TAX POLICY REFORM: THE ROLE OF EMPIRICAL EVIDENCE

Richard Blundell
University College London and Institute
for Fiscal Studies

Abstract
To understand the role of evidence in tax policy design, this paper organizes the empirical analysis of reform under five loosely related headings: (i) key margins of adjustment, (ii) measurement of effective tax rates, (iii) the importance of information and complexity, (iv) evidence on the size of responses, and (v) implications from theory for tax design. The context for the discussion is the recently published *Mirrlees Review* of tax reform. Although the Review focused on all aspects of tax reform, this paper highlights the taxation of earnings. It also comments on earnings taxation in the context of VAT base-broadening reforms and the taxation of capital (JEL: H2, H3).

1. Introduction

How should evidence be used in the study of tax design? What is the appropriate balance between theory and empirics? These questions lay at the heart of the *Mirrlees Review*. Motivated by the aim to develop a broad set of principles for what makes a *good tax system*, the *Review* was an attempt to base tax reform on the large body of economic theory and empirical evidence. It was inspired by the Meade Report (1978) with the idea to review tax design from first principles for modern open economies in general and for the UK in particular. The UK over the past thirty years would be the working laboratory.

The *Mirrlees Review* was published in two volumes: *Dimensions of Tax Design* (Mirrlees et al. 2010) bringing together expert evidence across a wide range of aspects of tax reform, and *Tax by Design* (Mirrlees et al. 2011) setting out the conclusions and recommendations. This paper examines the role of evidence used in the derivation of the recommendations for reform. It also examines the linkages between theory and empirical evidence. To maintain consistency and coherence in the discussion, the focus

Acknowledgments: I would like to thank the Editor, referee and participants at the JEEA Lecture at the AEA meetings and at the FBBVA Lecture in Madrid, for helpful comments. I am also grateful to my co-authors and co-editors on the Mirrlees Review for the many discussions over the course of the Review. All figures and tables not directly sourced are from the second volume of the Mirrlees Review, Tax by Design (2011).

E-mail address: r.blundell@ucl.ac.uk

Journal of the European Economic Association February 2012 10(1):43–77
 DOI: 10.1111/j.1542-4774.2011.01054.x

here will be on the taxation of earnings although the *Review* itself concerned all aspects of the tax system. The discussion is organized loosely under five related headings:

(i) Key margins of adjustment.
(ii) Measurement of effective tax rates.
(iii) The importance of information and complexity.
(iv) Evidence on the size of responses.
(v) Implications from theory for tax design.

The first of these headings highlights the importance of establishing empirical facts about key aspects of behavior where we think taxes could have an impact. The second reinforces a pervasive theme of the *Review* which was to consider the tax system as a whole and examine the *wedge* created by all aspects of the tax system, including the implicit tax rates in the benefit and tax-credit systems. This also naturally motives the third heading which relates to the understanding of the incentives implicit in the tax and benefit system by the individuals, households, and firms themselves and the stigma and hassle costs involved by those accessing the system. The forth heading is the core of any rigorous empirical analysis and concerns the robust measurement of the causal impact of tax reforms. Here I suggest the use of a mix of (quasi-)experimental and structural approaches with the experimental approaches acting as a 'reality check' on the structural model. Structural models allow the study of behavior in counterfactual environments and it is difficult to envisage a complete empirical analysis of tax design analysis that does not draw on such counterfactuals.

Under the final heading, these empirical relationships are brought together with the structure of mechanism design from economic theory to determine efficiency costs, overall optimality, and improvements to tax design. There are three key ingredients to any optimal tax analysis: the accurate measurement of response elasticities, the detailed description of the distribution of income, and some view of social welfare weights. The first two of these are positive and can be learned from a careful evidence-based analysis. The last is normative and therefore something over which reasonable people may differ. The aim here is to draw broad evidence-based conclusions while making fairly weak assumptions on social welfare weights, perhaps assuming no more than that they are declining in some measure of equivalized income.

Why the focus on earnings taxation? Earnings taxation is ideally suited for examining the role of evidence in tax design. There are substantial empirical results on labor supply responses to tax reform for individuals and families, see Blundell and MaCurdy (1999) and Meghir and Phillips (2010) for surveys. This research has emphasized the need to distinguish between the intensive and extensive margins of labor supply—that is between the decision of whether to work or not and how much to work, respectively. It has also shown clear differences in responses by age, gender, and family composition. Both of these observations are central to tax design. Further, tax return information provides additional evidence on taxable income elasticities, highly relevant for the design of earnings taxation, see Gruber and Saez (2002), for example. We will argue that this evidence naturally supplements and extends work on employment and hours of work responses to tax reform.

The next five sections of this paper reflect these five aspects of the empirical analysis of reform. This is not meant to imply that the taxation of earnings should stand separately from the design of the rest of the tax system. As the *Review* recommendations volume *Tax by Design* makes clear, any comprehensive reform must bring together all aspects of taxation. Indeed, the taxation of earnings bears the brunt of much of the tax reform proposals through the need to adjust for changes in redistribution and work incentives induced by other aspects of the reform package. Therefore, to round off this paper, the discussion turns to the interplay between earnings tax design and base-broadening reforms to VAT, as well as to the taxation of capital and reforms that seek to align effective tax rates across all sources of income.

2. Key Margins of Adjustment

With the focus on earnings tax reforms, our analysis begins with the key changes in lifetime employment patterns over the last three decades. This sets the scene for understanding where, over their working life, individuals and families are most likely, and most able, to respond to tax reform.

The recent history of variation in hours and employment has been made up of three key trends which we will argue also point to the three key margins where responses to tax reform are most likely to occur: a decline in employment among men especially at older ages, a strong rise in employment and total hours of work for women, and a decline in employment among those in their late teens and early 20s reflecting the increase in educational attainment over this period.

As has already been noted, an important distinction in analyzing labor supply responses is between the extensive (whether to work) and intensive (how much to work) margins of labor supply. Although it is the case that hours of work are often found to respond less than employment decisions, Blundell, Bozio, and Laroque (2011a) show that the intensive and extensive margins both matter in explaining the broad changes in total hours over the last three decades in the UK, France and the United States. But they matter in different ways for different age and demographic groups. For men, variations in the extensive margin occur mainly at the beginning and at the end of their working lives. These are the schooling–work margins and the early retirement margins: Figure 1, from Blundell, Bozio, and Laroque (2011b), provides a broad view of employment rates by age for the UK, France and the United States in 2007 (just before the onset of the most recent recession). The similarity of average employment rates in 2007 for men aged 30–54 in these three economies is striking. It suggests that differences in employment are concentrated at early and later points in the working life. Heckman (1993), Prescott (2004), Ohanian, Raffo, and Rogerson (2008), and Gruber and Wise (1999) have all pointed to the importance of the extensive margin at these points of the life-cycle.

The extensive margin is not the end of the story. Figure 2 points out that hours differences, conditional on employment, matter too for men and they matter across the working life. Although it is unlikely that tax and benefit systems alone explain all

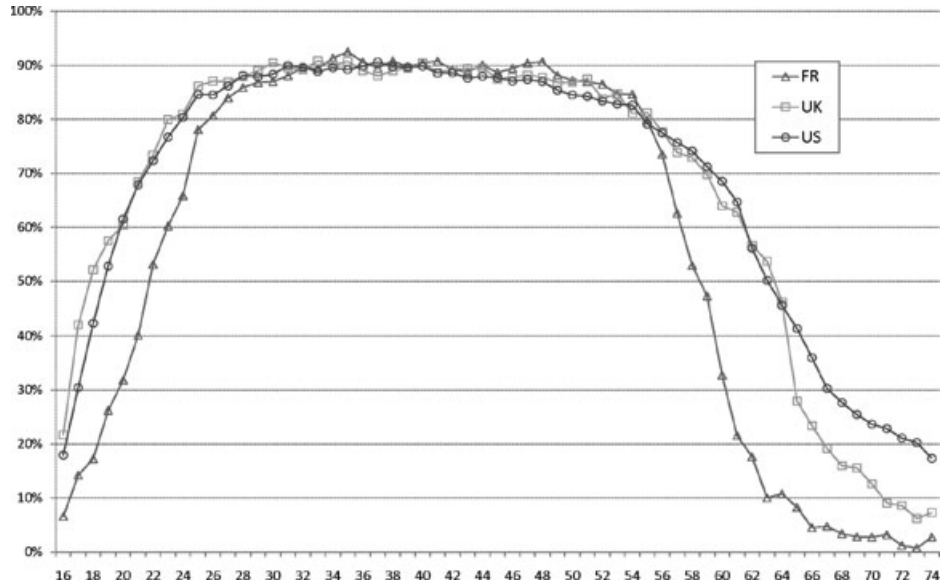


FIGURE 1. Male employment in the UK, 2007. Source: Blundell, Bozio, and Laroque (2011b).

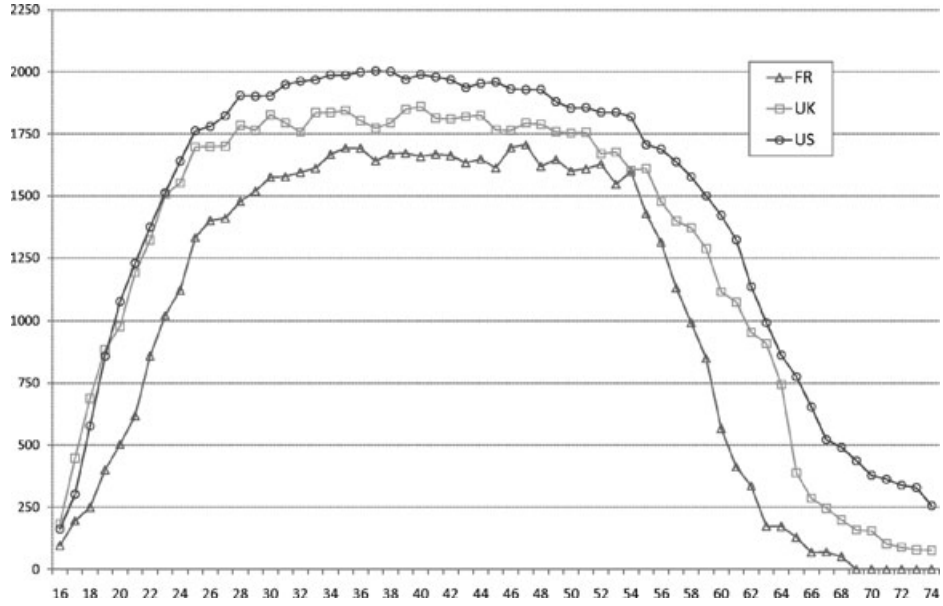


FIGURE 2. Male total hours worked in the UK. Source: Blundell, Bozio, and Laroque (2011b).

these differences, in any discussions of tax reform it would seem unwise to play down the intensive margin too much. [1]

1. See also Chetty 2009.

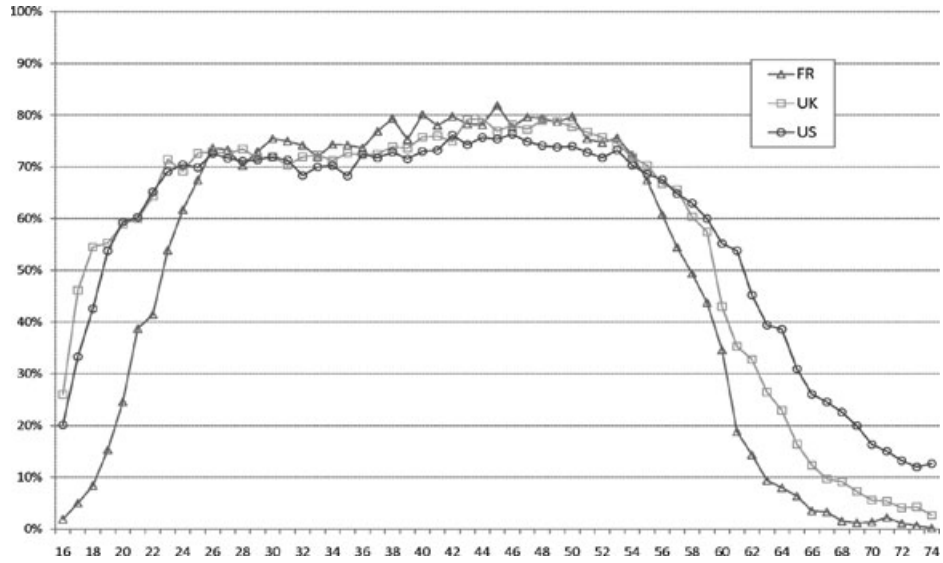


FIGURE 3. Female employment in the UK, 2007. Source: Blundell, Bozio, and Laroque (2011b).

