AVOIDING THE PITFALLS IN TAXING FINANCIAL INTERMEDIATION

Patrick Honohan
1. Introduction

Because it keeps systematic accounts and acts as a gatekeeper of liquid resources, the financial sector provides many useful tax ‘handles’ for the fiscal authorities. In many countries, a distorted structure of financial sector taxation has evolved, reflecting both inertia and opportunism. Inertia, in that some very old taxation and administrative practices have survived in the financial sector reflecting the convenience of collection, and the fact that those liable to comply are a small and privileged group of regulated intermediaries who are unlikely to be politically vocal and who can, in any event, pass on much of the tax to their customers. Opportunism, in that a sudden need for budgetary revenue can trigger an increase in reliance on this quick and reliable source.

Growing awareness of the strategic importance of the financial sector in catalyzing economic growth, combined with the increasing global competition in financial services has led to a substantial reconsideration of the domestic financial sector’s potential as a cash-cow for the budget.\(^1\) Indeed, in some instances the pendulum may have swung to the other extreme, with special interests successfully arguing for exaggerated fiscal concessions in the name of improved financial sector performance.

Financial sector taxation is complex both in theory and practice and it needs to evolve in response to financial innovation and wider economic changes.\(^2\) No simple blueprint is proposed or attempted, but the discussion of theoretical, empirical and practical considerations does lead to a number of guidelines for developing what would constitute a good financial sector tax system.

Proposals for financial sector tax reform typically come from one or other of two powerful perspectives. Either the reformer is an enthusiast for a big simplification, usually some form of “flat tax” (including VAT on financial services, zero taxation on capital income, or a universal transactions tax) or she is the advocate of subtle corrective taxation designed to offset some of the many market failures to which the financial sector is prone or to achieve other targeted objectives.

In practice, just like the perennial conflict between simplicity in tax administration and economic efficiency of the tax rates (not to speak of equity), the two perspectives can conflict rather severely. Information and control requirements of much of corrective taxation tend to be poorly accommodated by the big simplifications. This tension remaining unresolved over the years, elements of each approach become embodied in the taxation, explicit and implicit, of the sector. At the same time, the ever-pressing demands of revenue intrude as a further influence on policy design. As a result, the tax systems in most countries often end up as a complex mixture defying any straightforward rationalization. The big flat-tax ideas are diluted and modified, the corrective taxes may misfire by conflicting with others introduced for different reasons.

---

\(^1\)To be sure, the enormous fiscal cost of banking crises means that in many countries the financial system has been a net charge on the budget in recent years.

\(^2\)The survey of recent developments by Levin and Ritter (2003) documents the considerable variety that continues to exist among industrial country systems.
Meanwhile, even as simplification and correction continue their tug-of-war, policy design can all too often neglect the two distinctive traps into which financial sector taxation can fall, namely the sector’s unique capacity for arbitrage and sensitivity to inflation and thus to non-indexed taxes. This paper argues that the practical design of financial sector taxation should be guided by a defensive approach in which proposed taxes are assessed relative to their ability to resist arbitrage and their degree of inherent indexation. Although the defensive approach does not provide an adjudication between simplification and correction it will protect against many of the worst distortions which have been observed.

If a good tax system for the financial system is one which, so far as possible, (a) minimizes the distortions it imposes (for a given amount of revenue collected), especially by causing the formal financial sector to be bypassed through disintermediation to untaxed or differently taxed competitors, (b) is corrective of known distortions (such as those which result from imperfect and asymmetric information, and (c) does not push tax collection from the sector beyond the point where marginal distorting costs exceed those elsewhere in the economy, then the theory and experience described in the remaining chapters of this study do suggest some key practical guidelines.

(i) While reformers should not expect to find a complete and practical solution in any of the ‘big ideas’, each has lessons for a good system.

− Even if practicalities impede its introduction as such, the notion of a VAT on financial services represents a useful benchmark against which existing and proposed indirect taxes can be compared for their burden and impact.
− Significant financial transactions taxes are hard to justify on theoretical grounds and should be resorted to only as a transitory device when fiscal revenue is under particular pressure.
− Heavy emphasis on the taxation of income from capital should be avoided.

(ii) Attempts at corrective taxation should be undertaken with extreme caution as history suggests that unintended side-effects or deadweight losses may dominate the results. This implies that special tax-based schemes to encourage stock exchange listing, household saving and the like, should be viewed with caution, bearing in mind the substantial opportunity cost in terms of lost revenue and the questionable gains.

(iii) While tax shifting is common throughout the economy, the potential for arbitrage is very high in finance. All financial sector taxes need to be designed in as arbitrage-proof a way as possible (our first defensive criterion).

(iv) Inflation generally has a more pervasive effect in finance on the impact of taxation. All financial sector taxes need to be designed to be as inflation-proof as possible (the second defensive criterion).

(v) Approximating taxation of the financial sector to that other sectors is a reasonable goal: the challenge lies in mapping the somewhat distinctive institutions and concepts of
financial intermediation to that of the remainder of the economy, distinguishing between
its role as an intermediary and manager of the funds of others from those of a true
principal. This allows one to map the various tax rates and tax bases affecting financial
firms to the more familiar concepts most closely corresponding: sales taxes, corporate
income tax, collection on account of customer income taxes etc.

The remainder of this paper discusses the background to each of these guidelines,
referring as appropriate to the chapters of a forthcoming edited volume on the topic
(Honohan, 2003b). Section 2 reviews the main forms of tax that are relevant. Section 3
describes the thrust of reform ideas. Section 4 examines the effectiveness of the several
corrective taxes that have been employed. Section 5 highlights the two most distinctive
relevant features of the financial system for designing tax structures. Section 6 comes to
the practical question of tax rates and how to calibrate these for comparability with non-
financial taxes. Section 7 concludes.

