Empirical Evidence and Tax Design[*]

Richard Blundell[†]

August 2010, Revised December 2011

Abstract

This paper examines the role of evidence in drawing up the recommendations for tax reform in the Mirrlees Review. The arguments are organised 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, and (v) Implications from theory for tax design. Although the Mirrlees Review focusses on all aspects of tax reform, the focus is this paper is on the taxation of earnings with some examples drawn from the taxation of consumption and savings.

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 build a base for tax reform from the large body of economic theory and empirical evidence. It was inspired by the Meade Report (1978) with 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 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 keep consistency and coherence in the discussion the focus here will be on the taxation of earnings although the *Review* itself concerned all aspects of the tax system. The discussion is organised loosely under five related headings:

(i) Key margins of adjustment

(ii) Measurement of effective tax rates

(iii) The importance of information and complexity

[*]I would like to thank the editor Manuel Arellano and participants at the World Congress panel session for comments. I am also grateful to Antoine Bozio, Andrew Shephard, Guy Laroque, and my collegues on the Mirrlees Review for their many discussions over the course of the Review and for allowing me to draw on joint work.

[†]University College London and Institute for Fiscal Studies, email: r.blundell@ucl.ac.uk.

(iv) Evidence on the size of responses

(v) Implications from theory for tax design.

The main sections of this paper reflect these five aspects.

The first of these headings highlights the importance of establishing empirical facts about key aspects of behaviour 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. We 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.

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 where 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 they are declining in some measure of equivalised 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 labour 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 labour 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.

All 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 many of the tax reform proposals. Through both the need to adjust for changes in redistribution *and* to adjust for changes in work incentives induced by other aspects of proposed reforms. Broadening the base of VAT is a good example of this, see chapter 9 of *Tax by Design*. Redistribution and work incentives are key aspects of delivering efficiency improvements

of tax-base broadening reforms while compensating potential losers. Nonetheless, we can derive overall directions for the taxation of earnings, leaving the precise tax rate schedule on earnings will depend on the shape of the whole tax system.

1 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 labour supply responses is between the extensive (whether to work) and intensive (how much to work) margins of labour 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 US. 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. To see this Figure 1, from Blundell, Bozio and Laroque (2011b), provides a broad view of employment rates by age for the UK, France and the UK 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), Rogerson and Ohanian (2008), 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 these differences, in any discussions of tax reform it would seem unwise to play down the intensive margin too much.[1]

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 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

---

[1]See also Chetty (2009).

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.
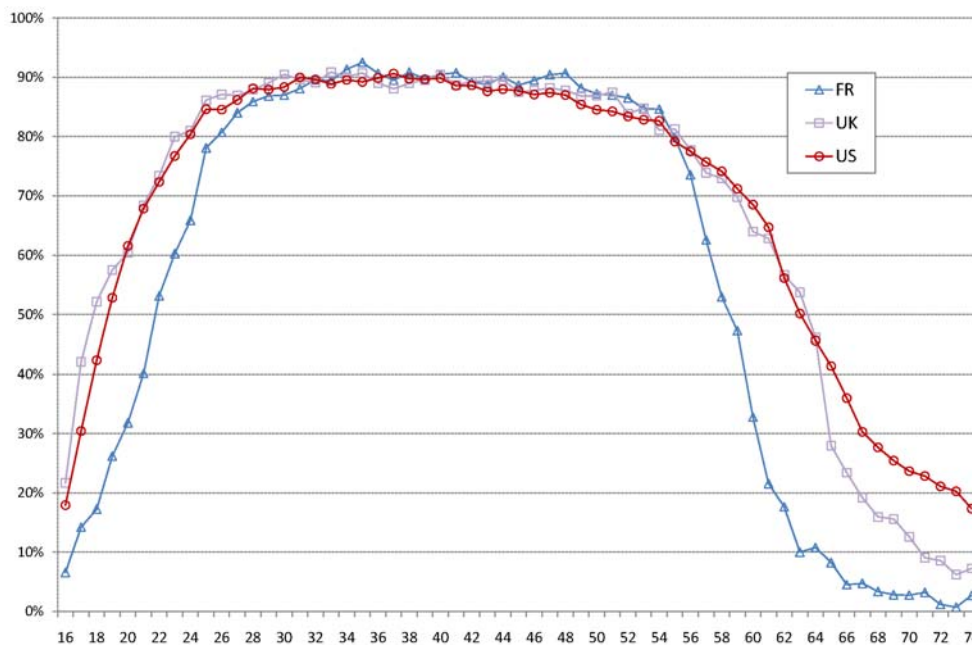

Figure 1: Employment: Men 2007

In the sections below we focus with a little more detail on what has happened to the labour 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 labour supply in France, the US 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 labour supply in the US 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 US 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
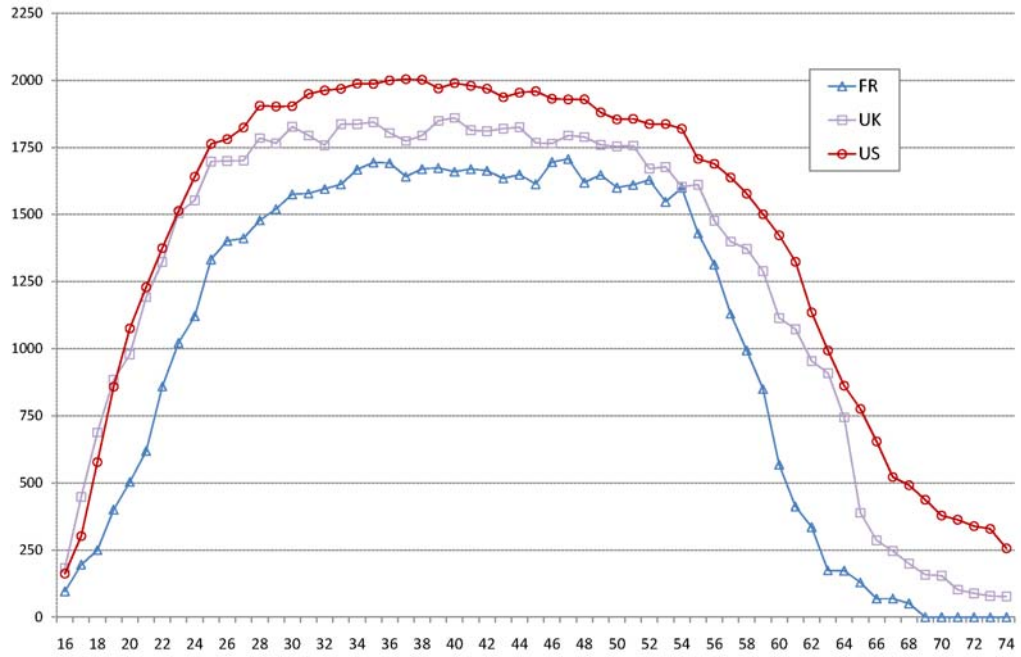
4

