

IZA DP No. 8897

## Modelling the Impact of Direct and Indirect Taxes Using Complementary Datasets

Michael Savage  
Tim Callan

February 2015

# **Modelling the Impact of Direct and Indirect Taxes Using Complementary Datasets**

**Michael Savage**

*ESRI, University College Dublin  
and Trinity College Dublin*

**Tim Callan**

*ESRI, University College Dublin  
and IZA*

**Discussion Paper No. 8897  
February 2015**

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **Modelling the Impact of Direct and Indirect Taxes Using Complementary Datasets**

Comprehensive modelling of the impact of taxes and tax policy options requires data on the impact at micro-level of both direct and indirect taxes. There are, however, limits on the amount of data that can be gathered by any one survey. With some exceptions, most notably the Living Costs and Food Survey (LCF) in the UK, most national expenditure surveys are not suitable for use in detailed modelling of the direct tax and welfare system. This makes approaches which impute expenditure data into detailed income surveys of considerable interest. In this paper, we assess the sensitivity of the distributional effects of indirect taxes to the choice between actual, estimated and imputed expenditure data. By doing so, the analysis here serves as an updated picture of the distributional effects of the indirect tax system in Ireland, as well as a base for future microsimulation analysis of simultaneous direct tax, indirect tax and welfare reform.

JEL Classification: D30, H22, H23, H24

Keywords: indirect tax, imputation, distribution, microsimulation

Corresponding author:

Michael Savage  
Economic and Social Research Institute  
Whitaker Square  
Sir John Rogerson's Quay  
Dublin 2  
Ireland  
E-mail: [michael.savage@esri.ie](mailto:michael.savage@esri.ie)

## 1. Introduction

Both direct and indirect taxes are important components of overall tax revenue in Ireland, as in most OECD countries. The balance between direct and indirect taxes changes over time, and varies across countries. The OECD reports that in 2012, for example, Value Added Taxes (VAT) and excise duties together represented between 18 per cent and 41 per cent of total tax revenue among its EU member states. Ireland was at the higher end of the scale, where VAT and excises represented 32 per cent of tax revenue<sup>1</sup>. While much attention has been given to examining the distributional effect of direct taxes, somewhat less is known about the distributional effect of indirect taxes.

Comprehensive modelling of the impact of taxes and of tax policy options requires data on the impact at micro-level of both direct and indirect taxes. From a research point of view, a single dataset combining detailed data on incomes, expenditures, labour market participation, and a range of socio-demographic variables would be ideal. There are, however, limits on the amount of data that can be gathered by any one survey. Recording expenditure data imposes a heavy burden on respondents: they are often required to keep expenditure diaries covering all purchases over a reasonably long period of time<sup>2</sup>. This limits the extent to which other data of interest can be gathered. Consequently, as noted in Sutherland et al. (2002), most countries do not have a single source of micro-data including high-quality disaggregated information on both incomes and expenditures.

With some exceptions, most notably the Living Costs and Food Survey (LCF) in the UK, most national expenditure surveys are not suitable for use in detailed modelling of the direct tax and welfare system. The main limitation of expenditure surveys in this context is the unit of observation in the data. In most cases, the unit of observation is the household, with some information on individuals within that household. For tax and welfare purposes, however, there is an intermediate unit - often called the family unit or tax unit - which is very relevant for policy purposes. Most tax and welfare policies do not operate at the level of the household, though household income and household welfare are of key concern to policy. Instead, tax and welfare policies tend to operate at either individual level (for example, contributions to social insurance) or at a family unit level. Detailed information on family and household composition is needed to ensure that it is possible to group individuals into family units, defined as an individual, together with his or her spouse, and dependent children. While the definition of a dependent child can vary across different taxes and benefits, the term "tax unit" often represents a married couple, cohabiting couple, or single person, together with all children aged under 15, and children aged less than 18 who are in full time education (Keane et al. 2014). Household expenditure surveys, by their nature, contain information at household level, and therefore omit the necessary detail to group people into family or tax-units, as well as some tax-relevant information on individuals within a household.

---

<sup>1</sup> Taxes on income and profits of individuals also represented 32 per cent of total tax revenue in Ireland in 2012. See <http://stats.oecd.org/> for details.

<sup>2</sup> In the Irish expenditure survey (see Section 3 for details), for example, respondents are required to keep an expenditure diary for a two week period.

In addition, the income data contained in expenditure surveys can often be assigned secondary importance, and as a result be of lower quality, compared to the expenditure data. Sutherland et al. (2002) argue that the income data in most Household Budget Surveys is usually very limited and of relatively low quality. When calculating taxes paid as a proportion of income, which is the most common means of assessing progressivity in a tax system, it is important to use the highest quality income data available. In the Irish case, income data collected in the Irish Household Budget Survey (HBS) covers a range of sources and types of income<sup>3</sup>, with even more detail gathered in the Survey on Income and Living Conditions (SILC). The Central Statistics Office (CSO, 2012) state that “although income data from the two sources are not directly comparable, calculation of key income distribution and poverty indicators using income data from both surveys resulted in similar values.” However, the CSO go on to say that “The SILC is recognised as the primary source of data on income in Ireland”, and elsewhere state that HBS income data “is used primarily for categorical purposes (e.g. for analysis of households according to different levels of income) rather than the provision of information on income levels”.

To be able to accurately simulate the distributional effects of the direct and indirect tax systems, as well as the social welfare system, detailed income information is needed for each individual in the household, as well as household expenditure information. In the absence of a single source of high quality income and expenditure data, such as the LCF in the UK, many researchers have used an imputation approach to add an estimate of household expenditure to the main income survey in their country studies. The expenditure imputation procedure allows detailed modelling of simultaneous reforms of direct and indirect taxes, as well as social welfare reforms.

Put simply, the imputation process involves two stages. First, expenditure on a range of goods and services is estimated in the expenditure survey based on a set of characteristics observable in both the expenditure survey and the income survey. Second, the coefficients from the estimation stage are used to impute expenditure into the income survey. A number of authors, including Decoster et al. (2010), O’Donoghue et al. (2004) and Pestel and Sommer (2013), have used this approach to measure the distributional effects of indirect taxes across a range of European countries. The procedure has been shown to produce plausible distributions of estimated expenditure in income surveys. However, it remains unclear how sensitive distributional analysis results are to the choice between the expenditure data produced at each stage of the imputation procedure. Ultimately, the question we wish to address is whether we can combine data from one survey with another via a predictive approach on common variables, in order to broaden our understanding of the distributional effects of the direct tax, indirect tax and welfare systems in Ireland. To do so, the first step is to assess the sensitivity of distributional effects of indirect taxes to the choice between actual, estimated and imputed expenditure data.

---

<sup>3</sup> The HBS contains information on Employees-wages/salaries, Self-employed income, Retirement pensions, Investment income, Property income, as well as direct taxes paid and state benefits received.

The contribution of this analysis is twofold. To assess the degree of sensitivity of the distributional results, we examine the distribution of indirect taxes in Ireland, as well as the distributional effects of a number of indirect tax reforms, using three different types of expenditure data. First we assess the distributional effects of indirect taxes using actual expenditure data, as recorded in the HBS. Second we base the distributional analysis on the estimated expenditure data in the HBS. This estimated expenditure data is the result of the first stage of the imputation process described above. Finally, having imputed expenditure data from the expenditure survey into the income survey, we use the imputed expenditure data as the basis for the distributional analysis. This approach allows us to identify the source of any sensitivity in the distributional results. The second contribution is more straightforward. Most recent analyses of the distribution of indirect taxes in Ireland are based on actual data in the HBS. They therefore do not allow for the possibility of simulation of reforms to the direct tax and welfare system, in combination with indirect tax reforms. The analysis here serves as an updated picture of the distributional effects of the indirect tax system in Ireland, as well as a base for future microsimulation analysis of simultaneous direct and indirect tax, and welfare reform.

The remainder of the paper proceeds as follows: in Section 2, we review a number of recent contributions to the distributional literature that use either the actual expenditure data approach, or the imputed expenditure data approach. Section 3 discusses the two main data sources used in the analysis, while Section 4 describes the estimation and imputation procedure used. Section 5 presents and discusses the sensitivity of the distributional analysis using the various forms of data. Finally, Section 6 concludes.

## 2. Previous Research on the Distributional Impact of Indirect Taxes

### 2.1 Modelling Indirect Taxes using Expenditure Data

Household expenditure data can be used to directly model the impact of changes in indirect taxation. Based on observed household expenditure, indirect tax contributions can be estimated and the distributive impact of changes to indirect tax rates can be modelled.