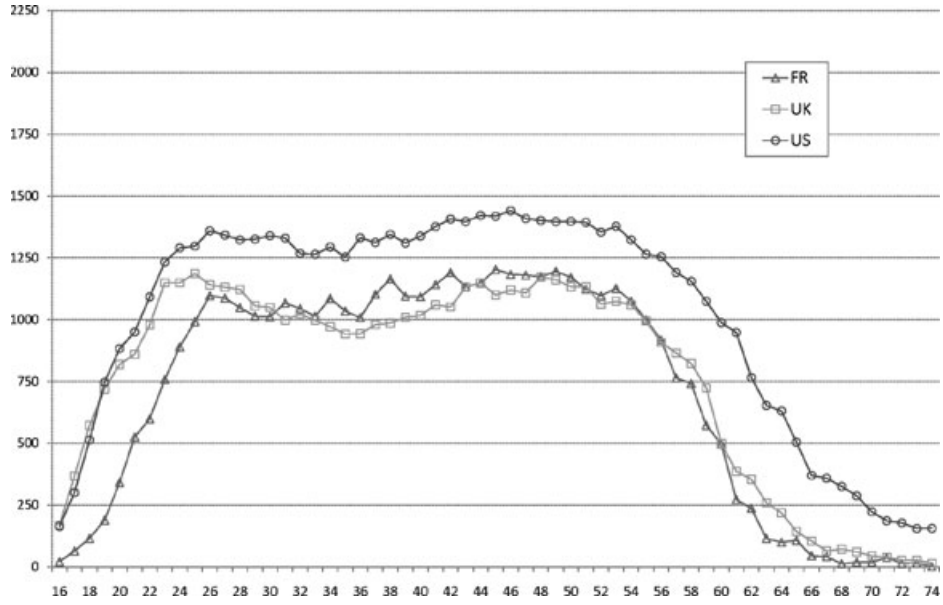


FIGURE 4. Female total hours worked in the UK, 2007. Source: Blundell, Bozio, and Laroque (2011b).

For women, Figures 3 and 4 show that hours conditional on employment and employment itself both vary across the working lives. As was the case for men, average employment rates in 2007 were surprisingly close at ages between the late 20s and early 50s. Again, it is at the early and later periods in the working life where the

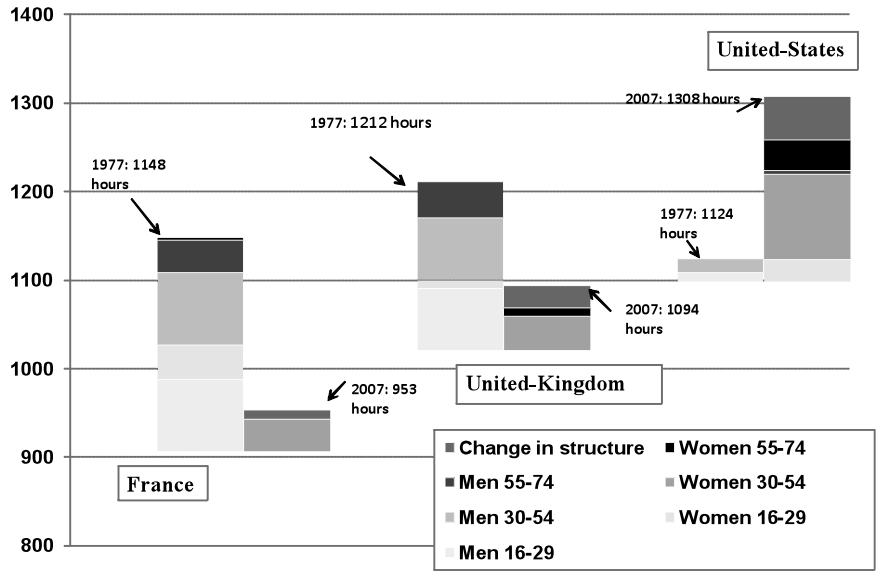


FIGURE 5. The change in total hours by age and gender. Source: Blundell, Bozio, and Laroque (2011a).

extensive margin choices become important. We will also point to important variation at the extensive margin for mothers with pre-school children and with lower levels of education. Hours of work conditional on employment for women show more variation over the life-cycle, especially in the UK where there still remains a dip around child-bearing ages. For women with younger children it is not usually just an employment decision that is important, it is also whether to work part-time or full-time. Some of this variation in the UK we will be able to attribute to the specific design of the tax and benefit system.

In the sections that follow we focus a little more in detail on what has happened to the labor supply of women over the recent past and relate it to some of the key changes in tax and benefit policy. To wrap up the descriptive discussion in this section it is worth examining the overall changes in labor supply in France, the United States and the UK over the three decades leading up to 2007. Figure 5 from Blundell, Bozio, and Laroque (2011a) presents such a breakdown of total hours worked by age and gender. The huge declines in total hours among men in the UK and France and the large rise in women's labor supply in the United States dominate the picture. These changes in total hours mask somewhat the key changes which have occurred at the extensive and intensive margin. For example, it turns out the expansion for women at the extensive margin (employment) over this period is quite similar in the United States and France—what differs is the distinctly different paths at the intensive margin.

To allocate total hours changes between the extensive and intensive margins is not a trivial exercise. While we observe the changes in hours per worker and employment, we do not know exactly how these changes contribute to the changes in total hours

worked. Blundell, Bozio, and Laroque (2011a) address this by developing bounds on the changes at the extensive and intensive margins which allow such a decomposition. They consider how the overall average hours worked H per person varies over time and across countries. Of course, this quantity differs across a person characteristics, age and gender for instance. Suppose there are $j = 1, \ldots, J$ broad categories. The overall statistic H_t is computed in any year t as an average of the total hours in category j, H_{jt}, with weights equal to the population shares q_{jt}:

$$H_t = \sum_{j=1}^{J} q_{jt} H_{jt}.$$

They then write total hours of work H_{jt} as the product of hours per worker h_{jt} and employment in the labor market p_{j_t}:

$$H_{jt} = p_{jt} h_{jt}.$$

When we observe a change in yearly hours worked per person, $H_t - H_{t-1}$, we would like to be able to know how much of the change is due to the intensive or extensive margins. First define a structural effect S_t due to the change in the composition of the population:

$$S_t = \sum_{j=1}^{J} H_{jt}[q_{jt} - q_{j,t-1}].$$

Then measure the change due to the behavior of category j, holding the population structure constant as in date $t - 1$, as in a Laspeyres index:

$$\Delta_{jt} = q_{j,t-1}[H_{jt} - H_{j,t-1}], \tag{1}$$

then the total change across all J categories of workers is simply

$$\Delta_t = \sum_{j=1}^{J} \Delta_{jt}, \tag{2}$$

and we have by construction

$$H_t - H_{t-1} = S_t + \Delta_t. \tag{3}$$

The change in total hours for any category of workers reflecting changes at the intensive margin (hours per worker), and at the extensive margin (employment) satisfies two polar exact statistical decompositions:

$$\Delta_{jt} = q_{j,t-1} \left\{ [h_{jt} - h_{jt-1}] p_{jt} + [p_{jt} - p_{jt-1}] h_{j_{t-1}} \right\} \tag{4}$$

or

$$\Delta_{jt} = q_{j,t-1}\left\{[h_{jt} - h_{jt-1}]p_{jt-1} + [p_{jt} - p_{jt-1}]h_{jt}\right\}. \quad (5)$$

The first term on the right-hand side is the intensive margin, weighted in the top formula (4) with the final employment rate (as in a Paasche index) and in the bottom formula (5) with the initial employment rate (as in a Laspeyres index). The second term is the extensive margin (Laspeyres in (4), Paasche in (5)). The empirical counterparts to these are given in Table 1.

The indices examine what part of any overall change in hours is attributable to changes at the extensive or intensive margin for any particular subgroup of the population. The row [I-L, I-P] shows the bounds on the intensive margin, *L* standing for Laspeyres (the change in hours being weighted by the initial employment rate), *P* for Paasche (final employment rate). Similarly, the Laspeyres index for the extensive margin (E-L) (resp. E-P), given by the second term in equation (4) (resp. (5)), is equal to the change in employment multiplied by average hours worked at the initial (resp. final) date.[2]

Turning first to prime-age workers, the steep decline at the *intensive* margin for prime aged men in France and the UK relative to the United States is striking. For this group the bounds are quite narrow and leave little room for ambiguity. These changes represent an enormous shift in the relative position of these countries. Table 1 tells us that the *extensive* margin for prime-age men in Britain and in France also falls more than in the United States, although there are declines in the United States too. As we have noted, for prime-age women it is the increase at the extensive margin that is so extraordinary, especially in the United States and in France where the bounds in Table 1 suggest a very similar change and one that is nearly twice the size of that experienced in the UK. Intensive margins provide an interesting picture here, falling back strongly in France, staying put in the UK while growing in the United States.

For older men and women there is a large decrease in hours per worker in France, similar in the UK, contrasting with an increase in the United States. There are falls at the extensive and intensive margins for UK men but increases at the extensive margin for UK women. This phenomenon is replicated to some extent across all countries and offsets the stronger incentives to retire earlier in the UK and in France. The contrast with the United States is stark, where at all margins and for both genders the bounds point to positive changes for older workers. The changes among the young are also sizable and predominantly negative. In France and the UK there are large falls for young men at both the extensive and intensive margin.

These changes inform us as to where labor supply is likely to be most responsive to reform. They also set up the key question in the analysis of tax incentives and labor supply: How well do structural economic models explain these changes in observed behavior? For this we have to turn first to the measurement of the effective tax rates in the tax and benefit system.

2. See Blundell, Bozio, and Laroque (2011a) for a more detailed analysis of the composition effects *S*.

TABLE 1. The extensive and intensive margins between 1977 and 2007.

		Youth (16–29)		Prime aged (30–54)		Old (55–74)		All
	Year	Men	Women	Men	Women	Men	Women	(16–74)
FR	Δ	−82	−38	−82	36	−36	−3	−195
	[I-L, I-P]	[−37,−28]	[−23,−19]	[−59,−56]	[−35,−49]	[−11,−8]	[−9,−10]	[−185,−183]
	[E-L, E-P]	[−54,−45]	[−19,−16]	[−27,−23]	[85,71]	[−28,−25]	[7,6]	[−12,−10]
UK	Δ	−71	−9	−70	39	−42	10	−118
	[I-L, I-P]	[−42,−36]	[−23,−26]	[−48,−45]	[−2,−3]	[−22,−19]	[−6,−8]	[−161,−167]
	[E-L, E-P]	[−35,−29]	[17,14]	[−25,−22]	[41,41]	[−23,−20]	[17,15]	[50,43]
US	Δ	−19	22	−19	90	6	38	165
	[I-L, I-P]	[−6,−6]	[1,1]	[−5,−5]	[14,19]	[3,3]	[3,5]	[15,17]
	[E-L, E-P]	[−13,−13]	[21,21]	[−14,−14]	[72,77]	[3,3]	[33,35]	[148,150]

Note: I-P designs the Paasche measure of the intensive margin, I-L the Laspeyres measure, with E-P and E-L respectively for the extensive margin, as described by equations (4) and (5).
Source: Blundell, Bozio, and Laroque (2011a).