Figure 2: Total Hours Worked, Men 2007
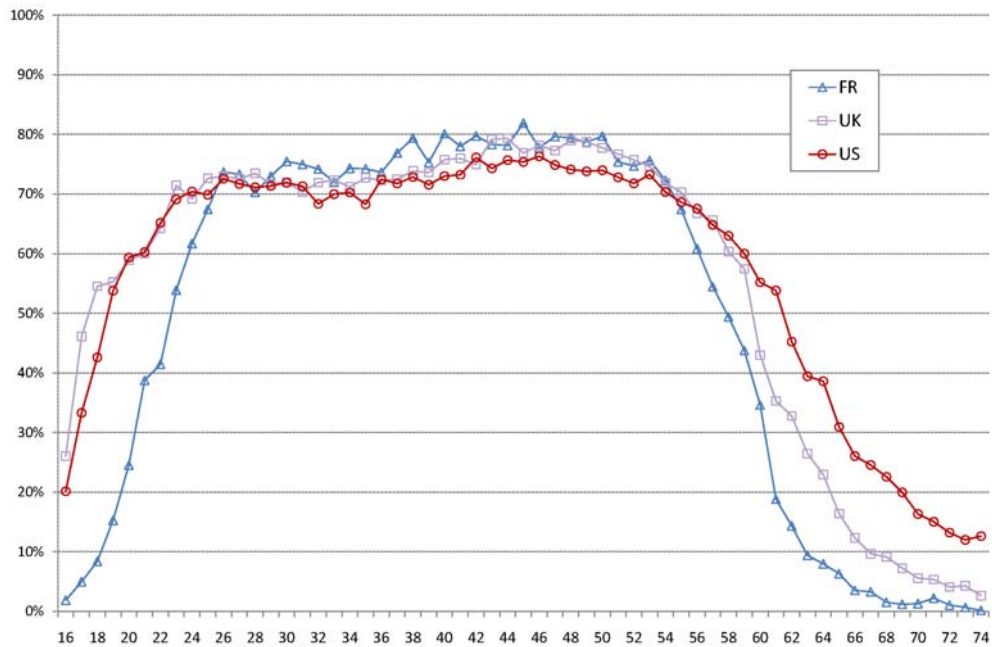

Figure 3: Employment: Women 2007

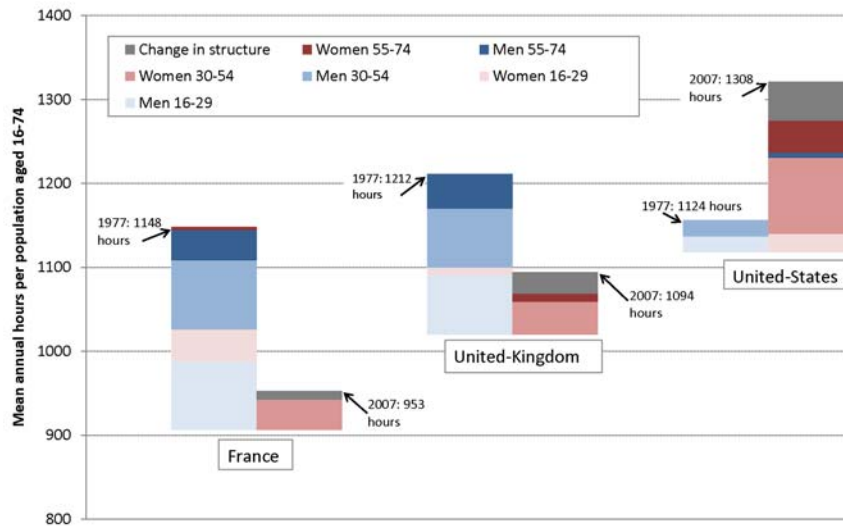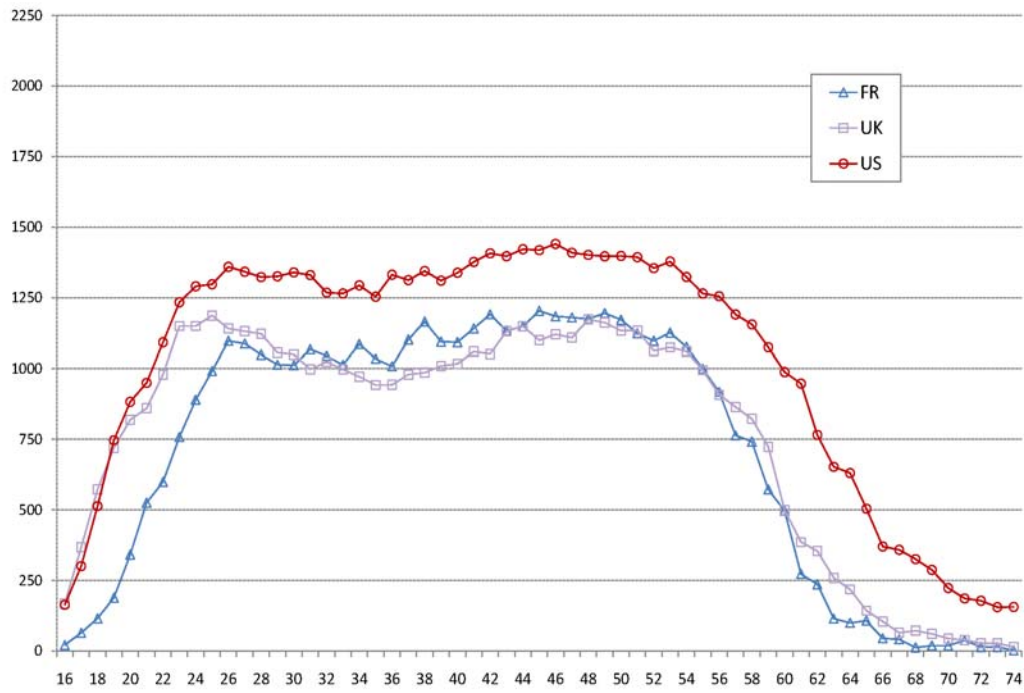Figure 4: Total Hours Worked: Women 2007

Figure 5: The Change in Total Hours by Age and Gender

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 $j = 1, ..., J$ broad categories. The overall statistic H_t is computed in any year t as an average of the category hours 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 participation to the labour 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

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

and 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:

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

or

$$\Delta_{jt} = q_{j,t-1} \{ [h_{jt} - h_{jt-1}]p_{jt-1} + [p_{jt} - p_{jt-1}]h_{jt} \} \tag{5}$$

The first term on the right hand side is the intensive margin, weighted in the top formula (4) with the final participation rate (as in a Paasche index) and in the bottom formula (5) with the initial participation rate (as in a Laspeyres index). The second term is the extensive margin (Laspeyres in (4), Paasche in (5)). The empirical counterpart 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]

Table 1: The extensive and intensive margins between 1977 and 2007

	Year	Youth (16-29)		Prime aged (30-54)		Old (55-74)		All
		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]

SOURCE: BLUNDELL, BOZIO AND LAROQUE (2011A). NOTE: I-P designs the Paasche measure of the intensive margin, I-L the Laspeyre measure, respectively E-P and E-L for the extensive margin, as described by equations (4) and (5).

shows the bounds on the intensive margin, L standing for Laspeyres (the change in hours being weighted by the initial participation rate), P for Paasche (final participation 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 participation multiplied by average hours worked at the initial (resp. final) date.

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 US 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 US, although there are declines in the US too. As we have noted, for prime age women it is the increase at the extensive margin that is so extraordinary, especially in the US 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 US.

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 US. There are falls at the extensive and intensive margins for UK men but increases at the extensive margin for UK women. This phenomena 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 US 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 labour supply is likely to be most responsive to reform. They also set up the key question in the analysis of tax incentives and labour supply: How well do

structural economic models explain these changes in observed behaviour? For this we have to turn first to the measurement of the effective tax rates in the tax and benefit system.

2 Effective Tax Rates

What of effective tax rates? To understand how taxes and benefits might affect labour supply choices, we need to measure the effective work incentives implicit in the tax and benefit system. 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 shows the complexity arising from the cocktail of taxes and benefits. This constraint assumes all eligible benefits are accessed.[2]

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 scheme - called the Working Tax Credit - has certain unique features. All are designed to enhance income in work for those facing low rates of pay and/or higher costs of work. In the UK eligibility to this tax credit 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 criteria 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 an income eligibility family net income below a certain threshold after which the credit is tapered 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.

