35 years of reforms: a panel analysis of the incidence of, and employee and employer responses to, social security contributions in the UK

STUART ADAM, DAVID PHILLIPS AND BARRA ROANTREE , INSTITUTE FOR FISCAL STUDIES[1]

This version: 31 May 2016

Abstract

We exploit variation in the National Insurance contributions (NICs) – the UK's system of social security contributions – and a large panel dataset to examine the effects of 35 years of employee and employer NICs reforms on employer cost (gross earnings plus employer NICs), hours of work and employer cost per hour, both immediately (0–6 months) after reforms are implemented and in the slightly longer term (12–18 months). We consider assumptions under which the estimated coefficients on net-of-marginal and net-of-average tax rates in a panel regression can be interpreted as behavioural elasticities or as reflecting incidence. We find a compensated elasticity of taxable earnings with respect to the marginal rate of employee NICs about 0.2–0.3, operating largely through hours of work, while that with respect to the marginal rate of employer NICs is not statistically significantly different from zero. We also find that employer cost falls approximately one-for-one when the average rate of employer NICs is reduced, but not when the average rate of employee NICs is reduced, which is consistent with the economic incidence of NICs following its formal legal incidence. Estimates from the hours and hourly employer cost regressions provide further support to this interpretation of the findings, and also suggest moderate-sized income effects. Each of these results remains true after 12–18 months (the effects of lagged changes in NICs rates are generally statistically insignificant), implying that any shifting of employer NICs changes to the individual employees concerned (and vice versa for employee NICs) does not begin over this time horizon. These results are very similar to those found by Lehmann et al. (2013) for France but represent an extension of that work by considering hours as well as employer cost responses and second-year as well as immediate effects.

1. Introduction

Social security contributions (SSCs) such as the UK's National Insurance contributions (NICs) are a key component of the labour tax wedge faced by workers. According to the OECD, the average labour tax wedge was 35.9% in 2015, of which 21.1 percentage

[1] The authors thank Richard Blundell, Raj Chetty, Jon Gruber, Henrik Kleven, Emmanuel Saez and seminar participants at the IFS for helpful comments. We especially thank colleagues at DIW Berlin, CPB Netherlands and IPP Paris School of Economics, working with us on the broader project on 'The impact of social security contributions on earnings' of which this paper is a part, for many helpful discussions throughout the course of the project. All errors and omissions are the responsibility of the authors. We gratefully acknowledge funding from the ESRC Centre for the Microeconomic Analysis of Public Policy based at the Institute for Fiscal Studies (ES/M010147/1) and ESRC Research Grant ES/K006185/1. Data from the New Earnings Survey Panel Dataset is produced by the Office for National Statistics and supplied by the Secure Data Service at the UK Data Archive.

points related to SSCs.[2] In the UK, NICs accounted for 16.7 percentage points of the overall 30.8% labour tax wedge for someone with average earnings. There is therefore good reason to investigate both the behavioural effects of SSCs and their economic incidence.

Unlike the income tax due on labour income, which is levied on the employee only, SSCs are levied on both employees and employers. In the UK, for instance, employee NICs are currently levied at a rate of 12%, and employer NICs 13.8%, of gross earnings above an exemption threshold (falling to 2% above a higher threshold for employees). Standard models of the labour market predict identical responses and economic incidence for employee and employer SSCs, at least in the long run. The effects of a tax on the amount of labour utilised and the amount paid by employers and received by employees reflects the relative elasticities of labour demand and supply (or the bargaining power of employees versus employers), not the formal legal incidence of the tax.[3] If bargaining and contracting relates to gross earnings however, as is typically the case, legal incidence may affect economic incidence, at least in the short term.

Empirical evidence on this matter is relatively limited. Much of the New Tax Responsiveness (NTR) literature on the response of income or earnings to taxation focuses on income tax rather than SSCs (see Saez et al (2012)), providing surprisingly scant evidence on the effects of SSCs specifically, and much of that literature implicitly assumes taxes are fully incident on the individual taxpayer. The literature on the incidence and employment effects of SSCs, meanwhile, tends to focus on responses to changes in employer or combined rates, perhaps reflecting the fact that independent variation in employee and employer rates is difficult to come by.

A series of reforms to the UK's system of NICs in the 1980s, 1990s and (to a lesser extent) 2000s does provide such independent variation. In this paper we use a panel regression approach to exploit this variation, using data from the New Earnings Survey Panel Dataset (NESPD) – which includes accurate panel data on the earnings and hours of a large, randomly selected sample of workers for up to 35 years – to examine the effects of employee and employer NICs on employer cost (gross earnings plus employer NICs), hours of work, and employer cost per hour. Data with this combination of large sample size, accurate measurement of earnings, long historical coverage and a panel dimension have never previously been used in the UK – and rarely in other countries – to examine such questions. In doing this we set out a series of assumptions that one can make in order to interpret the estimated coefficients on net-of-marginal and net-of-average tax rates as behavioural elasticities or as reflecting the incidence of the tax.

Our work builds on that of Lehmann et al (2013), who undertake a similar analysis of reforms to employer SSCs and income tax credits in France in the mid-2000s. They find

[2] https://stats.oecd.org/Index.aspx?DataSetCode=AWCOMP.

[3] This is true of standard versions of the competitive, union bargaining, search and matching, and efficiency wage models, as set out in Pissarides (1998).

evidence of compensated behavioural responses to income tax but not employer SSCs, and that economic incidence of taxes is affected by formal incidence, at least in the year following a reform. Our work differs from theirs in that we focus on comparing employee and employer NICs, which (unlike SSCs and income tax in France) do not differ in the linkages between liabilities and entitlements such as pensions and unemployment or disability assistance. We extend their work by considering not only responses of employer cost but also hours of work and employer cost per hour. This provides additional evidence that the results we obtain – which are remarkably similar to Lehmann et al – reflect differential incidence of the taxes, rather than differential income effects, for instance. In addition, as well as looking at very short-run responses (0–6 months), we look at slightly longer run responses (12–18 months) by including lagged changes in net-of-NICs rates as independent variables in our regressions.

In order to say something about both behavioural responses to NICs and the incidence of NICs, assumptions about the nature of behavioural responses need to be made. Our first approach follows directly that of Lehmann et al, and utilises information on employer costs (constructed from the observed earnings in our data) only. Motivated by the NTR literature, which emphasises responses to taxes other than hours of work, this approach allows for NICs to affect hourly effort and therefore wages directly via behaviour (rather than 'indirectly' via the effects of tax-incidence shifting). Responses of employer cost to changes in marginal NICs rates reflect substitution effects on both hours and non-hours margins. Responses of weekly earnings to changes in average tax rates can be interpreted as income effects (if the incidence of the tax is assumed) or else as reflecting the incidence of NICs (if we assume no income effects). Our second approach utilises data on hours of work and assumes that behavioural responses take the form of changes in hours of work while any changes in hourly employer cost reflect the incidence of NICs. In effect, this means ruling out the effort-per-hour margin that, in part, motivated the NTR literature.

Estimates from our employer cost regressions are remarkably similar to those of Lehmann et al (2013). The compensated elasticity with respect to the marginal employee NICs rate is statistically significant and positive (around 0.2–03), while that with respect to the marginal employer rate is not statistically significantly different from zero. The coefficient on net-of-average employee NICs is insignificant, while that on net-of-average employer NICs is large and negative, which is consistent with the economic incidence of NICs following its formal legal incidence. Estimates from the hours and hourly employer costs regressions provide further support to this interpretation of the findings, and also suggest moderate-sized income effects. Each of these results remains true after 12–18 months following a reform (the effects of lagged changes in NICs rate are largely statistically insignificant), implying that any shifting of employer NICs changes to the individual employees concerned (and vice versa for employee NICs) does not begin over this time horizon.

As with Lehmann et al (2013), and much of the micro-econometric literature on labour tax incidence, our estimates capture what could be called the 'local' incidence of a labour tax. Throughout this paper, we use 'incident on employers' as a shorthand to mean 'incident on someone other than the employee whose tax rate changed'. Of course, the ultimate burden of a tax must always fall on real people, not the businesses or organisations employing them; it may be passed on to the employers' owners, customers and/or suppliers, and thence perhaps more widely via general equilibrium responses. One important possibility is that a tax that is not incident on the employee whose tax rate changes may nevertheless be incident on a broader group of workers: the nature of market responses may be such that (say) a tax increase affecting one small group of employees results not in a large wage reduction for those employees but in a small wage reduction for all employees in the firm, or in an infinitesimal reduction in equilibrium wages in the wider market. All we attempt to discern in this paper is how an individual's wage is affected by the tax rate applied to that individual's earnings; insofar as their net wage is not reduced one-for-one then we refer to the tax as being at least partly incident 'on the employer', even though the burden may be felt by a wider group of employees rather than by (say) the employer's shareholders.

The paper proceeds as follows. In Section 2 we provide describe the UK's NICs regime and the reforms which we use to identify its effects. Section 3 briefly reviews the NTR and incidence literatures. Section 4 sets out the conceptual underpinnings of our analysis, focusing on the identification of the behavioural and incidence effects of NICs and other labour taxes. Section 5 describes the empirical specification, focusing on our strategy for addressing the endogeneity of changes in NICs rates and on our use of lagged changes in NICs rates to examine slightly longer-run effects on employer cost and hours of work. Section 6 describes our data source, the NESPD, and our results are set out in Section 7. Section 8 concludes.

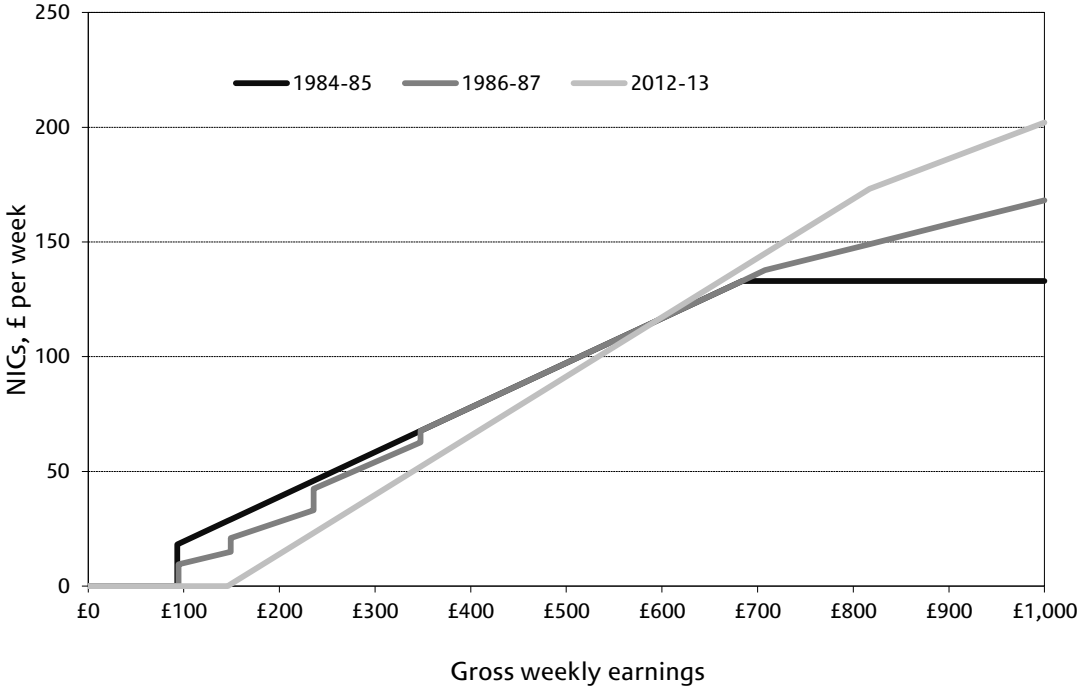
2. Institutional background

Like SSC systems in most countries, the UK's system of NICs consists of both employer and employee contributions. Contributions are a function of the employee's gross earnings (including employee, but not employer, private pension contributions) and are calculated separately in each pay period (typically a week or month): unlike income tax, NICs liability over the course of a year does not depend on earnings for the year as a whole. The NICs rate schedule changed markedly during the period we analyse (1978 to 2010), and it is those changes that we exploit in this paper.

Figure 1 shows the structure of the combined employee and employer NI system before and after the important 1985 reform and as it stood after the end of our period of

analysis, in 2012–13.[4] To facilitate comparison, thresholds from earlier systems have been uprated to April 2012 prices.

Figure1. The changing structure of National Insurance contributions (April 2012 prices)



Note: Cash values uprated to April 2012 prices using the retail prices index. Assumes employee contracted into State Earnings-Related Pension Scheme (SERPS) or State Second Pension (S2P).

No NICs are due below an exemption threshold.[5] Before 1985, contributions for those earning above this level were charged at a flat percentage rate on the entirety of earnings, including earnings below the threshold, up to a ceiling called the Upper Earnings Limit (UEL). This meant a jump, or notch, in contributions at the threshold: both marginal and average rates of NICs increased from zero to the main rates, which at the start of 1985 stood at 9% for employees and 10.45% for employers. The October 1985 reform replaced this single large notch and flat rate of tax with a series of smaller notches, reducing the jump in marginal and average rates at the exemption threshold to 5% each for employees and employers, and introducing a number of graduated steps instead, where higher (marginal and average) rates applied to the entirety of earnings once earnings exceeded higher thresholds. For employees, two additional notches were introduced at higher thresholds, so that the rates of employee NICs jumped to 7% and then 9%, with the highest rate applying to all earnings up to the UEL. For employers three additional notches were introduced at higher thresholds (at rates of 7%, 9% and 10.45%). At the same time, the cap on employer contributions was abolished so that the

[4] For simplicity, the rate schedule with describe in this schedule ignores the possibility of 'contracting out' of the State Earnings-Related Pension Scheme (SERPS) or State Second Pension (S2P), and ignores the 'married women's reduced rate'. These features are described at the end of Appendix A.