The most recent analysis of the distribution of indirect taxes in Ireland comes from Collins (2014), who used the 2009/10 wave of the Irish Household Budget Survey (HBS) to estimate the direct and indirect tax contributions of households in Ireland. To estimate VAT payments made by each household, Collins established a VAT classification for each of the expenditure items in the HBS. Excise and other indirect tax payments were estimated by assuming a representative price on relevant goods and the indirect tax payments on these goods were estimated accordingly<sup>4</sup>. The distributional effects of recent VAT reforms were also examined, assuming that demand does not

---

<sup>4</sup> Excise Duties and Carbon Taxes are applied to a certain quantum of a product (e.g., per litre) rather than as a percentage of the price..

change in response to the tax changes. Direct tax contributions were also included in the analysis using the income data contained in the HBS, although the authors recognise that the HBS data is secondary to the income data from the Survey on Income and Living Conditions. Indirect taxes, when expressed as a proportion of income, were found to be regressive.

With the 2004/05 wave of the Irish HBS, Leahy et al. (2011) also estimated the distributional effects of Value Added Tax in Ireland. VAT payments were measured by equivalised income decile, by households of different composition and by household of different sizes to assess the amount of VAT that various households pay as a proportion of weekly disposable income. The authors also estimated the impact of changing the VAT rate on certain groups of items and the associated change in revenue. With the observed household expenditures and the appropriate tax rates, the authors simulated the distributional impacts of various VAT rate changes. Based on the 2009 and 2010 VAT systems in Ireland, they found that households in the lowest equivalised income decile paid about 16 per cent of their disposable income in VAT while the richest households paid only about 6 per cent. Examining the impact of uniform VAT rates of 7 per cent, 13.5 per cent and 21 per cent, the authors found that in each scenario, the poorest households would pay the highest proportion of disposable income in VAT. Previously, Barrett and Wall (2006) used the 1999/00 wave of the HBS data to analyse Irish indirect taxes, and also found a high degree of regressivity inherent in the system.

Using a similar approach, Callan et al. (2009) estimated the distributional impact of the introduction of a €20/tCO<sub>2</sub> carbon tax<sup>5</sup>. As in Leahy et al., the authors used the Household Budget Survey 2004/05. They made use of the micro-data on energy used for home heating, electricity used, and expenditures on motor fuels. While the impact of a carbon tax of €20/tCO<sub>2</sub> was found to be regressive, the impact was small compared to pre-existing taxes and benefits. SWITCH, the ESRI tax-benefit model<sup>6</sup>, was used to show that one can compensate for the impact of a carbon tax with a relatively small increase in benefits, and a relatively small decrease in income taxes. The analysis undertaken by Callan et al. is something of a hybrid of the actual data and imputation approaches. The distributional impact of the direct tax and benefit reforms was examined using SILC data, while the impact of the Carbon Tax was examined using the HBS. To combine results, the analysis used a decile-matching approach, where the tax reform impacts were analysed using different datasets, and the results were aggregated by decile before being compared. The authors concluded that in a country like Ireland, distributional concerns surrounding the introduction of a carbon tax could be addressed.

In the absence of a single source of high quality income and expenditure data, the use of expenditure surveys can provide useful evidence on the distribution effects of indirect tax systems.

---

<sup>5</sup> A carbon tax (on carbon dioxide emissions from fossil fuel combustion) is a tax on energy use. Patterns of expenditure on such products mean that a carbon tax, of itself, would be regressive i.e. a carbon tax would be a higher proportion of income for those with low incomes.

<sup>6</sup> SWITCH is based on SILC data, which does not contain expenditure information.

However, most countries, including Ireland, do not have a single source of micro-data including high-quality disaggregated information on both incomes and expenditures. The income data in most Household Budget Surveys is usually very limited and of relatively low quality. The UK Living Costs and Food Survey (LCF)<sup>7</sup> is an exception<sup>8</sup>. The LCF is somewhat unusual in that it contains household expenditure information, as well as detailed income and household composition information which can form the basis of a tax-benefit microsimulation model. TAXBEN, the tax and benefit microsimulation model used in the Institute for Fiscal Studies (IFS), is based on the LCF. TAXBEN can therefore model the distributional impacts of direct taxes and social welfare, as well as indirect taxes based on actual expenditure data. Using the same survey, the Office for National Statistics (ONS, 2011) examined the distributional effects of taxes and benefits, including indirect taxes, on household income.

As shown by the literature reviewed in this section, household expenditure data can be used to estimate indirect tax payments made by households, as well as the distributional impacts of indirect tax reform. Using income data contained within household expenditure surveys, one can simultaneously estimate direct and indirect tax contributions, as per Collins (2014). With the exception of the UK, however, the use of expenditure surveys to assess distributional impacts is problematic in two ways. Firstly, the income data may be of a lower quality than the data contained in an income survey such as SILC, as pointed out by the CSO. Secondly, and possibly more importantly, regardless of the quality of the income data, budget surveys often do not contain enough information to accurately model the direct tax and benefit system, as is the case with the Irish Household Budget Survey. In order to do so, an income survey, such as the Survey of Income and Living Conditions (SILC), is required. However, such surveys often do not contain the expenditure data required to model the indirect tax system. This makes approaches which impute expenditure data into detailed income surveys of considerable interest. We turn next to some recent examples of such methods.

## 2.2 Modelling Indirect Taxes using Income data with Imputed Expenditure

The imputation of expenditure data to an income survey has become a commonly used approach to measure the distributive impact of indirect taxes<sup>9</sup>. A number of methodologies have been developed to combine expenditure information with income surveys. Pestel and Sommer (2013) categorise the methods into implicit and explicit techniques. Implicit methods, they state, seek to infer expenditures based on observations that are as similar as possible to the target observation. The main advantage of using implicit methods is that theoretical assumptions regarding the relationship

---

<sup>7</sup> Previously known as the Family Expenditure Survey (FES) and the Expenditure and Food Survey (EFS)

<sup>8</sup> Another exception is with Australia, where Stinmod-Statax is a combination of two microsimulation models that are based on a dataset containing both income and expenditure information. See Lambert and Warren (1999) for details.

<sup>9</sup> The imputation approach has been used widely in other fields of economics – Arellano and Meghir (1992) for example, combine the UK Labour Force Survey with the UK Family Expenditure Survey to estimate female labour supply and on-the-job search.

between common variables and expenditure are not required. Explicit methods, however, usually base expenditure imputation on some form of Engel Curve relationship, where a functional form of expenditure is assumed.

Decoster et al. (2007) carried out a detailed comparison of four methods to impute expenditure into an income survey. The four methods were: parametric and kernel estimation of Engel curves, Mahalanobis distance matching and grade correspondence. The results suggested that parametric estimation provided the most accurate results in terms of the imputed budget shares for a range of goods in the income survey. At least partly based on these results, parametric estimation of expenditure has become the most common method of imputing expenditure to an income survey for indirect tax reform analysis. To investigate the effect of increases in the Value Added Tax in Germany for example, Pestel and Sommer imputed expenditure from the German Sample Survey of Income and Expenditure (EVS) to the German Socio-Economic Panel Study (SOEP) which does not report expenditures. Based on the recommendations of Decoster et al. (2009), Pestel and Sommer used parametric estimation of Engel Curves for the imputation of expenditure data. The imputed expenditure in the SOEP survey was then compared to the actual expenditures in the EVS survey as validation for the imputation procedure. Using similar methods, Decoster et al. (2010), Figari and Paulus (2012), and O'Donoghue et al. (2004) separately examined the impact of indirect tax changes for range of European countries. Common to each study is the use of a tax-benefit microsimulation model based on an income survey which lacks expenditure data. In each case, parametric estimation of expenditure based on a range of common variables between the income and expenditure surveys was used.

While in each of these analyses the comparison of actual and imputed expenditure was encouraging, the limitations of the imputation approach must be recognised. Sutherland et al. (2002) discussed the imperfections of the imputation procedure and note that it is particularly difficult to predict expenditure patterns that capture micro-level diversity, as exhibited in the actual data. They correctly highlighted that imputed expenditure variables are only ever an adequate second-best substitute for actual data, and recommended that the imputed variables are used at a sufficient level of aggregation to mask differences that are not controlled by the imputation procedure. The first approach based on the expenditure survey alone is of interest in itself, they argued, and also provides a benchmark against which to assess approaches based on prediction of expenditure. Sutherland et al. conducted an evaluation of the imputation procedure by comparing results of indirect tax simulations using estimated and actual expenditures. They concluded that they could only be confident in the imputations when the dimensions that are important to the end analysis have been controlled for in the imputation procedure.