Figure 5: Source: Tax by Design, Mirrlees Review, 2011.

Figure 5: Notes: 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.

Figure 6 examines the budget constraint for an example low income single parent in the UK, but what does the tax and benefit system imply for effective tax rates across the distribution of earnings and different family types? 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

---

[2]See Adam, Browne and Heady(2010) for a more detailed overview.

Figure 6: The Interaction between Taxes and Benefits in the UK
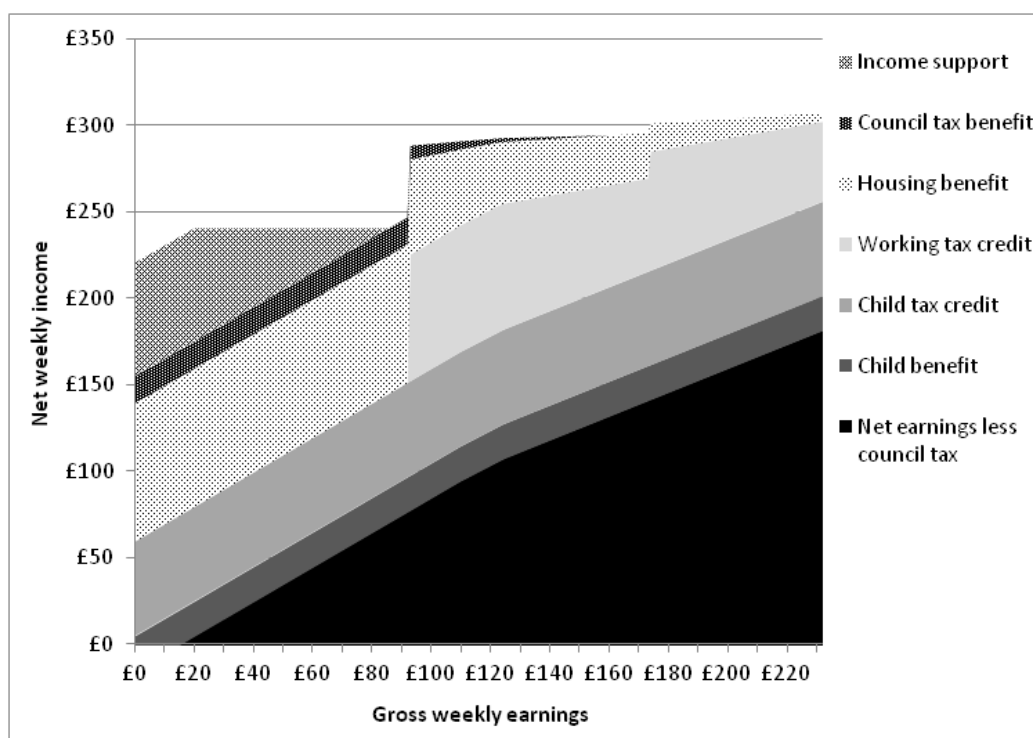
withdrawn benefits; and the participation tax rates (PTR) – the incentive to be in paid work at all – defined by the proportion of total earnings taken in tax and withdrawn benefits.

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.

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 below, on the responsiveness

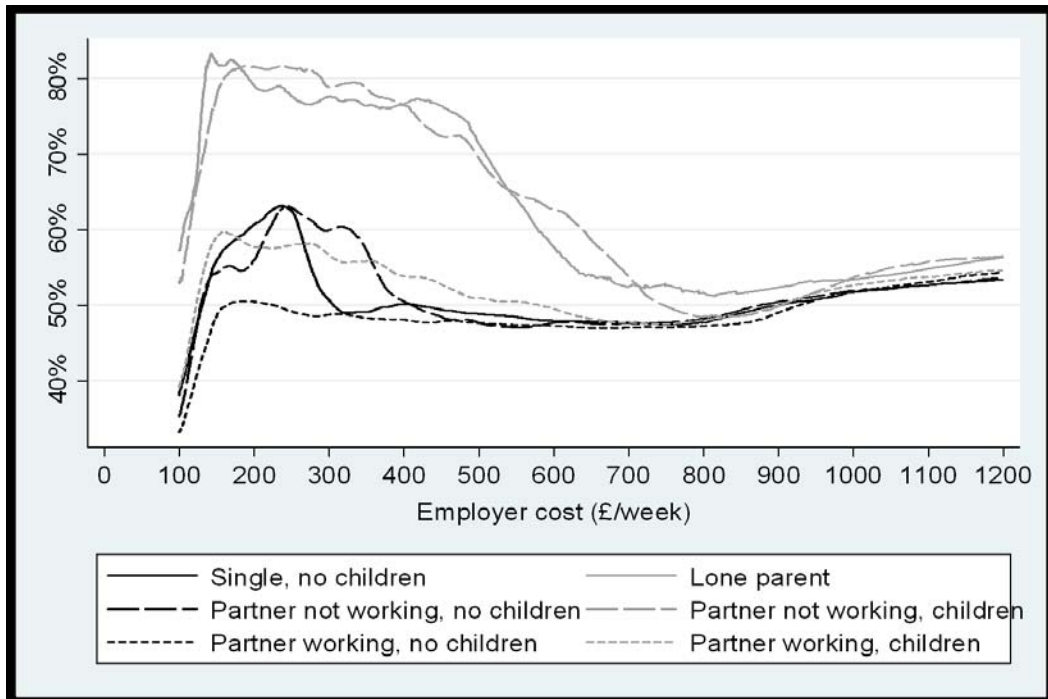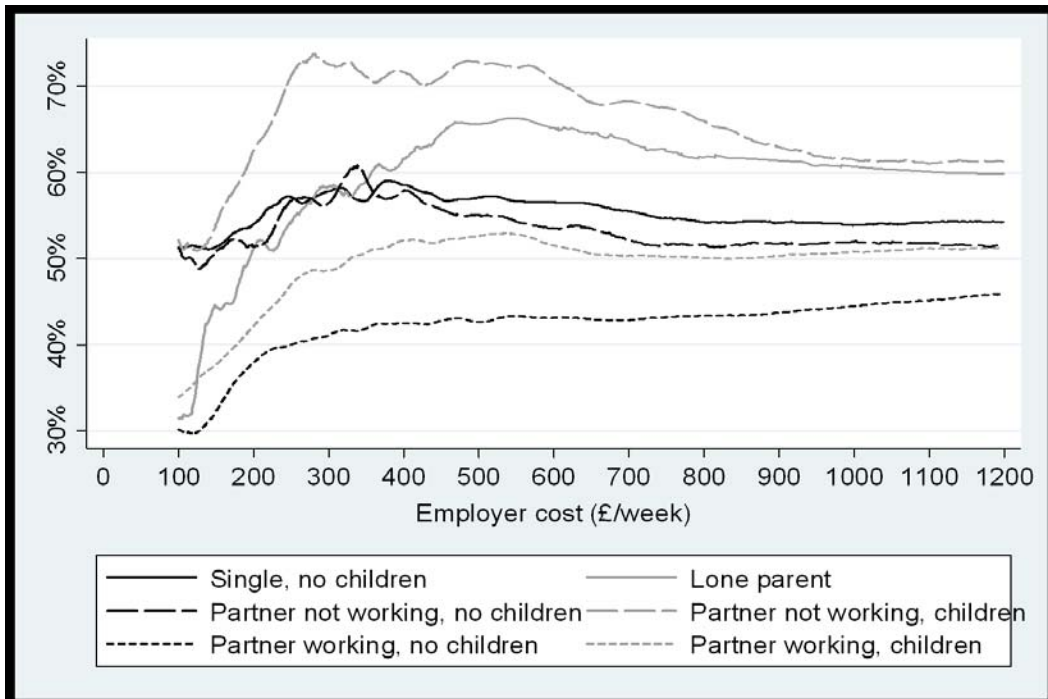Figure 7: Average EMTRs for different family types

Figure 8: Average PTRs for different family types

of labour 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.