2. The main types of explicit and implicit tax
Governments have used financial intermediaries to relieve their budgetary pressures in
three main ways. First, they have applied a variety of explicit taxes, some of which are
common to firms in other sectors of the economy; some of which are special to the
financial sector (such as financial transactions taxes, unremunerated reserve requirements
and deposit insurance premia) and some of which seem similar to those applied to other
sectors, but in practice have a qualitatively different impact even if imposed at the same
nominal rate. Additionally, differential application of mainstream explicit taxation
(including different rates of tax) to financial intermediaries can be important, as in the
treatment of loan-loss provisions in calculating taxable income, or in the application of
sales taxes to interest received by banks. Second, they have imposed reserve
requirements which have had the effect of boosting the net revenue of the central bank
and hence indirectly the government. Third they have made regulations channeling funds
to government or favored sectors and borrowers in ways that involve implicit subsidies,
notably by imposing interest rate ceilings.

Explicit Taxes
Taxes may be levied on many different elements of a financial intermediary’s business.
Net corporate income (profits), gross revenue (interest and fees) and the value of
payments made or received through the intermediary are the most important types.
Interest paid by the intermediary to its creditors are also often taxed, and the intermediary
may be obliged to withhold this tax, thus making only net-of-tax payments to the
creditors. Less commonly, elements of the balance sheet of the financial institution
(assets, liabilities or net capital) could also form tax bases.

Inasmuch as non-financial corporations are also liable to corporate income tax and to a
variety of sales taxes, it is important to identify whether, and in what way, taxation of the
financial intermediary often differs sharply from the standard situation. This can happen
either because the financial intermediary is subject to special rules or rates, or because the
way in which the standard tax is applied has a distinct incidence on financial
intermediaries because of characteristic ways in which their business differs structurally from that of non-financial businesses.

For instance, the total value of payments made and received by a bank (credits and payment to customer accounts) is a large multiple of the total value-added of a bank. Furthermore, the value of payments bears no very stable relationship to the value-added or profits of a bank. As with the value of goods carried by shipping or airline companies, a tax on such payments, even at a low rate, could not be regarded as an approximation to a value-added tax on other companies. The same would be true of taxes levied on securities market transactions.

On the other hand gross interest, insurance premium income and fee receipts in a non-inflationary environment could be of the same order of magnitude (perhaps twice) the value-added – not too far from the experience of many non-financial companies. However, in contrast to these, and unlike net interest, the gross interest is highly sensitive to the nominal level of wholesale interest rates and to expected inflation. In a volatile inflationary environment, this too becomes a rather arbitrary tax base.

The calculation of appropriate reserves against loan losses is an issue for the accounting of any company with receivables or other claims in its balance sheet. But it looms much larger for financial intermediaries, where annual loan-losses even in good years can often be much larger than the profits earned. Therefore the tax treatment of loan-loss provisioning is relatively much more important for financial intermediaries, in that the timing of very sizable tax payments can be at stake.

Reflecting the inertial element in explicit taxation of finance are stamp and registration duties which have a long history in taxation (having been applied to the formal registration of legal documents including those recording transfers of property ownership) and which have their legacy in taxes on payments transaction and transactions in securities exchanges. Modern tax systems depend to a large extent on approximations to a comprehensive income tax or expenditure tax. Stamp duties are poor approximations of either concept.

Withholding taxes on interest paid to depositors and other forms of special treatment of income received by the customers of financial intermediaries can also be distorting, especially when they apply at different rates to different categories of income (such as on local currency, and dollar-denominated deposits).³

Reserve requirements & seigniorage
The inflation tax and related taxes deserve a section by themselves because of their historical importance, the scale of potential revenue and the ease with which they can be collected.

---
³ Differential treatment of taxation of dividends of listed companies can also be seen as an implicit negative tax on the used of formal stock markets.
Requirements that banks should hold a certain fraction of their deposits in the form of liquid reserves whether in cash, at the central bank or at some analogous institution dates at least to the early part of the 19th century and represented initially a convenience to ensure the smooth completion of the daily clearing and to reduce the recourse of banks to central bank borrowing. Unremunerated as reserves placed with the central bank often were (though they would not have had to be to meet the above-mentioned requirements), reserve requirements boosted net income of the central bank, which is usually passed to the fiscal authority as a dividend payment in due course and recognized as a non-tax revenue item in the budget. In this way the banks were implicitly taxed and the budget relieved. The fiscal element was at first not considered especially important, but it became so as bank margins narrowed, especially where nominal interest rates were rising. Some central banks responded by introducing remuneration on required reserves; others tolerated avoidance through substitution by banks of non-reservable categories of instrument.

Nowadays, reserve requirements are generally seen as an extension of the base of seigniorage, inasmuch as substitution of deposits for cash holdings had reduced the base of seigniorage as a tax.

Secondary liquidity reserve requirements were also imposed in several countries, usually to be held in designated government securities, sometimes sold directly to the banks with off-market yields and as such embodying a fairly obvious implicit tax. Such requirements have often also been imposed on insurance companies and other non-bank intermediaries. These types of requirement thus shade into directed credit and interest ceiling arrangements.

**Directed credit and interest ceilings**

Control over where the loanable funds mobilized by the financial system will be applied is in principle a distinct motivation to that driving reserve and liquidity requirements, but it too has a clear fiscal dimension. This kind of mechanism has been operated in nearly all countries over the years and takes many forms. Sometimes there is a requirement to place a special deposit amounting to a specified proportion of the bank’s mobilized resources in the central bank or another public agency charged with onlending these to borrowers in preferred sectors. Sometimes there is a requirement to lend a certain fraction of the bank’s resources to specified sectors, or failing that, to deposit an equivalent amount with a specialized bank that can do the lending. Whether or not there is an explicit interest rate ceiling on these sectoral requirements, the diversion of funds has the effect of lowering the market-clearing rate for them and this will act as if there were a tax on the interest income from this part of the lending (partly compensated by a higher market-clearing rate on non-favored sectors. Except where the government is the borrower, the benefit of this tax does not directly go to it, but it is appropriate to see the budget as a hidden beneficiary, in that, absent the directed credit, subsidization of the preferred borrowers would have to have been done through other means, including direct budgetary allocations.
Systemwide interest rate ceilings, much rarer now than in the past, and capital controls having the effect of lowering local interest rates and this too can be seen as a tax affecting financial intermediation. The government’s budget is almost always the largest single borrower, and as such the biggest direct beneficiary of system-wide interest ceilings and their equivalents.