3. Effective Tax Rates

What of effective tax rates? To understand how taxes and benefits might affect labor supply choices, we need to measure the effective work incentives implicit in the tax and benefit system. To describe the distribution of incentives implicit in the tax and benefit system, there are two summary measures that are useful to document: the effective marginal tax rate (EMTR, that is the proportion of a small increase in earnings taken in tax and withdrawn benefits) and the participation tax rates (PTR, the incentive to be in paid work at all) defined by the proportion of total earnings taken in tax and withdrawn benefits.

Perhaps the main (perceived) defects in current welfare/benefit systems is that participation tax rates at the bottom remain very high. This is certainly the case in the UK where effective marginal tax rates are well over 80% for some low-income working families. As we will see, this is mainly due to the phasing-out of means-tested benefits and tax credits. But high implicit tax rates at low incomes can be optimal for welfare functions that place a high weight on redistribution.

Consider a typical budget constraint for a single mother. A complete analysis of the effective tax rate will combine the implicit tax rates in the benefit system, the tax credit system and the income tax system. Figure 6 provides such a case study and

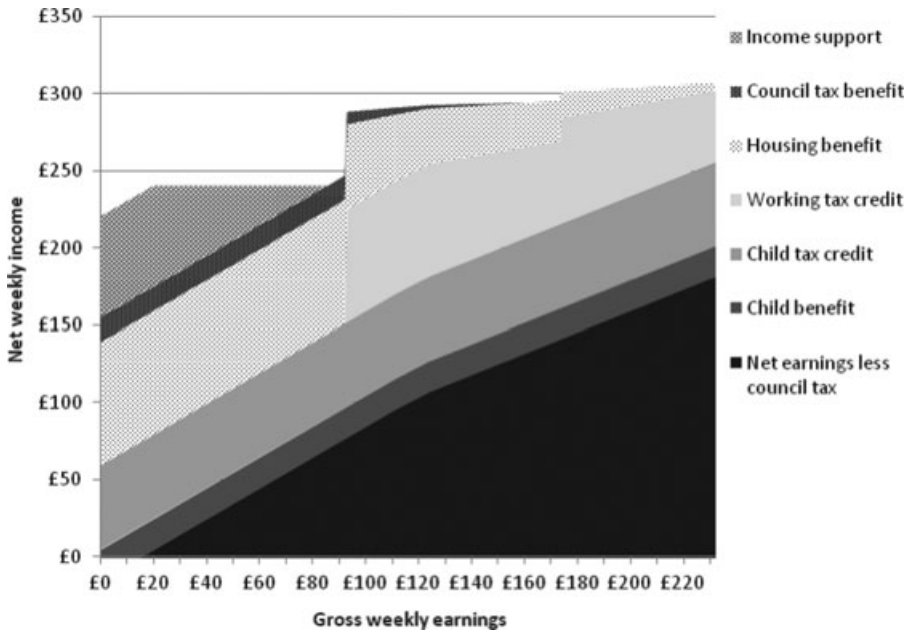


FIGURE 6. The interaction between taxes and benefits in the UK. Lone parent, with one child aged between one and four, earning the minimum wage (£5.80 per hour), with no other private income and no childcare costs, paying £80 per week in rent to live in a council tax Band B property in a local authority setting council tax rates at the national average. Source: Blundell and Shephard (2010).

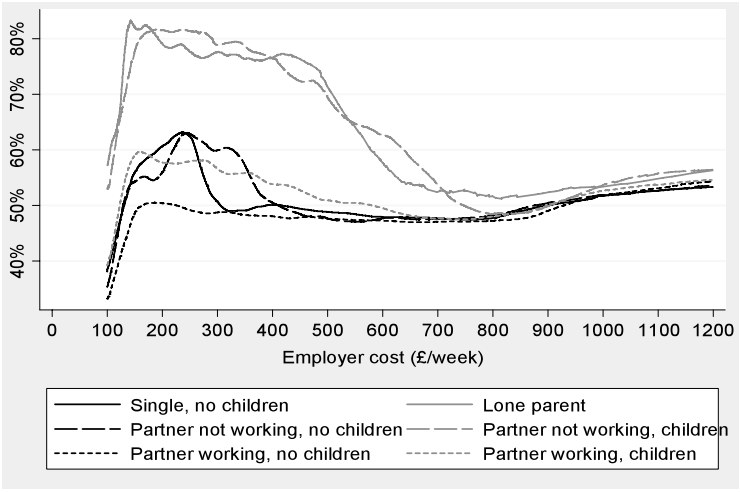


FIGURE 7. Average EMTRs for different family types.

shows the complexity arising from the cocktail of taxes and benefits. This constraint assumes all eligible benefits are accessed.[3]

One component of particular interest in the taxation of earnings is the tax-credit system which has become an increasingly important part of the effective tax system facing low earning families in many countries. In the UK, the earned income tax credit (called the Working Tax Credit, previously the Working Families Tax Credit) scheme has certain unique features. As with other tax credit systems, the UK system is designed to enhance income in work for those facing low rates of pay and/or higher costs of work. Figure 6 provides a case-study budget constraint for an example low-income single parent. In the UK eligibility depends on an hours of work condition which consists of a minimum hours rule at 16 hours per week with an additional hours-contingent payment at 30 hours. There is also a family eligibility criterion which requires children in full-time education or younger. The tax credit then consists of an adult credit plus amounts for each child. There is a family net income eligibility threshold, above which the credit tapers away at 55%. Taken together with Income Support and other benefits, low-income earners in the UK can face a complex rate schedule with relatively high effective tax rates. Indeed, families in receipt of other benefits would gain less from the WTC than otherwise equivalent families not receiving these benefits.

The distribution of these tax rates by income and family type in the UK is presented in Figures 7 and 8. In an important sense it is the participation tax rate that is relevant for the employment margin, and the marginal tax rate for the effort margin. The EMTRs and the PTRs can be negative as well as positive, but they are typically positive and often high at lower incomes.

3. See Adam, Browne, and Heady (2010) for a more detailed overview.

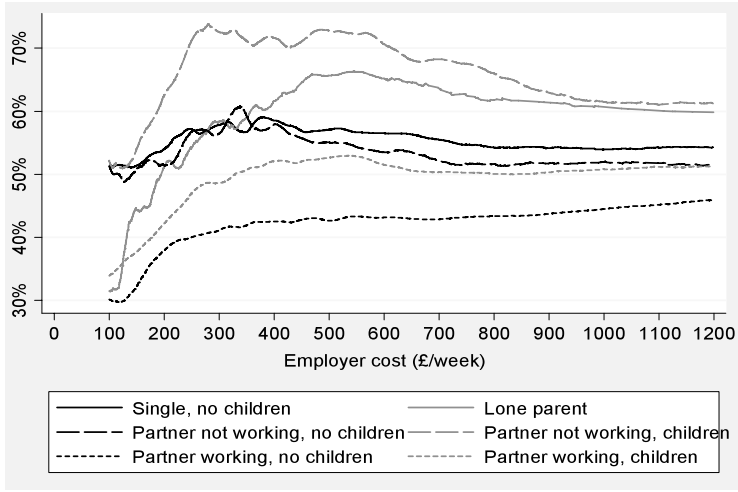


FIGURE 8. Average PTRs for different family types.

Couples with one earner and lone parents are the two distinct groups in the picture of average marginal tax rates by gross earnings in Figure 7. They face high effective marginal tax rates when their earnings are low. High tax rates at low earnings are a distinctive feature of many tax systems and have led some commentators to question why lower-earning individuals face the highest tax rates. But any system that redistributes income by targeting benefits towards families with low earnings and high needs will induce high effective tax rates as a natural by-product.

The effective tax rates in Figures 7 and 8 indicate the strong redistribution towards low-income families with children in the current UK tax system. Indeed, the more accurately the tax system targets low income, the higher the effective marginal tax rate on low earnings is likely to be. Not surprisingly therefore, tax schedules can easily possess the feature of high effective marginal tax rates at low earnings. It is simply the result of means-testing which is the flip-side of targeted redistribution. Whether it is optimal or not will depend, as we shall see in what follows, on the responsiveness of labor supply to these implicit tax rates, on the distribution of income and on the desire to redistribute to low-income families of a particular composition.

4. The Importance of Information and Complexity

The EMTRs and PTRs in Figures 7 and 8 are just *local* averages at each gross earnings level. As evidenced in our discussion of the single-parent budget constraint in Figure 6, the current structure in the UK, and elsewhere, of multiple benefits with an array of overlapping means-tests leaves some people facing effective marginal tax rates of over 90% (see Chapter 5 of *Tax by Design* for further details). This degree of complexity

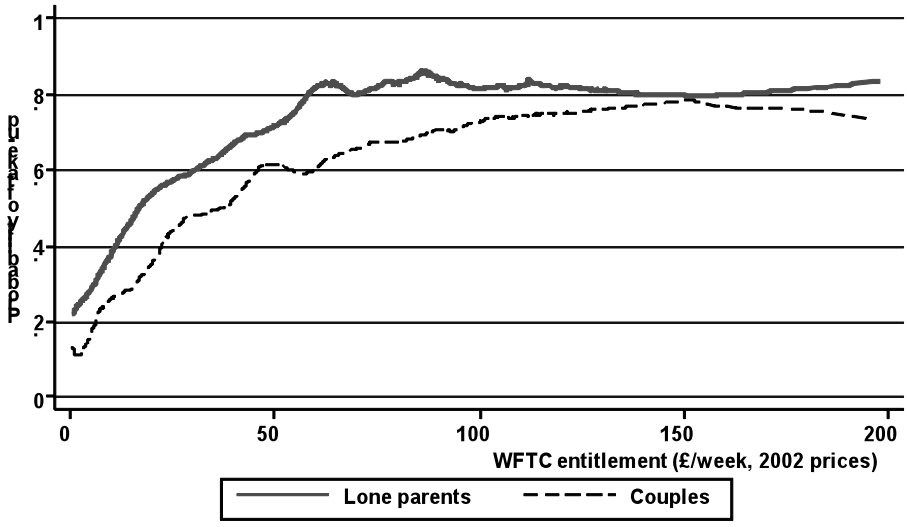


FIGURE 9. The take-up of tax credits in the UK.

can leave some individuals unwilling and/or unable to access all the benefits and tax credits to which they are eligible.

One way to formalize some of the issues surrounding information and complexity in earnings taxation is to allow individuals who are eligible for certain benefits and tax-credits not to participate or not to *take-up*. This reflects the idea that individuals may not understand the system or they may find the stigma or hassle costs involved in participating in the benefit or tax credit program too high to be worthwhile.

Typically, take-up is an increasing function of the eligible amount of benefit or tax credit but full take-up is rarely achieved, and rates of take-up can be quite low, especially for families eligible to small amounts of benefit. In Figure 9 the take-up rate for married couples and lone parents in the UK is plotted according to the eligible amount of Working Families Tax Credit (or its predecessor, Family Credit). This measures the proportion taking-up among those we estimate to be eligible for some benefit. It is plotted against the amount we estimate the family is eligible to. The figure suggests that the stigma or information costs are increasingly overcome as the value of take-up increases. This provides some insight into how to model take-up decisions.

To rationalize incomplete take-up of a benefit or tax credit program we follow Moffitt (1983), and the subsequent developments in Blundell et al. (2000) and Brewer, Duncan, and Shephard (2007), and assume the presence of some *stigma* or *hassle* cost. This will provide an interpretation of Figure 9 and will also feed into our structural econometric specification. In turn it will help separate preferences from information and stigma costs.

As an example, we can imagine representing the take-up cost as an amount subtracted from the overall utility of families who claim tax credit. Suppose η represents

this utility cost, then we could write

$$\eta = X'\beta_\eta + \varepsilon_\eta, \tag{6}$$

where X represents a vector of demographic and other household characteristics, β_η unknown parameter vector and ε_η unobserved heterogeneity. If the individual takes-up then his utility would be reduced by η. Families who are entitled to the credit or benefit will claim if the utility gain derived from the higher income exceeds the utility cost η.