3   The Importance of Information and Complexity

The *EMTRs* and *PTRs* in the last two Figures 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 *Tax by Design* (chapter 5 for further details). This degree of complexity 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 to 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 programme 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 tax credit (or its predecessor, Family Credit). It measures the proportion taking-up among those we estimate are eligible to 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 in to how to model take-up decisions.

To rationalize incomplete take-up of a benefit or tax-credit programme we follow Moffitt (1983), and the subsequent developments in Blundell, Duncan, McCrae and Meghir (2000) and Brewer, Duncan and Shephard (2007), and assume the presence of some 'stigma' or 'hassle' cost. This will provide an explanation of Figure 9 and will also feed in to 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 reflects 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.
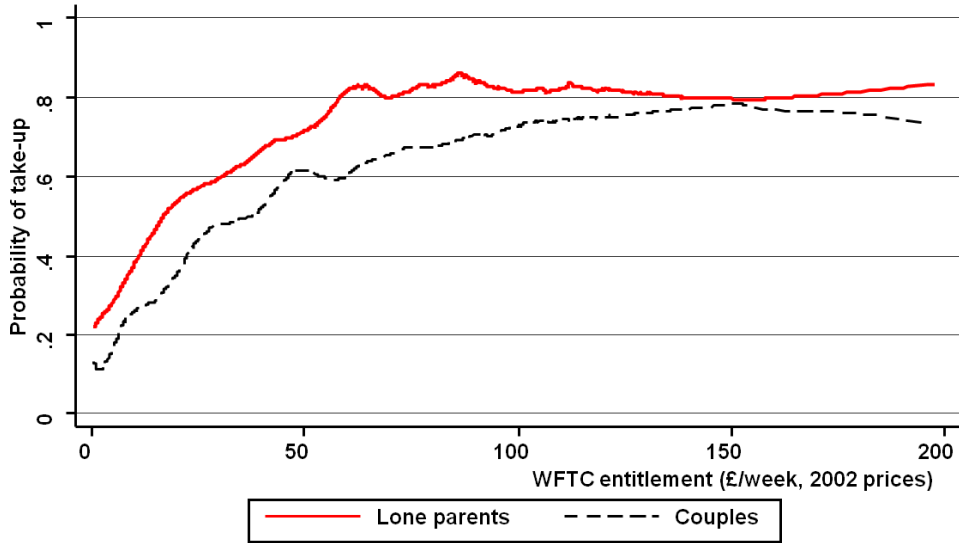
Figure 9: Take-Up of Tax-Credits in the UK

Suppose an individual faces an hourly wage rate w and works h hours, then the net taxes paid will only be a function of hours of work h and total earnings wh but also whether $P = 1$ or 0. The tax function becomes

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

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 below. 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.[3]

It is difficult to argue against any policy reform that makes it clear what benefits and tax credits individuals are eligible, and what the effective tax rates in the system are. Stigma is a slightly different matter. It maybe that a certain level of stigma cost could be optimal. However, one clear implication for reform drawn therefore from these analyses is that, to be effective, the tax and benefit system in the UK requires some simplification and some integration of the various benefits and tax credits so as to make them understandable and coherent.

Take-up and the integration of the benefit system may reduce the complexity of the tax and benefit system but it does not tell us about the rate structure itself. For this we need to know about the size of behavioural responses so that work incentives are strengthened where individuals are most responsive to them. We now turn to the robust measurement of the impact of tax reform.

---

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

4 Evidence on the Size of Responses to Tax Reform

In section 1, we documented the growth of female labour supply, changes in youth employment and changes in 'early retirement' behaviour which must form the backdrop to any earnings reform agenda. To these we add changes in demography, including growth in single person and single parent households.

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

4.1 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 optimisation 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 labour supply which we will also briefly discuss below.

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 labour 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 this approach to structural estimation with the observed changes in incentives and policy reforms is an important step in deriving convincing structural models. There is an increasing recognition that

changes in incentives over time can, and should, be used directly to estimate labour 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 labour 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 reliable 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 under certain assumptions. For example, the design of top tax rates using taxable income elasticities as originally proposed by Feldstein (1995) and the design of the overall earnings tax rate structure using intensive and extensive elasticities as pioneered in Saez (2002). We return to these approaches in Section 5. In general, Heckman and Vytlyacil (2005) have shown the connection between the measurement of (marginal) treatment effects and population-wide policy impacts. 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.

4.2 Randomised 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 experiment undertaken in the late 1960s and early 1970s in the US. 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 labour supply effects of the negative tax experiment. Nonetheless, in their comprehensive review of field experiments in economics, Card, DellaVigna and Malmendier (2010) find very few experimental studies that are directed towards recovering structural economic parameters. Even fewer directed at the tax and benefit system. 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, see 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 week assumptions, provided the experiment is carried our 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 (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 spend 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 $36000 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 12 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 12 months on welfare receipt, the eligibility criteria (full-time employment) has to be satisfied within 12 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 12 month window, was substantial. It left no doubt that financial incentives could impact quite strongly on the employment behaviour of welfare mothers. However, by the end of the program the employment effect has all but disappeared. We return to this point in section 4.4 below in the discussion of the dynamic effects of tax and benefit programmes. .

4.2.1   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 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 WFTC tax credit 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
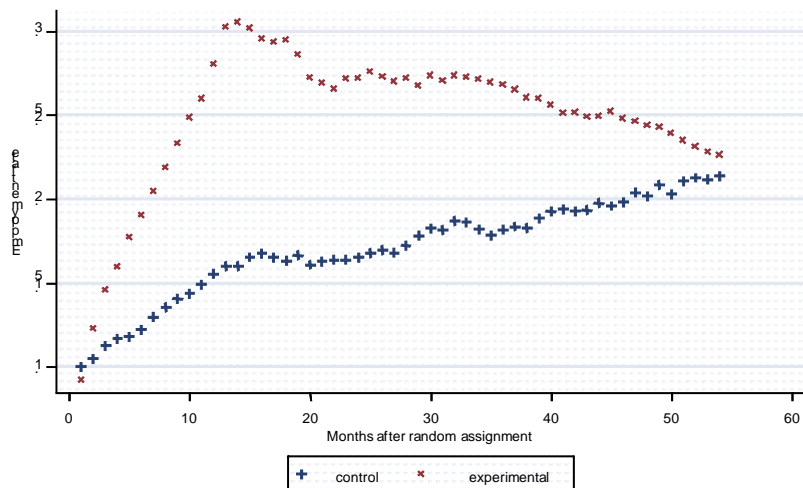
Figure 10: SSP: Employment Rate by months after RA

5, could face substantially improved work incentives from 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 rigorous 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 FRS is what we use in policy simulation analysis due to its comprehensive income measure. Both data sets point to a 4 to 5 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.