3. The big reform ideas

One general approach to financial sector taxation is to attempt a great simplification on the theory that low rates and a wide base with few exemptions is likely to generate relatively low distortions. This approach holds out the prospect not only of minimizing the incentive for complex schemes of financial engineering designed to avoid tax, but also of making such schemes relatively difficult to develop.

The three main handles for taxation: income, expenditure and transactions, have each been the subject of prominent and extensively discussed grand and simple schemes. These are, (i) the proposition that capital income should not be taxed at all, (ii) the proposal that value-added by the financial services industry should be subject to a uniform tax and (iii) the idea that a tax on all financial transactions at a very low rate could generate very large revenues with negligible distortion. We consider these one by one.

Capital income – should it be taxed at all?
The underlying basis for the argument that it might be optimal not to tax income from capital at all is the insight that this involves a form of double taxation on future consumption. By shifting the perspective from the statutory base of the tax – capital income – to a variable more closely relevant to economic policy, namely utility based on household consumption, this economic analysis of capital taxation shows that a constant nominal or statutory tax rate on capital income implies an effective rate on consumption that may increase without bound for consumption far into the future. Because future consumption depends on the reinvestment of after-tax capital income, the more remote is the date of future consumption, the higher the effective tax rate; and this effective tax rate may increase without bound. Evidently, optimal tax policy can improve on a situation with infinitely high effective tax rates; accordingly, this reasoning points to the optimality of capital income taxation converging to zero (cf. Boadway and Keen, 2003).

Many subtle qualifications can be made to the implicit models of utility, income and consumption underlying this analysis, and the precise prescription for zero taxation is not very robust, yet it retains some force and serves as an important counterweight to proposals for high rates of capital income taxation designed to achieve other goals. One such goal is that of ensuring the socially optimal rate of national saving (since private markets cannot generally be relied upon to do this and may result in oversaving). Another is redistribution. Yet even if households differ in their wage-earning capacity and tax policy is being used for redistributional goals, these can best be achieved by a tax on wage income alone – at least in simple models of intertemporal preferences. Once again the use of capital income taxation would be suboptimal because of the compound interest effect.
If income from capital is not to be taxed, then it might seem to follow that the income of financial intermediaries ought not to be taxed either. But in practice some corporate income – perhaps a large portion – represents pure profit or economic rent. Neglected in the models that generate the no-capital-income-tax result, pure profit may be taxed without distortion, and this argument is another important qualification. Where financial markets are uncompetitive – and the scale economies that are involved in parts of finance make this relevant, especially in financially closed economies – this could be an empirically important factor.4

A stronger line of attack on the zero capital income tax proposition comes from practical issues of enforcement and informational deficiencies. If capital income goes completely untaxed, this may provide an easy loophole for high earning households to camouflage their earnings by transforming or laundering them into capital income. A tax on capital income may be an important practical expedient to close such loopholes.5 If so, withholding the tax at source, or taxing corporate income as a form of implicit withholding may further help to overcome the tax authorities’ informational disadvantage and administrative collection costs. However, these considerations tend to be swamped in many developing countries by the need to consider the impact on foreign-owned firms. Given the fact that many capital exporting countries allow credit to their residents for tax paid in the host country, this can pave the way for host countries to tax foreign-owned companies in the knowledge that they will not be discouraged from investing to the extent that the tax paid to the host country simply reduces their home country tax.

The elegant simplicity of the theoretical argument against capital income tax thus ultimately fails, though it points to a need to justify such taxation – and the taxation of the income of financial and other companies – on grounds other than those of simple consistency with taxation of wage income.

Taxing financial services: can a VAT work?
About 70 per cent6 of the world’s population live in countries with a VAT and the tax is a key source of government revenue in more than 120 nations (Ebrill et al, 2001). So if a VAT is the way forward for the bulk of (indirect) taxation on expenditure, to what extent should it be the model for financial services also?

---

4 Caminal (2003) explores the implications for tax incidence of market power in banking. As he and others have noted, though, leaving banks with some untaxed economic rent (or franchise value as it tends to be called in the banking literature), can reduce the propensity, potentially strong among insured banks, to assume socially excessive risks (Stiglitz, 1994, Caprio and Summers, 1996). Indeed it has been argued that the credibility of banks as delegated monitors for the depositors depends on their being able to earn profits in most states of the world (Diamond, 1996).

5 Differentiating the rate of withholding tax as between income from high risk (equity) and low risk (debt, deposits) assets could help achieve progressivity even absent information on the income of the recipients, assuming diminishing risk aversion with wealth (Gordon, 2000).

6 The largest countries, by population, without a VAT are India, United States, Iran, Ethiopia, Congo DR, Myanmar, Afghanistan, North Korea, Iraq and Malaysia.
The first observation has to be that in practice, most financial services are “exempt” in virtually all countries employing a VAT. This does not mean that these financial services wholly escape the VAT, as the status of “exempt” does not allow financial service providers to recover VAT paid by their taxable suppliers and built into the price of their inputs. Indeed, taxable firms who use financial services as inputs cannot recover the VAT paid by the suppliers of financial service firms either, with the result that there is tax “cascading”. But value which has been added by the exempt financial sector firms is not captured in the tax. Whether aggregate tax receipts would increase or fall if the exemption were removed is an unresolved empirical issue (which depends not only on the degree to which financial services are used by tax-liable firms, but also on the different rates of VAT that may be in effect.

The exemption of most financial services from VAT appears to be a historical inheritance without much political or economic rationale. While it is not difficult to measure the value-added of a bank (profits plus wages), there is the practical difficulty of deciding how much credit taxable firms which use financial services would be entitled to claim, seeing that the charge for many financial services is an implicit one bundled with others in, for example, the spread between deposit and lending rates. Determining how much of the spread should be attributed to depositor services and how much to borrower services is not straightforward. Thus it is not obvious how much credit each should receive for VAT already paid on inputs.