The actual amount of taxes and benefits any individual worker receives (or pays) T will now depend on whether take-up occurs or not: we write $P = 1$ if take-up occurs and $P = 0$ if not. Suppose an individual faces an hourly wage rate w and works h hours, then the net taxes paid will not only be a function of hours of work h and total earnings wh but also whether $P = 1$ or 0. The tax function becomes

$$T = T(wh, \boldsymbol{h}, P; X). \tag{7}$$

The net financial gain in work will depend on take-up, wages, and the choice of hours.

This framework will be further developed when we return to the structural econometric specification in Section 4.3. For now it is worth noting the practical difficulty of incorporating many different benefits. Moreover, because benefits and tax credits are based on family income, incorporating several workers in each family is also challenging.[4]

It is difficult to argue against any policy reform that clarifies which benefits and tax credits individuals are eligible to, and what the effective tax rates in the system are. One clear implication for reform is that, to be effective, the tax and benefit system requires some integration of the various benefits and tax credits. This discussion of take-up and the integration may indicate how to reduce the complexity of the tax and benefit system but it does not tell us about the appropriate tax rate structure. For this we need to know about the size of behavioural responses. We now turn to the robust measurement of the impact of tax reform.

5. Evidence on the Size of Responses to Tax Reform

In Section 1, we documented the growth of female labor supply, changes in youth employment and changes in *early retirement* behavior which must form the backdrop to any earnings tax reform agenda. To these we add changes in demography, including growth in single person and single-parent households.

An important distinction in analyzing labor supply responses for the purposes of earnings tax design is between the extensive (whether to work) and intensive (how much to work) margins of labor supply. The microeconometric literature highlights difference between extensive/intensive responses (e.g. Heckman 1993; Blundell and MaCurdy 1999; Blau and Kahn 2007). Knowing precisely where the largest labor

4. Brewer, Duncan, and Shephard (2007) provide an excellent guide to the issues involved.

supply responses to incentives are to be found is a key ingredient in achieving a good earnings tax design.

Of course, other responses affecting taxable income matter for earnings tax design, certainly for the rich and self-employed. Although for many workers the employment and hours margins are the key measures of their labor supply, for other workers it is the level of effort, for any hour of work, that they can use to respond to tax incentives. For others still there will be exemptions and deductions which will allow them to change their taxable income with little change in their overall earnings. Acknowledging this, in Section 5, we will use the impact of taxation on taxable income to examine tax rate reform for top earnings.

5.1. Alternative Approaches to Measuring the Size of Responses

There are three dominant empirical approaches to the measurement of responses and all can prove useful in understanding earnings tax reform: the *experimental* approach using randomized control trials (RCTs), the *quasi-experimental* approach using historic reforms, and the *structural* approach based on a formal optimization model of individual and family choices. There are many comprehensive reviews of quasi-experimental approaches, see Blundell and Costa-Dias (2009) and references therein. Although few in number, there are also some influential control-trial experiments on labor supply which we will also briefly discuss.

It is difficult to envisage a full-fledged tax reform analysis that does not draw on a structural model. Policy simulation, and understanding the impact of particular rate structures, requires a model of decision-making and of the budget constraint. Herein though lies the difficulty; to fully specify the choice problem and the budget constraint inevitably requires assumptions for which we have relatively little empirical foundation.

In much of the literature on structural labor supply models, the complexity of the budget constraint has led researchers to approach estimation as a general discrete choice model with (unobserved) heterogeneity, see Blundell and MaCurdy (1999) for a review. For the most part these models have been estimated on cross-section surveys and do not attempt to directly model key *changes* in employment and hours, such as those documented in Section 1. Integrating the structural approach to estimation with the observed changes in incentives and policy reforms is an important step in deriving reliable structural models. There is an increasing recognition that changes in incentives over time can, and should, be used directly to estimate labor supply parameters, see Chetty (2009). In an early example of this approach, Blundell, Duncan, and Meghir (1998) use exogenous changes in work incentives to estimate structural parameters for female labor supply preferences at the intensive margin. They deal directly with endogenous selection and unobserved heterogeneity and, in doing so, are able to bring to bear the insights on transparent identification from the broader microeconometric and evaluation literature.

As we will argue, deriving convincing response effects from structural models is the key to tax design analysis. Experimental and quasi-experimental analyses have an important role to play in gauging the overall size of responses and in validating structural models. Further, we will see that some mechanism design problems in taxation can be expressed in terms of a small number of sufficient statistics, see Chetty (2008), some of which may be recoverable from quasi-experimental or experimental analysis. We suggest, at a minimum, quasi-experimental analyses of policy reforms should be used as a method of validating structural models. Inevitably, the more comprehensive and the more robust is the empirical evidence, the better we can address the tax design problem.

5.2. Randomized Control Trials

Experimental evaluations of tax and benefit policies are relatively rare, although the active use of RCTs in tax policy has existed for some time. For example, the Seattle–Denver Income Maintenance Experiment (SIME/DIME) was one of a small number of large-scale income maintenance/ negative income tax experiments undertaken in the late 1960s and early 1970s in the United States. The idea was to measure the disincentive effects of cash transfers on the market work of those eligible for them, see Hurd (1976), Johnson and Pencavel (1980) and West (1979).

Ashenfelter and Plant (1990) use the SIME/DIME experiment to estimate a nonparametric labor supply effects of the negative tax experiment. Nonetheless, in their comprehensive review of field experiments in economics, Card, DellaVigna, and Malmendier (2011) find very few experimental studies that are directed towards recovering structural economic parameters. Even fewer directed at tax and benefit reform. They note that

> as a result of the frustrations in dealing with the complex designs of the negative income tax experiments (and with the confusing message that emerged from such designs) many respected analysts adopted the view that social experiments should be designed as simply as possible. This shift away from designs that explicitly attempt to model response variation to multiple treatments and toward a single manipulation has led to a new round of criticism that the social experiments are often "black boxes" that "...contribute next to nothing to the cumulative body of social science knowledge..." (Heckman and Smith 1995, p. 108).

Even so, experiments that are simply a single control treatment contrast can be useful even in respect of a more theory-based analysis such as the tax design problem addressed here. Experiments are typically designed to test a hypothesis of no effect of some particular policy against a one-sided alternative of a positive effect. They estimate an average effect of the policy and do so under relatively weak assumptions, provided the experiment is carried out correctly. They can therefore be used to assess the reliability of theoretical prediction and/or gauge the overall size of incentive effects in some tax or benefit reform.

Of the more recent randomized control trials in tax and welfare policy, perhaps the most successful and most influential has been the Canadian Self Sufficiency Program

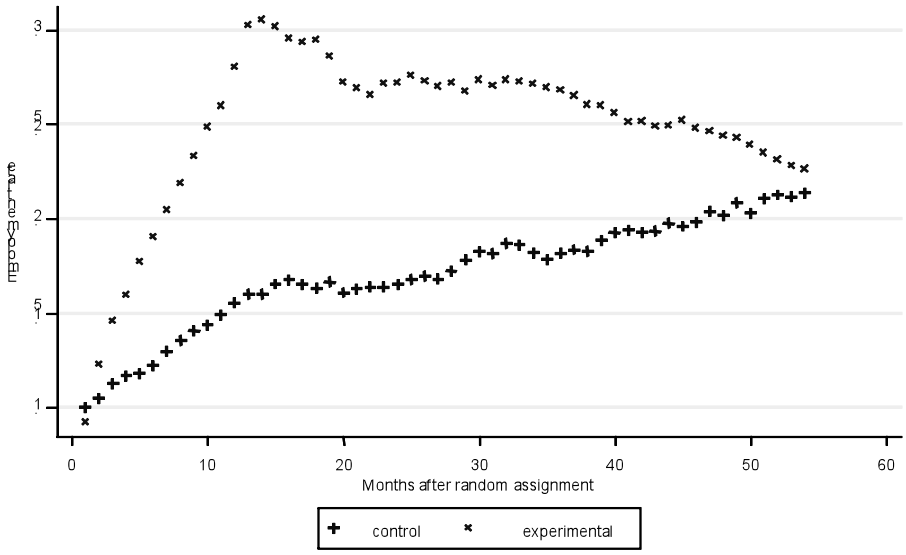


FIGURE 10. SSP, employment rate by months after RA.

(SSP), see Card and Robins (1998). This was designed to answer the question as to whether financial incentives could encourage work among low-skilled lone parents who had spent time on welfare. The aim of the SSP was to encourage employment among this group. The reform consisted of a 50% earnings supplement—effectively a tax credit—for acquiring a job with at least 30 hours per week. This was paid on earnings up to an annual limit of $36,000 and, as in the case of earned income tax credits, the SSP award was provided directly to the individual. The individual had to have at least twelve months on welfare before they could be eligible.

The SSP has many design aspects that are similar to the British tax credits and the US EITC. There were however three big differences—eligibility required at least twelve months on welfare receipt, the eligibility criterion of full-time employment has to be satisfied within twelve months of program entry, and the program only lasted for 36 months. All of this information was fairly well disseminated to the treatment group.

The impact on the employment rate for eligibles and controls is presented in Figure 10 – see Card and Hyslop (2005) for more details and for further references. The increase in monthly employment, at least at the twelve-month window, was substantial. It left no doubt that financial incentives could impact quite strongly on the employment behavior of welfare mothers. However, by the end of the program the employment effect has all but disappeared. We return to this latter point in Section 4.4 in the discussion of the dynamic effects of tax and benefit programs.

5.2.1. Quasi-experiments Quasi-experiments are by definition ex-post analyses. They evaluate the impact of an historic policy through the comparison with a control group who are ineligible but in many other aspects similar to the group targeted by the

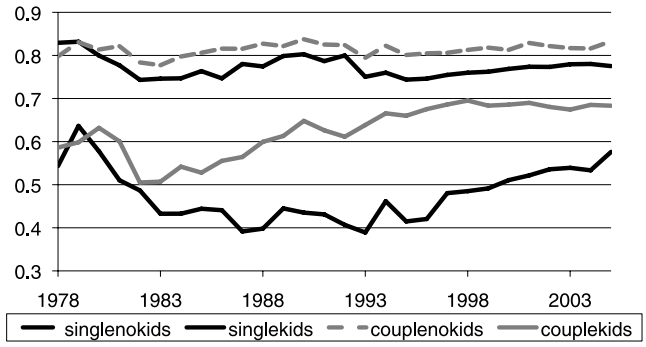


FIGURE 11. The female employment rate in the UK by demographic type. Source: Blundell and Shephard 2010.

policy. Typically, this takes the form of a difference-in-differences analysis, comparing outcomes of eligibles and non-eligibles before and after the reform, estimating *average* impact of the reform on the targeted group—treatment on the treated. As in the case of an experiment, what is learnt is typically only indirectly related to what is needed for optimal design. But again, quasi-experiments form an excellent base from which to (partially at least) validate structural models.

As an example of quasi-experimental validation analysis consider the analysis of earnings taxation among low-wage workers. In particular, the Working Families Tax Credit (WFTC) reforms in the UK described in Section 2. These were aimed at improving net income in work for low wage parents. Our running example will again be tax reform for single mothers. This was a group who, as we saw in Figure 5, could face substantially improved work incentives in WFTC, certainly if their housing benefit and council tax payments were not too large.

The WFTC policy reform was enacted at the end of 1999 and Figure 11 shows the employment rates for working-age women 1978 through 2005. A comparison between single mothers (who are eligible) and single women without children (who are not eligible) is the contrast of direct interest. The large fall in employment in the early 1980s is finally turned around by the late 1990s, somewhat coincident with the WFTC reform. A more systematic analysis would control for observable differences in age, education, etc., between these two groups. This is what a matched difference-in-differences does (see Blundell and Costa-Dias 2009, for example) and is what we will use to validate our structural model in the next section.

Table 2 shows the average marginal impact of the WFTC reform for two data sets: the Labour Force Survey (used in Figure 11) and the smaller, but more detailed, Family Resources Survey. The Family Resources Survey is what we use in policy simulation analysis due to its comprehensive income measure. Both data sets point to a four to five percentage point increase in employment on a base of around 45% for single mothers in comparison to women without children. As we will see, this lines up quite well with the predictions of our preferred structural model specification.

TABLE 2. Average impact of WFTC reform: Percentage employment.

