The Distortionary Costs of Taxation

John Creedy
The University of Melbourne

Abstract

This paper provides a non-technical review of concepts, measurement problems and estimates of the efficiency costs of direct and indirect taxation. Emphasis is placed on work most closely related to New Zealand.

1 Introduction

Direct and indirect taxes distort individuals' behaviour.[1] This means that there is an intangible efficiency cost, or distortionary cost, imposed on individuals in addition to the direct burden of a tax. The direct burden is measured simply by the amount of tax paid. The 'excess burden', or distortionary cost, of a tax is thus the extent to which the total money measure of the welfare loss from the tax exceeds the tax paid. Taxes may be imposed deliberately to influence expenditure patterns (as with 'sumptuary' taxes and those imposed for externality reasons), or to achieve redistribution, or raise revenue to finance public expenditure in ways which policy makers deem to be desirable. The benefits thought to arise from taxation and government expenditure need to be compared with this 'excess burden' created by the distortion.[2]

The present discussion concentrates on distortions to labour supply arising from income taxes and benefits (transfer payments), and distortions to expenditure patterns arising from indirect taxes. Income taxation affects the net wage rate and therefore the implicit price of leisure. Indirect taxes distort expenditures by affecting the price of consumption goods.[3] There are of course other types of distortion. The concentration on these two types here should not be taken to imply that the other types are less important.

In planning tax increases, it would clearly be useful to have information

[1] The exception to this statement is where tax systems are entirely uniform and no form of expenditure or type of income is exempt, and labour supply is fixed. Furthermore, in cases where a previously untaxed good is taxed, this may introduce uniformity into the tax structure and so reduce the extent of distortions.

[2] It should be stressed that there can be a substantial distortion even if there is no change in the demand for a good whose price has increased, as explained further below.

[3] In practice it is very difficult to know the precise extent to which indirect taxes are shifted to final consumers. The assumption is usually made that they are fully shifted. Furthermore, large changes which affect the allocation of resources are also likely to give rise to changes in factor prices (wage rates and capital rentals, depending on the intensity with which factors are used in different industries). However, most analyses are 'partial equilibrium' and are unable to allow for such 'general equilibrium' effects.

on the expected excess burden from the change (for a specified population group) in relation to the extra revenue to be collected. This gives rise to the concept of the marginal excess burden per dollar of extra revenue: it is referred to as the 'marginal welfare cost' of the tax change.[4]

These efficiency costs of tax changes are not directly observable and are consequently perhaps too easy to neglect in policy debates. Despite the central importance of the concept of excess burden, and the fact that it has been understood for over 150 years, there have been relatively few empirical studies reporting excess burdens and marginal welfare costs at the level of aggregation that is helpful for policy advice. This paper gives a brief summary of information relating to the New Zealand economy.

First, the excess burden, and the closely related concept of the marginal welfare cost of taxation, are explained in more detail in Section 2. Alternative approaches to measuring these concepts are discussed in Section 3. Labour supply elasticities are considered in Section 4. Income and consumption taxes are then examined in turn in Sections 5 and 6. Section 7 contains brief conclusions.

2 Welfare Change and Excess Burden Concepts

Measures of welfare changes of individuals, and hence the excess burden of taxation, are defined in money units in terms of changes in the cost of achieving specified levels of wellbeing. Two measures are available, depending on whether the perspective is from the point of view from before or after the tax (or change in the tax). The 'equivalent variation' is the maximum amount that an individual would be prepared to pay, after the tax change, to return to the old prices. The 'compensating variation' is the minimum

[4] When the price of one good changes as a result of an increase in the tax imposed on that good, the relevant change in revenue must allow for the fact that the revenue collected from other goods is likely to change, as a result of the reallocation of budgets.

amount which, at the new prices, would be needed to compensate for the tax change, such that the individual would be indifferent between the old and new choices (that is, after compensation). If these money welfare changes are available for different individuals or population groups, an overall evaluation can then be made by making explicit value judgements about the way in which the changes for different groups are to be compared.