In the present analysis, in common with many other distributional analyses based on imputed expenditure data, we group households into income deciles. This step is motivated by a number of factors. First, as argued by Sutherland et al., we do not expect the imputation procedure to provide perfect estimates of expenditure on a household-by-household basis. Rather, the imputation

procedure is used to provide results which are as accurate as possible at the level of tabulations sought: in this case, income deciles. In addition, Decoster et al. (2007) address this issue when discussing the level of aggregation at which to present results in their analysis. A problem they identified was the effect observed for most countries that the tails of the disposable income distributions tend to differ in income and expenditure surveys. However, since the number of observations in the “diverging tails” is often observed to be very small<sup>10</sup>, they argued that it is more likely that these values can be considered as outliers rather than as systematic measure differences. Therefore grouping by decile minimises the risk that these outliers would have a significant effect on the outcome.

### 2.3 Contribution to the Literature

Sutherland et al. concluded that the use of imputed expenditure is not a perfect substitute for using actual data. This is confirmed by a number of papers that impute expenditure into an income survey, which show that, while largely comparable, imputed expenditure and budgets shares exhibit somewhat different distributions and budget shares when compared to actual expenditure data. In this paper, our first aim is to examine the source of these differences, and examine the sensitivity of distributional results to the choice between actual, estimated and imputed expenditure. As detailed in Section 4, the imputation procedure can be split into two broad steps. First, expenditure is estimated in the expenditure survey. Second, the coefficients from the estimation phase are used to impute expenditure into an income survey. While many of the studies discussed here compared imputed expenditure in the income survey to actual expenditure in the expenditure survey, it remains unclear whether any sensitivity in the distributional analysis is driven mainly by the estimation step, or by the imputation step in imputing expenditure.

To assess the degree of sensitivity, we examine the distribution of indirect taxes in Ireland, as well as the distributional effects of a number of indirect tax reforms, using three different types of expenditure data. First we assess the distribution of indirect taxes using actual expenditure data in the HBS. Second we use estimated expenditure data, remaining in the expenditure survey. Finally, we use imputed expenditure data in the income survey. This approach allows us to identify the source of any sensitivity in the distributional results.

Our second contribution is more straightforward. As discussed above, most recent analyses of the distribution of indirect taxes in Ireland are based on the HBS. They therefore do not allow a combined analysis of the direct and indirect taxes, along with social welfare, using the primary source of income data available. In addition, using the HBS does not allow the possibility of using microsimulation techniques to simultaneously model direct and indirect tax and social welfare reforms. The most recent analysis of this nature that we are aware of was done by O’Donoghue et

---

<sup>10</sup> Less than 1 per thousand in the Belgian Budget Surveys analysed in Decoster et al (2007).

al. using 1994 expenditure data as the source for the imputations. Therefore, the analysis done here serves as an updated picture of the distributional effects of the indirect tax system in Ireland<sup>11</sup> using the primary income survey. In addition, the methods and results found in this analysis can form a base for future microsimulation analysis of simultaneous direct and indirect tax, and welfare reform.

### 3. Data

The expenditure survey used in this paper is the 2009/10 wave of the Irish Household Budget Survey (HBS)<sup>12</sup>. The main purpose of the HBS is to gather detailed information that is used in the construction of the Consumer Price Index. In a nationally representative sample of 5,891 households<sup>13</sup>, responding households are asked to maintain a detailed diary of household expenditure over a two-week period. Some expenditure items, such as durable expenditure, are collected over a longer time frame. The survey also contains information on the characteristics of the households such as numbers in the household and the ages of household members, as well as some income information. Using the income information, we assign each household to a decile of equivalised disposable income<sup>14</sup>. Distributional impacts are presented using these deciles.

The indirect taxes included in this analysis are VAT, excise duties and the carbon tax. To estimate the VAT payments made by each household using the actual expenditure data, we assign the appropriate VAT rate to each of the expenditure items in the HBS data. Using these VAT rates, the VAT contribution of each household can be measured. Given that VAT is applied as a percentage of producer price, the VAT element of expenditure can be calculated as:

$$VAT_i^j = x_i^j * \frac{t^j}{1 + t^j}$$

where  $x_i^j$  household  $i$ 's expenditure good  $j$ , and  $t^j$  is the VAT rate on good  $j$ . The distributional impact of VAT reforms can then be estimated by changing  $t^j$ . We assume demand does not respond to relative price changes and that the full indirect tax is passed on to the consumer, as is common in the literature (see Pestel and Sommer (2013), Picos-Sanchez (2011) for example).

To estimate excise and carbon tax payments, an assumption is required on the price of the relevant goods<sup>15</sup>. SEAI (2013) report average prices for the goods on which carbon taxes apply<sup>16</sup>, while the

<sup>11</sup> Of course, the more recent contributions using the HBS data alone also provide valuable information in this regard.

<sup>12</sup> Survey conducted by the Central Statistics Office (CSO) of Ireland

<sup>13</sup> In the main analysis of the paper, we keep the entire sample of households. Sensitivity analysis was carried dropping certain proportions of the sample. The main conclusions of the paper did not change.

<sup>14</sup> Equivalisation of income and expenditure uses a scale of 1 for the first adult, 0.66 for additional adults, and 0.33 for children

<sup>15</sup> Excises and Carbon Taxes are charged per quantity. Only total expenditure is observed in the HBS data. Therefore, a representative price is required to estimate quantities purchased, and so excises paid.

Revenue Commissioners (2012) report national average retail prices for goods on which excise duties apply<sup>17</sup>. Using these price estimates, we can estimate excise payments as

$$Ex_i^j = \frac{x_i^j}{\tilde{p}^j} * Excise\ Duty^j$$

where  $\tilde{p}^j$  is the assumed price of good  $j$ , and  $Excise\ Duty^j$  is the excise duty applied per unit of good  $j$ . Carbon tax payments can be estimated in a similar manner. With the estimates of VAT, excise duties and carbon taxes paid by each household, we can use the income data contained in the HBS to assess the distributional impact of the indirect tax system, as well as the affects of reforms to that system.

In order to estimate household expenditure on different goods and services, expenditure items in the HBS are aggregated into a smaller number of representative goods. Hicks' Composite Commodity Theorem suggests that commodities among which the relative prices are constant can, in a natural way, be treated as a single commodity (Bradford 1974). Similarly, the expenditure items we group together should be as homogenous as possible. A trade-off exists between the desire to estimate expenditure for as many different goods and services as possible, with the need to be able to estimate the parameters of the model with precision. A review of the previous literature suggests aggregating expenditure items into 10 to 18 representative goods. Here we group the expenditure items in the HBS into 15 aggregated commodities<sup>18</sup>.

The income survey used in the paper is the 2010 wave of the Survey of Income and Living Conditions (SILC). The SILC is an annual survey conducted by the Central Statistics Office (CSO) in order to obtain information regarding the income and living conditions of Irish households. It is the Irish component of an EU wide survey which aims to capture information on poverty and social exclusion across Europe. The 2010 SILC surveyed 4,642 households and 11,587 individuals between January 2010 and January 2011, with an income reference period being the 12 months prior to interview. Crucially, it contains information on household composition so that individuals within a household can be grouped into tax-units. SWITCH, the tax-benefit microsimulation model of the ESRI, is based on this survey, while EUROMOD<sup>19</sup>, an EU-wide tax-benefit model, is based on the EU-SILC data.

Weber and Tonkin (2013) described the importance of ensuring similar distributions of matching variables between the income and expenditure surveys, while also arguing that having a smaller

<sup>16</sup> Carbon Tax on Kerosene, Marked Gas Oil, Fuel Oil, LPG (Other), Auto LPG and Natural Gas commenced in May 2010. In 2013, it was extended to include solid fuels.

<sup>17</sup> Various forms of alcohol, tobacco and energy products.

<sup>18</sup> Food, alcohol, tobacco, fuel and light, clothing, housing, household non-durables, health, public transport, private transport, communications, education, recreation and culture, durables, and other goods. See Appendix Table B for a sensitivity analysis where we group items into 10 aggregated goods.

<sup>19</sup> EUROMOD is developed at the Institute of Social and Economic Research (ISER) in the University of Essex

number of variables increases the risk that any model fitted to the data will be misspecified and the results of the matching will not be reliable.