[5] This threshold was formerly the Lower Earnings Limit, now the Earnings Threshold.

highest rate of employer NICs applied even above the UEL. The effect of this reform on the combined employee and employer NICs schedule can be seen in the black and dark grey lines in Figure 1.

The system of graduated steps did not last. In October 1989, the system of graduated employee contributions was replaced by a single small notch at the threshold (equivalent to 2% of the threshold) and a single 9% rate of employee NICs that applied to earnings between the threshold and the UEL. However, the graduated system of employer NICs with four notches remained in place at that stage. In April 1999 the remaining notch in the employee NICs schedule, and all of the notches in the employer schedule, were removed, so that the schedule now contains only kinks (that is, changes in marginal contribution rates).

In addition to these main structural reforms, there were many (mostly small) changes in NICs rates and thresholds throughout the period we analyse, culminating in the light grey line in Figure 1. Among these changes, a one percentage point increase in both employee and employer NICs introduced in 2003 is notable as, for the first time, the increase in employee rates applied above as well as below the UEL, so that the UEL – already abolished for employer contributions in 1985 – no longer acted as a complete cap on employee contributions either. Table 1 shows the rates and thresholds that applied in each year.

All these changes contribute to the variation that provides econometric identification in our model. The combination of changes in both thresholds and rates, the move from a notch-based system to a kink-based system via a series of smaller notches, and the extension of NICs (particularly employer NICs) above the UEL, gives us a range of sources of variation across the earnings distribution. Sometimes employer and employee NICs rates changed together, but sometimes not; in some cases individuals' marginal and average NICs rates changed in parallel, but sometimes differentially. This allows us to separately identify earnings responses to changes in both marginal and average rates of both employee and employer NICs.

National Insurance was originally envisaged as a 'true' social insurance scheme, with a broadly actuarial link between contributions paid and benefit entitlements for each individual. Insofar as there is – or, perhaps, is perceived to be – such a link, National Insurance may not have the same disincentive effects as a simple tax on earnings (Summers, 1989). Increasing earnings is made less attractive by the NICs that must be paid on the additional earnings, but simultaneously made more attractive by the increased entitlements it generates; the extent to which these offset each other depends on how much I value the increased entitlements. However, the link between contributions and benefits – particularly at the margin – had already been significantly weakened by 1978, and had all but disappeared by 2010.[6] There was still some link,

[6] For details of contributory benefits, see Hood and Oakley (2014a). The decline of the 'contributory principle' is discussed by Adam and Loutzenhiser (2007) and Hood and Oakley (2014b), among others.

particularly in the early part of the period we analyse. But crucial for our purposes is that none of the *changes* in contribution rates over this period was associated with a corresponding change in entitlements. When an individual in our data sees their NICs

Table 1. Rates and thresholds of National Insurance contributions, 1978–79 to 2012–13

Year	Threshold	Upper earnings limit	Employee contributions			Employer contributions		
	(£ p.w.)	(£ p.w.)	Rate at threshold (%)	Main rate(s) (%)	Rate above UEL (%)	Rate at threshold (%)	Main rate(s) (%)	Rate above UEL (%)
1978-79	£17.50	£120.00	6.5	6.5	0	10	10[g]	0
1979-80	£19.50	£135.00	6.5	6.5	0	10	10[g]	0
1980-81	£23.00	£165.00	6.75	6.75	0	10.2	10.2[g]	0
1981-82	£27.00	£200.00	7.75	7.75	0	10.2	10.2[g]	0
1982-83	£29.50	£220.00	8.75	8.75	0	10.2	10.2[g]	0
1983-84	£32.50	£235.00	9	9	0	10.45	10.45[g]	0
1984-85	£34.00	£250.00	9	9	0	10.45	10.45[g]	0
1985[a]	£35.50	£265.00	9	9	0	10.45	10.45	0
1985-86[b]	£35.50	£265.00	5	5-9	0	5	5-10.45	10.45
1986-87	£38.00	£285.00	5	5-9	0	5	5-10.45	10.45
1987-88	£39.00	£295.00	5	5-9	0	5	5-10.45	10.45
1988-89	£41.00	£305.00	5	5-9	0	5	5-10.45	10.45
1989[c]	£43.00	£325.00	5	5-9	0	5	5-10.45	10.45
1989-90[d]	£43.00	£325.00	2	9	0	5	5-10.45	10.45
1990-91	£46.00	£350.00	2	9	0	5	5-10.45	10.45
1991-92	£52.00	£390.00	2	9	0	4.6	4.6-10.4	10.4
1992-93	£54.00	£405.00	2	9	0	4.6	4.6-10.4	10.4
1993-94	£56.00	£420.00	2	9	0	4.6	4.6-10.4	10.4
1994-95	£57.00	£430.00	2	10	0	3.6	3.6-10.2	10.2
1995-96	£58.00	£440.00	2	10	0	3	3-10.2	10.2
1996-97	£61.00	£455.00	2	10	0	3	3-10.2	10.2

1997-98	£62.00	£465.00	2	10	0	3	3-10	10
1998-99	£64.00	£485.00	2	10	0	3	3-10	10
1999-00	£66.00[e]	£500.00	0	10	0	0	12.2	12.2
2000-01	£76.00[f]	£535.00	0	10	0	0	12.2	12.2
2001-02	£87.00	£575.00	0	10	0	0	11.9	11.9
2002-03	£89.00	£585.00	0	10	0	0	11.8	11.8
2003-04	£89.00	£595.00	0	11	1	0	12.8	12.8
2004-05	£91.00	£610.00	0	11	1	0	12.8	12.8
2005-06	£94.00	£630.00	0	11	1	0	12.8	12.8
2006-07	£97.00	£645.00	0	11	1	0	12.8	12.8
2007-08	£100.00	£670.00	0	11	1	0	12.8	12.8
2008-09	£105.00	£770.00	0	11	1	0	12.8	12.8
2009-10	£110.00	£844.00	0	11	1	0	12.8	12.8
2010-11	£110.00	£844.00	0	11	1	0	12.8	12.8

Notes: Assumes employee contracted into the State Earnings-Related Pension Scheme (SERPS) or State Second Pension (S2P).

(a) Rates apply from April to October 1985.

(b) Rates apply from October 1985 to April 1986.

(c) Rates apply from April to October 1989.

(d) Rates apply from October 1989 to April 1990.

(e) £83 for employer contributions

(f) £84 for employer contributions

(g) Between April 1977 and October 1984 a National Insurance Surcharge applied in addition to the main employers' contribution. The rate was 2% from April 1977 to October 1978, then 3.5% to August 1982, then 2% to April 1983, then 1.5% to July 1983, and 1% to the end of September 1984. For example, the total employers' contribution rate for the first half of 1984-5 was 11.45%, not 10.45%.

Source: IFS *Fiscal Facts*, http://www.ifs.org.uk/tools_and_resources/fiscal_facts/.

rate rise, it simply changes their current budget constraint, just like a tax; they do not acquire additional implicit savings or insurance which they might value. The only reason we might expect people to respond to changes in NICs differently from another tax on earnings is if they (wrongly) perceive it differently.

3. Literature Review

Interest in the extent to which labour supply and demand respond to taxation, and the economic incidence of labour taxes, is longstanding in applied economics (Blundell and Macurdy (1999), Fullerton and Metcalf (2002), Hamermesh (1996)). The scale of these supply and demand responses are a key driver of the efficiency costs of labour taxation (and, for instance, the revenue generated by a given tax or tax reform). Of course, the issues of behavioural response and incidence are intimately linked: it is the relative responsiveness of labour supply and demand that determines labour tax incidence in a classical labour market. Starting in the mid-1990s the growing availability of micro-level administrative or pseudo-administrative data on individuals' and employers' taxes, taxable income, and earnings, spurred a reinvigoration of the study of both the behavioural effects of taxation and incidence.

First, following the seminal work by Feldstein (1995, 1999), the NTR literature shifted the focus of the analysis of the behavioural effects of taxation from the estimation of hours-of-work elasticities to taxable income elasticities. Changes in total taxable income may reflect not only changes in hours of work, but also changes in hourly wages, changes in non-labour income, and changes in tax avoidance (such as shifting income into untaxed forms) and tax evasion (such as not reporting income). Under certain conditions, Feldstein (1995) demonstrated that this overall response of total taxable income to the net-of-marginal tax rate is a summary statistic of the deadweight losses due to taxation.

A large literature has subsequently developed, mostly in the US, but more recently in Western Europe (Saez et al (2012)). In general, these studies find that: the elasticity of taxable income is greater than the elasticity of overall gross income,[7] suggesting tax avoidance and the use of tax shelters is an important component of overall taxable income responses; relatedly, that taxable income elasticities are lower when there are fewer avoidance opportunities;[8] and, that taxable income elasticities are generally higher for those with the highest incomes, perhaps reflecting the greater avoidance opportunities available to such individuals.

[7] Overall gross – or 'broad' – income includes income from sources which can be deducted from income tax, such as that used for charitable donations or pension contributions, and in many countries, income used to pay mortgage interest expenses.

[8] Saez et al (2012) report that, on average, studies in the US find lower estimated elasticities for tax changes in the 1990s than those in the 1980s, when the tax base was narrower. Kopczuk (2005) explicitly examines how the taxable income elasticity changes when the share of income that is deductible changes, and finds that decreases in deductibility reduce estimated taxable income elasticities.

A number of papers examine the responsiveness of labour income specifically to taxation, finding a lesser degree of responsiveness than for overall taxable income. Blomquist and Selin (2010) find a taxable earnings elasticity of 0.2 for Swedish men; Saez (2003) finds a statistically insignificant taxable earnings elasticity of 0.1 for the US; and Kleven and Schultz (2014) find an elasticity of 0.05 – 0.12 in Denmark.[9] Relatively few studies have attempted to estimate the elasticity of earnings subject to social security with respect to the social security tax rate in a similar manner (instead, analysis of SSCs tend to focus on tax incidence, as discussed below). Lehmann et al (2013) is one that does, using reforms to France's income tax credits and employer payroll taxes, to separately estimate the elasticity of employer cost (gross earnings plus employer SSCs) to the net-of-marginal and net-of-average income tax and social security tax rates. They find a statistically significant elasticity with respect to the net-of-marginal income tax rate of 0.2, but virtually no effect of the net-of-marginal social security tax rate on employer cost. They find an elasticity with respect to the net-of-average income tax rate of -0.44 (although this is statistically insignificant), and an elasticity with respect to the net-of-average social security tax rate of -0.866. They interpret this latter finding as indicating that, at least in the short term, the statutory incidence of employers' SSCs on employers' matters: employers bear the cost of these in the form of higher costs, rather than workers in the form of lower wages.[10]

An implicit maintained assumption in most of the rest of the NTR literature is that the incidence of taxes is on the individual in question, rather than, for instance, their employer. This is what allows estimated coefficients on net-of-marginal and net-of-average tax rates to be interpreted as individual responses to taxation. In their review of the taxable income elasticity literature, Saez et al (2012) discuss the implications of such incidence effects (which they term 'classical general equilibrium' effects). They point out that the shifting of tax incidence represents, in effect, a transfer of the burden of a tax from one factor of production (e.g. labour) to another (e.g. capital). If these different factors are taxed at different rates (due to, for example, the levying of SSCs on labour but not capital income), then such shifts may have revenue effects. But, following Diamond and Mirrlees (1971), such transfers do not affect the overall efficiency cost of taxation, nor optimal tax rates. Nevertheless, estimated responses to net-of-marginal tax rates would reflect responses by both sides of the market (not just the individual taxpayer in question), and estimated responses to net-of-average tax rates would at least in part reflect the incidence of a tax, and not only income effects.

The literature on the incidence of labour taxation, on the other hand, largely ignores the issue of non-hours-of-work labour supply responses that are at the heart of the NTR literature. Seminal papers by Gruber (1997) and Anderson and Meyer (1997, 2000), for

[9] Kleven and Schultz (2014) indeed find that taxable income elasticities are greater for capital income than labour income, and for deductions than positive income.

[10] As discussed below, assuming away income effects, an 'elasticity' of employer cost with respect to the net-of-average social security tax rate of -1 would indicate full incidence on the employer. The estimate of -0.866 is not statistically different from this.

instance, examine the impact on earnings and employment of changes in SSCs that affect different firms differently in Chile and the United States, respectively. Gruber (1997) finds that establishment-level earnings increased effectively one-for-one with reductions in establishment-level employers' social contributions, while establishment-level employment was unaffected. Anderson and Meyer (1997, 2000) find broadly similar results for earnings for industry-level changes in employers' social contributions (although the link between establishment-level changes in employers' social contributions is much weaker).

In both instances, these authors interpret these earnings effects as evidence that the incidence of employer SSCs in question is largely, or fully, on workers. But, it could be the case that part of the change in earnings when SSCs change instead reflects increases in productivity that follow from changes in effort induced by the tax change. In other words, there is the risk that such behavioural responses confound estimates of tax-shifting.

