on VAT revenues and design (other than the basics of rate structure), especially outside the OECD. There is also, perhaps, a third reason. Tolstoy has it that all happy families are alike, but every unhappy family is unhappy after its own fashion.³¹ Much the same is true of the VAT. If it functions as it should, any VAT is indeed simply a uniform tax on final consumption, to which all our simple textbook models in principle apply. But a less than perfectly functioning VAT is an analytical mess, with particularities of production relationships and compliance behavior immediately becoming key. And it is imperfectly functioning VATs that we observe in practice and need to understand better. Tolstoy, of course, chose to write about an unhappy family. Similarly, I hope this talk will have helped made the case that the study of imperfect VATs is not only worthwhile, but interesting too.

³⁰ With the presidential panel having raised, and not ruled out, the possibility of VAT adoption in the U.S., this may be changing: see for instance the symposium on VAT in the December 2006 *National Tax Journal*.

³¹ In the opening line of *Anna Karenina*.

References

- Aizenman, Joshua and Yothin Jinjarak, 2005, “The Collection Efficiency of the VAT,” (mimeo, Princeton University).
- Anderson, James E., 1999, “Trade Reform with a Government Budget Constraint,” in *International Trade Policy and the Pacific Rim*, ed. by John Piggott and Alan Woodland (Macmillan Press: London).
- Besley, Timothy and Michael Smart, 2007, “Fiscal Restraints and Voter Welfare,” *Journal of Public Economics*, Vol. 91, pp. 755–73.
- Baungsgaard, Thomas, and Michael Keen, 2005, “Tax Revenue and (or?) Trade Liberalization” IMF Working Paper 05/112 (Washington: International Monetary Fund).
- Brennan, Geoffrey, and James M. Buchanan, 1977, “Towards a Tax Constitution for Leviathan,” *Journal of Public Economics*, Vol. 8, pp. 255–73.
- de Paula, Áureo, and José A. Scheinkman, 2006, “The Informal Sector” (mimeo, Princeton University).
- Desai, Mihir A., and James R. Hines Jr., 2005, “Value Added Taxes and International Trade: The Evidence” (mimeo, University of Michigan).
- Diamond, Peter, and James A. Mirrlees, 1971, “Optimal Taxation and Public Production I-II,” *American Economic Review*, Vol. 61, pp. 8–27 and 261–78.
- Ebrill, Liam, Michael Keen, Jean-Paul Bodin, and Victoria Summers, 2001, *The Modern VAT* (Washington: International Monetary Fund).
- Emran, Shahe M., and Joseph E. Stiglitz, 2005, “On Selective Indirect Tax Reform in Developing Countries,” *Journal of Public Economics*, Vol. 89, pp. 599–623.
- European Commission, 2004, *Report from the Commission to the Council and the European Parliament on the Use of Administrative Cooperation Arrangements in the Fight Against VAT Fraud*, COM(2004) 260 final.
- Harrison, Graham, and Russell Krelove, 2005, “VAT Refunds: A Review of Country Experience,” Working Paper 05/218 (Washington: International Monetary Fund).
- Hatzipanayotou, P., M.S. Michael, and S.M. Miller, 1994, “Win-Win Indirect Tax Reform: A Modest Proposal,” *Economics Letters*, Vol. 44, pp. 147–51.
- Jenkins, Glenn P., Jenkins Hatice P., and Chun Yan Kuo, 2006, “Is the VAT Naturally Progressive?” (mimeo, Queens University).

- Keen, Michael, 2006, "VAT, Tariffs and Withholding: Border Taxes and Informality in Developing Countries" (mimeo, International Monetary Fund).
- _____, and Jenny Ligthart, 2001, "Coordinating Tariff Reductions and Domestic Tax Reform," *Journal of International Economics*, Vol. 56, pp. 407–25.
- _____, and Ben Lockwood, 2006(a), "Is the VAT a Money Machine?" *National Tax Journal*, Vol. LIX, pp.905–28.
- _____ and _____, 2006(b), "The Value Added Tax: Its Causes and Consequences" (mimeo, (Washington: International Monetary Fund).
- _____, and Stephen Smith, 2006, "VAT Fraud and Evasion: What Do We Know and What Can be Done?" *National Tax Journal*, Vol. LIX, pp. 861–87.
- _____ and _____, 1996, "The Future of Value-Added Tax in the European Union," *Economic Policy* Vol. 23, pp. 373–411 and 419–20.
- _____ and _____, 2000, "Viva VIVAT!," *International Tax and Public Finance*, Vol. 7, pp. 741–51.
- Munk, Knud J., 2004, "Tax-Tariff Reform with Costs of Administration," forthcoming in *International Tax and Public Finance*.
- Newbery, David M., 1986, "On the Desirability of Input Taxes," *Economics Letters*, Vol. 20, pp.267–70.
- Piggott, John and John Whalley, 2001, "VAT Base Broadening, Self Supply, and the Informal Sector," *American Economic Review*, Vol. 91, pp.1084–94.
- Ruffles, David, Geoff Tily, David Caplan, and Sandra Tudor, 2003, "VAT Missing Trader Intracommunity Fraud; The Effect on Balance of Payments Statistics and U.K. National Accounts." *Economic Trends* No. 597: pp. 58–70.
- Schneider, Freidrich, 2002, "Size and Measurement of the Informal Economy in 110 Countries Around the World" (Mimeo, University of Linz).
- Sinn, Hans-Werner, Andrea Gebauer, and Rüdiger Parsche, 2004, "The Ifo Institute's Model for Reducing VAT Fraud: Payment First, Refund Later," *CESifo Forum*, February, pp. 30–34.
- Slemrod, Joel, 2006, "Cheating Ourselves: The Economics of Tax Evasion," Forthcoming in the *Journal of Economic Perspectives*.
- Stiglitz, Joseph E., 2003, "Development Oriented Tax Policy," PowerPoint presentation to Congress of International Institute of Public Finance.

Varsano, Ricardo, 2000, "Sub-National Taxation and Treatment of Interstate Trade in Brazil: Problems and a Proposed Solution." In *Decentralization and Accountability of the Public Sector*, edited by Shahid Javed Burki and Guillermo Perry, et al., Proceedings of the Annual Bank Conference on Development in Latin America and the Caribbean, pp. 339–55. (Washington: The World Bank).

Zee, Howell, 2006, "VAT Treatment of Financial Services: A Primer on Conceptual Issues and Country Practices," *Intertax*, Vol. 34, pp. 458–74.