Single mothers	Average effect	Standard error	Sample size
Family Resources Survey	4.5	1.55	25,163
Labour Force Survey	4.7	0.56	233,208

* Notes: Data for Spring 1996 to Spring 2002. Matching covariates: age, education, region, and ethnicity.

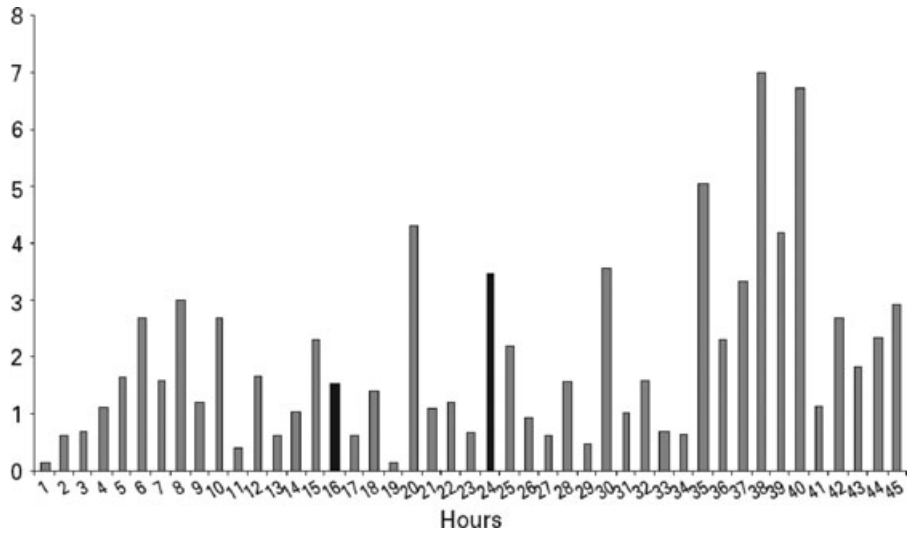


FIGURE 12. The hours distribution for single mothers, before the 16 hour rule.

Of course, the difference-in-differences estimator of the average impact does not come for free. It relies on two key assumptions: a common trends assumptions which states that time-varying unobservable differences must be common across the two groups, and a no-composition-shift assumption which states that cross-section differences in the composition of the two groups should not vary systematically. Both are strong assumptions, even given the matching covariates. But at least these are useful descriptive statistics from which to judge the predictions of any structural model.

Before moving to the structural approach it is worth noting that tax credit reforms do not just change the incentives for employment, they also change incentives for hours of work. In the UK this is especially the case given the minimum weekly hours of work requirement for eligibility. This allows a further quasi-experimental contrast. First we note that the minimum weekly hours eligibility condition was moved from 24 hours to 16 hours in 1992. Figures 12 and 13 provide the histogram of hours before and after reform and show a strong shift towards a spike at 16 hours. The incentive at 16 hours clearly has an impact on behavior (for further comparisons see Blundell and Shephard 2010). Indeed the change in hours of work over this period is clear from Figure 14, and can be seen to follow the hours condition in the tax credit system moving from

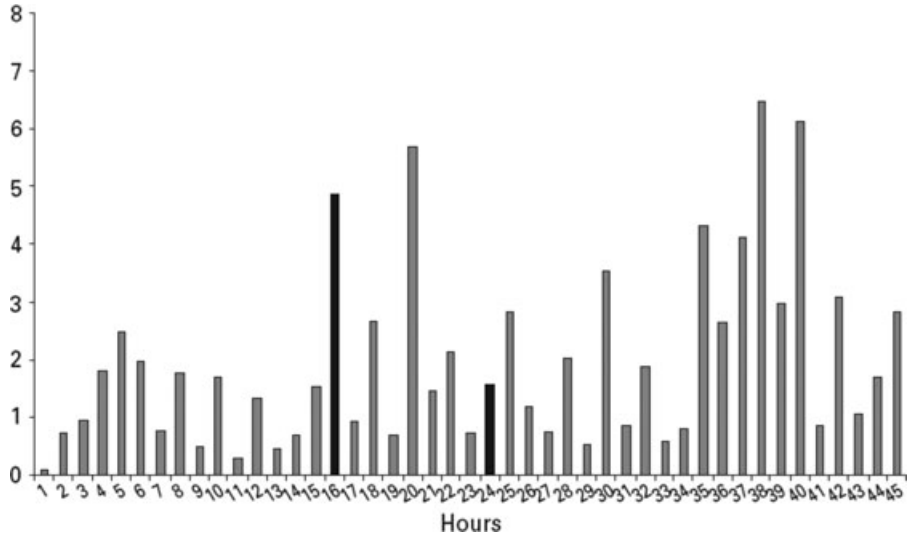
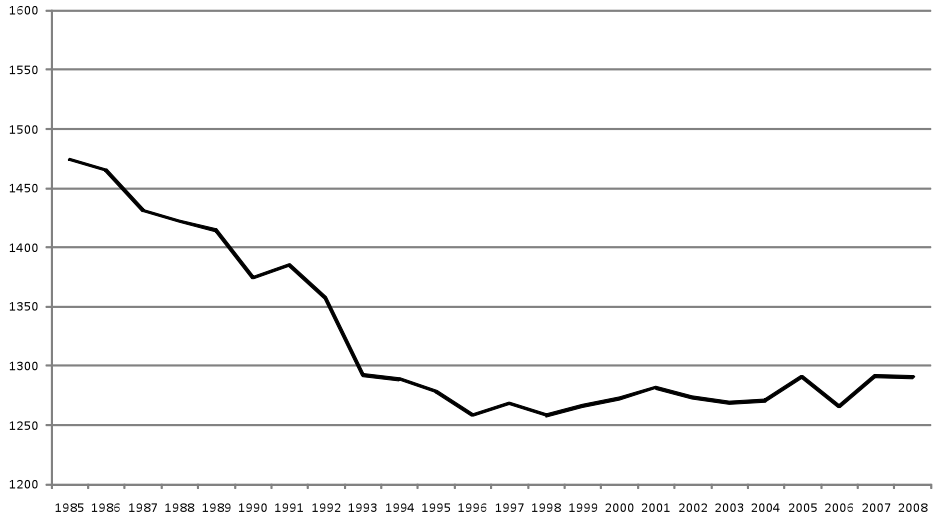


FIGURE 13. The hours distribution for single mothers, after the 16 hour rule. Source: Blundell and Shephard 2010.



Source: Blundell and Shephard (2010)

FIGURE 14. Average annual hours of work: single mothers. Source: Blundell and Shephard 2010.

30 to 24 in the mid-1980s, and then to 16 in 1992. Average hours for working single mothers falls systematically over this period.

We would be unhappy with any structural model that could not reproduce these important contrasts before and after the various policy reforms. This places a stronger

requirement on the model than is often adopted in structural studies. The model not only has to fit the cross-section distribution of hours and employment, it also has to be able to explain the time series *changes* in the distribution of hours and employment induced by changes in the rules of the tax and benefit system.

5.3. Structural Models

The structural modeling approach is useful because it fully specifies all the elements of the choice problem and the budget constraint. It therefore allows us to simulate the impact of actual or hypothetical reforms. It also allows the comparison of reforms in terms of their predicted behavior and deadweight loss. It will be a key component of any optimal tax design too.

As we have already noted, these advantages are also the potential undoing of the structural approach. The specification of constraints and choice probabilities will typically be built on strong assumptions—about the distribution of unobserved heterogeneity, about the budget constraint for each worker, and about the optimizing framework from which choices are made. A complete structural model for hours and employment choices will have to allow for unobserved work-related fixed costs, childcare costs, observed and unobserved heterogeneity, program participation *take-up* costs.

There is now a long history in the specification of such models, see Hoynes (1996), Keane and Moffitt (1998), and van Soest (1995). There are many alternative specifications for preferences, see Blundell and MaCurdy (1999). As an illustration, I follow the model developed in Blundell and Shephard (2010).[5] Section 2 has already detailed the complexity of the budget constraint and Section 3 has stressed the importance of stigma and hassle costs in modeling benefits and tax credits. Following the earlier discussion we focus on the employment and hours decision of a single worker with children and the tax credit reform in the UK. As before, we represent the *heterogeneity* across individuals in observed characteristics by X and in unobserved characteristics by ε. In Section 3 above, we let P be a binary indicator for the take-up of tax credit. The participation cost of taking up is $P\eta(X, \varepsilon_\eta)$ which is subtracted from utility.

We need now to be more specific about the form of *utility* over working hours h and consumption c for each individual. Consider a utility specification:

$$u(c, h, P; X, \varepsilon) = \alpha_y(X, \varepsilon_y)\frac{c^{\theta_y} - 1}{\theta_y} + \alpha_l(X, \varepsilon_l)\frac{(1 - h/H)^{\theta_l} - 1}{\theta_l} - P\eta(X, \varepsilon_\eta), \quad (8)$$

where θ_y and θ_l describe the shape of the marginal utility over consumption and hours respectively. The set of functions $\alpha_y(X, \varepsilon_y)$, $\alpha_l(X, \varepsilon_l)$, and $\eta(X, \varepsilon_\eta)$ capture observed and unobserved preference heterogeneity. H is the total hours available for work and leisure.

5. Background programs and data are avaiable at http://www.ucl.ac.uk/ũctp39a/lect.html, under Section III. Labour Supply and Tax Policy Simulation.

To make estimation tractable we need to make some assumptions over the various definitions and functions in (8). Blundell and Shephard (2010) assume hours of work h are chosen from some finite set $\mathcal{H}$, which in our main empirical results will correspond to the discrete weekly hours points $\mathcal{H} = \{0, 10, 19, 26, 33, 40\}$. These hours points correspond to the empirical hours ranges 0, 1–15, 16–22, 23–29, 30–36 and 37+ respectively.[6] They also set $\ln \alpha_y(X, \epsilon) = X_y' \beta_y + \epsilon_y$ and $\ln \alpha_l(X, \varepsilon_l) = X_l' \beta_l$. This is clearly restrictive but, as we shall see in what follows, it appears to provide a reasonably accurate description of observed employment and hours for single parents in the UK.

As we noted in Section 3, the function $\eta(X, \varepsilon_\eta)$ is included to reflect the possible disutility associated with claiming in-work tax credits ($P = 1$), and its presence allows us to rationalize less than complete take-up of tax credit programs. We denote $P^*(h) \in \{0, E(h; X, \varepsilon)\}$ as the optimal choice of program participation for given hours of work h, where $E(h; X, \varepsilon) = 1$ if the individual is eligible to receive tax credits at hours h, and zero otherwise. Assuming eligibility, it then follows that $P^*(h) = 1$ if and only if

$$u(c(h, P = 1; T, X, \varepsilon), h, P = 1; X, \varepsilon) \geq u(c(h, P = 0; T, X, \varepsilon), h, P = 0; X, \varepsilon).$$

The choice of hours of work h affects consumption c through two main channels. Firstly, through its direct effect on labor market earnings and its interactions with the tax and transfer system; secondly, working mothers purchase childcare for their children which varies with maternal hours of employment.

Individuals face a budget constraint, determined by a fixed gross hourly wage rate (generated by a log–linear relationship of the form $\log w = X_w' \beta_w + \varepsilon_w$) and the tax and transfer system $T(wh, \boldsymbol{h}, P; X)$. Non-labor income, such as child maintenance payments, enter the budget constraint through the dependence of the tax and transfer schedule T on demographic characteristics X. To arrive at a measure of consumption c, in the absence of saving, we subtract both childcare expenditure and fixed work-related costs, $\alpha_f(X, \varepsilon_f) \times \mathbf{1}_{\{h>0\}}$, from net income, $wh - T(wh, \boldsymbol{h}, P; X)$.

The hours of childcare h_c is assumed to vary stochastically with hours of work and demographic characteristics. Total weekly childcare expenditure is then given by $p_c h_c$, with p_c denoting the hourly price of childcare. Empirically, this is modeled by assuming that p_c follows some distribution $p_c \sim F_c(\cdot; X_c)$ which varies with demographic characteristics.