Subsequent papers making use of differential impacts of changes in labour taxes across workers adopt a similar approach Kubik (2004), Leigh (2010), Rothstein (2010), for instance examine the incidence of the personal income tax and Earned Income Tax Credit (EITC) in the US, and interpret changes in hourly earnings as indicative of tax incidence. Using this methodology, they find that a significant proportion of income tax or tax credits are shifted from employees to employers: wages fall for groups affected by income tax cuts or EITC increases. Again these estimates of incidence could be confounded by effort-induced changes in productivity, and hence wages.

Notwithstanding these issues, Melguizo and González-Páramo (2013) undertake a review and meta-analysis of the incidence literature,[11] and find that, on average, studies find around one-third of the burden of labour taxes is borne by employers – or more correctly their shareholders or customers – and two-thirds by employees, but that there is a very wide range of estimates.[12] Some studies such as Gruber (1997), find almost full incidence of employers' SSCs on workers, while others find the incidence remains on employers. Most of these examine short term incidence (Lehmann et al (2013), Bunel and L'Horty (2012)), but a few suggest longer-run incidence on employers (Saez et al (2012b)).

In a classical labour market model, the formal incidence of a tax should not matter for its economic incidence. Wages should adjust, based on the relative elasticities of labour supply and demand, irrespective on which side of the market a tax is formally levied. Relatively few studies directly examine this Invariance of Incidence Proposition (IIP), however. This likely reflects the fact that many reforms to SSCs affect employee and employer contributions simultaneously and highly co-linearly, making identification of

[11] Which also includes studies making use of cross-country time series analysis, relating changes in average earnings or labour-income shares to changes in labour taxes.
[12] A more narrative review is available in CPB et al (2015).

separate effects difficult. Among those studies that do, which are mainly based on cross-country regressions (such as OECD (1990) and Arpaia and Carone (2004)), there is evidence that the IIP is violated in the short-term, perhaps reflecting short-term stickiness of nominal wages. Whether the IIP holds in the longer-run is less clear, in part because of weak statistical power in the long-run analyses contained the papers that examine this issue (CPB et al (2015)).

By making use of a panel covering a long time period in which there were separate reforms to employee and employer SSCs, we can examine the IIP. As discussed further below, we also examine slightly longer-run effects of changes in SSCs and income taxes on earnings and hours of work by including lagged changes in the relevant tax rates.

Bingley and Lanot (2002) is a rare example of a paper which unites the NTR and tax incidence literatures. In order to identify both incidence and behavioural response, they make use of the fact that income tax rates vary across municipalities in Denmark, and the tax rate levied depends upon where an employee lives, as opposed to where the employer is based. This means that a given employer (or more precisely, establishment) will often employ people who face a range of different income tax rates. At the same time, the mean tax rate faced by a firm's employees will differ across employers based on where they are located, as employees tend to be drawn from a fairly local pool. Bingley and Lanot assume that establishment- or employer- level wage setting means that employers cannot vary wage rates paid to employees based solely on differences in the income tax rates they face. Any *within-employer* variation in the earnings of 'similar employees' associated with differences in the income tax rates they face, are therefore assumed to be the result of behavioural response (e.g. working longer hours, or working harder per hour). Information on the incidence of the income tax is then backed out using information on *between-employer* variation in earnings and mean income tax rates, and what is known about behavioural response from the *within-employer* variation. Results suggest there is partial (40%) shifting of the burden of income tax from workers to employers.

NICs – the UK's system of SSCs – are uniform across the UK, and vary only by level of income/earnings (although benefits and tax credits also vary by family composition). This precludes adopting a similar approach in our analysis. In the absence of the kind of policy variation, we instead use two approaches and sets of assumptions to analyse incidence and behavioural responses to taxation. The first, following the approach generally taken in the incidence literature, involves examining changes in hourly employer cost (defined as wages plus employers' social security) and hours of work, and requires assumptions about productivity/effort responses to changes in tax rates in order to separately identify behavioural effects and incidence. The second involves examining overall changes in a workers' cost to their employer (e.g. per week or per month). Assumptions on income effects are required to separately identify behavioural effects and incidence.

4. Approaches to identifying incidence and behaviour

More precisely, our approach builds on the work of Lehmann et al (2013), who examine the short-term effects on employer cost (i.e. gross earnings plus employer SSCs) of income tax and employer SSCs in France using reforms during the early 2000s. We start with the same simple behavioural function determining employer cost:

$$Z = Z(\tau^R, \tau^E, R^R, R^E) \qquad (3.1)$$

Where τ^R and τ^E are the net-of-marginal employer and employee NICs rates respectively, and R^R and R^E are the virtual net-of-NICs and net-of-employer-NICs incomes. Lehmann et al show that differentiating with respect to the various features of the tax system and rearranging, you can obtain the following expression:

$$\frac{\Delta Z}{Z} = \beta^R_{Z,\tau} \frac{\Delta \tau^R}{\tau^R} + \beta^E_{Z,\tau} \frac{\Delta \tau^E}{\tau^{RE}} + \beta^R_{Z,\rho} \frac{\Delta \rho^R|}{\rho^R} + \beta^E_{Z,\tau} \frac{\Delta \rho^E|}{\rho^E} \qquad (3.2)$$

where ρ^R is the net-of-average employer NICs rate (i.e. gross earnings divided by employer cost), and ρ^E is the net-of-average employee NICs rate (i.e. net earnings divided by gross earnings). $\Delta\rho^R|$ and $\Delta\rho^E|$ are the changes in these net-of-average NICs rates calculated holding earnings fixed at their initial (pre-reform) level, which differs from the actual changes ($\Delta\rho^R$ and $\Delta\rho^E$) if the NICs schedule is not linear and employees or employers respond to changes in NICs by changing their labour supply or labour demand. Lehmann et al (2013) show that using the actual change in average tax rates or virtual income (as is usually done in the NTR literature following Gruber and Saez (2002)) may lead to inconsistent estimates, even if instrumented for.

The coefficients $\beta^R_{Z,\tau}$ and $\beta^E_{Z,\tau}$ denote elasticities of employer cost with respect to compensated changes in net-of-marginal employer and employee NICs, respectively. The standard interpretation of these coefficients in the NTR literature is as compensated labour supply or other individual responses (i.e. pure substitution effects). If one allows for less than fully elastic labour demand – and therefore the prospect of at least partial incidence of NICs on employers – the coefficients will pick up a combination of compensated labour supply and labour demand effects.

Similarly, the coefficient on net-of-average NICs rates ($\beta^R_{Z,\rho}$ and $\beta^E_{Z,\rho}$) would generally be assumed to capture income effects, since it is the change in average tax rate, irrespective of the marginal tax rate, that captures how much income has been lost/gained and might prompt more/less work effort to make up the difference. However, these coefficients will also capture the incidence of NICs changes: shifting of the burden of

NICs means employer costs rising/falling as the total amount of NICs charged, reflected in the average rate, changes.[13]

So (how) can we separate out incidence and income effects? It is impossible to do so purely by looking at employer cost responses to tax rate changes without making additional assumptions. For example, if employer cost increases in response to an increase in the average rate of employee NICs, that could reflect employees' working harder to make up the loss of income (a standard income effect) or some shifting of the burden of the tax increase onto employers (without necessarily any change in the amount of work being done). As already discussed, most of the existing NTR and incidence literatures implicitly assume away one or other of these possibilities.

We take two approaches.

The first is simply to analyse what alternative assumptions would imply for the interpretation of our results. Specifically, we could (a) assume that labour demand is perfectly elastic and the incidence of NICs is therefore fully and immediately on employees, so responses to average tax rates reflect income effects, and then instead (b) assume that income effects are negligible so that earnings responses to average NICs rates reflect the incidence of the change.[14] (Of course, the reality may be somewhere between these two extreme assumptions, meaning $\beta_{Z,\rho}^{R}$ and $\beta_{Z,\rho}^{E}$ pick up some combination of income effects and incidence). Under assumption (b), in principle, the estimate of NICs incidence – which in a classical labour market would reflect the relative elasticities of labour supply and demand – could be used to back-out the labour supply and demand elasticities from the overall compensated elasticities with respect to marginal rates, $\beta_{Z,\tau}^{R}$ and $\beta_{Z,\tau}^{E}$. Table 2 shows the values the various coefficients would be expected to take under these different assumptions.

Row (1) shows what we would expect if labour demand is fully elastic and the incidence of both employee and employer NICs is on the employee. $\beta_{Z,\tau}^{R}$ and $\beta_{Z,\tau}^{E}$ pick up compensated labour supply responses and should therefore be greater than or equal to 0, and in most standard models of the labour market should be equal to each other. The coefficients on net-of-average tax rates would pick up income effects and should also be equal to each other, but in this case should be less than or equal to 0. If coefficients for employee or employer differ from each other non-standard features such as differing salience of the two taxes may play a role (Chetty et al (2009), Lehmann et al (2013)).

The following rows rule out income effects. Row (2) shows what would happen if labour demand were less than fully elastic and the incidence of NICs was shared between employees and employers. $\beta_{Z,\tau}^{R}$ and $\beta_{Z,\tau}^{E}$ would pick up a combination of compensated

[13] In other contexts the average tax rate should also capture extensive margin (i.e. employment) responses, but since we condition our sample on being observed in employment we do not estimate extensive margin responses in this paper.

[14] It would be possible to take this further to consider alternative, more complicated assumptions, most obviously assuming some particular non-zero size of income effects or some particular degree of shifting.

labour supply and labour demand responses and should therefore be greater than or equal to 0. The coefficients on net-of-average tax rates would pick up the shifting of employee NICs partly on to employers and vice versa, and would therefore lie between -1 (full incidence on employers) and 0 (full incidence on employees). This is similar to what would be observed if incidence was fully on employees but there were income effects. It is therefore difficult to distinguish between income effects and the sharing of the burden of SSCs between employers and employees. Indeed, coefficients on net-of-average tax rates could pick up both income effects and incidence.

Table 2. Coefficient values under various assumptions about incidence and income effects, employer cost

		Net-of-marginal rate coefs.		Net-of-average rate coefs.	
		$\beta^R_{Z,\tau}$	$\beta^E_{Z,\tau}$.	$\beta^R_{Z,\rho}$	$\beta^E_{Z,\rho}$
(1)	Full incidence on employee[a]	≥0	≥0	≤0	≤0
Assuming away income effects:					
(2)	Sharing of incidence[b]	≥0	≥0	-1<β<0	-1<β<0
(3)	Full incidence on employer[c]	≥0	≥0	-1	-1
(4)	Statutory incidence[d]	≥0	≥0	-1	0

Notes: (a) In standard models, furthermore $\beta^R_{Z,\tau} = \beta^E_{Z,\tau}$ and $\beta^R_{Z,\rho} = \beta^E_{Z,\rho}$.
(b) In standard models, furthermore $\beta^R_{Z,\tau} = \beta^E_{Z,\tau}$ and $\beta^R_{Z,\rho} = \beta^E_{Z,\rho}$.
(c) Unless labour supply was fully elastic, full incidence on employers requires $\beta^R_{Z,\tau} = \beta^E_{Z,\tau} = 0$.
(d) Statutory incidence requires models with at least temporary gross wage stickiness.

Row (3) shows the values the coefficients would take if taxes were fully incidence on employers. Employer cost would move one-for-one with changes in net-of-average NICs rates ($\beta^R_{Z,\rho} = \beta^E_{Z,\rho} = -1$). Coefficients on net-of-marginal rates would in general be expected to be 0, as in standard models, full incidence on workers would require completely inelastic labour demand unless labour supply were perfectly elastic (however these coefficients could be positive).

Row (4) shows what we would expect if economic incidence followed statutory incidence. Changes in average rates of employee NICs would be borne by employees and leave employer cost unaffected ($\beta^E_{Z,\rho} = 0$), while changes in average rates of employee NICs would be borne by employers and affect employer cost one-for-one ($\beta^R_{Z,\rho} = -1$). $\beta^E_{Z,\tau}$ would pick up compensated labour supply responses to employee NICs and would be greater than or equal to 0.

Lehmann et al (2013) states that $\beta^R_{Z,\tau}$ would equal 0 in such circumstances. This may reflect the fact that in standard labour market models, employers would only bear the burden of NICs if their labour demand was perfectly inelastic. However, for economic incidence to follow statutory incidence in this manner requires a degree of gross earnings stickiness. If employers and employees (or their representatives) contract on

the basis of gross wages/earnings, which is generally the case, it may take some time for wages/earnings to adjust following changes in employee and/or employer NICs rates. Thus it may not be surprising to find economic incidence following formal incidence in the short term. Indeed, if employers and employees also bargain over these gross-earnings levels, in some models of the labour market, formal incidence may matter for economic incidence on a more long-term basis (perhaps explaining the findings of Saez et al (2012b)). But in such models, it need not be the case that labour demand is completely inelastic for employer NICs to be incident on employers. Thus we might expect $\beta^R_{Z,\tau} \geq 0$ rather than $\beta^R_{Z,\tau}=0$.