Yet it is not impossible to devise simple rules of thumb which can provide a reasonable approximation. Thus, for example, the cash flow method where VAT is paid on all net cash receipts (including capital amounts), could be adequate in a static environment. However, start-up problems and treatment of risk may not be adequately resolved by this method, and changing tax rates also presents difficulties for the approach. A variant of the cash-flow method, using suspense accounts and an accounting rate of interest to bring transactions at different dates to a common standard, could help ease the transition problems and has been shown to be workable by detailed pilot studies in the EU (Poddar, 2003).

The lack of any clear potential revenue gain, and fears about the practical complexity and possible hidden distortions or loopholes, have inhibited any significant move to bringing financial services into the VAT net.\(^7\) The resulting distortions are quite serious in some cases. For one thing there is a clear incentive to self-supply inputs. Second, there are distortions at the margin, with financial services such as factoring, which can represent a particularly effective form of lending to SMEs – low cost and low risk – severely tax-disadvantaged by falling within the VAT net in many jurisdictions for which other forms of lending are exempt.

---

\(^7\) Though a few countries have introduced substitute taxes based on applying a rate to the estimated value-added of banks obtained by summing the wage and profits.
The grand simplification offered by the VAT thus fails, not on theoretical grounds, but on the grounds on administrative and practical difficulties or uncertainties. Nevertheless, it does point in the direction of what might be desirable for substitute indirect taxes.8

Transactions taxes: panacea or Pandora’s box?
Because of their loose connection with consumption and utility, and their potential for generating significant distortions in the organization of production and distribution, transactions taxes (including trade taxes) have lost favor as a tool of general tax policy over the years relative to income and expenditure taxes. But the vast scale of financial sector transactions has presented itself to some scholars and some governments as a convenient base for rapidly generating substantial revenue.

There is a paradox here in that critics of transactions taxes point to the potentially seriously distortions that is causes, while advocates argue that, because of the large base, very sizable revenues can be realized with low nominal tax rates. To the extent that the deadweight cost of a tax is often supposed to be proportional to the square of the tax rate, introducing a low-rate financial transactions tax in order to allow a reduction in the much higher rates of labor income or other taxes might be supposed to reduce total deadweight in the tax system as a whole.

At the most extreme, a recent proposal suggests that what seems at first sight to be an administratively trivial and quantitatively tiny 0.15% rate of tax on all automated payments could raise enough revenue (in the United States) to replace the entire existing tax system (Feige, 2000). Feige shows that existing automated payments amounted (in 1996) to somewhere in the region of US$300-500 trillion, or of the order of 50 times the value of GDP. How, he asks, could anyone argue that a tax rate of 0.15%, even applied to such a large base, be considered seriously distorting by comparison with the existing tax regime?

Analysis of the payments that would be affected reveal that about 85% relate to financial transactions (purchase or sale of stocks, bonds, foreign exchange or other money changing transactions, etc.). To a large extent, then, the initial burden of an universal payments tax would fall on the financial sector.

Of course, if we proceed (as before with the capital income tax) to transform our perspective from the statutory or nominal base to the more economically relevant concept of consumption, we see that the average good or service in the typical consumption bundle must be ‘hit’ by the tax not once, but dozens of times, as it works its way through financing, design, production and distribution.

8 Compare the Chinese “business tax,” which has recently applied at the rate of 8 percent to the income from a bank’s trading portfolio, but is not applied to the income from the bank’s investment portfolio. This strongly discourages trading by banks in government securities, thereby sharply diminishing liquidity in that market. A VAT-like tax would not create such a distortion. (A further distortion in China comes from the exemption from income tax of interest from government bonds, but the inclusion of capital gains from trading such bonds at a rate of 33 percent. Such problems are avoided in some other countries by treating capital gains from trading activities of financial intermediaries on par with income for the purposes of direct taxation.)
Criticisms of this proposal fall into three main groups.

First that the tax would not collect as much revenue due to the sizable elasticities involved. Financial sector transactions in particular would be arbitrated in such a way as to drastically reduce the number of recorded transactions. What are now sequences of linked transactions carried out for little more than book-keeping convenience at negligible cost would be collapsed into a single more complex transaction. Portfolio readjustments would be made with reduced frequency without substantially altering expected return and risk. Microeconomic studies of the precise mechanisms that are at work to generate gross transactions of such a high multiple of GDP in wholesale financial markets are not plentiful (but see Lyons, 2001, for the foreign exchange market) so that reliable estimates of these effects are not yet available. Furthermore, the scope for avoiding such a tax through offshore financial transactions has to be taken seriously.

The second main objection is that, even if the tax did collect the expected revenue, the distortion costs would not necessarily be any smaller than with the existing system. This objection relies either on (a) the observation that the financial system would bear the main brunt, and as such that the tax would in fact be more concentrated, not less; or (b) the observation that, in terms of final consumption, the tax would effectively cascade to cumulative rates comparable to those observed at present.

No country has seriously considered replacing its tax system with an universal payments tax, but there are numerous examples of partial transactions taxes, applying for example to bank debits or to securities transactions. Bank debit taxes introduced in half a dozen Latin American countries in the past 15 years or so in a bid to raise revenue have been successful in that goal, at least for a while, with revenues ranging from about ½% of GDP to as much as 3½% in one case for one year. It is fair to say that revenue from these taxes held up unexpectedly well over 3-4 years. That revenue would fall off after the first year was predicted by many, and it did occur on average, though the effect did not prove to be statistically significant in regression of the available data. Nevertheless, many of the schemes had to be adapted administratively in the course of their operation, to exempt some transactions that would otherwise have been too distorting (and probably also to capture others that had escaped the net). The distortions of these and of securities transactions taxes have been discussed in the literature: they certainly are distorting, yet applied in moderation, these transactions taxes have been less distorting than many observers expected (Kirilenko and Summers, 2003; Habermeier and Kirilenko, 2003).

---

9 This consideration needs to be kept in mind by those who would see the proposal as socially progressive in that payments in which they are directly or indirectly involved likely represent a much higher multiple of the income of prosperous people than of the poor. After all, if such a tax did not raise the hoped-for revenue, the consequence might have to be cutbacks in public services which disproportionately benefit the poor.