Table 1: Sample Descriptives – HBS data and SILC data

	HBS	SILC
Mean Disposable Income (€)	885.7	827.6
Mean N in Household	2.8	2.7
Proportion with Married Head of Household (%)	51.4	52.7
Mean Proportion Females in Household (%)	52.2	51.3
Highest Education Level in Household (%):		
Junior Cert or Still in Education	30.8	27.9
Leaving Cert, Higher Certificate or Ordinary Degree	45.6	46.4
Honours Degree or Higher	22.7	22.9
Other	0.9	2.8
Location (%):		
Dublin	28.3	26.7
Other Urban	35.1	34.0
Rural	36.7	39.2
Household Tenure (%):		
Own	33.6	44.6
Mortgage	34.3	27.8
Rent	20.1	14.1
Other	12.0	13.5
Household Reference Person LFS (%)		
Employed	53.6	45.6
Unemployed	11.5	8.8
Retired	14.5	14.5
In Education	3.8	2.0
Home Duties	10.4	22.3
Other	6.2	6.8
Predominant Labour Force Status of Adults in HH (%):		
Employed	40.3	34.2
Unemployed	7.1	4.9
Retired	9.8	7.6
In Education	4.9	4.8
Home Duties	4.0	7.1
Households with Equal N of Adults across LFSs	20.3	23.6
Other	13.6	17.7

Table 1 compares a range of common variables between the HBS and SILC data. For the majority of variables, the statistics are very comparable between the two datasets. For example, approximately 23 per cent of households contain at least one person with an honours degree or higher in both

surveys, while the regional split of households is also very similar in the surveys. Some reasonably large differences do emerge between the surveys, however. Mean disposable income is almost €60 per week higher in the HBS compared to SILC. As explained by the CSO (CSO, 2012), these differences emerge due to different reference periods in the two surveys. In addition, a higher proportion of households outright own their house in the SILC, while more households have a mortgaged property or rent their property in the HBS. The labour force status (LFS) distribution also differs between the datasets. In particular, the HBS contains a higher proportion of households where the household reference person (HRP) is employed, and a lower proportion where the HRP is engaged in home duties, compared to SILC. To minimise differences in the LFS distribution between the surveys, we use an alternative specification of this variable in the estimation equations, which indicates the predominant LFS in a household, rather than the LFS of the HRP. As can be seen in Table 1, while some differences remain, this alternative specification brings the LFS distribution closer between the surveys.

## 4. Estimation and Imputation of Expenditure

The method of estimating expenditure in the HBS proceeds in two steps. First, we estimate total household expenditure. We then estimate budget shares for the range of aggregated commodities. To get household level estimated expenditure on each of our aggregated commodities, we simply multiply the estimated total expenditure by the estimated budget shares. Using the coefficients of each of these estimation steps, we impute total expenditure and budget shares from the HBS into SILC, and estimate expenditure on each of the goods in the same way as above. The following sections examine in more detail the estimation methods, and present results from these estimations.

### 4.1 Total Expenditure

Total household expenditure is estimated in the expenditure survey using the following specification:

$$\ln(E_i) = \alpha + \sum_{k=1}^3 \gamma_k \ln(y_i)^k + x_i' \beta + \varepsilon_i \quad (1)$$

for households  $i = 1, \dots, H$ .  $\ln(E_i)$  is the logarithm of total expenditures at the household level.  $\ln(y_i)$  denotes the logarithm of disposable household income and  $x_i$  denotes a vector of household-specific characteristics, as detailed in

**Table 1.** The error term  $\varepsilon_i$  is assumed to be independent and identically distributed. As the variance of the predicted expenditure is lower than actual expenditure, an additional error,  $\mu_i$ , is added to

the predicted expenditure. This error term is mean zero, and is specified so that the variance of actual and predicted expenditure is equal in the HBS data.

The coefficients from the total expenditure equation are used to impute expenditure into the income survey, as in Equation 2<sup>20</sup>. Total expenditure estimations and imputations are corrected for retransformation bias via use of the smearing estimator (see Duan et al. (1983) and Manning (1998) for details). Similarly to the estimated total expenditure in the HBS data, imputed total expenditure in the income survey has lower variance than actual total expenditure in the expenditure survey. Therefore, the additional error term  $\mu_i$  as before, is added to the imputed expenditure in the SILC data.

$$\ln(E_{SILC}) = \hat{\alpha}_{HBS} + \sum_{k=1}^3 \hat{\gamma}_{k_{HBS}} \ln(y_{SILC})^k + x'_{SILC} \hat{\beta}_{HBS} + \mu \quad (2)$$

As in Sutherland et al. (2002), our validation of the imputation procedure relies on comparison of the distribution of expenditure (and later indirect taxes) when we use actual, estimated or imputed data. Table 2 compares the distribution of actual total expenditure with estimated total expenditure in the HBS, and imputed total expenditure in SILC. Survey weights are applied in each case.

Table 2: Comparison of actual, estimated and imputed total expenditure – % in each expenditure category and expressed as percentage of actual figure

	HBS	SILC	
Expenditure per Week	Actual	Estimated	Imputed
<€200	6.6	8.2	9.5
<€400	16.5	12.9	14.6
<€600	19.5	16.7	17.0
<€800	16.0	17.2	16.7
<€1000	12.7	14.0	13.5
<€1250	11.6	11.4	12.5
<€1500	7.0	8.8	7.2
<€1750	4.4	4.9	4.2
<€2000	2.2	2.4	2.3
>€2000	3.6	3.5	2.5
Mean (€)	811	839	792
Median (€)	683	738	700

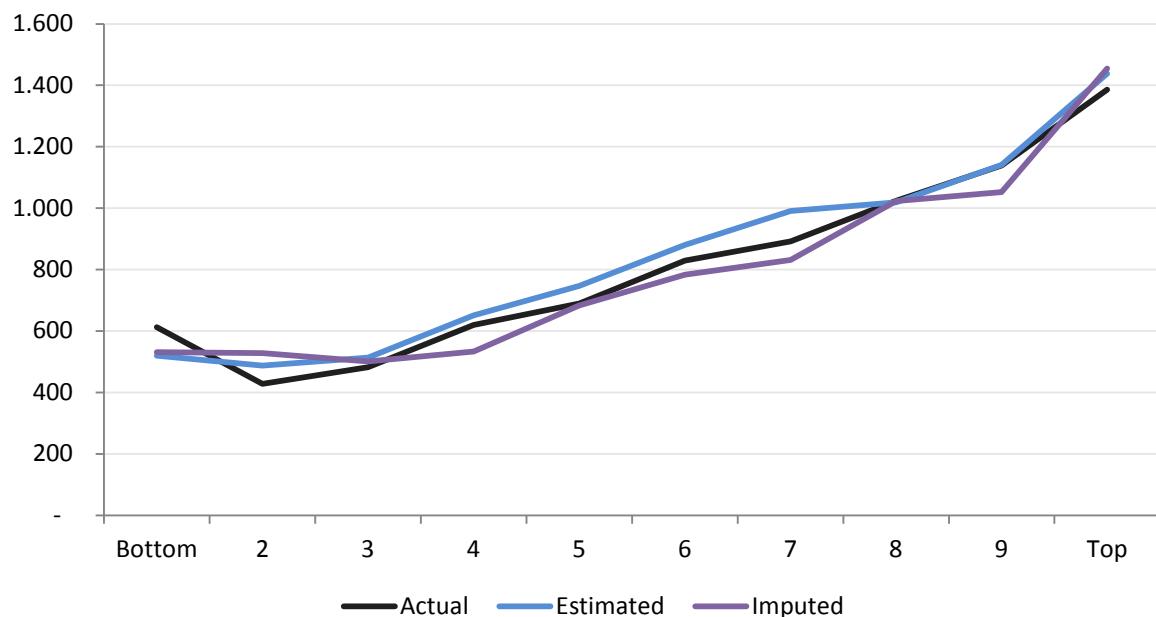
*NOTE: Survey weights applied here, and in all other tables*

<sup>20</sup> Household subscripts dropped for clarity

The proportion of households in the highest expenditure categories is very similar across all three expenditure data types. Between 4 and 5 per cent of households have expenditure between €1,500 and €1,750 a week, approximately 2 per cent of households have expenditure between €1,750 and €2,000 a week, while around 3 per cent of households have expenditure about €2,000 per week. The mean and median expenditure in the estimated and imputed data also compare favourably to the actual data. The largest discrepancies arise in the lowest expenditure category. According to the actual HBS data, 6.6 per cent of households have expenditure of less than €200 per week. The estimated data results in 8.2 per cent of households in this category, while the imputed data results in a figure of 9.5 per cent. The importance of including the error term in the estimated and imputed expenditure is highlighted here. Without including the error term, less than 0.5 per cent of households are in this category in both the imputed data in SILC and the estimated data in the HBS.