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 assumption 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.
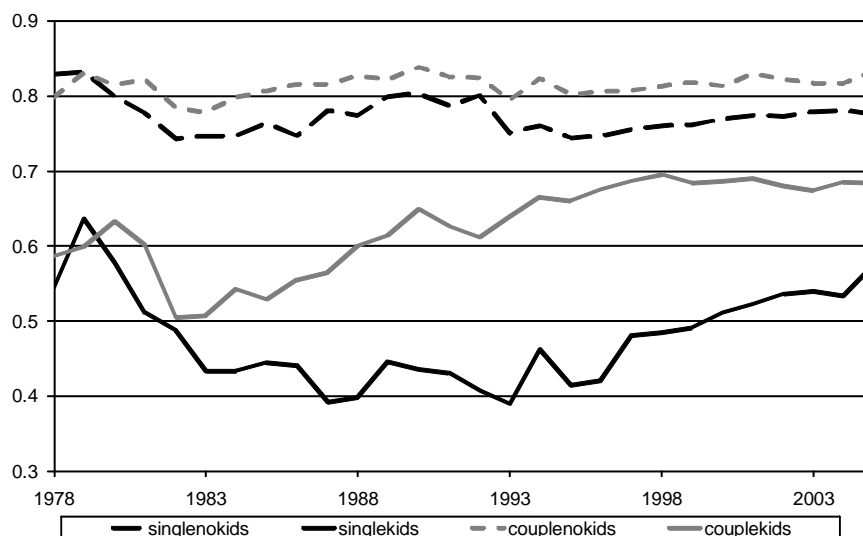
17

Figure 11: Female Employment in the UK

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: Spring 1996 – Spring 2002.

Matching Covariates: age, education, region and ethnicity.

Before moving to the structural approach it is worth noting that tax credit reforms do not just change the incentives for employment they changes incentives for hours of work too. 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 before and after reform hours and show a strong shift towards a spike at 16 hours. The incentive at 16 hours clearly has an impact on behaviour (see Blundell and Shephard (2011) for further comparisons). 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 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 specification than is often the case in structural analyses. 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
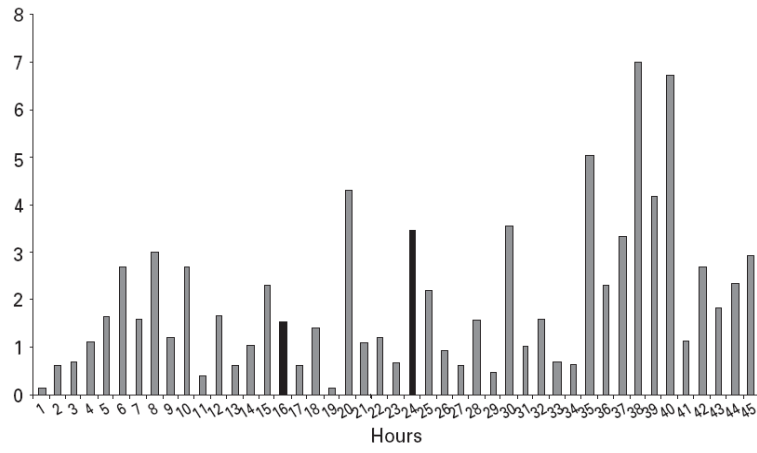
Figure 12: Hours' distribution for lone parents, before 16 hour rule
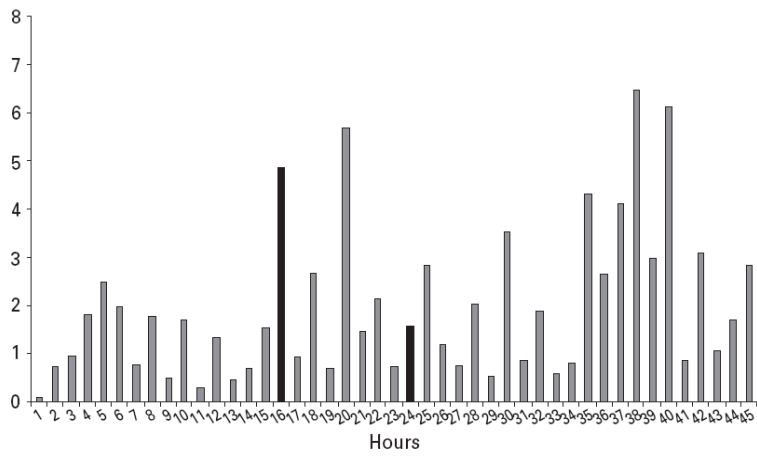

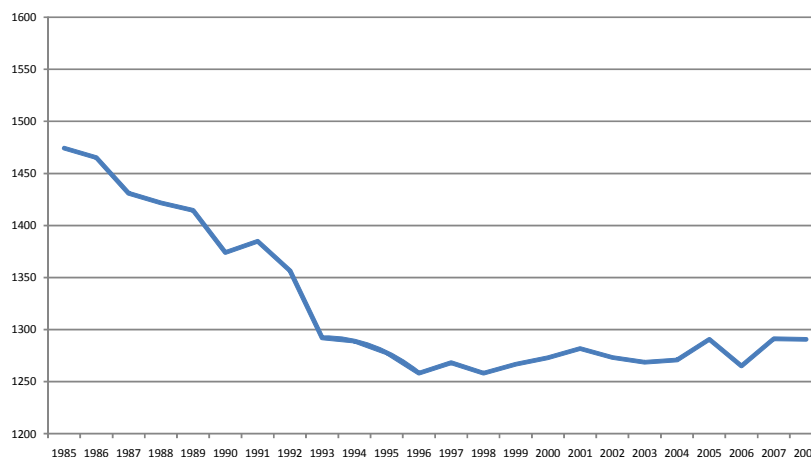Figure 13: Hours' distribution for lone parents, after 16 hour rule

Figure 14: Average Annual Hours of Work: Single Mothers

and benefit system.

To accurately gauge the impact on hours of work and employment of the incentives implicit in these reforms it is important to specify exactly what happened to the incentives to work at different hours levels over this period. This is the first step in any structural analysis to which we now turn.

4.3 Structural models

The structural modelling 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 behaviour and deadweight loss. It will be a key component of any optimal tax design too.

As we noted above the 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 optimising 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, programme participation 'take-up' costs.

There is now a long history of 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).[4] Section 2 has already detailed the complexity of the budget constraint and section 3 has stressed the importance of stigma and hassle costs in modelling 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

---

[4]Background programmes and data are avaiable at http://www.ucl.ac.uk/~uctp39a/lect.html , under sectionIII. Labour Supply and Tax Policy Simulation.

'heterogeneity' across individuals observed characteristics X and unobserved characteristics ε. In section 3, 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:

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

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 of potential hours available for work.

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.[5] 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 see below, 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 rationalise less than complete take-up of tax credit programmes. We denote $P^*(h) \in \{0, E(h; X, \varepsilon)\}$ as the optimal choice of programme 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 labour 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 $T(wh, \mathbf{h}, P; X)$. Non-labour 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 we subtract both childcare expenditure and fixed work related costs from net income, $wh - T(wh, \mathbf{h}, P; X)$.

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

---

[5]In the Supplementary Material to their paper the sensitivity of the results to a finer discretisation of weekly hours are examined; the main results appear robust.

price of childcare. Empirically, this is modelled by assuming that p_c follows some distribution $p_c \sim F_c(\cdot; X_c)$ which varies with demographic characteristics.

From net income we obtain consumption c by subtracting fixed-costs of work: $\alpha_f(X, \varepsilon_f) \times \mathbf{1}_{\{h>0\}}$ and childcare expenditure. Incorporating endogenous take-up of tax credits through cost $\eta(X, \varepsilon_\eta)$, it then follows that the optimal choice of hours $h^* \in \mathcal{H}$ maximises

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

subject to the various constraints detailed above.