These concepts of welfare change and excess burden are fundamental in the study of welfare economics, but their empirical measurement faces a number of difficulties. Two basic approaches, discussed in the following section, may be used. In one approach, the nature of preferences must somehow be estimated directly, allowing costs of achieving a specified level of wellbeing at different prices to be evaluated. In the second approach, in the case of indirect taxes information is needed about consumers' responses to price changes and, in the case of direct taxation, the effects of tax changes on labour supply are required.[5]

Faced with these problems, an approximation to excess burdens is often exploited. For small tax (and thus price) changes, the excess burden is equal to one half of the 'compensated price elasticity of demand' multiplied by the product of the amount spent on the good and the square of the tax rate.[6] The compensated elasticity of demand that is relevant here is a measure of the responsiveness of demand to price changes, in the hypothetical situation where 'real income' is held constant; that is, the income effect of a price change is not included. This means that there can be a significant excess burden even if individuals do not appear in practice to respond to price

[5] Although welfare changes are defined in terms of the cost of reaching specified welfare levels (indifference curves), it is in fact sufficient to have information about only Marshallian demand or supply curves. The necessary information can then be obtained by a process of numerical integration; see Creedy (2006).

[6] For the equivalent variation, the relevant expenditure level is that before the tax, while the tax rate is the tax-inclusive rate and the elasticity is that along the post-tax indifference curve. For the compensating variation, the appropriate tax rate is the tax-exclusive rate, the expenditure level is the pre-tax expenditure and the elasticity is that along the pre-tax indifference curve. These differences are often neglected in popular discussions.

changes, depending on the nature of those income effects. The approximation has frequently been used to make 'back of the envelope' calculations regarding excess burdens at a high level of aggregation but these need to be treated with much caution.[7] The approximation has more pedagogic than practical value.

This approximation also shows the important result that the excess burden increases disproportionately with the tax rate. If the rate doubles, the excess burden quadruples. This means that any erosion of the tax base, which requires a higher tax rate than otherwise if revenue is to remain unchanged, creates higher burdens. This is one important reason why the often-cited rule of thumb, that a special case needs to be made when departing from a 'broad base and low rate', receives so much support.

This approximation may also suggest a generalisation regarding comparisons between income taxation and consumption taxes. A broad based goods and services tax (like the GST in New Zealand) which has very few exemptions, and has a relatively low rate, may be expected to give rise to lower efficiency costs than income taxation, which is imposed at higher rates and is not as broad based.

3 Approaches to Measuring Excess Burdens

The fact that welfare changes cannot be measured directly means that some type of economic modelling must be carried out in order to determine how individuals are likely to respond to price changes (including the implicit 'price of leisure'). This modelling is much more complicated in the case of labour supply compared with commodity demands. In the latter case it is reasonable

[7] They usually 'borrow' elasticities from other studies (often other countries), are at a very high level of aggregation, use the 'wrong' elasticity concept (and sometimes the wrong tax rate measure – the choice is between a tax-inclusive or tax-exclusive rate). More importantly they are often carried out for changes which cannot be considered as sufficiently small for the approximation to work: indeed, most tax policy changes are 'non marginal'.

to suppose that all individuals pay the same price irrespective of the amount they purchase, but in the former case the complexity of the tax and transfer system means that there are many prices (net of tax wage rates), even for a single individual.

In trying to model responses to indirect and direct taxation, the majority of empirical studies do not attempt to allow for any changes in wage rates which might arise from consequent changes in the allocation of resources among industries: they are 'partial equilibrium' in nature. Allowing for the additional inter-relationships requires a type of 'general equilibrium' model. A model of this kind was used by Diewert and Lawrence (1994) in their study of excess burdens in New Zealand. However, it has not been possible to allow for much heterogeneity on the household side of these models. In fact they model only one household – a 'representative household' – and distinguish four consumer goods in addition to leisure. Thus, their results are for a very high level of aggregation and, in addition relate to approximations which apply to small changes. Furthermore, such models cannot deal with the complexity of actual tax and transfer systems; thus a single tax rate is used to describe taxation on labour income. Hence the method cannot capture the fact that the income tax structure in the 1970s had a high degree of rate progression, and was subject to considerable simplification in the 1980s. Similarly, the simplification of indirect taxation associated with the introduction of a GST was not modelled. The results reported by Diewert and Lawrence give marginal welfare costs for taxes on consumption goods increasing from 5 cents in 1972 to 14 cents in 1991, and marginal welfare costs for taxation on labour income increasing from 5 cents in 1972 to 18 cents in 1991.[8] .

In the context of partial equilibrium studies, a distinction can be drawn

[8] They use the term marginal excess burden to refer to the marginal welfare cost (the marginal excess burden per dollar of extra revenue). The single consumer is assumed to have a quadratic expenditure function, which gives rise to simple forms for compensated demand curves.

between 'reduced form' and 'structural' models. In the former case, a relationship between, say demand and price, is specified and estimated without direct reference to the precise behaviour of individuals. However, under certain conditions it is possible to work backwards from the results in order to obtain the precise welfare measures necessary.[9] This approach is much more successful in the case of goods, largely because of the point made above regarding the fact that a single price applies to all units sold (that is, the budget constraints are nonlinear). Early studies of burdens in the context of income taxation used reduced-form labour supply functions which suffered from the problem of endogeneity (net wages and hours worked are jointly determined) and the fact that the estimated relationships were at a high level of aggregation. They were not capable of capturing the precise (or even an approximate) nature of the nonlinear budget constraints faced by individuals (for example a crude measure of 'average marginal tax rate' was often used).