10 Tobin taxes are much more focused and do not typically have revenue as the main objective, but instead are seen as corrective taxes intended to reduce volatile speculative capital flows. They have generated an enormous literature and I am not going to summarize that here.
Thus, despite expectations that they would not only distort financial markets and drive out capital, but would quickly lose their revenue-raising ability, such transactions taxes have been surprisingly resilient. But they are far from being a panacea, and indeed have little to recommend them beyond their ability to deliver revenue speedily and with low direct administrative costs.

4. Corrective taxes

It is not just taxation that distorts financial markets. Information deficiencies, monopoly power and other factors push most financial markets away from the ideal of the atomistic market with fully informed participants competing on a level basis. Under these circumstances, the non-revenue side-effects of taxes and tax-like measures can be turned to advantage and form part of the corrective policy structure in this area.

Indeed, many measures of this type may have regulation and market efficiency as their primary objective, with revenue seen as a side-effect. But as we will see, the effectiveness of many such measures in their supposedly corrective role has been challenged and remains controversial.

Deposit insurance: supposedly helps stability, but maybe outweighed by the induced moral hazard

The most complex and contentious of these debated corrective quasi-taxes is deposit insurance.

That it is a tax is fairly clear from the contributions or levies that are generally imposed on participating banks, especially given that these are typically compulsory and that the rate of tax usually bears at best an imperfect relation to the “fair premium”.

Indeed, the anticipated gross revenue from the levy is typically small and in many cases is calculated to be insufficient to cover even the expected pay-out costs as calculated using option-pricing formulae (Laeven, 2002). Furthermore the probability distribution of net payout costs is severely skewed: systemic banking crises entailing fiscal costs of up to 50% of a year’s GDP are never matched by a corresponding deposit insurance fund accumulation in lucky, crisis-free, countries.

For many advocates, the perceived corrective role of deposit insurance is essentially one of reducing the likelihood of a depositor panic. By protecting depositors against the risk that their deposits will be unpaid if a bank proves to be insolvent, it is hoped that a self-fulfilling panic, including contagion to other banks triggered by the insolvency of one bank can be avoided. On the other hand, by lowering the vigilance of potentially informed depositors, the moral hazard of heightened risk-taking by the bankers,

---

11 Not always explicitly accounted for, as when unremunerated reserve requirements augment the central banks net revenue but are nowhere accounted for explicitly as a revenue source.
12 Even the relatively much smaller fiscal costs of the US banking crises of the 1980s were more than enough to empty the insurance funds.
13 Protection of the small depositor is another goal. This is quite a distinct role, of course, as runs by small depositors only do not threaten systemic liquidity.
unpunished by market discipline, could in theory result in heightened risk to the system as a whole.

Although early deposit insurance schemes entailed a uniform insurance premium per dollar of deposit, there have been moves in several countries to differentiate the rate of premium in accordance with some measure of the perceived riskiness of the participating bank’s portfolio. This dimension of such taxes is designed to reduce the moral hazard potential but it depends to some extent on the information available to the deposit insurer as to the accuracy of the ex ante risk assessment (Honohan and Stiglitz, 2001). About a quarter of schemes have some risk-differentiation, but the differentials are small and are not always systematically imposed (Demirgüç-Kunt and Tobaci, 2001).14

Econometric estimates of how financial system performance varies across countries with the existence and characteristics of deposit insurance systems suggest that, in countries whose socio-political institutions are generally rated as strong need not fear that the moral hazard side-effect will outweigh other beneficial effects. Although deposit insurance weakens market discipline even in such countries, the effects seem to be offset by better official oversight. However, for countries with less well-developed institutions (along the dimensions of rule of law, governance and corruption), the establishment of a formal deposit insurance scheme15 does appear to present a heightened risk of crisis (Demirgüç-Kunt and Detragiache, 2000; Demirgüç-Kunt and Kane, 2002) and does not even promote deposit growth (Cull, Senbet and Sorge, 2000). Having risk-based deposit insurance premia does not appear to mitigate the systemic risk, so that the potential for introducing a corrective structure of the deposit insurance tax may be limited.

Deposit insurance, with or without risk-based premia, may not be a very effective corrective mechanism. It clearly needs to be supplemented in this role by strong administrative or other controls, including supervision of minimum capitalization ratios (Brock, 2003). And it may interact with other taxes. For instance, a tax on bank gross receipts16 will reduce the expected after-tax return to a risky investment, though Brock shows that there would be some offset to this inasmuch as the government (deposit insurer) is coinsuring the risk to a greater extent in the presence of such a tax. On the other hand, he also shows that a marginal reserve requirement (see below) could be more likely to reduce the moral hazard effect on bank risk-taking behavior. All in all, though, the uncertain strength and reliability of such effects argue for blunter and more reliable

---

14 Some examples: the US premia currently vary according to two criteria, capitalization and supervisory assessment, from zero (for a well-capitalized bank that is highly rated by the supervisors) to 0.27 per cent of deposits (for an undercapitalized bank which is seen by supervisors as posing a substantial probability of loss to the insurer unless corrective action is taken). Argentina has charged a basic rate of 0.36 per cent subject to being doubled where banks were paying high interest rates for deposits (and those with very high rates were not covered at all). Cameroon and other francophone African countries impose 0.15 per cent plus 0.5 per cent of net non-performing loans. Other risk-based formulations, including ex post assessments are levied in other countries.

15 This, despite the consideration that a degree of implicit protection may be assumed by depositors even when no formal scheme exists.

16 Such as has been in effect in China at a rate of 8 per cent (7 per cent since 2000), cf. Langlois, 2001 (though he rather overstates the evils of this tax).
instruments in restraining bank risk-taking, a matter which lies beyond the scope of the present exercise.

_Provisioning and capital adequacy_

The amount of loan-loss provisioning which is allowable to banks as a deduction against income for tax purposes can be a very significant factor in arriving at the net tax liability, -- often sufficient to shelter the entire tax bill. By the same token, this can be a matter of considerable revenue significance for the authorities. But it has long been acknowledged that there is a potential corrective role for the treatment of loan-loss provisions. This argument hinges on the arbitrariness which inevitably exists in arriving at a reasonable provision that would result in the banks’ accounts representing a true and fair picture of the business. If the fiscal rules have the effect of biasing company accounting, this could be damaging for the transparency of the financial system and for good decisions on risk management. Recent accounting scandals have focused attention on the difficulty of seeing through valuation procedures used in non-financial company reporting procedures and bank accounts can be arguably even less clearcut especially in times of economic turbulence or change.