We can also compare the distribution of expenditure by decile of equivalised disposable income. Figure 1 shows that in the actual HBS data, mean expenditure is higher in the bottom decile than either of the second or third deciles. From the second decile onwards, mean expenditure is increasing in income. The estimated data in the HBS captures this pattern to a certain degree, as mean expenditure decreases between the first and second deciles, although by less than observed in the actual data. The pattern is less pronounced in the imputed data, where mean expenditure remains relatively constant between deciles one, two and three. Given the different income reference periods between the surveys, this pattern is not surprising. Households with low current income in the HBS may be temporarily ranked in the bottom decile yet maintain their higher expenditure levels through savings. As SILC uses an annual income measurement, households are less likely to have a temporary fluctuation in their income, and so expenditure may be more likely to rise in the income deciles.

Figure 1: Mean Expenditure by Decile of Equivalised Disposable Income



## 4.2 Estimating Budget Shares of Aggregated Commodities

Having estimated total expenditure for each household, and decided upon the level of aggregation of expenditure items, the next step in the process is to predict expenditure shares for each of the aggregated goods. Here the aggregated goods are split into two groups. The first group contains goods for which all households record a positive expenditure<sup>21</sup>. Household expenditure for this group of goods is estimated using the following specification, using the full sample of households in the expenditure data:

$$w_i^j = \alpha^j + \gamma_1^j \ln(E_i) + \gamma_2^j \ln(E_i)^2 + x_i' \beta^j + \eta_i^j \quad (3)$$

for households  $i = 1, \dots, N$  and commodities  $j = 1, \dots, J$ .  $x_i$  contains the same set of variables as in the specification for total expenditure. Again the error term,  $\eta_i^j$ , is assumed to be independent and identically distributed.

The second group contains goods for which a significant proportion of households record zero-expenditures. For example, a significant proportion of households will commonly not have any expenditure on goods such as alcohol and tobacco. For this group of goods, a two-stage estimation procedure is required<sup>22</sup>. In the first step, a Probit model is estimated for the probability of positive demand for the respective good:

$$\Pr(D_i^k = 1) = \phi(\alpha^k + \delta_1^k \ln(Y_i) + \delta_2^k \ln(Y_i)^2 + x_i' \beta_0^k + v_i^k) \quad (4)$$

for households  $i = 1, \dots, N$ , where  $D_i^k$  denotes a dummy variable that is 1 for household  $i$  if demand for good  $k$  is positive, and zero otherwise. The term  $\alpha^k$  denotes a commodity-specific constant, and  $v_i^k$  is an error term, assumed to be independent and identically distributed. Equation (3) is estimated separately for each of the goods in the ‘zero-expenditures’ group.

In the second step, budget share equations are estimated for each of the goods, conditional on the outcome of the first step. Following the approach taken in Decoster et al. (2013), predicted probabilities are set equal to zero if  $\Pr(D_i^k = 1)$  is lower than the mean probability for good  $k$ , and

---

<sup>21</sup> In practice, we include goods for which at least 99 per cent of households have positive expenditure.

<sup>22</sup> See Appendix Table C for a sensitivity test of this approach, where we simply use a full-sample OLS regression to predict expenditure for each good. The OLS approach significantly underestimates the proportion of households that do not report expenditure on this group of goods

set equal to one otherwise. For households for which demand is estimated to be positive ( $\widehat{D}_i^k = 1$ ), we estimate on a restricted sample

$$w_i^k = \alpha_o^k + \gamma_1^k \ln(E_i) + \gamma_2^k \ln(E_i)^2 + x_i' \beta^k + \eta_i^k \quad if \quad \widehat{D}_i^k = 1 \quad (5)$$

for goods  $k=1,\dots,K$ .

#### 4.3 Imputing Budget Shares into SILC

Using the results of the budget share equations, budget shares are imputed into the income survey, and expenditure on each representative good is estimated by multiplying the predicted budget shares by the predicted total expenditure. In some cases, negative budget shares are predicted for some households in the income survey. Negative imputed budget shares are set to zero, and the other budget shares are rescaled, to comply with adding-up conditions.

For the goods with significant proportions of zero-expenditures, imputation is based on the two-stage estimation procedure. To incorporate a degree of randomness in estimating which households have positive expenditure on these goods, we take the approach suggested in Decoster et al. (2013). After imputing the probabilities for positive expenditures, a number is drawn for each commodity from a uniform distribution between 0 and 1. If the number drawn is smaller than the predicted probability,  $\widehat{D}_i^k$  in the SILC data is coded 1, and 0 otherwise. Expenditure is then estimated using the second stage regression output for households with  $\widehat{D}_i^k = 1$ . This approach ensures that a portion of households with a relatively low predicted probability of having positive expenditure are estimated to have positive expenditure (as is observed in the HBS data).

Table 3 compares the actual, estimated and imputed budget shares for the 15 aggregated goods in the analysis. The estimated and imputed budget shares are very similar for the majority of goods. The mean and median budget share for food is close to 20 per cent for both, while the budget shares for most of the other commodities are within 0.4 percentage points of each other. One exception that emerges is the budget share for housing. In the estimated HBS data, the mean budget share for housing is 19.9 per cent, over two percentage points larger than the mean imputed housing budget share. Similarly, the median estimated budget share is over four percentage points larger than the imputed budget share. This difference is likely to be the result of the different distributions of housing tenure between the two datasets, as highlighted in

Table 1. A higher proportion of households own their home in SILC, and therefore will face lower housing costs than in the HBS. Interestingly, the imputed housing budget share, rather than the estimated share, compares more favourably to the actual budget share for housing.

Table 3: Budget Shares – Actual, Estimated and Imputed

	HBS			SILC		
	Actual		Estimated		Imputed	
	Mean (%)	Median (%)	Mean (%)	Median (%)	Mean (%)	Median (%)
<b>Food</b>	18.4	16.9	20.1	19.4	20.9	20.0
<b>Alcohol</b>	3.2	1.8	2.3	2.6	2.3	2.6
<b>Tobacco</b>	2.3	0.0	1.0	0.0	1.0	0.0
<b>Clothing</b>	4.5	2.5	3.3	3.9	3.3	4.0
<b>Fuel and Light</b>	5.8	4.4	6.4	5.8	6.8	6.2
<b>Housing</b>	18.2	16.3	19.9	21.4	17.6	17.0
<b>Household Non-Durables</b>	2.2	1.7	2.4	2.4	2.4	2.3
<b>Health</b>	2.2	0.4	1.4	1.0	1.6	1.2
<b>Public Transport</b>	1.1	0.0	0.6	0.0	0.5	0.0
<b>Private Transport</b>	9.2	8.0	9.2	10.0	9.4	10.2
<b>Communications</b>	3.8	3.1	4.1	4.0	4.2	3.9
<b>Recreation and Culture</b>	4.2	3.0	4.5	4.4	4.7	4.7
<b>Education</b>	2.0	0.0	1.1	0.0	0.9	0.0
<b>Durables</b>	6.4	3.5	5.8	6.5	6.0	6.9
<b>Other</b>	16.5	14.5	17.8	18.2	18.5	18.7

Table 4 compares the frequency of zero budget shares among the eight goods in the ‘zero-expenditure’ category. Tobacco, public transport and education are the goods with the highest proportion of zero-expenditures; more than 60 per cent of households record zero expenditures on each of these goods. Almost 40 per cent of households record no expenditure on health, while between 25 and 30 per cent of households have no expenditure on alcohol or clothing. The two-step estimation and imputation technique results in a similar distribution of zero-expenditures, in both the estimated and imputed figures. In most cases, the two-step procedure results in a slight over-estimation of the frequency of zero-expenditures, although for the majority of goods the bias is small. The proportion of households recording zero expenditure for alcohol, for example, increases from 29 per cent in the actual data to 31 per cent in the estimated HBS data, and 33 per cent in the imputed SILC data. Similarly, health zero-expenditures are over-estimated by about 6 percentage points in the estimated and imputed data.

Table 4: Frequency of Zero Budget Shares – Actual, Estimated and Imputed

	HBS	SILC	
	Actual (%)	Estimated (%)	Imputed (%)
<b>Alcohol</b>	29.0	30.7	32.8
<b>Tobacco</b>	66.0	68.0	69.8
<b>Clothing</b>	26.3	29.8	31.5
<b>Health</b>	38.0	44.1	44.4
<b>Public Transport</b>	60.9	63.7	65.4
<b>Private Transport</b>	14.3	13.8	13.3
<b>Education</b>	62.3	65.7	69.5
<b>Durables</b>	10.1	13.7	16.3

#### 4.5: Estimating Indirect Tax payments using Estimated and Imputed Data

The final pieces of information we require to assess the distributional affects of indirect taxes, given our estimates of expenditure on each of the aggregated goods, are the VAT rates, excise duties and carbon taxes applicable to each of the aggregated goods.