The second approach we take exploits the availability of hours of work in our data. In particular, we estimate the empirical counterparts to:

$$\frac{\Delta H}{H} = \beta^R_{H,\tau}\frac{\Delta\tau^R}{\tau^R} + \beta^E_{H,\tau}\frac{\Delta\tau^E}{\tau^E} + \beta^R_{H,\rho}\frac{\Delta\rho^R|}{\rho^R} + \beta^E_{H,\tau}\frac{\Delta\rho^E|}{\rho^E} \qquad (3.3)$$ and

$$\frac{\Delta(Z/H)}{(Z/H)} = \beta^R_{Z/H,\tau}\frac{\Delta\tau^R}{\tau^R} + \beta^E_{Z/H,\tau}\frac{\Delta\tau^E}{\tau^E} + \beta^R_{Z/H,\rho}\frac{\Delta\rho^R|}{\rho^R} + \beta^E_{Z/H,\tau}\frac{\Delta\rho^E|}{\rho^E} \qquad (3.4)$$

where H is hours of work, and Z/H is hourly labour cost.

If we assume that any income effects operate entirely through hours of work, the coefficients on net-of-average tax rates in the hourly employer cost regression ($\beta^R_{Z/H,\rho}, \beta^E_{Z/H,\rho}$) would pick up the incidence of a tax only. The coefficients from the hours of work regressions would pick up the combined effects of demand and supply responses ($\beta^R_{H,\tau}, \beta^E_{H,\tau}, \beta^R_{H,\rho}, \beta^E_{H,\rho}$).

If one rules out income effect operating via employer cost per hour, it would also be natural to rule out substitution effects too. Doing this means ruling out the kinds of non-working-hours behavioural responses to taxation that, in part, motivated the NTR literature: notably effort per hour, but also other behavioural responses that operate through the observed hourly wage, such as shifting to/from unobserved forms of remuneration that are not subject to the tax change. We cannot test this assumption *per se*, but note that if substitution effects operate entirely through hours of work then we would expect marginal tax rates to have no effect on hourly employer costs (since incidence should be a function of the average rate, not the marginal rate), so we can test for this by looking whether marginal rates are significant in our hourly employer costs regression (i.e. test whether we can reject that $\beta^R_{Z/H,\tau}$ and $\beta^E_{Z/H,\tau}=0$). In principle it is possible that income effects operate through the hourly wage but substitution effects do not, or vice versa, but this evidence should at least be suggestive. Tables 3 and 4 shows the values the various coefficients would be expected to take in the hours and hourly employer cost equations, respectively, under different assumptions about incidence and whether behavioural effects operate via hourly employer cost.

Table 3. Coefficient values under various assumptions about incidence and income effects, hours

		Net-of-marginal rate coefs.		Net-of-average rate coefs.	
		$\beta^R_{H,\tau}$	$\beta^E_{H,\tau}$.	$\beta^R_{H,\rho}$	$\beta^E_{H,\rho}$
(1)	Full incidence on employee[a]	≥0	≥0	≤0	≤0
Assuming away income effects:					
(2)	Sharing of incidence[b]	≥0	≥0	0	0
(3)	Full incidence on employer[c]	≥0	≥0	0	0
(4)	Statutory incidence[d]	≥0	≥0	0	0

Notes: (a) In standard models, furthermore $\beta^R_{H,\tau} = \beta^E_{H,\tau}$ and $\beta^R_{H,\rho} = \beta^E_{H,\rho}$.
(b) In standard models, furthermore $\beta^R_{H,\tau} = \beta^E_{H,\tau}$ and $\beta^R_{H,\rho} = \beta^E_{H,\rho}$.
(c) Unless labour supply was fully elastic, full incidence on employers requires $\beta^R_{H,\tau} = \beta^E_{H,\tau} = 0$.
(d) Statutory incidence requires models with at least temporary gross wage stickiness.

With full incidence of NICs on employees, coefficients in the hours equations (Table 3) would reflect standard labour supply substitution effects (row 1). In the absence of income effects, coefficients on net-of-average NICs rates would be expected to be zero (incidence effects would operate via hourly employer cost). Coefficients on net-of-marginal NICs rates would pick up combinations of labour supply and demand elasticities. Coefficients in hourly employer cost equations (Table 4) could pick up both incidence and behavioural effects. If income effects are assumed not to affect hourly employer cost (e.g. via effort/productivity margins), incidence can be inferred from coefficients in net-of-average NICs rates in the same way as in the overall employer cost equations.

Table 4. Coefficient values under various assumptions about incidence and income effects, hourly employer cost

		Net-of-marginal rate coefs.		Net-of-average rate coefs.	
		$\beta^R_{Z,\tau}$	$\beta^E_{Z,\tau}$.	$\beta^R_{Z,\rho}$	$\beta^E_{Z,\rho}$
With					
(1)	Full incidence on employee[a]	≥0	≥0	≤0	≤0
Assuming away income effects:					
(2)	Sharing of incidence[b]	≥0	≥0	-1<β<0	-1<β<0
(3)	Full incidence on employer[c]	≥0	≥0	-1	-1
(4)	Statutory incidence[d]	≥0	≥0	-1	0
Also assuming away substitution effects:					
(5)	Sharing of incidence[b]	0	0	-1<β<0	-1<β<0
(6)	Full incidence on employer[c]	0	0	-1	-1
(7)	Statutory incidence[d]	0	0	-1	0

Notes: (a) In standard models, furthermore $\beta^R_{Z/H,\tau} = \beta^E_{Z/H,\tau}$ and $\beta^R_{Z/H,\rho} = \beta^E_{Z/H,\rho}$.
(b) In standard models, furthermore $\beta^R_{/HZ,\tau} = \beta^E_{Z/H,\tau}$ and $\beta^R_{Z/H,\rho} = \beta^E_{Z/H,\rho}$.
(c) Unless labour supply was fully elastic, full incidence on employers requires $\beta^R_{Z/H,\tau} = \beta^E_{Z/H,\tau} = 0$.
(d) Statutory incidence requires models with at least temporary gross wage stickiness.

5. Econometric methodology

As discussed above, the objective of our analysis is to identify the responsiveness of labour cost, hours of work, and labour cost per hour to employer and employee SSCs. Furthermore by placing certain restrictions on assumed behaviour – such as the absence of income effects, or of non-hours labour supply responses –, it is possible to interpret findings as indicating the extent of incidence-shifting and underlying behavioural (labour supply and demand) responses.

We take our lead from Lehmann et al (2013), and estimate the following empirical counterparts to equations 3.2, 3.3 and 3.4:

$$\Delta lnZ_{i,t} = \alpha_Z + \beta^R_{Z,\tau}\Delta ln\tau^R_{i,t} + \beta^E_{Z,\tau}\Delta ln\tau^E_{i,t} + \beta^R_{Z,\rho}\Delta ln\rho^R_{i,t}| + \beta^E_{Z,\rho}\Delta ln\rho^E_{i,t}| + +\boldsymbol{\gamma_Z X_{i,t}} + \varepsilon_{i,t,Z} \quad (5.1)$$

$$\Delta lnH_{i,t} = \alpha_H + \beta^R_{H,\tau}\Delta ln\tau^R_{i,t} + \beta^E_{H,\tau}\Delta ln\tau^E_{i,t} + \beta^R_{H,\rho}\Delta ln\rho^R_{i,t}| + \beta^E_{H,\rho}\Delta ln\rho^E_{i,t}| + \boldsymbol{\gamma_H X_{i,t}} + \varepsilon_{i,t,H} \quad (5.2)$$

$$\Delta ln(Z/H)_{i,t} = \alpha_{Z/H} + \beta^R_{Z/H,\tau}\Delta ln\tau^R_{i,t} + \beta^E_{Z/H,\tau}\Delta ln\tau^E_{i,t} + \beta^R_{Z/H,\rho}\Delta ln\rho^R_{i,t}| + \beta^E_{Z/H,\rho}\Delta ln\rho^E_{i,t}| + \boldsymbol{\gamma_{Z/H} X_{i,t}} + \varepsilon_{i,t,Z/H} \quad (5.3)$$

Where changes in employer cost, etc., are calculated for periods of 1 year in length. $\boldsymbol{X_{i,t}}$ is a vector of controls, including time period dummies (to pick up, for instance, the effect of inflation, in 5.1 and 5.3), and controls for differential trends in different parts of the employer costs or hours distributions. $\varepsilon_{i,t}$is an error term that captures unobserved and time-varying heterogeneity.

It is well known from the labour supply and NTR literatures that various issues arise with estimation of such equations. The first is a potential simultaneity bias. Because of the nonlinearity of the employee and employer NICs schedules, the marginal net-of-tax rates $\tau^E_{i,t}$ *and* $\tau^R_{i,t}$ are functions employer cost, hours or hourly employer cost. Thus, to identify the effect of NICs on labour cost, hours, etc, we need instruments.

In the literature, the standard approach, proposed by Auten and Caroll (1999), and used in such influential studies as Gruber and Saez (2002), is to use the change in the log net-of-tax rate should earnings remain unchanged at their t-1 level (adjusted for inflation, of average earnings growth, say). Thus, by construction, the instrument captures changes in tax rates that strip out the effect of any behavioural response. More formally one

instruments for $\Delta ln\tau_{i,t}^{N}$ using $\Delta ln\tau_{i,t}^{N}|= ln\tau_{i,t}^{N}| - ln\tau_{i,t-1}^{N}$ where $\tau_{i,t}^{N}|= \tau_{i,t}^{N}|\mathrm{W_{i,t-1}}$, and n = E or R.[15] These are the 'type-I' instruments used by Lehmann et al (2013).

A further issue arises with such instruments though: non-tax related changes in employer cost, hours, or employer cost per hour. These changes may be correlated with the level of these variables, and hence with changes in tax rates that affect different parts of the earnings or hours distributions. For instance, during the 1980s, earnings inequality was increasing significantly in the UK (Blundell and Etheridge (2010)). At the same time, changes in NICs (most notably in 1985) reduced average NICs rates at the bottom of the earnings distribution, and increased them at the top. The risk is that one would inappropriately attribute those changes in employer cost, for instance, that relate to these non-tax factors to the tax reform instead, biasing estimated coefficients.

When using panel data, mean reversion constitutes another significant identification problem. An individual with an unusually low (respectively high) employer cost in period t−1 due to a temporary earnings shock is likely to have higher (lower) earnings in period t, when the shock has (at least partly) worked its way out. Again, these changes in employer cost could be erroneously attributed to a tax change affecting individuals at particular levels of employer cost.

Again, since Auten and Carroll (1999), the standard procedure to control for differential secular trends and mean reversion has been to include a function of income in period t-1 in the set of controls $\boldsymbol{X_{i,t}}$. Auten and Carroll (1999) use a simple linear control ($lnZ_{i,t-1}$), while Gruber and Saez (2002) propose a 10-piece spline. Kopczuk (2005) points out that mean reversion and differential secular trends across earnings groups are two separate issues. Mean reversion is about predictable changes in earnings following previous *changes* in earnings. Thus to control for this, Kopczuk (2005) uses a 10-piece spline of the log difference between income in period t-2 and t-1 (for example, $\Delta lnZ_{i,t-1}$). Differential secular trends relate to changes in earnings that differ across different parts of the earnings distribution. For this, a 10-piece spline of the log of income in period t-2 is used (for example, $lnZ_{i,t-2}$). Lehmann et al (2013) utilises these two sets of controls.

Weber (2014) offers a critique to these 'standard' approaches, arguing that such controls cannot properly address the problem of mean reversion, and that instruments based on income/earnings in period t-1 will not be exogenous. Instead, she proposes to instrument for changes in the net of tax rate using changes in these tax rates calculated holding income/earnings fixed at its level in period t-1-k, (rather than period t-1). The lag k should be chosen so that it is far enough before the period in question so that income/earnings in period t-1-k are unaffected by any transitory shocks affecting income/earnings in period t-1, but not so far that the instrument is a poor predictor of

[15] Note that in the specification, we include average tax rates calculated holding earnings held fixed at t-1 levels directly: $\Delta ln\rho_{i,t}^{I}|$. This does not need to be instrumented if all one is concerned about is simultaneity.

the actual change in tax rate observed. Further, Weber argues that one can test for whether a particular instrument is exogenous, conditional upon the assumption that other excluded instruments (for example, based on longer-lags) are exogenous using a Difference-in-Sargan test. Using US income tax data from the 1980s she finds that she cannot reject that instruments based on income in period t-2 (i.e. k=1) are exogenous (p-value: 0.229), but her preferred specification is based on instruments in period t-3 (p-value: 0.858).

Lehmann et al (2013) use changes in net-of-tax rates calculated at earnings held fixed at their levels in period t-2 (i.e. k=1) as their type-II instruments. In this paper, we use such instruments in our main specification, following testing of them using Difference-in-Sargan tests. As with Weber, we reject the exogeneity of (type-I) instruments based on earnings in period t-1, even when controls are included (p-values of 0.00 – 0.01, depending on precise specification). We cannot reject the exogeneity of our preferred type-II instruments based on earnings in period t-2 (p-values of 0.15 – 0.7). We also report estimates based on instruments based on earnings in period t-3 to examine the robustness of our main specification. To control for differential trends in employer cost and hours at different parts of the distribution, we include controls based on $lnZ_{i,t-2}$. (but not $\Delta lnZ_{i,t-1}$, as our use of lagged earnings in the instrument addresses mean-reversion concerns).

Behaviour and incidence in the longer-run

Examining changes in employer cost, hours and employer cost per hour between period t-1 and t, and relating this to changes in tax rates between period t-1 and t allows one to pick up only very short-run behavioural effects and incidence shifting given many tax reforms take place at the start of April, the point at which our data is collected (see below). Frictions in behaviour and earnings may mean that it takes some time for individuals and employers to respond to tax policy changes.