More recent studies of labour supply have used structural models which enables preference functions to be estimated directly. This approach uses a discrete hours specification in which only a small number of hours levels (such as 'full time' and various degrees of part time work) are regarded as feasible. This approach can deal with the full detail of the tax and transfer structure as well as the considerable population heterogeneity that is found in cross-sectional data. The estimated preference functions are then embedded in a behavioural microsimulation model, though further innovations are needed in order to obtain the required welfare measures. The standard method of computing welfare changes needs to be modified to deal with the nonlinear budget constraints.[10] The number of studies reporting excess burdens is very

[9] Technically, 'working backwards' refers to a process of integration from demand to preference functions. The results must satisfy certain 'integrability conditions' to ensure that the actual demands can be regarded as arising from utility maximisation; see Creedy (2006).

[10] Furthermore, these discrete hours models contain a random component to the specified preference functions. This means that the models generate, for each individual, a probability distribution of hours worked rather than a deterministic level of labour supply.

much smaller than those examining labour supply behaviour. Furthermore, the reliance on the use of cross-sectional data means that any process of adjustment over time cannot be modelled.

In New Zealand, the modelling of consumption and income taxes has been severely constrained by the difficulty of obtaining the datasets needed, although recent changes mean that these can now be accessed by academics as CURFs (confidentialised unit record files). Furthermore, the construction and use of behavioural microsimulation models requires teamwork, high fixed costs and continuous maintenance, all of which involve special funding. This kind of research funding is very limited in New Zealand.

4 Labour Supply Responses

Although there are few detailed excess burden calculations, many estimates of labour supply elasticities have been produced. These are usually based on calculating the effects of a specified small change in a wage rate. However, care needs to be taken in viewing reported values, which are usually at a very high level of aggregation (such as 'all males' or 'all females'), and it is necessary to clarify precisely how the elasticity is computed.[11] Importantly, the reporting of a single elasticity value for a population group can tends to detract from the point that such elasticities are not constant, even along the labour supply curve of a single individual. Elasticities with respect to changes in wage rates also depend crucially on the nature of the tax structure

Methods of dealing with this complication have only recently been developed: see Creedy and Kalb (2005b) and Creedy, Hérault and Kalb (2007).

[11] Questions needing to be considered are: do the values relate to changes at the average wage, or are they average values over the population group; if the latter case applies, what treatment is given to those initially not working (for whom the elasticity is not defined); if a discrete choice approach is used whereby individuals have a probability distribution of hours worked for each wage rate, precisely how is the elasticity computed (is it obtained from expected labour supply for each wage, or is a calibration approach used whereby labour supply at the pre-change wage is always set at the observed value).

itself, not simply the preferences and characteristics of individuals.[12] Hence aggregate elasticities obtained from one country, and tax structure, cannot simply be taken as indicative of responses in another country.

An examination of results from a range of studies, such as those reported in the survey by Meghir and Phillips (2009), indicate substantial variations depending on the type of model, estimation method and data used. However, one broad generalisation which is often made is that elasticities, particularly for men, are low and in nearly all cases are less than 1: an increase in the wage rate produces a less than proportionate increase in labour supply. But this sweeping generalisation is not particularly helpful, given the heterogeneity involved and the wide range of possible tax changes which may be considered as parts of a policy reform. Extensive tabulations of labour supply elasticities for different types of individual in Australia are reported in Kalb and Ghantous (2006). They make the important point that elasticities, 'provide the most information at the disaggregated level, since at the aggregate level a small elasticity can hide a wide variation of responses for subcategories in the population'. Indeed for various sole parent categories, and some married women, they obtained elasticities exceeding 1. This point is worth bearing in mind when broad comments are being made about labour supply responses.

While aggregate elasticities, particularly for males, are low, the point was made above that even if elasticities are low, the excess burden can be high because the relevant elasticity concept is in fact the one which allows only for substitution effects of wage changes. Not all studies report income effects, but again results show substantial variations.[13] The question arises of how

[12] For example, at kinks in budget constraints where the marginal tax rate increases at an income threshold, labour supply falls (along a rectangular hyperbola) as the wage increases, thereby leaving the individual 'stuck' at the income threshold for a range of wages. Ranges where the marginal rate falls, when a means-tested benefit is exhausted, give rise to large 'jumps' in labour supply for very small wage changes.