Incorporating endogenous take-up of tax credits through cost $\eta(X, \varepsilon_\eta)$, it then follows that the optimal choice of hours $h^* \in \mathcal{H}$ maximizes

$$U(c(h, P^*(h); T, X, \varepsilon), h, P^*(h); X, \varepsilon),$$

subject to the various constraints already detailed.

This brief outline of the key features of a structural model illustrates some of the key ingredients and assumptions required, and this for a single worker decision.

6. In the Supplementary Material to the Blundell and Shephard paper the sensitivity of the results to a finer discretization of weekly hours are examined; the main results appear robust.

Family labor supply models require further assumptions, in particular the modeling of joint hours and employment choices, see Blundell et al. (2000) and Brewer, Duncan, and Shephard (2007). Blundell and MaCurdy (1999) and Meghir and Phillips (2010) give further insights in to structural modeling of labor supply.

Blundell and Shephard (2010) take this structural model specification to the UK Family Resources Survey data and argue that it does a good job of describing observed behavior. For example, the model is used to simulate the WFTC reform. This is then compared with the simulated average response with the quasi-experimental estimate described in Table 1. The simulated difference-in-differences parameter from the structural evaluation model is precise and does not differ significantly from the difference-in-differences estimate itself.

5.4. Dynamics and Frictions?

Finding that a structural model does a reasonable job of predicting the changes of reforms over time does not imply that it is the correct model, simply that it is not rejected. One area where we might expect the model to perform poorly is in capturing the dynamics of labor supply. In particular, experience effects and adjustment frictions. In their study of the hours and employment changes around the tax credit reforms in the UK, Blundell, Brewer, and Francesconi (2008) already note that changes seem to take place over a relatively short time, within a year or two, suggesting relatively small adjustment costs. Of course, for the most part, these reforms occurred in a period of economic expansion and we would not expect our structural model to provide a good description of labor supply choices during a recession. Moreover, these reforms were large and well announced. Smaller tax reforms may have less impact, see Chetty et al. (2009).

But what about experience effects? That is the dynamic pay-off in terms of earnings and employment of being in the labor market. For some evidence on this for low-skill workers we turn to the Canadian SSP. Figures 15 and 16 from the previously described Canadian SSP control-trial experiment tell an interesting story. Recall that the SSP program is complete after 36 months from eligibility, and that eligibility can take up to twelve months. By 50 months from baseline all the treatment group would have completed the program. Indeed we saw, in Figure 10, that the treated had, on average, a sizable increase in employment. But did this translate into higher wages or earnings after the program was finished? Not according to these figures. There is little noticeable difference after 50 months in hourly wages or earnings, see also Card and Hyslop (2005). It seems that experience effects, for this group at least, are negligible. In general, they may well be small for low-skill workers, explaining why relatively simple models that ignore experience effects explain labor supply adjustments quite well for these types of workers.

6. Implications from Theory for Tax Design

We have argued earlier that there are three key ingredients to any optimal tax analysis: the accurate measurement of response elasticities, the detailed description of the

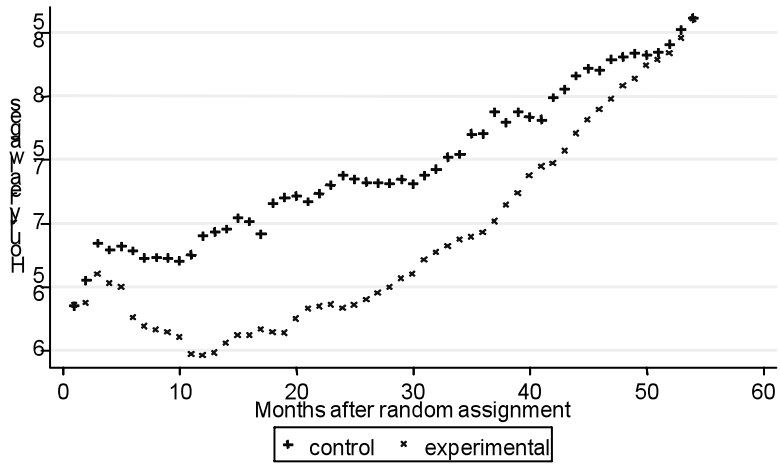


FIGURE 15. SSP, hourly wages by months after RA.

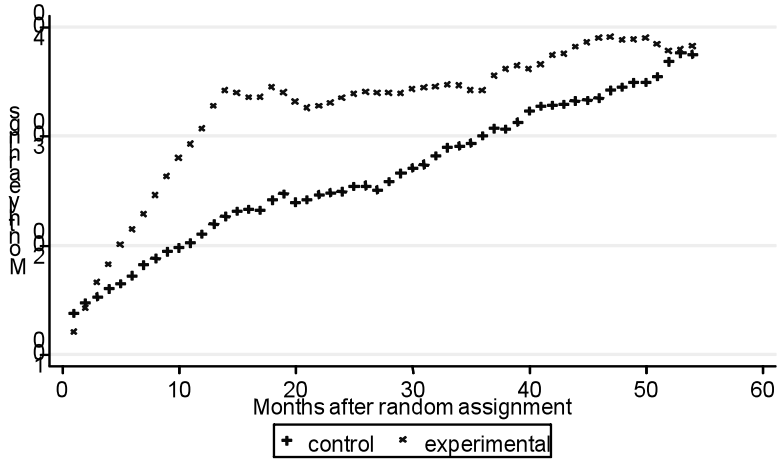


FIGURE 16. SSP, monthly earnings by months after RA.

distribution of income, and some view of social welfare weights. The first two of these are positive and can be learned from a careful evidence-based analysis. The last is normative and therefore something where opinions may differ. In this section we will see how these come into play in the design of earnings taxation.

An important distinction we took from our discussion of labor supply responses in sections 2 and 5 is between the extensive and intensive margins of labor supply. Why is this also important for tax design? The optimal taxation literature explores the consequences for tax design (e.g. Diamond 1980; Saez 2002; Laroque 2005). Once individuals are allowed to respond to changes in the tax schedule by deciding whether or not to work, as well as how hard to work, then the optimal tax schedule can change dramatically. In particular, when this employment decision becomes relatively more

important, optimal marginal tax rates can be lower (and perhaps even negative) for those with low earnings capacity. As Brewer, Saez, and Shephard (2010) note:

> a striking implication is that, if the government values redistribution then the participation tax rate should be negative for low earnings—in other words, low-income workers should receive an earnings subsidy. Hence, in sharp contrast to the intensive model, the extensive model implies that earnings subsidies or work-contingent credits (such as the earned-income tax credit or the working tax credit) should be part of an optimal tax system.

This is one of the key lessons from recent optimal tax design. A *large* extensive elasticity can *turn around* the impact of declining social weights, implying a higher transfer to low-wage workers than those out of work and a role for earned-income tax credits.

A further key consideration in tax design is the way in which responses differ across individuals of different characteristics. Unless there are good redistributive reasons to do otherwise, tax rates will generally be lower on those types of individuals with more elastic responses.

Finally, the degree of inequality and of income uncertainty will also matter for earnings tax design. The past three decades or more has seen strong growth in earnings inequalities and a change in the nature of earnings risks. The redistributive element of the earnings tax and benefit system acts, in part, as an insurance to earnings risks. As the nature of these risks changes and as underlying inequality grows, the balance between inequality and work incentives gets harder to balance. Designing an efficient structure to the earnings tax and benefit system such that it achieves the desirable distributional objectives becomes ever more salient.

6.1. Optimal Design for Low Income Workers

How should we think about an optimal design? We will assume that the government seeks to maximize *social* welfare subject to revenue constraints. Following on from our running illustration of tax design for low-income single mothers, assume we want to redistribute '£R' to this group. What is the *optimal* way to do this?

Our aim will be to recover optimal tax/credit schedule in terms of earnings. There are two related approaches. The first to use the Diamond–Saez approximation in terms of extensive and intensive elasticities at different earnings. The second approach involves a *complete* Mirrlees optimal tax computation requiring a complete specification of choices and constraints.

The first approach is exemplified in the work of Saez (2002). He provides an intuitive expression for the tax rate schedule in terms of the extensive and intensive elasticities, and simple summary measures of the distribution of earnings and the social welfare weights. The formula is only approximate and assumes away income effects, however as a guide to the setting of tax rates it is extremely informative, see Immervoll et al. (2005). We return to these optimal tax formulas in our discussion of the optimal top rate of tax in what follows.

The latter approach might be labeled the *structural microeconometric approach* to tax design—effectively, a stochastic mechanism design problem. In this case the optimal tax model *is* the labor supply model. Consequently, all of the assumptions concerning behavior are also required for this analysis. The distribution of earnings, fixed costs of work, childcare, demographic differences, and unobserved heterogeneity described in the previous section all influence choice of tax rate schedule.

Suppose we assume earnings (and certain characteristics) are all that is observable to the tax authority.[7] Social welfare is represented by the sum of transformed utilities. Social welfare W is represented by the sum of transformed utilities:

$$W = \int_{X,\epsilon} \int_{\varepsilon} \Upsilon(U(c(h^*; T, X, \varepsilon), h^*; X, \varepsilon)) dF(\varepsilon) dG(X, \varepsilon),$$

where function Υ captures redistributive preferences.[8]

The government seeks to maximize W subject to revenue constraint:

$$\int_{X,\epsilon} \int_{\varepsilon} T(wh^*, \boldsymbol{h^*}; X) dF(\varepsilon) dG(X, \varepsilon) \geq \bar{T}(\equiv -R).$$

Blundell and Shephard (2010) control the preference for equality in social welafre by the transformation function

$$\Upsilon(U|\theta) = \frac{1}{\theta} \times \{(\exp U)^{\theta} - 1\} \tag{9}$$

where $\theta < 0$ favors equality of utilities.

The objective is to find robust tax rate schedules for fairly general social welfare weights. Given the structural parameter estimates, we can solve for optimal schedules. In their application Blundell and Shephard restrict to piecewise linear tax schedule (out-of-work income, nine marginal rates at breaks of £50 up to £400), with possible hours contingent payments.

The key findings of the Blundell–Shephard analysis, under range of values for θ (i.e. allowing different degrees of preference for inequality), are that optimal marginal rates are broadly increasing in earnings for all groups. The results also point to a shift of out of work support towards families with younger children. This suggests an optimal tax schedule with tagging according to age of children. Moreover, pure *tax credits* are found to be optimal at low earnings but only for those with school-aged children.

The analysis also found that hours contingent payments can improve design: if hours are accurately observed a full-time bonus is desirable for low wage mothers with older children. But measurement error and the possibility of hours manipulation are found to weaken the argument for hours rules.

In the *Review* recommendations we not only stress reforms for lone parents, but also for married parents, and older workers pre-retirement. From the examination of response elasticities we recommended 'tagging' tax rates by age of (youngest) child for mothers/parents and also at pre-retirement ages.

7. Blundell and Shephard (2010) relax this to allow for *partial* observability of hours to capture the minimum hours conditions in the British tax credit system.

8. In this design problem we assume full take-up and do not include the stigma model used in estimating the preference parameters.

6.2. Tax Rates at the Top

At the top of the income distribution we take a rather different approach. Hours and employment may not be the only, nor the most important, ways to change earnings in response to tax changes. When it comes to the taxation of top incomes, concerns about the tax base come back into play. Feldstein (1995, 1999) makes a convincing case for looking directly at taxable income. The more opportunities for exemptions and deductions and the possibility to pass income through other lower tax jurisdictions, the more difficult it is to raise revenue from the top income earners. Consequently, we require a more general elasticity measure that captures these other avenues for response. The taxable income elasticity does just that.

A higher tax rate on a smaller base will raise less revenue and will probably be harder to sustain. To quote Slemrod and Kopczuk (2002):

> When personal tax rates on ordinary income rise, evasion may increase, businesses may shift to corporate form, there may be a rise in the consumption of deductible activities such as charitable giving, and individuals may rearrange their portfolios and compensation packages to receive more income as tax-preferred capital gains. These responses to higher taxes, and all others, will show up in declines in taxable income, and there is a growing body of evidence, that, at least for high-income individuals, the elasticity of taxable income to the marginal tax rate is substantial.