This brief outline of some of the key features of a structural model, shows some of the key ingredients and assumptions required, and this for a single worker decision. Family labour supply models require further assumptions, in particular the modelling of joint hours and employment choices, see Blundell et al (2000) and Brewer et al (2007). Blundell and MaCurdy (1999) and Meghir and Phillips (2010) give further insights in to structural modelling of labour supply.

Blundell and Shephard (2010) take this structural model specification to the UK Family Resources Survey data and argue it does a good job of describing observed behaviour. 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 above. 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.

4.4 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, just that it is not rejected. One area where we might expect the model described above to perform poorly is in capturing the dynamics of labour 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 out structural model to provide a good description of labour 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 labour market. For some evidence on this for low skill workers we turn to the Canadian SSP. Figures 15 and 16 from the Canadian SSP tell an interesting story. Recall that the SSP program is complete after 36 months from eligibility and eligibility can take up to 12 months. By 50 months from baseline all the treatment group would have completed the programme. Indeed we have see that the treated had, on average, a sizable increase in employment. But did it translate in to 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
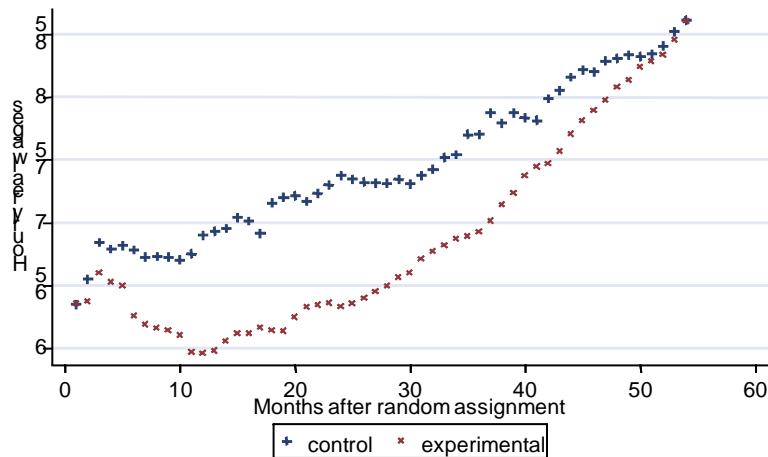
Figure 15: SSP: Hourly wages by months after RA

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 labour supply adjustments for quite well for these types of workers.

5   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 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 reasonable people may differ. In this section we will see how these come in to play in the design of earnings taxation.

An important distinction from our discussion of labour supply responses is between the extensive and intensive margins of labour 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 participation 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
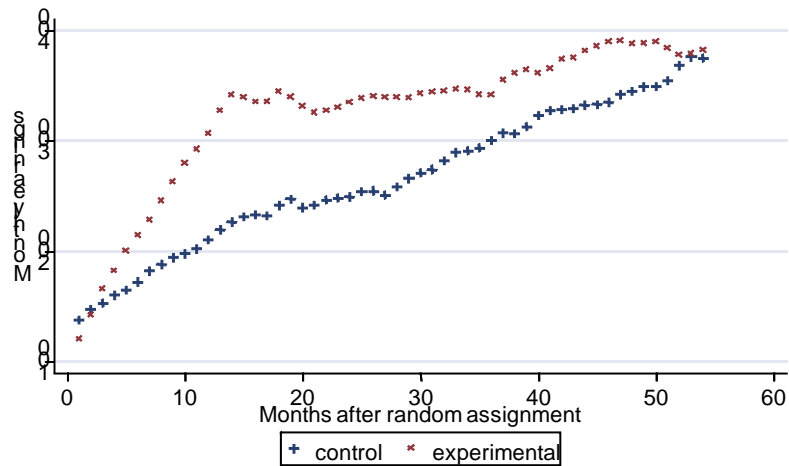
Figure 16: SSP: Monthly earnings by months after RA

system.'

This is one of the key lessons from recent 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.

5.1 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 constraint. 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? 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 distribution of income and, some view of social welfare weights.

Our aim will be to recover optimal tax/credit schedule in terms of earnings. There are two

24

related approaches. First to use Diamond-Saez approximation in terms of extensive and intensive elasticities at different earnings. Second, 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 extremely intuitive expression for the tax rate schedule in terms of the extensive and intensive elasticities, simple summary measures of the distribution of earnings and the social welfare weights. The formula is only approximate and assume away income effects, however as a guide to the setting of tax rates it is extremely informative, see Immervoll, Kleven, Kreiner, and Saez, (2007). We return to these optimal tax formula in our discussion of the optimal top rate of tax below.

The latter approach might be labelled the structural microeconometric approach to tax design. Effectively a stochastic mechanism design problem. In this case the optimal tax model *is* the labour supply model. Consequently all of the assumptions concerning behaviour 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.[6] Social welfare is represented by the sum of transformed utilities. Social welfare W is represented by the sum of transformed utilities

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

where function Υ captures redistributive preferences.

The government seeks to maximize W subject to revenue constraint:

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

Blundell and Shephard (2010) control preference for equality by transformation function:

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

where $\theta < 0$ favours 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 optimal schedules. In this analysis we typically make no distinction between tax and transfer system, assuming full integration. 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 θ) were that marginal rates are broadly increasing in earnings for all groups, although much as in the original Mirrlees study, effective marginal tax rates are fairly constant.

They also document 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, they

---

[6]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.

find pure *tax credits* but only at low earnings and for those with school aged children. They also consider where *optimal* hours rule may be placed. The analysis suggests increasing hours where bonus is paid. Typically part-time hours rules were not found to be optimal.

Overall heterogeneity is found to be important for tax design: tax credits may be optimal but only for those with school aged children. Hours contingent payments can improve design: if hours accurately observed full-time bonus desirable for mothers with older children. But measurement error and the possibility of hours manipulation do appear to weaken argument for hours rules.

5.2 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 in to 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 bare in mind is that the responsiveness itself will be affected by the tax base. This elasticity can be expected to be larger the broader the tax base. Given the need to capture all these margins and the fact that effort is very hard to quantify, means that the behavioural 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. 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. Assuming, that is, that the top of the taxable income distribution is well approximated by a Pareto distribution. Figure 17 shows this to be the case. Indeed for the UK a is approximately 1.67, see Brewer and Browne (2009).
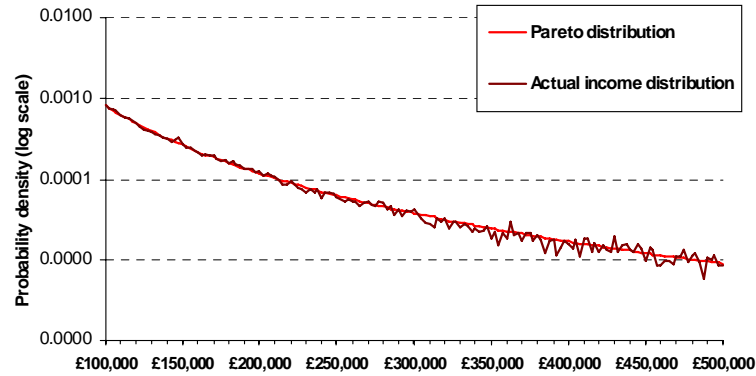
Figure 17: Pareto distribution as an approximation to the income distribution

To estimate e reliably is much harder. 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 *Review* suggested a preferred estimate of e of .46 with a standard error of .13.

Exploring various formulations of the differences in differences specification for the UK, the estimate of e remains in the .35 - .55 range with a central around .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.