To calculate the proportion of expenditure on each aggregated good that was paid in VAT, we categorise the expenditure items into groups according to whether the item had a zero, reduced, second reduced or standard VAT rate applied. We then calculate the share by decile of each aggregated good that was spent on each VAT category. For example, in the bottom decile 64 per cent of expenditure on food was on food items that had a zero VAT rate applied in 2010; the corresponding figure in the top decile was 51 per cent. With this information we can simply estimate the VAT element of expenditure on each aggregated good as:

$$\widehat{VAT}_i^j = \sum_{v=0}^V \hat{x}_i^j * \pi_d^{jv} * \frac{t^v}{1+t^v} \quad (6)$$

where  $v=1,\dots,V$  are the different VAT rates in the indirect tax system, and  $\pi_d^{jv}$  is the share of total expenditure on good  $j$  in decile  $d$  ( $d=1$  to 10) that was spent on expenditure items taxed at VAT rate  $v$ . In the 2010 Irish system, goods could be subject to a zero, reduced or standard VAT rate, so in this case  $V=3$ . From July 1<sup>st</sup> 2011, a second reduced rate of VAT of 9 per cent was introduced in Ireland, resulting in 4 VAT rates in total, so  $V$  increased to 4.

Excise duty and carbon tax payments can be estimated in a similar manner. As before, an assumed price per unit is required in this setting.

$$\widehat{Ex}_i^j = \sum_{n=1}^N \frac{\widehat{x}_i^j \pi_d^{jn}}{\tilde{p}^n} * Excise\ Duty^n \quad (7)$$

where  $n=1,\dots,N$  are the component expenditure items of each estimated good on which excise duty applies,  $\pi_d^{jn}$  is the share of aggregated good  $j$  in decile  $d$  that is spent on expenditure item  $n$ ,  $\tilde{p}^n$  is the assumed price per unit of good  $n$ , and  $Excise\ Duty^n$  is the excise duty per unit of good  $n$ . Carbon taxes can be estimated in a similar manner.

Table 5: Share of expenditure on aggregated commodity spent on goods/services subject to each VAT rate ( $\pi_d^{jn}$ ) - HBS 2009/10 – Food and Clothing

Decile	Food			Clothing	
	Zero	Reduced	Standard	Zero	Standard
1	64	22	14	15	85
2	69	17	13	16	84
3	68	18	13	13	87
4	67	18	15	20	80
5	64	21	15	15	85
6	62	24	14	18	82
7	61	25	14	14	86
8	58	30	12	13	87
9	54	34	12	13	87
10	51	39	10	11	89
State	62	25	13	15	85

*Note: No Clothing items are subject to the reduced VAT Rate*

Allowing  $\pi_d^{jn}$  to vary by decile has one implication worth discussing. For each of the aggregated goods, it is possible to calculate an implicit tax rate given information on household expenditure and the appropriate tax rates on the component goods. With a constant  $\pi^{jn}$ , the implicit tax rate on each aggregated good would also be constant across the income distribution. However, by allowing  $\pi_d^{jn}$  to vary by decile, the implicit tax rate on each good will also vary by decile. A simple example will illustrate this. Say 50 percent of recreation expenditure was on zero-rated recreation items in the bottom decile, and the remaining recreation expenditure in the bottom decile was on goods subject to the standard rate (21 per cent) of VAT. In the top decile, say only 25 percent of recreation expenditure was on zero-rated items, with the remainder on standard-rated items. In this case, by allowing  $\pi_d^{jn}$  to vary by decile, the implicit tax rate on recreation in the bottom decile is 10.5 percent<sup>23</sup>, and the same tax rate for the top decile is 5.25 percent. If  $\pi^{jn}$  was held constant, say at

<sup>23</sup> 50%\*Zero-VAT Rate + 50%\*Standard-VAT rate

40 per cent, then the implicit tax rate on recreation would be 8.4 percent. The same reasoning can be used for other goods, and other indirect tax instruments.

So does this matter? Actual indirect tax rates do not vary according to income. Allowing implicit indirect tax rates to vary by income may therefore seem counterintuitive. However it is important to use all available information in the imputation procedure. Indeed, Sutherland et al. (2002) explicitly warn that we can only be confident in the imputations when the dimensions that are important to the end analysis have been controlled for in estimation. In presenting results by income decile, it therefore seems reasonable to use all information that varies by decile in estimating indirect tax payments. Ultimately, we believe that the counterintuitive result of varying implicit tax rates on aggregated goods by income decile is more than compensated for by increasing the precision of our estimates<sup>24</sup>.

## 5. Distributional Effects of Indirect Taxes and Indirect Tax Reform

Imputation procedures, such as the one used in this analysis, cannot be expected to accurately predict expenditure levels on a household by household basis. In this section, we analyse the distribution of indirect tax payments in Ireland, comparing the results using the estimated and imputed expenditure data with the results using the actual HBS data. As argued by Sutherland et al (2002), the imputed expenditure should be used at a sufficient level of aggregation to mask differences that are not controlled by the imputation procedure. Therefore, like much distributional analysis (see Callan et al. (2013), Decoster et al. (2011) among others), the results are shown by income decile. We examine if this level of aggregation masks any imperfections that emerge in the expenditure imputation procedure.

Table 6 shows the proportion of disposable income paid in indirect taxes across the income distribution under the 2010 indirect tax system. The results portray the regressive nature of the indirect tax system in Ireland. The actual HBS data suggests that the bottom decile paid almost a quarter of its disposable income in indirect taxes, while this falls to less than a tenth of income in the 8<sup>th</sup> and 9<sup>th</sup> deciles. The top decile paid just over 6 per cent of its income in indirect taxes; just above a quarter of the proportion paid by the bottom decile. The estimated and imputed expenditure data suggest a similarly regressive pattern in indirect tax payments. Encouragingly, in 8 out of the 10 deciles, the estimated and imputed indirect tax payments are within 1 percentage point of the actual data. Similarly, the state average indirect tax payment is approximately 10 to 11 per cent in each case. The largest difference emerges in the bottom decile. The estimated data suggests indirect taxes were under 18 per cent of disposable income, while the imputed data figure of almost 21 per cent is closer to the actual figure of 24.6 per cent.

---

<sup>24</sup> See Appendix Table D for a sensitivity analysis where we fix  $\pi$  across income deciles, by using the state average expenditure shares.

Table 6: Indirect Taxation by decile of equivalised disposable income, 2010 as % Disposable Income

<b>Decile</b>	<b>HBS</b>	<b>SILC</b>	
	<i>Actual (%)</i>	<i>Estimated (%)</i>	<i>Imputed (%)</i>
<b>Bottom</b>	24.6	17.6	20.5
2	15.3	14.4	15.2
3	14.6	12.6	12.7
4	12.6	11.7	11.4
5	12.1	11.4	11.6
6	11.7	10.9	10.7
7	10.0	10.1	9.7
8	9.7	9.0	9.8
9	9.0	8.2	8.6
<b>Top</b>	6.4	6.5	6.6
<b>State</b>	10.5	9.7	10.0

Table 7 and Table 8 decompose the indirect tax payments into VAT payments and excise and carbon tax payments by decile, again shown as a proportion of disposable income. Again the results are encouraging, with VAT payments representing 7 per cent of total disposable income, and excise and carbon taxes making up the remaining 3 per cent of income, in each case. The results show that the regressive nature of the indirect tax system is not driven solely by either tax component. Rather, both components contribute to the regressive pattern. The proportion of income paid in VAT is approximately 3 times larger in the bottom decile than the top decile, for example, while excises and carbon taxes are also 2 to 3 times larger at the bottom than the top of the income distribution.

Table 7: VAT by decile of equivalised disposable income, 2009/10 as % Disposable Income

<b>Decile</b>	<b>HBS</b>	<b>SILC</b>	
	<i>Actual (%)</i>	<i>Estimated (%)</i>	<i>Imputed (%)</i>
<b>Bottom</b>	16.4	13.1	14.9
2	9.9	10.4	10.9
3	9.3	9.1	9.3
4	8.4	8.4	8.4
5	8.0	8.2	8.4
6	7.9	7.9	7.8
7	6.8	7.3	7.0
8	6.9	6.4	7.0
9	6.4	6.0	6.3
<b>Top</b>	4.8	4.9	5.0
<b>State</b>	7.2	7.1	7.3

Lower predicted excise payments are the main cause of the imperfections in the comparison of indirect tax payments in the bottom decile. While the actual data suggests excises made up approximately 8 per cent of disposable income in the bottom decile, with the estimated and imputed data this figure is just above 4 per cent and 5 per cent respectively. The largest discrepancy in the imputation procedure occurring in the bottom decile in relation to excise duties is unsurprising. First, expenditure on the goods subject to excise duties are all estimated via the two-stage zero-expenditure imputation procedure. With an extra estimation step, it is unsurprising to find slightly more error in the results. Secondly, the relationship between income and expenditure in the bottom decile can be particularly difficult to model. Meyer and Sullivan (2011) suggested that that income is likely to be poorly measured for households with low resources and, in particular, likely to be under-reported. Similarly, Brewer and O'Dea (2012) found evidence from UK data of under-reporting of income among households with low resources. They argue that household expenditure may be more likely to be accurately measured. Therefore, given the extra estimation steps in relation to excise duties, and the possibility of some data error in the bottom decile, it is unsurprising to see some element of discrepancy in the results here. Nonetheless, the distributional results shown in Table 6 are an encouraging basis from which to undertake some indirect tax reform analysis.