It is hardly surprising therefore to find that the responsiveness of taxable income to the tax rate is a key parameter on which the setting of top tax rates depends. What we have to bear in mind is that the responsiveness itself will be affected by the tax base. This elasticity can be expected to be larger the narrower the tax base. Given the need to capture all these margins and the fact that effort is very hard to quantify, the behavioral effect will require a different kind of measurement from that used to gauge hours and employment responses.

Consider an *optimal* top tax rate and suppose the welfare weight on top bracket incomes is negligible. The optimal rate will be the revenue maximizing rate—the Laffer rate. We first note that the top of the taxable income distribution is well approximated by a Pareto distribution. Figure 17 shows this to be the case. Suppose we write e as the taxable income elasticity, then Brewer, Saez, and Shephard (2010) show that the revenue maximizing rate is given by

$$t = 1/(1 + a \cdot e),$$

where a is the Pareto parameter. For the UK, a is approximately 1.67, see Brewer and Browne (2009).

To estimate e reliably is fraught with difficulties. Typically a difference-in-difference methodology is used, see Gruber and Saez (2002). When this approach is applied to past changes in tax rates among the top 1% in the UK, using the 2%–5% group as a control, the evidence to the *Mirrlees Review* suggested a preferred estimate of e of 0.46 with a standard error of 0.13.

Exploring various formulations of the differences-in-differences specification for the UK, the estimate of e remains in the 0.35–0.55 range with a central value around

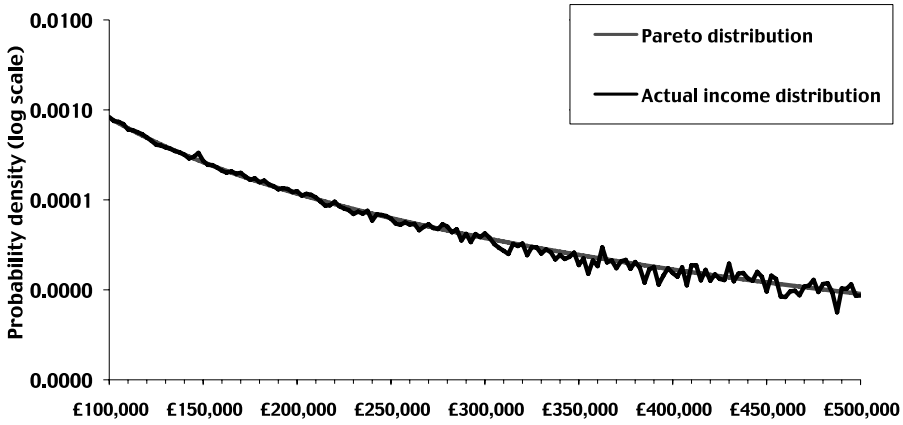


FIGURE 17. The Pareto distribution and the taxable income distribution at the top.

0.45, but is clearly quite fragile. An estimate of this magnitude would suggest the optimal top 1% bracket rate of around 57%, close to the current top rate (including indirect taxes). This analysis suggests little room for any further raising of the top rate of income tax in the UK without changes to the tax base for earned income itself.

7. Work Incentives, Redistribution and Base-Broadening Reforms

Earnings taxation plays a major role in getting the structure of tax design right across the whole tax system. For example, a major recommendation from the *Review* is a broadening of the VAT tax base. This is especially important in the UK which has one of the narrowest bases for VAT and has also just raised the standard rate to 20%. With many goods (food at home, children's clothing, financial services) being zero rated or exempt, the potential welfare cost created by differential commodity taxation is large. Indeed, empirical evidence suggests that current indirect tax rates do not line up with any reasonable justification. But broadening the base of VAT is not trivial. On its own it would be regressive and would harm work incentives. To reap welfare gains from base-broadening, distributional concerns have to be addressed and reduced work incentives redressed. Both of these require the careful redesign of earnings taxation. Earnings taxation becomes an integral part of the VAT reform.

Three key empirical observations about consumer behavior come into play when thinking about justifying differential rates of VAT. First, some commodities are luxuries and some are necessities. Differential commodity taxation can act as a redistributive mechanism. But they are an inefficient way of delivering redistribution given the other direct (earnings) tax instruments available. Secondly, nonseparabilities with labor supply are important. Certainly this is a key justification in the optimal tax literature for differential commodity taxes with goods that are complementary to work bearing a heavier tax. Empirically these relate mainly to childcare and work-related expenditures.

Current VAT rates are quite different from these, see Crawford, Keen, and Smith (2010) for example. Finally, price elasticities differ with total expenditure and wealth. That is responses, and therefore welfare costs, differ across the income distribution. To value welfare losses and calculate compensation these microeconomic differences matter a lot.

The welfare gain from broadening the VAT is based on assumptions that are unlikely to hold in reality: weak separability between goods and leisure, common preferences, and competitive pricing by suppliers. Nonetheless, a broad uniform base seems likely to be a good baseline from which to judge reform. Nonseparabilities are clearly evident but mainly in relation to work-related expenses. Preferences are heterogeneous but often differ by characteristics that are in the tax system, like family composition and family income. Differentials in tax rates across commodities seem to be more motivated by redistributive concerns and the power of certain pressure groups: with food and children's clothing belonging to the first and financial services to the second.

On its own, the base broadening of VAT in the UK would be regressive and weaken work incentives. Can a practical package avoid this? *Tax by Design* provides an illustration of how this can be done, implementing a reform package that achieves compensation while also avoiding significant damage to work incentives. Working with the existing set of UK direct tax and benefit instruments, the *Review* simulated removing almost all zero and reduced rates in the UK. This raises £24bn (with a 17.5% VAT rate) if no behavioral response. With responses, the empirical results suggest (in principle) every household could be compensated and still leave a £3bn to £5bn welfare gain.

Summary results from the *Review* show the key interaction between earnings taxation and the base-broadening reform. These are summarized in Figures 18, 19 and 20. Turning first to distributional concerns, Figure 18 shows that average percentage rises in the cost of living by income decile are more than compensated for by increases in income. But what about work incentives? Figures 19 and 20 show that effective tax rates on earnings are also left largely unchanged at both the extensive and intensive margin.

Other base-broadening reforms also require care in their interactions with earnings taxation. For example, in the discussion of capital taxation, the *Review* suggests moving towards an expenditure tax treatment of saving by providing an allowance for income saved which effectively eliminates the tax on the normal return to saving. This provides a framework for the integration of capital income taxation with corporate taxation. Exempting the normal rate also gives neutrality between debt and equity capital gains and dividends are treated in the same way and the resulting capital tax structure overcomes the *lock-in* incentive from Capital Gains Tax. However, this also requires alignment between tax rates on corporate income, shareholder income, and labor income. With progressive tax rates on labor income, progressive rates are also required on shareholder income to avoid differential tax treatments of incorporated and unincorporated firms—a lower progressive rate structure on shareholder income than on labor income reflecting the corporate tax already paid.

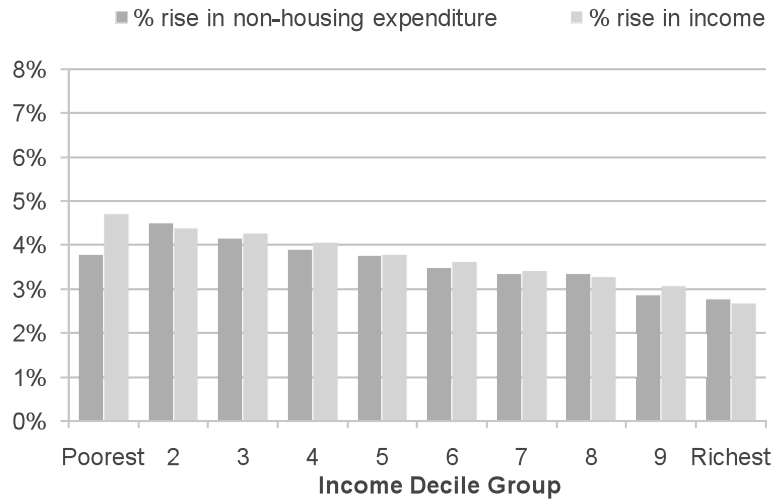


FIGURE 18. VAT reform, effects by income decile.

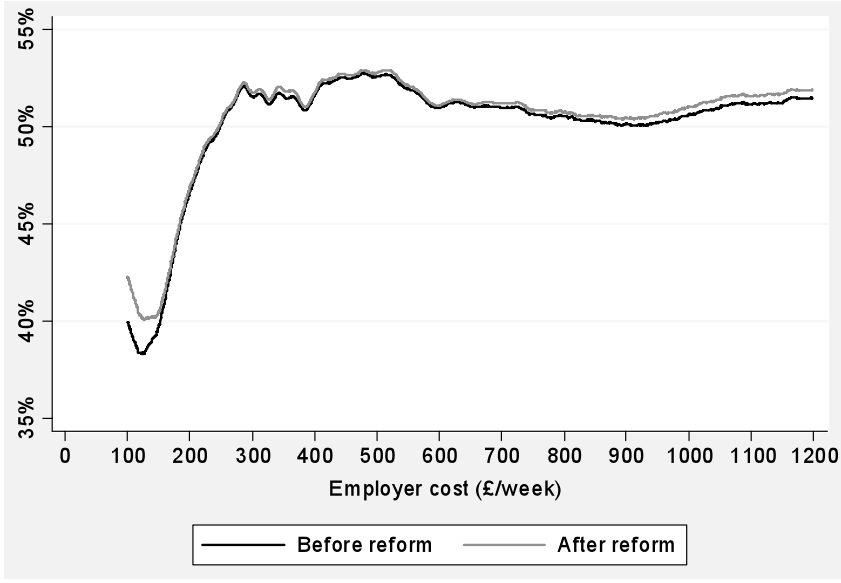


FIGURE 19. VAT reform, impact on participation tax rates.

8. Summary and Conclusions

In this paper I have used the *Mirrlees Review* as an illustration of the use of evidence in the development of a tax reform program. In developing the recommendations for the reform, the *Review* attempted to draw on empirical evidence wherever possible.

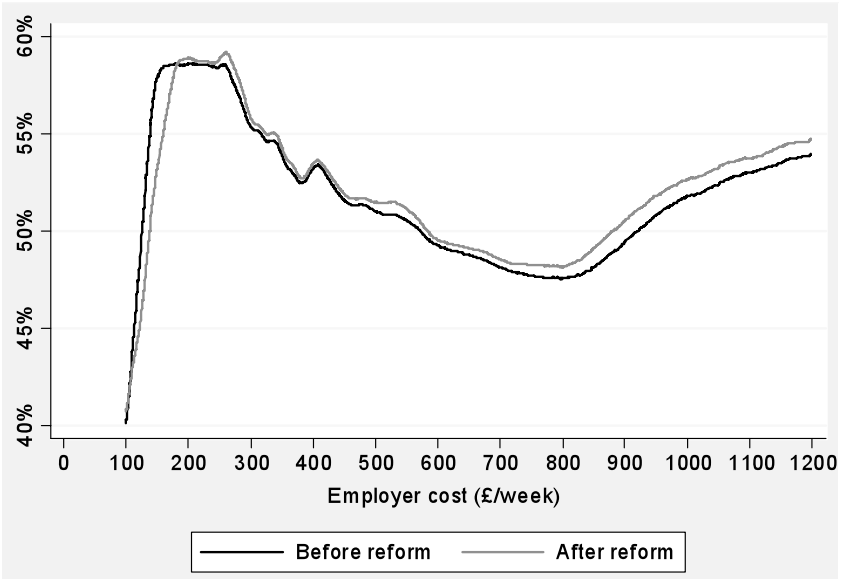


FIGURE 20. VAT reform, impact on effective marginal tax rates.

The aim here was to show how to make the best use of all available evidence, from broadly descriptive evidence to that gleaned from quasi/experimental evaluations and also from structural model estimation.