To the extent that equity capital represents a cushion protecting depositors and other claimants against the consequences of a decline in the value of the bank’s loan portfolio and other assets, the equity holders (and the directors to the extent that they are acting as the equity holders’ agents) of a lightly capitalized bank at risk of failure will have an incentive to minimize the amount of capital which they have truly at risk (thereby transferring risk to other claimants), provided they can do this without inducing an increase in the required return on their other liabilities. If the fiscal authority disallows the deductibility of reasonable loan-loss provisions (for example provisions that can be justified on the basis of a reasonable objective forecasting model), that reinforces the incentive to understate provisions and thereby to overstate capital, potentially misleading regulators and the market.

On the other hand, a well-capitalized bank may be more attracted by the advantages of advancing tax deductibility, and may use the range of uncertainty to increase loan-loss provisioning thereby reducing revenue.

Balancing the pressures of revenue needs with the risk of losing transparency is thus a constant tug-of-war and different countries adopt different rules (Laurin and Majnoni, 2003; Sunley, 2003). A move away from mechanical rules (such as disallowing general provisions but allowing specific provisions) towards a more realistic, forward-looking accounting that allows predictable but not yet identified losses to be adequately provisioned, so long as these are accepted by the institutional regulator, would seem to be the preferred goal here.\(^{17}\)

---

\(^{17}\) Various formulations are possible. One is to have the tax authorities accept provisioning that had been agreed by the prudential regulators. A tougher approach would be to require the bank to defend its provisioning on the basis of an objective forecasting method – though it needs to be recognized that such forecasts are often more an art than a science.
Promoting Saving

A very widespread explicit goal of corrective tax measures affecting the financial sector is the promotion of saving. The goal is driven partly by fiscal needs, in an attempt to ease the financing of government deficits, partly by a perception (colored by an earlier generation of macroeconomic theories and, because of new research findings, no longer generally accepted by economists) that aggregate economic growth is, in the long-run, driven by national saving and partly by a desire to ensure that households do not undersave, particularly for retirement, but also for housing and education.\(^{18}\)

In practice, such measures tend not to affect all savings media equally, hence their sometimes substantial impact on the structure and performance of the financial system, which, in certain cases at least, can far outweigh the net impact of the policy on the goal of increasing household saving (OECD, 1994, Honohan, 1996).

For practical reasons, measures that operate by modifying income tax schedules tend to be relevant only in middle-income countries, or at least in countries which have achieved a certain minimum level of the effectiveness of the income tax system. Furthermore, there is widespread skepticism among experts as to the effectiveness of mandatory saving for housing in achieving the goal of improving access to housing for the targeted low-income groups. On the other hand mandatory retirement saving programs to appear to increase national saving by a significant amount on average, especially perhaps where they are tax-advantaged (Jappelli-Pistaferri, 2003)

Other dimensions of corrective financial taxation

In other cases, supposedly corrective financial sector taxation comes more in the form of a vague and unthinking encouragement of what are seen as social “goods”. This is not unique to the financial sector: finance ministers are typically bombarded with proposals to exempt from taxation items or activities thought to be meritorious. Except where tax relief appears to be the most effective way of correcting some market distortion that is resulting in an undersupply of the item or activity in question, the ministers are usually advised to resist such special pleading. But lobbying of this type does appear to be notably successful in finance. For example, consistent with the observation that most countries feel that their financial system is unduly bank-dominated, there is constant advocacy of tax concessions targeted at companies with a stock exchange listing.\(^{19}\) This is at best a crude instrument, especially if the underlying reason for the underdevelopment of the stock exchange lies in an insufficiently developed information and legal infrastructure, as is often the case. Much better to direct policy attention to correcting these infrastructural deficiencies.

\(^{18}\) Tax incentives are also widely used to favor health and life insurance (Jappelli-Pistaferri, 2002).

\(^{19}\) In Egypt, very favorable tax treatment of listed firms from the 1980s induced a widespread listing of firms on the Cairo and Alexandria Stock Exchange, to which the lenient listing requirements provided no discouragement. Over 500 companies were listed by 1990; ten years later, the figure had grown to more than 1,000. However, most of the firms were closely held and the free-float of shares is little more than one-tenth of the total capitalization. The market capitalization of the actively traded free-float was only about 1 percent of GDP. In this case the supposedly corrective tax exemption failed to achieve the presumed objective of a deep and liquid equity market. Recent listing rule reforms are expected to dramatically cut the number of listed firms in 2003.
Another much used quasi-tax often thought of as, in a sense, corrective, is the unremunerated reserve requirement. The sense in which this might have been thought of as corrective is that it provides a lever on which monetary policy can operate. Actually, as is now acknowledged by authorities on monetary policy, the perceived need for unremunerated reserve requirements was based on a misconception. Monetary policy does not require unremunerated reserve requirements or any other quasi-tax for its effectiveness. (cf. Brock, 2003)

5. Two distinctive elements for financial sector tax design

If there are two key features of the financial sector which distinguish it from other sectors when it comes to designing taxation, these must surely be the system’s capacity for arbitrage and its sensitivity to inflation and thus to non-indexed taxes.

The system’s capacity for arbitrage

Whether mainly flat or mainly corrective, the impact in practice of most financial sector taxes depends crucially on the extent to which they have been constructed in such a way as to be insulated from the high elasticities that prevail in the sector. Arbitrage between functionally equivalent contracts or institutional forms bedevils tax design in this area.