The traditional response to this problem is to use panel lengths of longer than one year. For instance, rather than calculating $\Delta lnZ_{i,t} = lnZ_{i,t} - lnZ_{i,t-1}$, instead calculating $\Delta lnZ_{i,t} = lnZ_{i,t} - lnZ_{i,t-2}$ or $\Delta lnZ_{i,t} = lnZ_{i,t} - lnZ_{i,t-3}$. This is the approach taken in Gruber and Saez (2002), who stack these 3-year changes (so that their estimates are based on changes in incomes and tax rates between 1979 and 1982, 1980 and 1983, 1981 and 1984, etc). However such an approach does not actually estimate responses after 3 years.

To see this, consider a policy change occurring in period T. Estimates based on the changes between period T and T+3 will pick up the effects of the policy change after 3 periods; estimates based on the change between period T-1 and T+2 will pick up the effects after 2 periods; and estimates based on the change between period T-2 and T+1 will pick up the effects after 1 period. Estimates based on stacked 3-year changes would therefore, at best pick up an average of responses over 1 to 3 periods.

If there are multiple reforms, however, estimation may be confounded. To see this consider a tax increase in period T that is followed by a tax decrease in period T+1. Changes in earnings (or hours) between period T and T+3 will pick up the effects of the period T tax increase after 3 periods, and period T+1 tax decrease after 2 periods. However, the entire change will be attributed to the period T tax increase. In such circumstances, estimated elasticities would not represent even an average of shorter- and longer- run responses.

Thus, rather than adopt this approach, in this paper we instead include lagged changes in marginal and average tax rates as regressors. We estimate regressions of the form:

$$\Delta lnZ_{i,t} = \alpha_Z + \sum_{n=0}^{1}(\beta_{Z,\tau}^{R,n}\Delta ln\tau_{i,t-n}^{R} + \beta_{Z,\tau}^{E,n}\Delta ln\tau_{i,t-n}^{E} + \beta_{Z,\rho}^{R,n}\Delta ln\rho_{i,t-n}^{R}| + \beta_{Z,\rho}^{E,n}\Delta ln\rho_{i,t-n}^{E}|) + \boldsymbol{\gamma_Z X_{i,t}} + \varepsilon_{i,t,Z} \quad (5.4)$$

In this example, $\beta_{Z,\tau}^{R,0}$ picks up the short-run effect of changes in the net-of-marginal employer NICs rate between period t-1 and t on the changes in employer cost, Z, between period t-1 and t. $\beta_{Z,\tau}^{R,1}$ picks up the effect of changes in the net-of-marginal employer NICs rate between period t-2 and t-1 on the changes in employer cost, Z, between period t-1 and t, *on top* of any initial impact on employer cost, Z, between period t-2 and t-1. Thus the effects of tax changes after two periods (approximately 12-18 months) can be calculated by adding the coefficients on the contemporaneous and lagged changes in tax rates (e.g. $\beta_{Z,\tau}^{R,0} + \beta_{Z,\tau}^{R,1}$).

As with contemporaneous changes in tax rates, it is important to instrument $\Delta ln\tau_{i,t-n}$ and $\Delta ln\rho_{i,t-n}|$ appropriately. We do this using instruments based on earnings held fixed in period t-n-1 (i.e. setting the lag k=1) as in our basic specification examining short-run responses. We again test the robustness of results to using instruments based on earnings held fixed at period t-n-2 levels.

6. Data

The data used to estimate these regressions come from the New Earnings Survey Panel Dataset (NESPD), a mandatory survey of employers' payroll records collecting data on employees' earnings and basic characteristics for a pay period each April.[16] The target sample frame of the NESPD is civilian employees in Great Britain whose National Insurance (NI) number ends with a specific pair of digits. Since the last digits of NI numbers are allocated randomly to all adults and the NESPD sample uses same pair of digits each year, in principle this should deliver a random 1% panel sample of employees; and since we have data from 1978 to 2010, we can follow the same individuals in any year in which they are employed (including if they change employer or region, say) for up to 35 years – a period during which there was a great deal of reform to the NICs schedule, as we describe in Section 2. At around 165,000 individuals

[16] The NESPD is in fact the result of joining together the old New Earnings Survey and the similar Annual Survey of Hours and Earnings which replaced it from 2004.

per year, the NESPD contains a much larger sample than is available in other datasets of hours and earnings (such as the Labour Force Survey and the Family Resources Survey) and does not suffer from the same degree of measurement error, as responses are provided by employers with reference to their payroll and employment records.

In practice, the NESPD is not quite a random 1% sample of employees. In fact it includes around 0.7% of employees on average over the period (1% of employees in Britain would be around 235,000 per year, not the 165,000 we actually observe). The main reason for this is that, despite supposedly being mandatory, the survey suffers from significant non-response. The valid response rate fell from over 75% in the 1980s to around 60% by 2012.[17] Non-response reduces sample size and therefore the precision of our estimates, though as noted above our sample remains large. More troubling is that non-response could lead to bias in our estimates if it is correlated with labour market behaviour. But again, we note that non-response is typically at least as big a problem in these alternative household survey datasets.[18]

We do not observe people when they are not employed, and cannot distinguish whether an employee who is absent from a particular year of data was not working, was self-employed, or was working for an employer who failed to respond to the survey. Our estimation uses only people who we observe in employment in successive waves of the NESPD (though not necessarily in the same job, or even in employment throughout the year in between waves), and we are estimating earnings responses among that subset of the population. Thus the estimates of employer cost and hours responsiveness to NICs we obtain are purely intensive margin responses.

In order to calculate changes in net-of-NICs rates, and construct instruments based on earnings held fixed in period t-2, our estimation sample includes only those individuals observed for at least 3 consecutive years in the NESPD. When including lagged changes in NICs rates to examine slightly longer-run effects – or sensitivity testing our results using instruments based on earnings held fixed in period t-3 –, we further restrict our sample to those observed for 4 or 5 consecutive years. This means that our sample consists of those who are working in employers who respond to the survey in 3 – 5 consecutive survey waves (although they may be out of work between waves). The behaviour of such employees and their employers may differ to that of the rest of the population, and it is important to note our results relate to this sub-sample of the population only. The achieved sample size in each set of regressions is between 1.5 and 2 million (exact sample sizes are reported with the relevant results), which is approximately 33-40% of the overall NESPD sample for the years in question.

Key variables

[17] Source: authors' correspondence with the Office for National Statistics.

[18] There are a number of more minor reasons that our sample may not be completely random, although we do not expect these to have a significant effect on the validity of results (see Appendix A: Data).

The main earnings variable recorded in the NESPD measures total cash earnings (including pay for overtime, shift premiums, commission, performance-related pay, etc.), excluding benefits in kind and employer pension contributions but without deducting employee pension contributions, relating to a particular pay period (typically a week or month, but in all cases converted to a weekly equivalent by the data provider). This corresponds closely to the tax base for NICs, which is levied on a similar definition of earnings and is charged separately in each pay period. Thus the behavioural parameters we estimate are close to true taxable earnings elasticities for NICs purposes, and the elasticities capture all the corresponding behavioural responses: not only labour supply and demand but also shifting to make more/less use of forms of remuneration such as employer pension contributions which are not subject to NICs.

One slight wrinkle to this relates to benefits in kind:

- Some things we might think of as benefits in kind (broadly those that can be exchanged for cash or are equivalent to cash, such as goods or services bought by the employee but paid for by the employer) are treated like cash in tax law and subject to NICs in full. It is difficult to know whether employers are including those things when they provide earnings measure in the NESPD; if they are not then our earnings measure underestimates taxable earnings. Note that this need not bias our estimates unless these taxable benefits respond more or less to changes in NICs rates than cash earnings do.

- Other benefits in kind – the principal ones being company cars and fuel and private medical insurance – were not subject to NICs at all until the 1990s. But, the NICs base was gradually broadened to bring more benefits in kind within the scope of employer NICs (employer NICs were applied to company cars and fuel from 1991, and to most other benefits in kind from 2000),[19] though benefits in kind remain outside the scope of employee NICs. Thus from 1991 our earnings measure will be a slight underestimate of low-paid workers' earnings for employer NICs purposes (though not for employee NICs purposes). As with any other reforms or economic shocks, these extensions to the NICs base can confound our estimation if they are correlated with the changes to NICs rates on which our estimates are based.[20] However, since these benefits in kind account for only around 2% of total remuneration, we do not expect the effect on our estimation to be large.[21]

[19] Except for employees earning less than £8,500 per year, for whom benefits in kind remained outside the scope of NICs – and indeed income tax – until April 2016.
[20] In fact 1991 saw a small reduction in employer NICs rates, and 2000 an increase in the earnings level at which employee NICs became payable, so if (say) employers reduced employees' earnings in those years to pass on the cost of the extension of employer NICs to more benefits in kind then we will be wrongly attributing those earnings reductions to the effects of the changes in NICs rates that happened in those years.
[21] Tolley's (2014).

The survey also includes compatible measures of hours of work (including and excluding overtime, etc.). Hourly wages are derived by dividing earnings by hours of work, which creates some potential division bias insofar as hours are measured with error (there should be little measurement error in earnings as they come from employers' payroll records).

Item non-response is low. Before 2004 observations with missing earnings were removed from the dataset (and are included in the survey non-response rate discussed above), and we also remove observations with missing hours (between 4% and 12% of observations). Since 2004 missing values have been imputed (and such cases are not identified), but the number of such cases is small: around 0.5% of observations have missing earnings imputed and the figure for hours is just over 1%.

Calculating tax rates

The principal independent variables of interest for our analysis are functions of people's marginal and average NICs rates. Since the NESPD measures the tax base – gross earnings – well (see above), we can essentially apply the relevant year's rate schedule to that tax base to calculate marginal and average rates of NICs.

Note that we calculate NICs rates, not overall tax rates. The elasticities we calculate are therefore elasticities with respect to net-of-NICs rates, not net of the overall tax wedge. From 1990 onwards we can also calculate the income tax rates individuals face on their earnings, and in Appendix B we present results using income tax rates as well as NICs rates for the period since 1990. Since most of the biggest reforms to NICs happened in the 1980s, however, restricting attention to the later period sacrifices much of our variation, making identification difficult. Before 1990 income tax in the UK was assessed on married couples' joint income (albeit with an option to elect to be taxed separately, with some loss of allowance); since we do not observe whether the employee is married, or the income of any spouse, we cannot account for this in our analysis, so our main results for the full period use NICs rates alone. Similarly, means-tested benefits and tax credits may also contribute to the effective tax wedge on an individual's earnings, but such entitlements always depend on couples' joint income and other characteristics (such as housing costs and the number and presence of children) that are not observed in our data, so we cannot account for those components of the tax wedge. Ignoring jointly-assessed income tax and means-tested support will chiefly be a problem for our estimation if changes to these other elements of the tax wedge were correlated with changes to the NICs schedule, much like other omitted influences on earnings can confound our estimates if they are correlated with changes in NICs.

Timing

One feature of the data which complicates our analysis is the proximity of the earnings we observe to the turn of the fiscal year. This means that we typically observe earnings

at a time very near changes in NICs rates, and we cannot even be certain which NICs schedule applies to the earnings in question.

The fiscal year in the UK runs from 6 April to 5 April, and changes in NICs rates and thresholds usually take effect at the start of the fiscal year (though not always: two of the biggest reforms, in 1985 and 1989, took effect in October). The NESPD collects information each year about earnings and hours of work in the particular pay period that includes the 'survey reference date', a specific date in April. The precise date varies from year to year, ranging from 4 April to 29 April. Hence the earnings level reported by the employer in the NESPD will refer to the pay period containing the survey reference date, but the applicable NICs rate will generally depend on whether the amount in question is paid before or after 6 April.

Earnings in respect of the pay period containing a particular date in April may be paid before or after 6 April, so we cannot be certain which fiscal year's NICs schedule applies to the earnings in our data. For example, if the employee's pay period is the calendar month then the employer will record their April earnings in the survey; but if the employee is paid on the first day of each month then those April earnings will be subject to the NICs schedule for the old fiscal year (ending on 5 April), whereas if they are paid on the 15th day or the last day of each month then their April earnings will be subject to the NICs schedule for the new fiscal year (starting on 6 April). Similar ambiguities can arise for employees with other pay periods, depending on the relationship between the survey reference date, the lengths and dates of pay periods, and the point in the pay period at which earnings are actually paid.[22]

For the large majority of observations in our dataset, the earnings we observe will be subject to the NICs schedule of the fiscal year just beginning, but this will not be the case for all observations (particularly in years when the survey reference date is near the start of April) and we cannot identify those for which it is not true.

In what follows we proceed as if the earnings we observe are subject to the NICs schedule of the fiscal year just beginning. Under that assumption, we are typically estimating very short-run responses to changes in NICs rates – the effect on earnings of reforms implemented earlier in the same month – although note that (i) two of the biggest reforms to the NICs schedule (in 1985 and 1989) were implemented in October, not April, so a significant part of our identifying variation comes from reforms implemented around six months before the earnings outcomes we observe, and (ii) changes to the NICs schedule are invariably announced at least a few months in advance

[22] Some employers align their pay periods with 'tax periods'. Tax weeks and tax months are always counted from 6 April, so in such cases (for example, pay periods running from the 6th of each month to the 5th of the next month) then the earnings recorded in our data will always be subject to the NICs schedule of the fiscal year just beginning unless the survey reference date is before 6 April (as it was in 1979, 1990 and 1995). We have no way to identify such cases, however.