On face value this analysis suggests little room for raising further revenue from any further raising of the top rate in the UK. However, there are two key caveats. First, the elasticity is not a constant of nature. A broadening of the tax base and/or an alignment of tax rates across different sources of capital and labour income could significantly reduce the elasticity by eliminating 'escape routes' from earnings taxation. Consequently, any reform to the taxation of earned incomes should been seen in the context of reforms to the whole capital and labour income tax base, as strongly argued in the *Mirrlees Review.* Second, the resulting policy response depends on the nature of the changes to income shares that underlie the tax revenue elasticties on top incomes. This partly reflects the design of the tax base but also more directly the degree of tax avoidance and of rent seeking behaviour, as argued in Piketty, Saez and Stantcheva (2011).

6   Summary and Conclusions

In the second volume of the Mirrlees Review, *Tax by Design,* the authors of the Review (of which I was one) set out a full programme of reforms across all aspects and areas of taxation. These are summarised in Chapter 20 of that volume. The focus in *this* paper has been on the central role of empirical evidence in developing the reforms. In particular, for the proposals on the *taxation of earnings.*

The aim has been to assess the best use of available evidence, from broadly descriptive evidence to that gleaned from quasi/experimental evaluations and from structural econometric models. The paper has identified five steps in building up a reliable evidence base for tax policy design. Starting with a descriptive analysis of the key margins of adjustment, followed by a detailed measurement of effective tax rates, then an investigation of information issues and complexity. Based on these three foundations the forth step is a systematic causal analysis of the size of responses. Finally, a structural econometric framework is brought to bear allowing the measurement of deadweight loss and the characterisation of an optimal design. The main sections of this paper have reflected these five building blocks.

Even with respect to the taxation of earnings, the *Review* laid out a broad set of recommendations that extended well beyond the sole use of empirical evidence. For example, the proposals related to the delivery as well as to the schedule and the base of earnings taxation. They included arguments for moving to a much more transparent and coherent rate schedule with a single integrated benefit. The reforms also emphasized the primary role played by the schedule of personal taxes and benefits in achieving progressivity, reflecting the inevitable trade-off between redistribution and work incentives. Although taxation of the rich and topping up the incomes of the poor will inevitably affect behaviour to some extent, the careful deign of the tax system can minimize the efficiency loss associated with achieving progressivity. This means having a rate schedule that reflects knowledge of the shape of the income distribution and the responsiveness of people to taxes and benefits at different income levels. It is in this last aspect of the design of earnings taxation where empirical evidence played the most important role.

In this paper, as in the Review, one key aspect of empirical evidence for earnings taxation was the distinction between labour supply responses at the extensive and intensive margins. That is the separation of the impact of earnings taxation on decisions over whether to work (including when to retire) and how much to work. Both were found to matter but differ in importance by gender, age, education and family composition. For example, empirical evidence shows that mothers of school-age children and people around retirement age are particularly responsive to work incentives. It makes sense therefore to design the rate schedule to take into account these observable characteristics.

There are, of course, limits to how tax and benefit payments might 'tagged' according to observed characteristics, with some constituting unfair and illegitimate discrimination. And being more generous to people with certain characteristics can create an undesirable incentive to acquire those characteristics. There is also some tension here with seeking a neutrality in the tax system and with the design of the taxation of life-cycle savings. So the hurdle for such departures should be high.

Nonetheless, the results of our empirical analysis do point to lower marginal rates at the bottom of the earnings distribution, at least for some key groups. Tax credits could be better targeted to lower incomes and to the families where labour supply is most responsive. From the examination of response elasticities we suggested tagging effective tax rates in the tax, tax-credit and benefit system by age of (youngest) child for mothers/parents and also by 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.

Much of the detailed empirical discussion in this paper has used tax design for one small group - single mothers. Partly for illustrative reasons but also because for this group there is large body of evidence on responses to tax/benefit reform and to policy design. In the *Review* we used a similar design framework for family labour supply and early retirement. For some responses we often had less in the way of detailed structural econometric models to draw on and typically combined knowledge of specific groups with more general reduced form information on elasticities, following closely the lead in Saez (2002), Brewer, Saez and Shephard (2010) and Chetty (2008). There are many areas where the evidence base remains sadly incomplete.

The evidence presented here also suggests limits to tax rises at the top of the income distribution at least in the absence of reforms to the tax base to address avoidance and revenue shifting. But the Review also stressed the fragility of this evidence and the importance of understanding and measuring tax avoidance and evasion. In its discussion of the interaction between earnings taxation and capital taxation, it recommended that income from all sources should be taxed according to the same rate schedule. Applying different rates to different income sources and different assets complicates the system, unfairly favours those taxed more lightly, distorts economic activity towards lightly taxed forms, and facilitates tax avoidance. Taxing income from all sources equally does not just mean taxing fringe benefits in the same way as cash earnings. It also means applying that same rate schedule to, inter alia, self-employment income, property income, savings income, dividends, and capital gains. However, unlike a standard income tax, the *Review's* approach, mimicking a generalised progressive expenditure tax, would allow all costs of generating that income, including income saved, to be deducted. The proposals present a coherent and integrated tax system for earnings, savings, capital gains and business incomes.

In any tax reform, earnings taxation is also called upon to undo the impact on distributional and work incentives of the rest of the package. For example, part of the *Review's* recommendations was a broadening of the VAT base. In particular, applying standard rates of VAT to the financial services industry, to food and to clothing. This requires very special care in terms of correcting for the impact of these reforms on work incentives and redistribution. Therefore, and as emphasised. in the *Review*, it is impossible to separate the taxation of earnings from indirect taxation and the taxation of savings, etc. Nonetheless, broad proposals for earnings tax reform can, and the *Review* argues *should*, be made that are be based on a careful empirical analysis of earnings responses to earnings taxation.

References

[1] Adam, S., Browne, J. and Heady, C. (2010), "Taxation in the UK", in Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G. and Poterba, J. (eds), *Dimensions of Tax Design: The Mirrlees Review*, Oxford University Press for Institute for Fiscal Studies, (Oxford).

[2] Atkinson and Piketty (2007), Top Incomes over the Twentieth Century: A Contrast between European and English-Speaking Countries, OUP.

[3] Ashenfelter, Orley and Mark W. Plant. "Nonparametric Estimates of the Labor Supply Effects of Negative Income Tax Programs. Journal of Labor Economics 8 (January 1990, part 2), pp. S396-S415

[4] Besley, T. and S. Coate (1992), "Workfare versus Welfare: Incentive Arguments for Work Requirement in Poverty Alleviation Programs", American Economic Review, 82(1), 249-261.

[5] Blau, F. & Kahn, L. (2007), 'Changes in the Labor Supply Behavior of Married Women: 1980-2000', Journal of Labor Economics 25(3), 393-438.

[6] Blundell, R. (2006), "Earned income tax credit policies: Impact and Optimality", The 2005 Adam Smith Lecture, Labour Economics, 13, 423-443.

[7] Blundell, R., Bozio, A. and Laroque, G. (2011a), 'Labour Supply and the Extensive Margin', *American Economic Review,* Volume 101, Issue 3, May, 482-486.

[8] Blundell, R., Bozio, A. and Laroque, G. (2011b), 'Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France', *IFS Working Paper 11-01*, June.

[9] Blundell, R., M. Brewer and M. Francesconi (2008), 'Job changes, hours changes and the path of labour supply adjustment', *Journal of Labor Economics*, Volume 26, Issue 3, Page 421–453, July.

[10] Blundell, R. and M. Costa-Dias, (2009), 'Alternative Approaches to Evaluation in Empirical Microeconomics', *Journal of Human Resources*, 44(3), Summer, 565-640.

[11] Blundell, R.W., Duncan, A. and Meghir, C. (1998), "Estimating Labour Supply Responses using Tax Policy Reforms", Econometrica, 66, 827-861.