Incidence-shifting of bank taxes

Because of substitutability and the possibility of arbitrage and near-arbitrage, the full incidence of taxation imposed on one component of the intermediation process (deposits, loans, intermediary profits), may very well be fully shifted to another component. Ramon Caminal has recently developed a formal model of intermediation, taking account of the provision of liquidity as well as intermediation services by banks in order to examine the influence of various bank taxes on volumes and cost of intermediation, those provided to depositors by banks. Several striking results are obtained. For instance, the ability of at least some borrowers to substitute alternative sources of funding implies a tendency for the imposition of a VAT on banking services to be passed back to depositors.20 Furthermore, the conditions under which a tax on bank loans falls not on the cost of funds, but instead on the return to bank shareholders are also plausible, including a range of assumptions on competitive conditions. (However, if regulatory capital requirements are likely to be binding in the sense that banks hold more capital than they would freely choose to, a tax on banks’ profits may in contrast fall wholly on lending interest rates). In contrast to general models of production, then, plausible modeling of the degree of substitutability in banking involves such high elasticities that predicting the incidence of a tax to fall wholly on a class of agents not directly the subject of the taxation can be plausibly predicted. On the other hand, recognizing that the services provided to savers by investment funds may be highly substitutable for some of the services obtained from bank deposits, Caminal has also shown how, under reasonable circumstances, the presence of untaxed investment funds implies that taxation of deposits...

---

20 At least under the plausible assumption that the marginal borrower is VAT-liable while they marginal depositor is not (cf. Caminal, 2003).
will affect only the monitoring and transaction service provision by banks, and not the provision of liquidity.

These contrasting cases suggest the heightened risks involved in imposing taxes under the assumption that the taxpayer who is liable will be the one incurring the incidence of the tax. Just what the incidence will be can be worked out in theoretical cases (to a greater extent than is the case for taxes on non-financial sectors), though the task of matching these theoretical cases to the real world is a striking challenge for the empirical policy analyst given the difficulty of estimating many of the relevant behavioral relationships, as is evident from their relative absence from the literature, even for industrial countries.

Along with the shifted incidence can be a very large behavioral effect. This may not be socially costly in equilibrium (if the substitute truly is functionally equivalent) but short-term disruption and costly incurring of new sunk capital to support the substitute activity could be quite severe.

New financial instruments
At the heart of financial innovation is, in the words of Boadway and Keen (2003), the creation of new instruments by repackaging the cash-flows generated by others. Arbitrage is here the mechanism, not just an outcome. The reasons for this repackaging are manifold – to better align the instruments with the liquidity and maturity preferences of different classes of investors, to shift particular risks between investors who have different appetites for them, whether based on information or on correlations with the remainder of their portfolio. If the rebundled instruments are differently treated by taxation, this can block the repackaging and inhibit the risk-sharing that is involved.21 Furthermore, of course, differential tax treatment (for example of debt and equity, or of income and capital) can be a powerful driver of innovation designed for no better reason than to repackage cash flows into a less heavily taxed form.

Boadway and Keen note that many of these issues have been dealt with on a piecemeal and ad hoc way by tax authorities in advanced economies. Theoreticians have been exploring ways of rationalizing the taxation of new financial instruments, both by devising unambiguous decompositions of the instruments into fundamental components, and by determining the timing (accrual versus realization) at which the taxable amounts are crystallized. But no general agreement among theoreticians, let alone practitioners in advanced economies, has yet emerged. This rules out, for the present, the possibility of developing country tax authorities’ piggy-backing on a pre-packaged solution. Indeed, for market participants, the tax situation is even less satisfactory in developing countries where the likely tax treatment of new instruments is often undetermined or disputed.

Sensitivity to inflation
Although inflation has pervasive effects throughout the economy and in particular has been shown to be negatively correlated with growth, at least for sufficiently high rates, it is evident that banking and other parts of the financial sector which extensively employ

---

21 For example, the existence of withholding taxes on gross interest receipts can stifle the market in interest rate swaps.
nominal financial contracts can be more directly and deeply affected than most. High and variable rates of inflation induce significant substitution away from non-interest-bearing monetary assets in favor of assets offering higher real returns and inflation hedges. This can, on the one hand, shrink the size of the banking system’s intermediation. But, on the other hand, the financial system’s capacity to provide the instruments to insulate economic agents from the inflation will tend to expand this side of its activities. Indeed, empirically, the balance-sheet size of the banking system is found to shrink with inflation, whereas inflation is found to be positively associated with profitability and the value-added of the banking system (Honohan, 2003a).

Inflation also has a strong influence on the government’s finances, and, the term “inflation tax” is well chosen, even though there is no perfect correspondence between the implicit inflation tax rate as measured by the opportunity cost of holding interest-free base money (which will be related to the expected inflation rate), and the flow of financing to the budget from money creation (Honohan, 1996).

The interaction between inflation and a non-indexed tax system can have sizable and unexpected effects even in a country with single digit inflation (Feldstein, 1983, 1999). As inflation increases, the double distortions of inflation and taxation can be multiplicative rather than additive, with severe consequences. For financial sector firms, the impact of inflation on the scale and activity of financial services firms needs to be considered alongside its impact on their tax-inclusive cost structures. The effective tax rate of several commonly employed financial sector taxes, such as taxes on gross interest receipts of banks, or unremunerated reserve requirements rise almost in proportion to the rate of inflation. And in the case of nominal interest rate ceilings the effective rate of tax rises more than in proportion to the rate of inflation. Given that inflation rates can be high, volatile and unplanned, this degree of sensitivity to inflation in the effective rate of tax is generally quite undesirable (Honohan, 2003a).

Good tax policy needs to take account of these sensitivities. A defensive approach is needed which ensures that the tax structure is not vulnerable to severe distortion along these lines.

6. Calibrating different types of tax

Where a defensive approach has not been followed, poorly constructed tax systems – whether the consequence of a drive for revenue, or of misdirected sophistication – have often had sizable unexpected side-effects.

Part of the problem in many difficult cases has been that the financial sector taxes and implicit or quasi-taxes have not been seen for what they are. Thus very high effective tax rates have emerged in cases where legislators would not have conceived of imposing comparable nominal tax rates.22

---

22 The case of Russia in the 1990s, described by Granville (2003), is a cautionary example.
On the other hand, lobbyists are prone to finding ways of exaggerating the tax burden on financial intermediaries by adding-up taxes which touch the sector only slightly and expressing these as a percentage of the sector’s profits.

Is there some simple way of approximating the burden of a given tax, or better the impact of reform in a particular tax? This section looks at how this question might be addressed in respect of the main types of tax or quasi-tax which most often raise such questions.