[13] The reference to an 'income elasticity' here refers to the effect of a change in non-wage (or virtual) income. Again, results given by Kalb and Ghantous (2006) show that these can be non-trivial for some categories.

useful such elasticities are when it comes to trying to say something about excess burdens. As mentioned above, they appear as a component of the familiar approximation for the excess burden, but this approximation is of little, if any, practical value in trying to obtain values which can contribute towards policy advice. In practice, not only do excess burden calculations needs to allow for non-marginal changes for individuals, and indeed for those whose elasticity may not even be defined (who move from not working to working when their wage is increased), but the effects can depend very much on the nature of the tax change (given the complexity of tax functions).

5 Welfare Costs of Income Taxation

An indication of potential welfare costs of income taxation can be obtained from an Australian study by Creedy, Hérault and Kalb (2007), who examined the implications of increasing all non-zero income taxation rates by 5 percentage points.[14] They found that, despite small changes for most individuals in their labour supply and net incomes (except for sole parents), there were large variations in the marginal welfare cost of taxation. These ranged from close to zero for some individuals, to over \$5 for others. For example, the marginal welfare cost for a low-income single parent (taken at random from the sample) was found to be \$2.70 per extra \$1 of tax raised, and for a medium-income single parent, it was found to be \$5.60. These high efficiency costs arise despite small expected reductions in labour supply. In some cases, the expected reduction in labour supply is such that there is actually a reduction in tax paid as a result of the tax rate increase. Although in general, high income individuals faced lower marginal welfare costs, examples were obtained of high income single men with marginal excess burdens per dollar of 50 cents. For low and middle income single women, examples of

[14] For welfare measures based on different policy changes, see Creedy, Hérault and Kalb (2009).

the marginal welfare cost were as high as 80 cents and $1.20 respectively.[15]

In looking at aggregate effects, which conceal the substantial variations discussed above, the average marginal welfare cost for couples, single men and women, and single parents were found to be 24 cents, 23 cents, $1.80 and $2.10 respectively.

6 Welfare Costs of Indirect Taxation

The only estimates of welfare changes associated with indirect taxes in New Zealand have been produced by Creedy (2004) and Creedy and Sleeman (2005, 2006), who obtained results for a range of household types, distinguished according to whether anyone in the household was a smoker.[16]

Here the way in which the welfare cost of the tax system varies, when expressed as a proportion of total expenditure, provides an initial indication of progressivity. The existing tax system was found to be regressive for multi-adult smoking households. However, among single adult smoking households, the system is progressive. Among the non-smoking households the effect of the current system was found to be ambiguous. The tax system is generally regressive over lower levels of total expenditure, but becomes increasingly progressive over higher levels, particularly among single adult households and those with multiple children. Hence, when examining separate demographic groups, there is no strong evidence of a regressive effect of indirect taxes.

For the majority of household groups, the welfare costs per dollar of tax raised, at mean total expenditure within the group, are small, suggesting that the indirect tax system as a whole is relatively efficient at raising revenue. For every dollar of indirect tax paid, the efficiency losses incurred by

[15] It was also found that different conclusions could be drawn with regard to how the different groups and individuals in the population are affected, depending on whether net income changes or welfare changes are considered. Hence the extra effort involving in attempting to produce measures of welfare changes is worthwhile.

[16] Substantial differences in expenditure patterns were observed between smoking and non-smoking households.

smoking household groups (at their respective average total expenditures) range between about 2 and 8 cents, while for non-smoking households, these losses range between only about 1 and 4 cents. These marginal welfare costs are thus much smaller than the orders of magnitude found for income taxes, as discussed above.[17] The small efficiency costs of indirect taxation are not surprising in view of the fact that indirect taxes in New Zealand are relatively 'flat', with the main exceptions arising from the excise taxes.

Several hypothetical reforms involving excise taxes were also examined. For reforms eliminating all excises, all smoking household groups experience substantial marginal welfare benefits. For every dollar reduction in tax paid, the revenue-neutral removal of all excise taxes leads smoking household groups to experience efficiency gains ranging between about 20 and 50 cents. However, for single adult households with three or more children these are much higher, despite the fact that they devote the smallest budget shares to alcohol, tobacco and petrol.[18] For both reforms, the efficiency gains of smoking household groups typically rise with the number of children, while at the same time the reductions in tax paid fall.