Table 8: Excise Duties and Carbon Tax by decile of equivalised disposable income, 2009/10 as % Disposable Income

	HBS	SILC	
Decile	<i>Actual (%)</i>	<i>Estimated (%)</i>	<i>Imputed (%)</i>
<b>Bottom</b>			
2	8.2	4.6	5.6
3	5.4	4.0	4.3
4	5.3	3.5	3.4
5	4.2	3.3	3.0
6	4.1	3.2	3.2
7	3.8	3.0	2.9
8	3.1	2.9	2.7
9	2.9	2.5	2.8
Top	2.6	2.2	2.3
<b>State</b>	1.6	1.6	1.6
	3.3	2.6	2.7

Between 2010 and 2014 in Ireland, two major VAT reforms occurred. The first was the introduction of a second reduced rate of VAT. A new rate of 9 per cent was applied to a range of goods and services in the tourism sector, such as certain food and drink products, recreational activities such as admission to museums and cinemas, hotel lettings and printed material such as newspapers and magazines. The second major VAT reform was an increase in the standard rate of VAT from 21 per cent to 23 per cent. Table 9 compares the distributive impact of these reforms, again using the three

sources of expenditure data. Taken together, these reforms to the VAT system were regressive, costing the bottom decile between 0.6 and 0.8 per cent of disposable income. The second decile experiences the second largest impact of the reform, costing between 0.4 and 0.5 per cent of income. The impact in the middle deciles was smaller, at about 0.3 per cent of income, while the top decile lost just 0.1 to 0.2 per cent of income. The results show that for a marginal indirect tax change, such as those in Ireland between 2010 and 2014, the actual and imputed data produce very similar results, with the largest difference, of just 0.15 percentage points, occurring in the bottom decile between the actual and imputed expenditure.

Table 9: The Distributive Impact of Two Recent VAT reforms, % Disposable Income

	HBS	SILC	
Decile	<i>Actual (%)</i>	<i>Estimated (%)</i>	<i>Imputed (%)</i>
<b>Bottom</b>	-0.6	-0.6	-0.8
<b>2</b>	-0.4	-0.5	-0.5
<b>3</b>	-0.3	-0.3	-0.4
<b>4</b>	-0.3	-0.4	-0.4
<b>5</b>	-0.3	-0.4	-0.4
<b>6</b>	-0.3	-0.4	-0.4
<b>7</b>	-0.2	-0.3	-0.3
<b>8</b>	-0.2	-0.3	-0.3
<b>9</b>	-0.2	-0.3	-0.3
<b>Top</b>	-0.1	-0.2	-0.2
<b>State</b>	-0.2	-0.3	-0.3

The final question we examine is how the results perform for a non-marginal change in indirect taxes. To examine this we examine the distributional impact of moving to a uniform VAT rate of 23 per cent. It must be remembered that the figures here represent a first round effect as they do not take account of any behavioural responses that would be likely in response to such a non-marginal change. As shown in Table 10, again the actual, estimated and imputed data produce very similar results. Such a reform would cost the bottom decile between 19 and 21 per cent of disposable income. This cost would fall to between 12 and 14 per cent in the second decile, and to just 8-9 per cent in the top decile. The state average cost would be between 10 and 11 per cent of disposable income.

Table 10: The Distributive Impact of a uniform VAT rate of 23 per cent, % Disposable Income

<b>Decile</b>	<b>HBS</b>	<b>SILC</b>
<b>Bottom</b>	<i>Actual (%)</i>	<i>Estimated (%)</i>
2	-21.0	-18.8
3	-11.8	-14.1
4	-10.9	-12.3
5	-10.9	-11.8
6	-10.2	-11.5
7	-10.1	-11.1
8	-9.5	-10.8
9	-9.5	-9.8
<b>Top</b>	-9.2	-9.5
<b>State</b>	-8.3	-8.6
	-10.0	-10.6

## 6. Conclusion

Accurate appraisal of public policy should include as much of the relevant information as possible. In measuring the distributional impact of the tax and welfare system, it is therefore important to include not only the impact of the direct tax and welfare system, but also the impact of the indirect tax system. Comprehensive modelling of the impact of both direct and indirect taxes, and of tax policy options, requires data on the impact at micro-level – household and tax-unit level. Most countries, however, do not have a single source of micro-data including high-quality disaggregated information on both incomes and expenditures. This makes approaches which impute expenditure data into detailed income surveys of considerable interest.

In this paper, we investigated the feasibility of imputing expenditure information from a household expenditure survey to an income survey. By examining the expenditure data, and the subsequent distribution of indirect taxes, at each stage of the imputation process, we highlighted the sources of imperfections in the imputation process. The distributions of estimated expenditure data in the HBS and the imputed expenditure data in SILC were highly comparable to the distribution of actual HBS expenditure data. When we used each of the expenditure data sources to estimate indirect tax payments, again the results were highly comparable across datasets. Despite some differences emerging in the distributions of common variables between the HBS and SILC, in most cases the estimated and imputed expenditure distributions were very similar, with some slight discrepancies emerging when compared to the actual data. Clearly, it is vital that a well-fitted model of household expenditure is used in the first stage of the estimation process. If not, the quality of the imputed expenditure is certain to suffer.

The results for the bottom income decile were most sensitive to the choice between the actual, estimated or imputed expenditure data. A number of factors give rise to this result. First, the income reference period in the HBS and SILC is different. Differences in the tails of the income distribution can result, particularly the bottom tail where temporary negative shocks to current income (as measured in the HBS) can push households temporarily into the bottom decile. This is also true of positive income shocks which can temporarily push households higher up the income distribution. With annual income (as measured in SILC), these temporary shocks are less likely. Imputed expenditure in the SILC survey was found to increase in income, while actual expenditure was higher in the bottom decile than the second or third deciles. This suggests that there may be a higher proportion of households in the HBS survey with temporarily low income that support a higher level of expenditure through savings or credit. Another reason why we observed more variability in expenditure at the bottom of the income distribution may be due to the quality of the income data in the HBS. As the CSO point out, the Survey of Income and Living Conditions (SILC) is the primary source of data on household and individual income (CSO, 2012). Therefore, some of the observations in the HBS with the lowest income yet higher expenditure may simply be the result of measurement error in income.

The empirical findings in the analysis confirm the regressive nature of the indirect tax system in Ireland, on a snapshot basis at least. It is important to remember, however, that the overall distributional impact depends not just on the indirect tax system, but also on the direct tax system and the welfare system. An integrated analysis of all three components – direct tax, indirect tax and welfare – is needed to provide a comprehensive picture of the overall distributional impact of the tax and benefit system. A number of recent recommendations on the structure and analysis of the effects of taxation are particularly relevant here. In a comprehensive review of the tax system in the UK, Mirrlees et al. (2011) argued that a failure of government was not examining the tax system as a whole and recognising that the rate schedule of personal income taxes and benefits is the instrument most suited to achieving redistributive ends. Similarly, in a World Bank publication, Minh le's (2003) main recommendation was that regressivity be studied in an overall context of the applicable fiscal policies including direct, indirect taxes, and public expenditures. He also argued that regressivity can be addressed more effectively with sound income taxation and government expenditures. More formally, Sah (1983) analysed the amount of redistribution that was possible using the indirect tax system as a redistributive tool. In the framework of a balanced public budget, the real income gain to the worst-off was taken as the metric of redistribution. The maximum possible redistribution using the indirect tax system was found to be meagre. Assessing the distributional impact of the indirect tax system alone, without considering the distributional impact of other parts of the tax and welfare system, therefore omits some of the most important schemes used to target distributional concerns. The analysis here is the first step in the process which will allow such a comprehensive study of the tax and benefit system in Ireland.