The relevant taxes include: (i) unremunerated reserve requirements (ii) tax on intermediary interest receipts; (iii) withholding tax on interest payments by intermediaries; (iv) stamp tax on bank debits; (v) stamp tax on bank loans.

One practical approach to calibrating these taxes and judging their appropriateness is to map each tax into its closest non-financial analog. Thus one decides whether the tax is more nearly an income or a sales tax. If an income tax, is it more a tax on the intermediary’s shareholders or on the intermediary’s fund-providing customers? If a sales tax, what is the product that is being taxed and what is its net-of-tax price?

As with most issues of incidence, these questions cannot always easily be answered. Nevertheless, even an approximate answer can clarify the issues significantly.

Market power and substitution possibilities are central. In many countries, the market power of banks is being eroded, both by international competition for depositor services and from alternative sources of industrial funding as well as by liberalization of entry. Taxes and quasi-taxes that might hitherto have been assumed to fall on the shareholders of banks in a manner analogous to an income tax may now be more likely to be passed on to those customers who have few alternatives, notably small borrowers whose creditworthiness is costly to determine. (Caminal, 2003, models these issues in some detail and Cardoso, 2003, presents interesting evidence that pass-through has been very high in Brazil).

Under such conditions, the taxes described fall into three groups: those that are best seen as a tax on lending services, those on transactions services and income taxes on suppliers of funds.

Both unremunerated reserve requirements imposed on banks and special taxes on interest receipts of banks are best seen (under these circumstances) as similar to sales taxes on the provision of lending services (e.g. credit appraisal and monitoring) to small borrowers. The effective tax rate can be approximated by comparing the tax paid (or, in the case of unremunerated reserve requirements, the opportunity cost of the reserved funds) per dollar lent to the net of tax cost of the service. High effective tax rates often result. Official estimates for Brazil in 2001 can be read, in this perspective, as implying an 85% effective tax rate on average for lending (Cardoso, 2003). Furthermore, because the tax

\[ \lambda = \frac{\varphi \theta}{1-\theta} \]

\[ \lambda \text{ or reserves remunerated below market rate. A very simple break-even calculation implies that an addition of } \lambda \text{ to the loan interest rate will be required to recover an interest penalty of } \varphi \text{ applied to reserve requirements of } \theta \text{ where } \lambda = \frac{\varphi \theta}{1-\theta}. \]
base – the cost of intermediation services – is not sensitive to the nominal rate of interest, whereas the tax paid is, the resulting effective rate can be very sensitive to the nominal rate of interest and thus to the rate of inflation (Honohan, 2002, 2003a).

The stamp duty on bank loans, typically proportional to the loan size but not to its maturity can be analyzed in much the same way. In this case the effective tax rate may increase sharply as maturities shorten, allowing the methodology to reveal the obvious technical deficiency in such a tax.\(^{24}\)

Transactions taxes and the stamp tax on cheques likely fall mainly on the user of the transactions involved. The relevant tax rate is thus computed as if it were a sales tax on the relevant service. This allows us to compute the stamp tax on cheques in Chile in 2002 as a sales tax rate of 100\%, compared with a rate of about 20\% for the similar tax then imposed in Ireland.

Judging the appropriate treatment of the withholding of income tax on deposit interest requires careful consideration of the effectiveness of the remainder of income tax. If income tax on the revenue from competing capital assets is collected effectively, then the fact that tax due on deposit interest is withheld at source can best be thought of as chiefly an administrative convenience, rather than as an additional imposition affecting the withholding intermediaries and their other customers. The empirical judgment here will often depend crucially on the degree of international capital mobility (cf. Huizinga and Nicodeme, 2001)

7. Concluding remarks

It would be hard to justify a dogmatic approach to reforming financial sector taxation on the basis of the review here presented. None of the extreme blueprints that have been proposed for the optimal structure of financial sector taxation can be fully endorsed.

But there are important lessons. Both empirical experience and theoretical propositions suggest that some financial sector taxation can have unanticipatedly large and damaging effects. Therefore the main message proposed here is one of moderation. There is no reason why the financial sector should not pay its share of needed tax revenue, but design should be defensive in the sense of guarding against the two major vulnerabilities to which the sector is prone: sensitivity to arbitrage and sensitivity to inflation.

Starting with the overall question of capital income tax, the case for exempting capital income from all taxation (the first “big idea” floating in the field) is too narrow to justify action, and all that we would propose in this regard is to take the literature as a caution against excessively high capital income tax rates, both in absolute terms and in comparison with income tax. Attempting to “soak the rich” through a high capital income tax rate may be an unwise way to proceed.

\(^{24}\) In Egypt, the application of a constant stamp tax independent of loan maturity hampered the development of short-term bridging finance.
The second big idea, that of extending VAT to financial services, seems much better-based. It certainly seems a more desirable goal than attempts to rely heavily on transactions taxes (whose distorting effects must soon dominate even if they have shown their rough and ready merits as transitory revenue-spinners in times of fiscal pressure).

Yet here too excessive enthusiasm should be tempered. The VAT might be the way to go in a world where administrative collection costs were low, but there may be much less to gain from it than appears at first sight and pioneering its adoption could be a diversion for a developing economy unless the authorities are sure of their administrative capacity to be pioneers in the area. For most, the VAT should best, perhaps, be seen as a benchmark against which to compare other indirect taxes in regard to their neutrality and their burden (effective tax rate equivalent on relevant value added).

Given the propensity for inflation rates to be high and volatile, taxes and quasi-taxes which are super-sensitive to inflation should be avoided. Taxes with such features should be replaced with others that are more inflation-neutral.

Above all, each tax proposal should be considered in terms of the possible substitutes which exist for the activity, asset, return or transaction being taxed and which could both magnify the behavioral effect of the tax and place its incidence far from where it was originally expected. This must be the key defensive mechanism.

And, in making judgments on the overall burden of taxation on financial intermediation, which inevitably requires some assessment of incidence, particular efforts need to be made to isolate those taxes that, it can reasonably be assumed, are passed through to depositors. Income tax on deposit interest collected at source, for example, is not borne by the interest payer if income tax on other sources of capital income are also efficiently collected.
References