The revenue-neutral removal of the alcohol and tobacco excise taxes leads all non-smoking household groups to incur marginal welfare costs. However, these costs are small. Furthermore, the removal of the petrol excise tax leads two thirds of all non-smoking household groups to acquire small gains in efficiency. Thus, these household groups experience increases in efficiency despite paying more tax. In contrast to smoking household groups, non-smoking household groups with no children incur the smallest marginal welfare costs and largest marginal welfare benefits from the first and second reforms respectively. Non-smoking household groups with multiple children

[17] Care needs to be taken in making these comparisons, because the income tax results were for Australia and relate to changes in tax rates, whereas the indirect tax results relate to the difference between no tax and the existing structure.

[18] This counterintuitive result may be explained by the resulting small reduction in tax paid by these household groups, so that the denominator is close to zero.

fared worst, sustaining the largest efficiency losses. The marginal changes in welfare are more varied across the smoking household groups than the non-smoking household groups.

7 Conclusions

This paper has provided a brief review of the concepts relating to the efficiency costs of income and consumption taxes. These include the central idea of the excess burden and marginal excess burden, along with the marginal welfare cost per dollar of revenue (or extra revenue) raised. For small tax changes the excess burden can be approximated by a simple expression involving the square of the tax rate and the (compensated) elasticity of demand (in the income tax case, this is a demand for leisure). While this approximation has considerable pedagogic value, is it of little value in practical policy evaluations.

Considerable emphasis has been given, in the context of income taxation, to labour supply elasticities. A wide range of values have been produced in the literature, mainly relating to broad population aggregates. It is suggested that this kind of emphasis is greatly misplaced. The aggregates conceal considerable population heterogeneity, and they are certainly not fixed. Indeed they vary over individual supply functions and, importantly, the elasticity measures are themselves influenced by the nature of the tax structure in operation when the data were collected. It would not be wise to apply reported elasticity values for, say, all married men, obtained for one country or time period to excess burden approximations in another country.[19]

Given structural estimates of preference functions which can allow for the precise nature of the tax structure and deal with the considerable population heterogeneity found in practice in cross-sectional data, the calculation of

[19]Even the definitions of elasticity vary between studies, given the alternative ways it could be calculated. Furthermore, it has been found that the welfare effects of tax changes are often not symmetrical with respect to increases or decreases in tax rates.

excess burdens and marginal welfare costs associated with proposed policy reforms makes absolutely no use of elasticity values.

Estimates relating to Australia and New Zealand suggest that, on average, the marginal welfare costs of indirect taxes are substantially below those of income taxation. This is not surprising in view of the broad-based nature of the current structure of consumption taxes. Yet, this statement again conceals much variation within the population. In any practical policy discussion, information is needed about the way in which particular groups in the population are likely to be affected, so that judgements can be made about the losses and gains involved. Furthermore, in examining policy reforms, it is necessary to be explicit regarding the total revenue requirement. If revenue-neutral changes are not considered, then there is in fact another kind of policy change involved in addition to the tax policy.

References

[1] Creedy, J. (2006) Measuring welfare changes with Marshallian demand functions. *Australian Economic Review*, 39, pp. 340-346 .

[2] Creedy, J. and Kalb, G. (2005a) Discrete hours labour supply modelling: specification, estimation and simulation. *Journal of Economic Surveys*, 19, pp. 697-734.

[3] Creedy, J. and Kalb, G. (2005b) Measuring welfare changes in labour supply models. *Manchester School*, 73, pp. 664-685.

[4] Creedy, J., Hérault, N. and Kalb, G. (2007) Comparing Welfare Change Measures with Income Change Measures in Behavioural Policy Simulations. *Melbourne Institute of Applied Economic and Social Research Working Paper* No. 21/07. The University of Melbourne.

[5] Creedy, J., Hérault, N. and Kalb, G. (2009) Abolishing the tax-free threshold in Australia: simulating alternative reforms. *Fiscal Studies* (forthcoming).

[6] Creedy, J. and Sleeman, C. (2005) Excise Taxation in New Zealand. *New Zealand Economic Papers,* 39, pp. 1-35.

[7] Creedy, J. and Sleeman, C. (2006) *The Distributional Effects of Indirect Taxes: Models and Applications from New Zealand.* Cheltenham: Edward Elgar.

[8] Diewert, W.E. and Lawrence D.A. (1994) *The Marginal Costs of Taxation in New Zealand*. Canberra: Swan Consultants

[9] Kalb, G. and Ghantous, S. (2006) Tables with elasticities broken down by individual and household characteristics. Melbourne Institute, University of Melbourne.

[10] Meghir, C. and Phillips, D. (2009) Labour supply and taxes. Institute for Fiscal Studies.