Much of the discussion has focused on earnings tax design, partly for illustrative reasons but also because there is large body of evidence on labor supply and taxable income responses to tax/benefit reform and to policy design. Wherever possible we have argued for the use of well specified and carefully validated structural microeconometric models as the basis for design. In cases where we have less in the way of detailed structural models to draw on, we have to use more general information on likely size response elasticities. In this we follow closely the lead in Saez (2002) and Brewer, Saez, and Shephard (2010).

Across the board, we have documented a key role for labor supply responses at the extensive and intensive margins. Both matter but differ by gender, age, education, and family composition. We found that labor supply responses for families with children vary by age of the youngest child. We also found different responses for older workers in *pre-retirement* years.

The results of our analysis suggested changing the rate structure to match lessons from *new* optimal tax analysis. It pointed to lower marginal rates at the bottom of the earnings distribution. Means-testing should be less aggressive, at least for some key groups. Tax credits should be better targeted to lower incomes and to the families where labor supply is most responsive. We particularly stress reforms for lone parents, married parents, and older workers pre-retirement. From the examination of

response elasticities we recommended 'tagging' tax rates by age of (youngest) child for mothers/parents and also at pre-retirement ages.

When put together, this reform agenda can be interpreted through a life-time view of taxation, implying a *life-cycle* rearrangement of tax incentives and welfare payments to match elasticities and early years investments, effectively redistributing across the life-cycle, distinguishing by age of (youngest) child for mothers/parents and at pre-retirement ages. The simulation results in *Tax by Design* suggested significant employment and earnings increases from such a reform package. The evidence on taxable income elasticities implies limits to tax rises at the top of the income distribution. Tax reforms for this group are better directed towards base broadening to address tax avoidance and revenue shifting.

Finally, we noted how earnings taxation is also called on to undo the impact on distributional and work incentives of the rest of any tax reform package.

References

Adam, Stuart, James Browne, and Chris Heady (2010). "Taxation in the UK." In *Dimensions of Tax Design: The Mirrlees Review*, edited by James Mirrlees, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba. Oxford University Press for Institute for Fiscal Studies.

Atkinson, Anthony and Thomas Piketty (2007). *Top Incomes over the Twentieth Century: A Contrast between European and English-Speaking Countries*. Oxford University Press.

Ashenfelter, Orley and Mark W. Plant (1990). "Nonparametric Estimates of the Labor Supply Effects of Negative Income Tax Programs." *Journal of Labor Economics*, 8, S396–415

Besley, Tim and Steve Coate (1992). "Workfare versus Welfare: Incentive Arguments for Work Requirement in Poverty Alleviation Programs." *American Economic Review*, 82(1), 249–261.

Blau, Fran and Larry Kahn (2007). "Changes in the Labor Supply Behavior of Married Women: 1980–2000." *Journal of Labor Economics*, 25, 393–438.

Blundell, Richard (2006). "Earned income tax credit policies: Impact and Optimality." The 2005 Adam Smith Lecture. *Labour Economics*, 13, 423–443.

Blundell, Richard, Antoine Bozio, and Guy Laroque (2011a). "Labour Supply and the Extensive Margin." *American Economic Review*, 101(3), 482–486.

Blundell, Richard, Antoine Bozio, and Guy Laroque (2011b). "Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France." IFS Working Paper 11-01.

Blundell, Richard, Mike Brewer, and Marco Francesconi (2008). "Job Changes, Hours Changes And The Path Of Labour Supply Adjustment." *Journal of Labor Economics*, 26, 421–453.

Blundell, Richard and Monica Costa-Dias (2009). "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources*, 44, 565–640.

Blundell, Richard, Alan Duncan, and Costas Meghir (1998). "Estimating Labour Supply Responses using Tax Policy Reforms." *Econometrica*, 66, 827–861.

Blundell, Richard, Alan Duncan, Julian McCrae, and Costas Meghir (2000). "The Labour Market Impact of the Working Families' Tax Credit." *Fiscal Studies*, 21(1) 75–103.

Blundell, Richard and Hilary Hoynes (2004). "In-Work Benefit Reform and the Labour Market." In *Seeking a Premier League Economy*, edited by Richard Blundell, David Card, and Richard B. Freeman. University of Chicago Press.

Blundell, Richard and Thomas MaCurdy (1999). "Labour Supply: A Review of Alternative Approaches." In *Handbook of Labour Economics*, edited by Orley Ashenfelter and David Card. Elsevier.

Blundell, Richard, Costas Meghir, and Sarah Smith (2002). "Pension Incentives and the Pattern Of Early Retirement." *Economic Journal*, 112, C153–170.

Blundell, Richard, and Andrew Shephard (2010). "Employment, Hours of Work and the Optimal Taxation of Low Income Families." IFS Working Paper W08/01. Forthcoming in *Review of Economic Studies*.

Brewer, Mike (2003). "The New Tax Credits." IFS Briefing Note No. 25, www.ifs.org.uk

Brewer, Mike, Alan Duncan, and Andrew Shephard (2007). "Did Working Families' Tax Credit Work? The Impact of In-Work Support on Labour Supply in Great Britain." *Labour Economics*, 13, 699–720.

Brewer, Mike, Emmanuel Saez, and Andrew Shephard (2010). "Optimal Household Labor Income Tax and Transfer Programs." In *Dimensions of Tax Design: The Mirrlees Review*, edited by James Mirrlees, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba. Oxford University Press for Institute for Fiscal Studies.

Brewer, Mike and James Browne (2009). "Can More Revenue be Raised by Increasing Income Tax Rates for the Very Rich?" IFS Briefing Note No. 84.

Card, David, Stefano DellaVigna, and Ulrike Malmendier (2011). "The Role of Theory in Field Experiments." *Journal of Economic Perspectives*, 25(3), 39–62.

Card, David and Philip K. Robins (1998). "Do Financial Incentives Encourage Welfare Recipients To Work?" *Research in Labor Economics*, 17, 1–56.

Card, David and Dean Hyslop (2005). "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers." *Econometrica*, 73, 1723–1770.

Chetty, Raj (2008). "Sufficient Statistics for Welfare Analysis: a Bridge between Structural and Reduced-Form Methods." National Bureau of Economic Research Working Paper 14399.

Chetty, Raj (2009). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." National Bureau of Economic Research Working Papers 15616.

Chetty, Raj, John N. Friedman, Tore Olsen, Luigi Pistaferri, and Anders Frederiksen (2009). "The Effect of Adjustment Costs and Institutional Constraints on Labor Supply Elasticities: Evidence from Denmark." Working paper, Harvard University.

Cogan, John (1981). "Fixed Costs and Labor Supply." *Econometrica*, 49, 945–964.

Crawford, Ian, Michael Keen and Stephen Smith (2010). "Indirect Taxation and VAT." In *Dimensions of Tax Design: The Mirrlees Review*, edited by James Mirrlees, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba. Oxford University Press for Institute for Fiscal Studies.

Diamond, Peter (1980). "Income Taxation with Fixed Hours of Work." *Journal of Public Economics*, 13, 101–110.

Eissa, Nada and Jeffrey Liebman (1996). "Labor Supply Response to the Earned Income Tax Credit." *Quarterly Journal of Economics*, 111, 605–637.

Feldstein, Martin S. (1995). "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy*, 103, 551–572.

Feldstein, Martin S. (1999). Tax Avoidance and the Deadweight Loss of the Income Tax. *Review of Economics and Statistics*, 4, 674–680.

Gruber, Jon and David Wise (1999). *Social Security and Retirement around the World*. NBER/The University of Chicago Press.

Gruber, Jon, and Emmanuel Saez (2002). "The Elasticity of Taxable Income: Evidence and Implications." *Journal of Public Economics*, 84, 1–32.

Heckman, James J. (1974). "Shadow Prices, Market Wages, and Labor Supply." *Econometrica*, 42, 679–694.

Heckman, James J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica*, 47, 153–161.

Heckman, James J. (1993). "What Has Been Learned About Labor Supply in the Past Twenty Years?" *American Economic Review*, 83(2), 116–121.

Heckman, James J and Jeffrey A. Smith (1995). "Assessing the Case for Social Experiments." *Journal of Economic Perspectives*, 9(2), 85–110.

Hoynes, Hilary (1996). "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation under AFDC-UP." *Econometrica, Econometric Society*, 64(2), 295–332.

Hurd, Michael (1976). "The Estimation of Nonlinear Labor Supply Functions with Taxes from a Truncated Sample." Center for the Study of Welfare Policy Research Memorandum 36, Stanford Research Institute.

Johnson, Terry R. and John H. Pencavel (1980). "Utility-Based Hours Of Work Functions for Husbands, Wives, and Single Females Estimated from Seattle–Denver Experimental Data." Socioeconomic Research Center Research Memorandum 71, SRI International.

Immervoll, Herwig, Henrik Kleven, Claus Kreiner, and Emmanuel Saez (2005). "Welfare Reform in European Countries: A Microsimulation Analysis." *Economic Journal*, 117, 1–44.

Keane, Mike and Robert Moffitt (1998). "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review*, 39, 553–589.

Kopczuk, Wojciech (2005). "Tax Bases, Tax Rates and the Elasticity of Reported Income." *Journal of Public Economics*, 89, 2093–2119.

Laroque, Guy (2005). "Income Maintenance and Labor Force Participation." *Econometrica*, 73, 341–376.

Liebman, Jeff (2002). "The Optimal Design of the Earned Income Tax Credit." In *Making Work Pay: The Earned Income Tax Credit and Its Impact on American Families*, edited by Bruce Meyer and Douglas Holtz-Eakin. Russell Sage Foundation, New York.

Meade, James (1978). "The Structure and Reform of Direct Taxation: Report of a Committee Chaired by Professor J E Meade for the Institute for Fiscal Studies. George Allen and Unwin, London (http://www.ifs.org.uk/publications/3433).

Meghir, Costas and David Phillips (2010). "Labour Supply and Taxes." In *Dimensions of Tax Design: The Mirrlees Review*, edited by James Mirrlees, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba. Oxford University Press for Institute for Fiscal Studies.

Mirrlees, James A. (1971). "The Theory of Optimal Income Taxation." *Review of Economic Studies*, 38, 175–208.

Mirrlees, James, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba (eds) (2010). *Dimensions of Tax Design: The Mirrlees Review*. Oxford University Press for Institute for Fiscal Studies.

Mirrlees, James, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba (eds) (2011). *Tax by Design: The Mirrlees Review*. Oxford University Press for Institute for Fiscal Studies.

Moffitt, Robert (1983). "An Economic Model of Welfare Stigma." *American Economic Review*, 73(5), 1023–1035.

Moffitt, Robert (2005). "Welfare Work Requirements with Paternalistic Government Preferences." Working paper, Johns Hopkins University.

Ohanian, Lee, Andrea Raffo, and Richard Rogerson (2008). "Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956–2004." *Journal of Monetary Economics*, 55, 1353–1362.

Ed Prescott (2004). "Why Do Americans Work So Much More than Europeans?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 28, 2–13.

Rogerson, Richard and Johanna Wallenius (2009). "Micro and Macro Elasticities in a Life Cycle Model With Taxes." *Journal of Economic Theory*, 144, 2277–2292.

Saez, Emmanuel (2002). "Optimal Income Transfer Programs : Intensive Versus Extensive Labor Supply Responses." *Quarterly Journal of Economics*, 117, 1039–1073.

Stern, Nicolas (1986). "On the Specification of Labor Supply Function." In *Unemployment, Search and Labour Supply*, edited by R. Blundell and I. Walker. Cambridge University Press, pp. 121–142.

Slemrod, Joel and Wojciech Kopczuk (2002). "The Optimal Elasticity of Taxable Income." *Journal of Public Economics*, 84, 91–112.

Arthur van Soest (1995). "Structural Models of Family Labor Supply: A Discrete Choice Approach." *Journal of Human Resources*, 30, 63–88.

West, Richard W. (1979). "The Effects of the Seattle and Denver Income Maintenance Experiments on the Labor Supply of Young Nonheads." Center for the Study of Welfare Policy Research Memorandum 60, SRI International.