(so that payroll software can be ready in time to operate it, among other reasons),[23] so we are estimating the effect on earnings of reforms *announced* some time beforehand.[24] In those cases where the earnings we observe are in fact subject to the NICs schedule of the fiscal year just ending, our estimates will capture only earnings responses in anticipation of the reforms' implementation, and the reform's implementation will instead be reflected in the subsequent year's earnings.

The timing of observed earnings relative to the announcement and implementation of reforms is thus worth bearing in mind for the interpretation of our results, and makes our estimates for the second year following a reform (using lagged changes in tax rates on earnings, as discussed in the preceding section), particularly pertinent.

7. Empirical Results

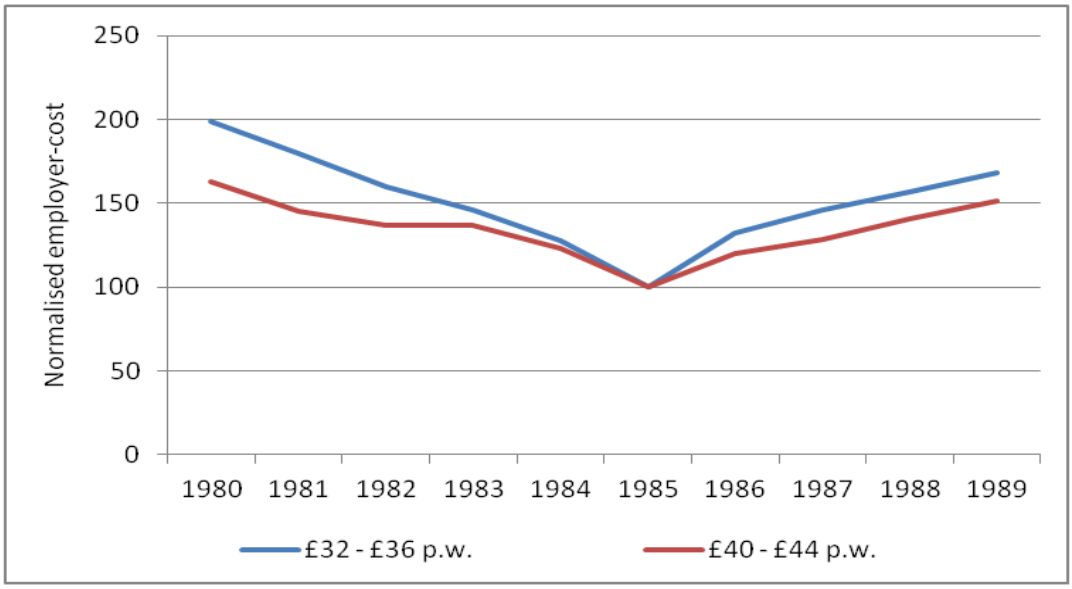
As discussed in Section 2, there were significant reforms to NICs in 1985, 1989 and between 1998 and 2000. Because the first of these affected employee and employer NICs very similarly (and average and marginal rates identically), and because the second two affected only one of employee or employer NICs, it is not possible to use a single reform to identify the separate effects of changes in average and marginal employee and employer NICs. Furthermore, as shown in Figures 1 to 3, mean reversion in earnings (and thus employer cost) is a significant issue when estimating the impact of reforms to NICs that affect those with low earnings.

Figure 2, for instance compares the employer cost of two groups of employees (normalised by adjusting employer cost in each year according to the changes in population-average employer cost and set equal to 100 for both groups 1985): those whose earnings were £32–£36 in 1985, and those whose earnings were £40–£44 in 1985. The latter range of income was subject to a reduction in the marginal and average rates of employee and employer NICs in 1986,[25] while the former was not subject to NICs in either 1985 or 1986. A naive difference-in-difference style analysis of this reform would attribute the relatively slower growth in the employer cost of the £40–£44 group following the reform to the reduction in NICs that this group likely faced, implying the (over)-shifting of NICs to employers and/or very large income effects. But employer cost trends for the two groups were different prior to the reforms as well: the normalised employer costs of those in the £40-44 group had fallen considerably less between 1980 and 1985, than those in the £32-36 group. The pattern observed is therefore likely to reflect mean reversion, with the earnings of those in the £32-36 (more) temporarily depressed than those in the £40-44 group, and therefore growing more quickly in the subsequent years. Figures 3 and 4 show similar (albeit less stark)

[23] This is different from income tax, for example, where rates are sometimes changed at short notice – and occasionally even retrospectively – because it operates annually and so incorrect tax deductions early in the fiscal year can be rectified by adjusting the amount of tax deducted later on in the fiscal year.
[24] Of course, when changes were announced (or expected) more than a year in advance, as occasionally happened, it is possible that our base period ('pre-reform') earnings might have responded in anticipation.
[25] From 9% to 5%, and from 10.XX% to 5%, respectively.

issues at the next NICs thresholds introduced in the 1985 reforms.[26] The use of multiple years of data and multiple reforms both increasing and decreasing NICs in our full panel estimates allows us to address mean reversion more satisfactorily than for a single reform. As discussed in Kleven and Schultz (2014), having reforms that both increase and decrease rates of a tax on a given part of the income/earnings distribution should make mean reversion somewhat less problematic (as its effects work in opposite directions for tax increases and decreases).

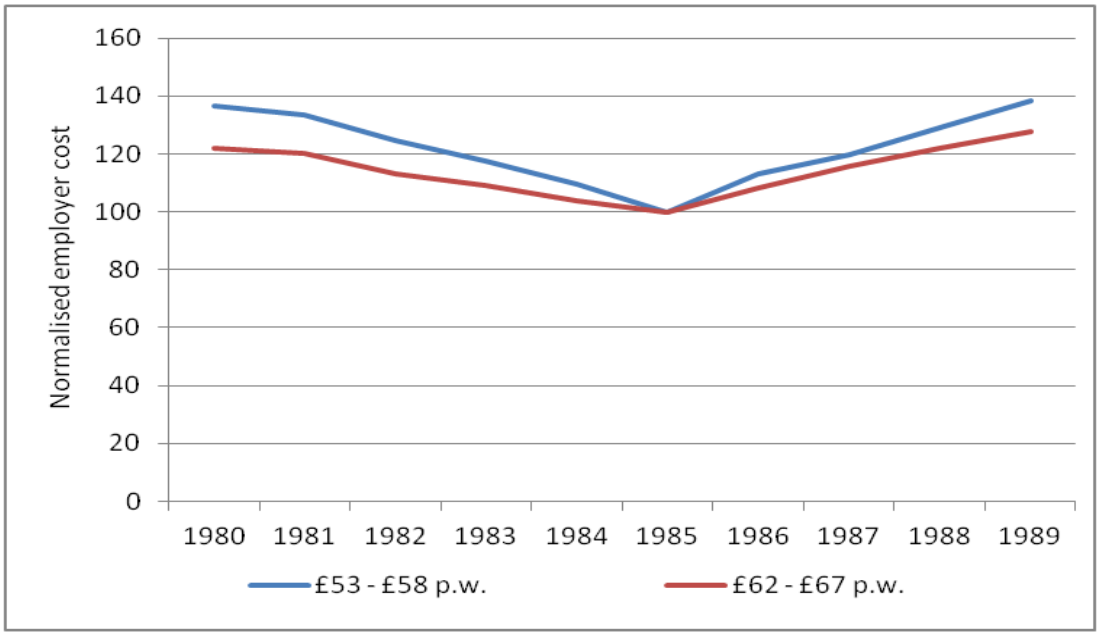
Figure 2. Normalised employer-cost between 1980 and 1989 for groups with earnings of £32–£36 and £40–£44 per week, respectively, in 1985



Source: Authors' calculations using NESPD.

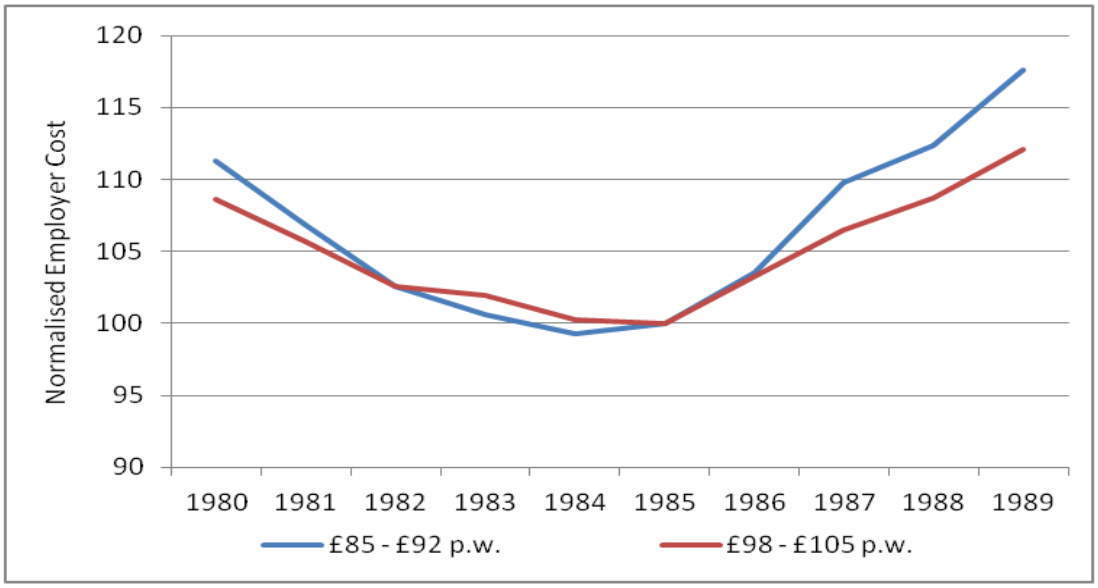
[26] Figure 2 shows differential growth in normalised employer cost for groups around the threshold where employee and employer NICs increased from 5% to 7% after the reform (rates were 9% and 10.45% respectively both sides of the threshold prior to the reform). Figure 3 relates to the threshold where employee and employer NICs increased from 7% to 9% after the reform (rates were 9% and 10.45% respectively both sides of the threshold prior to the reform)

Figure 3. Normalised employer-cost between 1980 and 1989 for groups with earnings of £53–£58 and £62–£67 per week, respectively, in 1985



Source: Authors' calculations using NESPD.

Figure 4. Normalised employer-cost between 1980 and 1989 for groups with earnings of £85–£92 and £98–£105 per week, respectively, in 1985



Source: Authors' calculations using NESPD.

We therefore turn to our panel estimates using the full set of data between 1978 and 2010. Before separately estimating the effects of employee and employers' NICs, we first estimate the responsiveness of employer cost to the overall rate of NICs levied on an individual's earnings, using changes in tax rates between 1978 and 2010. The results of these regressions are shown in Table 5. Column 1 shows estimates where our set of

controls $\boldsymbol{X_{i,t}}$ includes a cubic of $lnZ_{i,t-2}$; column 2 shows estimates based on a quintic of $lnZ_{i,t-2}$, and; column 3 shows estimates based on a 10-piece spline of $lnZ_{i,t-2}$. Estimated coefficients are broadly stable across coefficients.

Table 5. Estimated coefficient for regressions of $\boldsymbol{\Delta lnZ_{i,t}}$, overall NICs rate

	Cubic control	Quintic Control	Spline Control
$\beta_{Z,\tau}$	0.075	0.061	0.125*
	(0.0608)	*(0.0609)*	*(0.0615)*
$\beta_{Z,\rho}$	−0.796*†	−0.778*†	−0.909*
	(0.1002)	*(0.1005)*	*(0.1017)*
No. of Observations	1,933,397	1,933,397	1,933,397

Notes: * Signifies a statistically significant difference from 0. † Signifies a statistically significant difference from -1 (for coefficients on net-of-average tax rates (1-ρ) only).

Estimates of $\beta_{Z,\tau}$ are positive, albeit small and only statistically significant in the specification with controls based on splines of $lnZ_{i,t-2}$. This would imply employer cost is therefore relatively unresponsive to compensated changes in marginal NICs rates, suggesting a relatively small deadweight loss of the tax. On the other hand, estimates of $\beta_{Z,\rho}$ are negative, large, and highly statistically significant. These estimates can be interpreted in several ways. First, if one were to assume that the incidence of NICs was fully on employees, then this would imply large, negative income effects. Alternatively, if one were to assume that income effects were absent (or small), then the estimates would imply that the bulk of the incidence of NICs was on employers. Finally, results are consistent with moderate-sized income effects, and sharing of the burden of NICs between employees and employers.

Estimates based on the overall NICs rate include employer-cost responses to both employee and employer NICs. If the IIP does not hold – which is not necessarily likely in the very short term – effects of employee and employer NICs could differ significantly. Table 6 shows coefficients estimated separately for employee and employer NICs.

Estimates of $\beta^{E}_{Z,\tau}$ are statistically significantly and positive (0.2–0.3), while estimates of $\beta^{R}_{Z,\tau}$ are effectively zero. These results are very similar to those found by Lehmann et al (2013) for France, where they found positive compensated elasticities for net-of-marginal income tax rates, and zero elasticities for employers' SSCs. Estimates of $\beta^{E}_{Z,\rho}$ are not statistically significantly different from 0, which is evidence both that income effects are negligible, and that the incidence of employee NICs is on employees rather than employers. In contrast, estimates of $\beta^{R}_{Z,\rho}$ are large, negative, and highly statistically significant. The most natural interpretation of these is that incidence of employer NICs is, in effect, fully on employers in the very short term. Again, these results are consistent with those found by Lehmann et al (2013). Note that the combination of $\beta^{R}_{Z,\rho} \sim -1$ and $\beta^{R}_{Z,\tau} \sim 0$ would imply that employers do not reduce their demand for labour from the set

of individuals that remain employed. However, this does not preclude extensive margin responses by employers (changing the number of individuals employed when employer NICs change).