[12] Blundell, R, Duncan, A, McCrae, J and Meghir, C. (2000), "The Labour Market Impact of the Working Families' Tax Credit", Fiscal Studies, 21(1).

[13] Blundell, R. and Hoynes, H. (2004), "In-Work Benefit Reform and the Labour Market", in Richard Blundell, David Card and Richard .B. Freeman (eds) Seeking a Premier League Economy. Chicago: University of Chicago Press.

[14] Blundell, R. and MaCurdy (1999), "Labour Supply: A Review of Alternative Approaches", in Ashenfelter and Card (eds), Handbook of Labour Economics, Elsevier North-Holland.

[15] Blundell, R., Meghir, C., and Smith, S. (2002), 'Pension incentives and the pattern of early retirement', Economic Journal, 112, C153–70.

[16] Blundell, R., and A. Shephard (2010), 'Employment, hours of work and the optimal taxation of low income families', IFS Working Paper, forthcoming *Review of Economic Studies.*

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

[18] Brewer, M., A. Duncan, and A. Shephard, January 2007, 'Did working families' tax credit work? The impact of in-work support on labour supply in Great Britain,' *Labour Economics*, Vol. 13, No. 3, pp. 699-720, Elsevier, Journal Articles

[19] Brewer, M. E. Saez, and A. Shephard, (2010) "Optimal Household Labor Income Tax and Transfer Programs", in Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G. and Poterba, J. (eds), Dimensions of Tax Design: The Mirrlees Review, Oxford University Press for Institute for Fiscal Studies, (Oxford).

[20] Brewer, M. and J. Browne (2009), Can more revenue be raised by increasing income tax rates for the very rich?, IFS Briefing Note no. 84.

[21] Card, D., S. Dellavigna & U. Malmendier (2011) "The Role of Theory in Field Experiments," Jornal of Economic Perspectives, forthcoming.

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

[23] Card, D and D. Hyslop, 2005. "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," Econometrica, vol. 73(6), pages 1723-1770, November.

[24] Chetty, R. (2008), 'Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods', National Bureau of Economic Research (NBER), Working Paper 14399

[25] Chetty, R (2009), Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply, NBER Working Papers 15616, National Bureau of Economic Research, Inc.

[26] 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." Unpublished paper, Harvard University.

[27] Cogan, J. (1981), 'Fixed costs and labor supply', Econometrica 49(4), 945{964.

[28] Diamond, P. (1980), 'Income Taxation with Fixed Hours of Work', Journal of Public Economics 13(1), 101-110.

[29] Eissa, Nada and Jeffrey Liebman (1996), "Labor Supply Response to the Earned Income Tax Credit", Quarterly Journal of Economics, CXI, 605-637.

[30] Feldstein, M.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 (3), 551–572.

[31] Feldstein, M.S., 1999. Tax avoidance and the deadweight loss of the income tax. Review of Economics and Statistics 4 (81), 674–680.

[32] Gruber, J. & Wise, D. (1999), Social Security and Retirement around the World, NBER/The University of Chicago Press.

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

[34] Heckman, James J. (1974), 'Shadow Prices, Market Wages, and Labor Supply', Econometrica 42(4), 679{694.

[35] Heckman, James J. (1979), 'Sample Selection Bias as a Speci cation Error', Econometrica 47(1), 153{161.

[36] Heckman, James J. (1993), 'What Has Been Learned About Labor Supply in the Past Twenty Years?', American Economic Review 83(2), 116-121.

[37] Heckman, James J. and Edward Vytlacil (2005) 'Structural Equations, Treatment Effects, and Econometric Policy Evaluation', Econometrica, vol. 73(3), pages 669-738, 05

[38] Hoynes, H. (1996), 'Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation under AFDC-UP', Econometrica, vol. 64(2), pages 295-332, March.

[39] James J . Heckman; Jeffrey A. Smith (1995), Assessing the Case for Social Experiments, Journal of Economic Perspectives, Vol. 9, No. 2, Spring, 85-110.

[40] Hurd, Michael. The estimation of nonlinear labor supply functions with taxes from a truncated sample. 1976. 69 pp. Center for the Study of Welfare Policy, Stanford Research Institute. Research memorandum 36.

[41] Johnson, Terry R. and Pencavel, John H. Utility-based hours of work functions for husbands, wives, and single females estimated from Seattle-Denver experimental Data. 1980. v, 60 pp. Socioeconomic Research Center, SRI International. Research memorandum 71.

[42] Immervoll, H. Kleven, H. Kreiner, C, and Saez, E. (2007), 'Welfare Reform in European Countries: A Micro-Simulation Analysis' Economic Journal, 117 (517), 1-44.

[43] Keane, M.P. and Moffitt, R. (1998), "A Structural Model of Multiple Welfare Program Participation and Labor Supply", International Economic Review, 39(3), 553-589.

[44] Kopczuk, W. (2005), 'Tax bases, tax rates and the elasticity of reported income', Journal of Public Economics, 89, 2093–119.

[45] Laroque, G. (2005), 'Income Maintenance and Labor Force Participation', Econometrica 73(2), 341-376.

[46] Liebman, J. (2002), 'The Optimal Design of the Earned Income Tax Credit', in Bruce Meyer and Douglas Holtz-Eakin (eds.), Making Work Pay: The Earned Income Tax Credit and Its Impact on American Families, New-York: Russell Sage Foundation.

[47] Meghir, C. and Phillips, D., (2010), 'Labour Supply and Taxes', in Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G. and Poterba, J. (eds), Dimensions of Tax Design: The Mirrlees Review, Oxford University Press for Institute for Fiscal Studies, (Oxford).

[48] Mirrlees, J.A. (1971), "The Theory of Optimal Income Taxation", Review of Economic Studies, 38, 175-208.

[49] Moffitt, R. (1983), "An Economic Model of Welfare Stigma", American Economic Review, 73(5), 1023-1035.

[50] Moffitt, R. (2005), "Welfare Work Requirements with Paternalistic Government Preferences, mimeo, Johns Hopkins University, February.

[51] Ohanian, L., Rao, A. & Rogerson, R. (2008), 'Long-term changes in labor supply and taxes: Evidence from OECD countries, 1956-2004', Journal of Monetary Economics 55(8), 1353-1362.

[52] Piketty, Thomas, Emmanuel Saez and Stefanie Stantcheva (2011), 'Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities' NBER Working Paper 17616, November.

[53] Prescott, E. (2004), 'Why Do Americans Work So Much More than Europeans?', Federal Reserve Bank of Minneapolis Quarterly Review 28(1), 2-13.

[54] Rogerson, R. & Wallenius, J. (2009), 'Micro and Macro Elasticities in a Life Cycle Model With Taxes', Journal of Economic Theory 144(6), 2277-2292.

[55] Saez, E. (2002), 'Optimal Income Transfer Programs : Intensive Versus Extensive Labor Supply Responses', Quarterly Journal of Economics 117(3), 1039{1073.

[56] Stern, N. (1986), On the speci cation of labor supply function, in R. Blundell & I. Walker, eds, 'Unemployment, Search and Labour Supply', Cambridge University Press, pp. 121- 142.

[57] Slemrod, J. and W. Kopczuk (2002), 'The optimal elasticity of taxable income', Journal of Public Economics 84, 91 –112.

[58] A. van Soest (1995): Structural models of family labor supply: a discrete choice approach, Journal of Human Resources, 30(1), 63-88.

[59] West, Richard W. The effects of the Seattle and Denver income maintenance experiments on the labor supply of young nonheads. 1979. v, 54 pp. Center for the Study of Welfare Policy, SRI International. Research memorandum 60.