Table 6. Estimated coefficient for regressions of $\Delta lnZ_{i,t}$, employee and employer NICs rates

	Cubic control	Quintic Control	Spline Control
$\beta^{E}_{Z,\tau}$	0.214*	0.210*	0.294*
	(0.1041)	*(0.1045)*	*(0.1051)*
$\beta^{R}_{Z,\tau}$	0.003	-0.018	0.030
	(0.0607)	*(0.0606)*	*(0.0613)*
$\beta^{E}_{Z,\rho}$	-0.019	0.013	-0.181
	(0.1754)	*(0.1753)*	*(0.1776)*
$\beta^{R}_{Z,\rho}$	-1.244*	-1.233*	-1.315*†
	(0.1364)	*(0.1366)*	*(0.1381)*
No. of Observations	1,933,397	1,933,397	1,933,397

Notes: * Signifies a statistically significant difference from 0. † Signifies a statistically significant difference from -1 (for coefficients on net-of-average tax rates (1-ρ) only).

Of course, it is unsurprising that the effects of employee and employer NICs differ in the very short term: it may take time for wages to adjust to changes in tax rates. In Table 7, we therefore examine whether the effects of changes to NICs differ after an additional year, by including changes in net-of-NICs rates lagged one period as regressors. Overall effects after 1 further year can be obtained by adding the coefficient for the contemporaneous and lagged changes together (e.g. $\beta^{E,0}_{X,\tau}$+$\beta^{E,1}_{X,\tau}$).

Two things stand out. First, effects for contemporaneous changes are very similar to those in Table 6: adding lagged changes in NICs rates does not affect estimates. The second thing to note is that the coefficients on the lagged changes in net-of-marginal and net-of-average NICs rates are not statistically significantly different from 0. This would suggest that behavioural effects and the incidence of NICs are very similar after 12-18 months as after 0-6 months.[27] Not only does the IIP not hold after a further year, there is no evidence of any equalisation of the behavioural effects and incidence of employee and employer NICs.

Table 7 also includes estimated coefficients for regressions of $\Delta ln(H)_{i,t}$ and $\Delta ln(Z/H)_{i,t}$ on the same set of contemporaneous and lagged changes in net-of-NICs rates. Looking first at hours of work, the coefficients on contemporaneous changes in tax rates show a positive compensated hours-of-work elasticity for employee NICs of around 0.2, and a statistically insignificant (though negative) elasticity for employer NICs. The coefficients

[27] Although $\beta^{R,1}_{Z,\tau}$ is of moderate size and is close to significance in some specifications, suggesting that it may be the case that either income effects grow or more incidence is on employers after an additional year.

on the net-of-average tax rate indicate modest, although in general not statistically significant income effects.

As with employer cost, the coefficients on the lagged changes in net-of-tax rates are not statistically significantly different from 0: we cannot reject that the effect of NICs on hours of work is the same after 12-18 months as after 0-6 months. However, in each instance the insignificant effects of lagged changes in tax rates reinforce the direction of effects of the contemporaneous effects. This means income effects are statistically significant after 12-18 months, and the effect of reduced marginal employer NICs rates is to reduce hours of work after over the same time horizon. Note that this does not necessarily mean the overall response to reductions in marginal employer NICs rates is a reduction in labour input: as already mentioned, extensive margin responses are excluded from these estimates.

Hourly employer costs do not respond to compensated changes in net-of-marginal rates in either the very short term or after 12-18 months. The coefficient on the lagged marginal net-of-employee NICs rate is negative and statistically significant, but this just offsets the immediate positive (albeit not statistically significant) immediate impacts. Turning to net-of-average tax rates, the coefficient on employer NICs is close to -1 , and is not statistically significantly different after 12-18 months. This is further evidence that the splitting of the statutory incidence of NICs between employers and employees continues to matter as much after 12-18 months as after 0-6 months.

Table 8 includes the same set of regressions, albeit based on instruments calculated using incomes held fixed in period t-j (i.e. lagged by k=2 periods). The strongest findings in our preferred specification remain unaltered: the regressions suggest that the incidence of employer NICs remains on employers after 12-18 months; hours-of-work increase in response to reductions in marginal employee NICs rates, but if anything, reduce in response to reductions in marginal employer NICs rates. Some other results differ, however. Most notably, the lagged changes in employer NICs have a statistically significant effect on hourly employer cost: reductions in the marginal employee NICs rate reduce hourly employer cost, while reductions in the average employee NICs rate increase employer cost. It is difficult to interpret either of these findings in terms of standard labour supply or demand responses, or incidence shifting.[28]

Overall, our results are remarkably similar to those of Lehmann et al (2013) who examine changes in employer NICs and income tax that affect a similar set of low-middle earners in France. Our estimates suggest that if the incidence and behavioural effects of employee and employer NICs do begin to equalise – so that, for instance, eventually the IIP does hold –, this is a process which does not start for at least 12-18 months after a reform.

[28] Coefficients on net-of-marginal rates should be 0 or positive if substitution effects operate. Coefficients on net-of-average tax rates should be 0 or negative if there are income effects and/or incidence of NICs is partly on the employer.

Table 7. Estimated coefficient for regressions of $\Delta lnZ_{i,t}$, $\Delta lnH_{i,t}$ and $\Delta ln(Z/H)_{i,t}$, separate employee and employer NICs rates, and including changes in NICs rates lagged one period

	$\Delta lnZ_{i,t}$			$\Delta lnH_{i,t}$			$\Delta ln(Z/H)_{i,t}$		
	Cubic	Quintic	Spline	Cubic	Quintic	Spline	Cubic	Quintic	Spline
$\beta_{X,\tau}^{E,0}$	0.270*	0.278*	0.299*	0.201*	0.209*	0.227*	0.140	0.142	0.136
	(0.1399)	*(0.1399)*	*(0.1380)*	*(0.0995)*	*(0.0995)*	*(0.0982)*	*(0.0999)*	*(0.0999)*	*(0.0987)*
$\beta_{X,\tau}^{E,1}$ (Lag)	-0.060	-0.058	-0.022	0.096	0.101	0.105	-0.193*	-0.186*	-0.172*
	(0.1056)	*(0.1048)*	*(0.1065)*	*(0.0762)*	*(0.0756)*	*(0.0768)*	*(0.0735)*	*(0.0730)*	*(0.0742)*
$\beta_{X,\tau}^{R,0}$	-0.006	-0.015	-0.009	-0.083	-0.092	-0.073	0.020	0.020	0.004
	(0.0753)	*(0.0755)*	*(0.0739)*	*(0.0546)*	*(0.0548)*	*(0.0535)*	*(0.0484)*	*(0.0486)*	*(0.0475)*
$\beta_{X,\tau}^{R,1}$ (Lag)	0.025	0.008	0.033	-0.118	-0.107	-0.156	0.028	0.027	0.015
	(0.0791)	*(0.0787)*	*(0.0807)*	*(0.0603)*	*(0.0599)*	*(0.0616)*	*(0.0531)*	*(0.0528)*	*(0.0541)*
$\beta_{X,\rho}^{E,0}$	0.010	-0.003	-0.04	-0.185	-0.197	-0.235	0.201	0.194	0.219
	(0.2367)	*(0.2369)*	*(0.2324)*	*(0.1645)*	*(0.1647)*	*(0.1609)*	*(0.1598)*	*(0.1601)*	*(0.1570)*
$\beta_{X,\rho}^{E,1}$ (Lag)	-0.132	-0.120	-0.168	-0.118	-0.107	-0.156	0.055	0.053	0.074
	(0.1642)	*(0.1636)*	*(0.1675)*	*(0.1130)*	*(0.1124)*	*(0.1157)*	*(0.1142)*	*(0.1138)*	*(0.1165)*
$\beta_{X,\rho}^{R,0}$	-1.304*†	-1.304*†	-1.334*†	-0.173	-0.176	-0.219*	-0.986*	-1.000*	-0.951*
	(0.1408)	*(0.1409)*	(0.1465)	*(0.0922)*	*(0.0920)*	*(0.0953)*	*(0.0981)*	*(0.0981)*	*(0.1009)*
$\beta_{X,\rho}^{R,1}$ (Lag)	-0.222	-0.212	-0.275	-0.139	-0.135	-0.175	0.098	0.087	0.101
	(0.1546)	*(0.1536)*	*(0.1598)*	*(0.1088)*	*(0.1077)*	*(0.1129)*	*(0.1080)*	*(0.1074)*	*(0.1114)*
No. of Observations		1,777,732			1,697,556			1,697,556	

Notes: * Signifies a statistically significant difference from 0. † Signifies a statistically significant difference from -1 (for coefficients on contemporaneous net-of-average tax rates (1-ρ) only).

Table 8. Robustness check: estimated coefficient for regressions of $\Delta lnZ_{i,t}$, $\Delta lnH_{i,t}$ and $\Delta ln(Z/H)_{i,t}$, separate employee and employer NICs rates, and including changes in NICs rates lagged one period, using instruments based on period t-3 earnings

	$\Delta lnZ_{i,t}$			$\Delta lnH_{i,t}$			$\Delta ln(Z/H)_{i,t}$		
	Cubic	Quintic	Spline	Cubic	Quintic	Spline	Cubic	Quintic	Spline
$\beta_{X,\tau}^{E,0}$	0.096	0.096	0.117	0.187	0.194	0.195	-0.08	-0.076	-0.071
	(0.1529)	*(0.1535)*	*(0.1535)*	*(0.1072)*	*(0.1075)*	*(0.1075)*	*(0.1117)*	*(0.1121)*	*(0.1121)*
$\beta_{X,\tau}^{E,1}$ (Lag)	-0.285	-0.292	-0.253	0.202	0.203	0.210	-0.420*	-0.413*	-0.406*
	(0.1503)	*(0.1492)*	*(0.1537)*	*(0.1105)*	*(0.1099)*	*(0.1131)*	*(0.1072)*	*(0.1067)*	*(0.1097)*
$\beta_{X,\tau}^{R,0}$	-0.025	-0.036	-0.023	-0.147*	-0.161*	-0.133*	0.030	0.033	0.009
	(0.0693)	*(0.0693)*	*(0.0697)*	*(0.0516)*	*(0.0516)*	*(0.0520)*	*(0.0458)*	*(0.0458)*	*(0.0460)*
$\beta_{X,\tau}^{R,1}$ (Lag)	0.048	0.037	0.055	-0.006	-0.022	0.009	-0.023	-0.021	-0.042
	(0.0857)	*(0.0852)*	*(0.0881)*	*(0.0649)*	*(0.0644)*	*(0.0667)*	*(0.0575)*	*(0.0571)*	*(0.0591)*
$\beta_{X,\rho}^{E,0}$	-0.127	-0.128	-0.135	-0.202	-0.210	-0.216	0.138	0.133	0.152
	(0.2480)	*(0.2484)*	*(0.2463)*	*(0.1650)*	*(0.1653)*	*(0.1635)*	*(0.1720)*	*(0.1724)*	*(0.1708)*
$\beta_{X,\rho}^{E,1}$ (Lag)	0.252	0.262	0.238	-0.073	-0.063	-0.098	0.381*	0.376*	0.407*
	(0.2089)	*(0.2080)*	*(0.2141)*	*(0.1428)*	*(0.1420)*	*(0.1471)*	*(0.1470)*	*(0.1464)*	*(0.1508)*
$\beta_{X,\rho}^{R,0}$	-1.036*	-1.024*	-1.043*	-0.071	-0.064	-0.110	-0.843*	-0.853*	-0.793*
	(0.1632)	*(0.1632)*	*(0.1677)*	*(0.1080)*	*(0.1078)*	*(0.1117)*	*(0.1116)*	*(0.1115)*	*(0.1147)*
$\beta_{X,\rho}^{R,1}$ (Lag)	-0.387*	-0.369*	-0.417*	-0.285*	-0.275*	-0.328*	0.061	0.049	0.097
	(0.1749)	*(0.1739)*	*(0.1819)*	*(0.1225)*	*(0.1215)*	*(0.1282)*	*(0.1229)*	*(0.1222)*	*(0.1275)*
No. of Observations		1,615,596			1,543,590			1,543,590	

Notes: * Signifies a statistically significant difference from 0. † Signifies a statistically significant difference from -1 (for coefficients on contemporaneous net-of-average tax rates (1-ρ) only).

8. Conclusions

In this paper we estimate the responses of employment cost, hours and employment cost per hour to marginal and average rates of employee and employer NICs, using reforms during the 1980s, 1990s and 2000s as our source of identifying variation. Previous evidence on the responsiveness of earnings to social security contributions is sparse, in contrast to the voluminous literature on the responses of taxable income to income tax; this paper helps to fill that gap. Furthermore, by considering responses both in the context of behavioural responses and the (intimately related) incidence of a tax, this paper attempts to help link the New Tax Responsiveness and incidence literatures. We investigate the very short term effects using contemporaneous changes in NICs and the slightly longer-term effects using the change in NICs rates during the previous year. This is an improvement on the typical approach of using longer panel periods (e.g. 2 or 3 years rather than 1) in an attempt to examine such effects.

Our estimates show that responses to employee and employer NICs differ significantly both in the immediate period following implementation a reform (0-6 months) and the slightly longer term (12-18 months). We find positive statistically significant effects on employer cost, operating via hours of work, of reductions in marginal rates of employee NICs. In contrast we find zero or negative effects of marginal employer NICs rates on the employer cost and hours of work among those working, although this does not preclude the more usual positive effect operating via the extensive margin. We also find that employer cost falls approximately one-for-one when average employer NICs rates are reduced, but not when employee NICs rates are reduced, with most of the effect (and the discrepancy) operating via hourly employer cost. These differences are robust across specifications based on different instruments and different sets of controls for divergent earnings and hours trends.

Our interpretation of these findings is similar to Lehmann et al (2013), who find similar results for employer cost for France. That is that there is wage stickiness that does not begin to abate for at least 12–18 months, so the economic incidence of NICs reflects its legal incidence during at least this period; that low-middle earners respond modestly (on the intensive margin) to employee NICs changes that directly affect them; and that if firms respond to higher labour costs at the margin, they do so via the extensive rather than the intensive margin. Consideration of the effects of average tax rates on hours of work also suggests that income effects may be significant for the largely low-middle earners for whom our elasticities are estimated (most of the big reforms to NICs during this period affected low-middle earners, the uncapping of the UEL being the notable exception).

This work could be extended in several ways. First, we could disaggregate results according to the sex, age, earnings, sector, occupation, etc. of an employee. This would allow us to investigate the heterogeneous effects of NICs: many studies find that women are more responsive to taxation than men, for instance. Second, we could extend the

number of lagged changes in tax rates included to examine the effect of changes in NICs on employer cost and hours after longer time periods (for instance 24 – 30 months with one additional lag, or 36 – 42 months with two). This would allow us to test whether, for instance, the effects of employee and employer NICs begin to equalise, and if so, over what time horizon. In order to do this, we would probably need to use instruments based on earnings from longer before the period in question (although this could weaken their statistical power). Third, we could extend our analysis to consider the effects of income tax and of means-tested benefits and tax credits. The NESPD, on which this study is based, does not allow us to estimate the marginal and average tax rates associated with these parts of the tax and benefit system: they are assessed on broader, usually family-level, measures of income, rather than individual earnings. To do this, we would therefore need to make use in some way of alternative datasets (standard household surveys such as the Family Resources Survey, the Family Expenditure Survey and the British Household Panel Survey) which can provide information on these aspects but without all the other advantages of the NESPD.

References

Adam, S., J. Browne, D. Phillips and B. Roantree (2016), "Adjustment costs and labour supply: evidence from bunching at tax thresholds in the UK", Mimeo

Adam, S. and G. Loutzenhiser (2007), 'Integrating Income Tax and National Insurance: An Interim Report', IFS Working Paper no. W07/21, http://www.ifs.org.uk/publications/4101.

Anderson, P. M., and B. D. Meyer (1997), 'The effects of firm specific taxes and government mandate with an application to the US unemployment insurance program', Journal of Public Economics, Vol. 75, No. 2, pp 119 – 145

Anderson, P. M. and B. D. Meyer (2000), 'The effects of the unemployment insurance payroll tax on wages, employment, claims and denials', Journal of Public Economics, Vol. 78, pp. 119 – 145

Atkinson, A.B., Micklewright, J. and Stern, N.H. (1981), 'A comparison of the FES and NES 1971–1977: Part I. Characteristics of the sample', Social Science Research Council Programme on Taxation, Incentives and the Distribution of Income Working Paper no. 27.

Atkinson, A.B., Micklewright, J. and Stern, N.H. (1982), 'A comparison of the FES and NES 1971–1977: Part II. Hours and earnings', Social Science Research Council Programme on Taxation, Incentives and the Distribution of Income Working Paper no. 32.

Auten and Caroll (1999), 'The effect of income tax on household income', The Review of Economics and Statistics, Vol. 81, No. 4, pp 681 – 693

Bingley, P. and G. Lanot (2002), 'The Incidence of Income Tax on Wages and Labor Supply', Journal of Public Economics, Vol. 83, pp. 173 – 194

Blundell, R. and B. Etheridge (2010), 'Consumption, income and earnings inequality in Britain', Review of Economic Dynamics, Vol. 13, No. 1, p 76 – 102

Blundell, R. and T. Macurdy (1999), 'Labor supply: A review of alternative approaches', Handbook of Labor Economics Vol. 3, Part A, O. C. Ashenfelter and D. Card (eds)

Bunel, M. and Y. L'Horty (2012), 'The effects of reduced social security contributions on employment: an evaluation of the 2003 French reforms', Fiscal Studies, Vol. 33, No. 3, pp. 371 – 398

Chetty, R., A. Looney, and K. Kroft (2009), 'Salience and taxation: theory and evidence', The American Economic Review, Vol. 99, No. 4, pp. 1145 – 1177

Devereux, P. J and Hart, R. A. (2010), 'Forced to be rich? Returns to compulsory schooling in Britain', The Economic Journal, 120 (December), 1345–1364

European Commission (2013), "Study on the Impacts of Fiscal Devaluation", Working Paper No. 36–2013

European Commission (2015), "Study on the effects and incidence of labour taxation", Working Paper No. 56–2015

Feldstein, M. (1995), 'The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act', The Journal of Political Economy, Vol. 103, No. 3, pp. 551 – 572

Feldstein, M. (1999), 'Tax Avoidance and the Deadweight Loss of the Income Tax', The Review of Economics and Statistics, Vol. 81, No. 4, pp. 674 – 680

Fullerton, D. and G. E. Metcalf (2002), 'Tax incidence', Chapter 26, Handbook of Public Economics: Volume IV, A. Auerbach and M. Feldstein (eds)

Gruber, J. and E. Saez (2002), 'The elasticity of taxable income: evidence and implications', Journal of Public Economics, Vol. 84, pp 1 – 32

Hamermesh, D. (1996), *Labor Demand*.

Hood, A. and L. Oakley (2014a), 'A survey of the GB benefit system', IFS Briefing Note BN13, http://www.ifs.org.uk/publications/1718.

Hood, A. and L. Oakley (2014b), 'The social security system: long-term trends and recent changes', IFS Briefing Note BN156, http://www.ifs.org.uk/publications/7438.

Kleven, H., and E. Schultz (2014), 'Estimating taxable income responses using Danish Tax reforms', American Economic Journal: Economic Policy, Vol. 6, No. 4, pp 271–301

Kopczuk, W. (2005), 'Tax bases, tax rates and the elasticity of reported income', Journal of Public Economics, Vol. 89, No. 11-12, pp. 2093–2119

Kubik, J. D. (2004), 'The incidence of personal income taxation: evidence from the tax reform act of 1986', Journal of Public Economics, Vol. 88, pp. 1567 – 1588

Lehmann, E., F. Marical, and L. Rioux (2013), 'Labor income responds differently to income-tax and payroll-tax reforms', Journal of Public Economics, Vol. 99, pp. 66 – 84

Leigh, A. (2010), 'Who Benefits from the Earned Income Tax Credit? Incidence among Recipients, Co-Workers and Firms', The B. E. Journal of Economic Analysis and Policy, Vol. 10, No. 1, Article 45

Lhommeau, B. and V. Remy (2009), 'Les politiques d'allgements ont-elles un effet sur la mobilite salariale des travailleurs a bas salaires', Economie et Statistique, Vol, 429 – 430, pp 21 – 49

Melguizo, A. and J. M. Gonzalez-Paramo (2013), 'Who bears labour taxes and social contributions? A meta-analysis approach', SERIEs, Vol. 4, No. 3, pp 274 – 271

Mirrlees, J. et al (2011), "The Mirrlees Review: Tax by Design", Oxford University Press

Office of Tax Simplification (2015), 'Alignment of Income Tax and National Insurance: Terms of Reference', https://www.gov.uk/government/publications/ots-review-of-income-tax-and-national-insurance-tor

Pissarides, C. (1998), 'The impact of employment tax cuts on unemployment and wages; The role of unemployment benefits and tax structure', European Economic Review, Vol. 42, pp. 155 – 183

Saez, E., J. Slemrod, and S. H. Giertz (2012), 'The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review', Journal of Economic Literature, Vol. 50, No. 1, pp 1–50

Summers, L. H. (1989), 'Some Simple Economics of Mandated Benefits', The American Economic Review, Vol. 79 No. 2, pp 177–83

Weber, C. (2014), 'Towards obtaining a consistent estimate of the elasticity of taxable income using differences-in-differences', Journal of Public Economics, Vol 117, pp. 90 – 103

Appendix A. Data

As discussed in the main body of the text, the NESPD is supposed to be a random 1% sample of the population. Non-response reduces the same size by approximately 30% on average (to 0.7% of the population). In addition, there are a number of more minor reasons that our sample may not be completely random:

- As a cost-saving measure, the sample size was reduced by 20% in 2007 and 2008 before being restored in 2009. This reduction was not random but restricted to those employers – typically smaller ones – who completed the survey on paper rather than electronically (in order to maximise the saving in compliance effort for respondents) and concentrated in industries where earnings were less variable (in order to minimise the resulting increase in the coefficient of variation of earnings).[29] However, since the reduction applied in only two of the 35 years of our data, and those were not years of significant reform generating the variation we exploit, we do not believe this should significantly affect our overall estimation.
- There is potential for under-sampling of the employees with the lowest earnings. Specifically, those earning below the LEL (the earnings level at which NICs became payable, until the late 1990s – since then the NICs exemption level has been higher than the LEL). This is because employees are identified for inclusion in the NESPD using data from PAYE (Pay As You Earn), the UK's system for deduction of income tax at source by employers, and employers are only required to include those earning above the LEL in their PAYE scheme. However, this does not appear to be a significant problem in practice: employers seem to register all their employees on their PAYE scheme, even those they are not obliged to include. When we compare the distribution of earnings above and below the LEL in the NESPD data with those in other datasets that do not suffer from the same potential selection issues, we find that the two densities look broadly similar.[30]
- Before 2004, the NESPD sample was identified exclusively from PAYE records taken between January and March, and so excluded people starting a new job (whether previously working elsewhere or not working) between then and the survey reference date in April. Since 2004 this problem has been eliminated by taking a second, supplementary extract of PAYE records in April to pick up any employees missing from the initial sample. But for years before 2004 our data exclude employees starting a new job in the few weeks before the reference date in April.
- Since 2005, employees have been removed from the dataset if their earnings were below £10,000 per year (£11,000 since 2009) and either (a) their job title was 'Director', (b) they had the same first initial and surname as the employer completing the survey, (c) they 'fail the automated National Minimum Wage

[29] See Cotterell (2007) for details.

[30] We compared the NESPD with the Family Resources Survey and found no difference. Devereux and Hart (2010) compared the NESPD with the (smaller and less reliable but longer-running) General Household Survey and did find a somewhat lower proportion of observations were below the LEL in the NESPD than in the GHS (18% vs. 27% among women). Devereux and Hart (2010) also report that 'Atkinson et al. (1981, 1982) have compared the NESPD to a household survey, the Family Expenditure Survey, and found that the two surveys were fairly consistent in their hours and earnings patterns.' Looking within the NESPD, the Office for National Statistics report that there was little change in the observed earnings distribution in 2014 when the PAYE sampling frame moved to 'Real-Time Information' and larger employers were required to include all of their employees, not just those above the LEL, though any under-sampling was likely to be concentrated in smaller firms anyway.

check' or (d) their earnings were an outlier for their occupation.[31] This is an attempt to identify and remove company owner-managers who are manipulating their earnings – for example, taking dividends instead to reduce their tax liability – and are therefore perceived to be producing a distorted picture of the earnings distribution (though in practice these criteria may remove some others as well). However, for our purposes, such income shifting may be one of the kinds of response to taxation we might like to capture, and this procedure means that from 2005 onwards our estimation excludes these responses and this small but potentially highly responsive group.

It is also worth noting that there are two ways in which the tax rates we calculate are approximations of the true tax rates employees and employers face:

- Throughout the period we analyse, individuals contributing to a private pension could choose whether to 'contract in' or 'contract out' of the second pillar of the UK state pension system (initially the State Earnings-Related Pension Scheme, SERPS, and later the State Second Pension, S2P). Those 'contracting out' were charged slightly lower rates of employee and employer NICS in exchange for sacrificing future entitlement to SERPS/S2P.32 Our data do not reveal whether individuals were contracted in or out; we assume they are all contracted in. Note that the reforms we use for the purposes of identification applied to both the 'contracted in' and 'contracted out' schedules of NICs.

Perhaps more problematically, a special 'married women's reduced rate' of employee NICs – 2% at the start of our period, rising to 5.85% by the end – was (and remains) available, in exchange for reduced benefit entitlements, to married women who have been claiming it almost continuously since May 1977. This applied to a substantial number of women in the 1980s, but almost none by the end of the period. Since we cannot identify married women in the NESPD, let alone those choosing this option, we ignore it.

[31] Source: Authors' correspondence with the Office for National Statistics. It is not clear exactly what the 'automated National Minimum Wage check' entails, since we do observe people in our data receiving less than the national minimum wage. Nor is it clear what constitutes an 'outlier' for these purposes.

[32] For those contributing to a salary-related private pension, contracted-out rates of employee NICs were 2.5 percentage points lower and employer NICs 4.5 percentage points lower than contracted-in rates in 1978–79, falling to 1.4 percentage points and 3.4 percentage points respectively by 2012–13. The contracted-out rebate for those contributing to a defined-contribution pension varied by age.