Ultimately, the aim of this analysis is to assess the feasibility of using the imputation process to enhance the SILC data with expenditure data. This process allows simultaneous analysis of the direct tax, indirect tax and welfare systems. Taking the advice of Mirrlees and others, it is important to

examine the distributional impact of the indirect tax system in the context of the full tax and benefit system. The methods and results in this paper are encouraging in that regard. By showing that the imputation procedure used in this analysis produced results with the SILC data that are highly comparable to results produced with actual expenditure data in the HBS, we have opened the possibility of measuring the distributional impact of the full tax and benefit system in Ireland.

## Appendix

Appendix Table A: Estimation of Total Expenditure with HBS data – across a range of samples

	(1)		(2)		(3)		(4)		(5)	
	$\beta$	se								
<b>Intercept</b>	5.99	0.09	16.69	2.36	5.21	2.85	8.62	2.21	7.41	0.41
<b>Log(Income)</b>	-0.23	0.06	-5.40	1.08	-0.41	1.28	-2.01	1.01	-1.02	0.21
<b>Log(Income)<sup>2</sup></b>	0.02	0.02	0.84	0.16	0.12	0.19	0.36	0.15	0.16	0.04
<b>Log(Income)<sup>3</sup></b>	0.00	0.00	-0.04	0.01	-0.01	0.01	-0.02	0.01	0.00	0.00
<b>N in Household</b>	0.14	0.01	0.13	0.01	0.12	0.01	0.12	0.01	0.14	0.01
<b>N under 5 in HH</b>	-0.11	0.01	-0.09	0.01	-0.09	0.01	-0.08	0.01	-0.10	0.01
<b>N 5-13 in HH</b>	-0.09	0.01	-0.07	0.01	-0.07	0.01	-0.07	0.01	-0.08	0.01
<b>N 14-20 in HH</b>	-0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01
<b>LFS Employed</b>	0.09	0.02	0.05	0.02	0.02	0.02	0.05	0.02	0.07	0.02
<b>LFS Retired</b>	-0.21	0.03	-0.22	0.03	-0.24	0.03	-0.20	0.03	-0.21	0.03
<b>LFS In Education</b>	0.20	0.03	0.19	0.03	0.15	0.03	0.20	0.03	0.20	0.03
<b>LFS Home Duties</b>	-0.13	0.03	-0.14	0.03	-0.14	0.03	-0.11	0.03	-0.13	0.03
<b>LFS III_Disabled</b>	-0.22	0.04	-0.22	0.04	-0.23	0.04	-0.18	0.04	-0.22	0.04
<b>LFS Other</b>	-0.23	0.08	-0.24	0.08	-0.23	0.08	-0.20	0.08	-0.24	0.08
<b>LFS Equal - 1 Emp</b>	0.09	0.02	0.07	0.02	0.04	0.02	0.06	0.02	0.08	0.02
<b>LFS Equal - No Emp</b>	-0.02	0.03	-0.03	0.03	-0.07	0.03	-0.03	0.03	-0.02	0.03
<b>Mortgage</b>	0.21	0.01	0.21	0.01	0.20	0.01	0.20	0.01	0.21	0.01
<b>Tenant Purchase</b>	-0.12	0.10	-0.11	0.10	-0.12	0.09	-0.11	0.09	-0.11	0.10
<b>Rent Local Authority</b>	0.05	0.02	0.05	0.02	0.04	0.02	0.05	0.02	0.05	0.02
<b>Rent Free</b>	-0.13	0.05	-0.13	0.05	-0.08	0.05	-0.10	0.04	-0.13	0.05
<b>Other Urban</b>	-0.07	0.01	-0.07	0.01	-0.07	0.01	-0.06	0.01	-0.07	0.01
<b>Rural</b>	-0.08	0.01	-0.08	0.01	-0.07	0.01	-0.08	0.01	-0.08	0.01
<b>highed_1</b>	-0.25	0.02	-0.23	0.02	-0.21	0.02	-0.20	0.02	-0.24	0.02
<b>highed_2</b>	-0.21	0.02	-0.19	0.02	-0.18	0.02	-0.17	0.02	-0.20	0.02
<b>highed_3</b>	-0.12	0.02	-0.10	0.02	-0.10	0.02	-0.10	0.02	-0.11	0.02
<b>highed_4</b>	-0.10	0.02	-0.09	0.02	-0.09	0.02	-0.07	0.02	-0.10	0.02
<b>highed_5</b>	-0.05	0.03	-0.04	0.03	-0.06	0.03	-0.04	0.03	-0.05	0.03

<b>highed_6</b>	-0.03	0.02	-0.03	0.02	-0.03	0.02	-0.02	0.02	-0.03	0.02
<b>highed_8</b>	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.03	0.02
<b>highed_9</b>	-0.18	0.06	-0.16	0.06	-0.15	0.06	-0.14	0.06	-0.17	0.06
<b>married_hoh</b>	0.06	0.01	0.06	0.01	0.05	0.01	0.06	0.01	0.06	0.01
<b>Prop Female</b>	0.04	0.02	0.04	0.02	0.04	0.02	0.05	0.02	0.04	0.02
<b>R-Squared</b>	0.68		0.69		0.69		0.71		0.68	
<b>N</b>	5,891		5,838		5,715		5,796		5,867	

Notes:

- (1) Full Sample
- (2) Drop bottom 1% of Gross Income Observations
- (3) Drop bottom 3% of Gross Income Observations
- (4) Drop HHs with expenditure more than 4 times gross income (Pestel and Sommer 2013)
- (5) Trimming as per Decoster et al 2013 - if durable expenditure is greater than 2 times disposable income

Appendix Table B: Mean and Median Budget Shares with 10 Good Aggregation

	HBS				SILC	
	Actual		Estimated		Mean (%)	Median (%)
	Mean (%)	Median (%)	Mean (%)	Median (%)		
Food	18.4	16.9	19.4	18.8	20.3	19.4
Alcohol	3.2	1.8	2.3	2.6	2.3	2.6
Tobacco	2.3	0.0	1.0	0.0	1.0	0.0
Clothing	4.5	2.5	3.3	3.9	3.2	4.0
Fuel and Light	5.8	4.4	6.1	5.6	6.7	5.9
Housing	18.2	16.3	19.1	20.9	17.0	16.2
Household Non-Durables	2.2	1.7	2.3	2.3	2.3	2.3
Durables	3.7	2.4	3.3	3.6	3.3	3.8
Transport	13.3	11.3	13.9	14.0	13.7	14.2
Other	28.2	26.8	29.4	29.8	30.2	30.6

Appendix Table C: Frequency of Zero-Expenditures with Standard OLS Estimation for All Goods

	HBS	SILC	
	Actual (%)	Estimated (%)	Imputed (%)
Alcohol	29.0	0.7	1.9
Tobacco	66.0	5.2	4.5
Clothing	26.3	0.4	2.8
Health	38.0	2.9	5.8
Public Transport	60.9	4.4	3.0
Private Transport	14.3	0.3	2.7
Education	62.3	10.0	10.3
Durables	10.1	0.2	3.3

Appendix Table D: Distribution of Indirect Taxes with State Average Tax-Category Shares ( $\pi^{in}$ )

<i>Decile</i>	<i>Actual (%)</i>	<i>Estimated (%)</i>	<i>Imputed (%)</i>
Bottom	24.6	17.9	20.7
2	15.3	14.3	15.0
3	14.6	12.3	12.4
4	12.6	11.8	11.4
5	12.1	11.5	11.6
6	11.7	10.8	10.7
7	10.0	10.2	9.8
8	9.7	9.0	9.9
9	9.0	8.3	8.8
Top	6.4	7.0	7.1
State	10.5	9.9	10.1

## References

- Arellano, Manuel and Costas Meghir (1992) "Female Labour Supply and On-the-Job Search: An Empirical Model Estimated Using Complementary Data Sets," *Review of Economic Studies*, Wiley Blackwell, vol. 59(3), pages 537-59, July.
- Barrett, Alan and Caeman Wall (2006) "The distributional impact of Ireland's indirect tax system" Report for the Combat Poverty Agency.
- Bradford, D. F. (1974). Generalisation and proof of the Hicks composite commodity theorem without a utility function. Econometric research program Princeton university research memorandum no.161, Princeton University.
- Brewer, M. and C. O'Dea (2012). "Measuring living standards with income and consumption: evidence from the UK" IFS Working Papers W12/12, Institute for Fiscal Studies.
- Callan, Tim, Claire Keane, Michael Savage and John Walsh (2013) "Distributional Impact of Tax, Welfare and Public Service Pay Policies: Budget 2014 and Budgets 2009-2014" Special Article in the ESRI Quarterly Economic Commentary, Winter 2013.
- Callan, Tim, Sean Lyons, Susan Scott, Richard S.J. Tol and Stefano Verde, (2009) "The distributional implications of a carbon tax in Ireland," *Energy Policy*, Elsevier, vol. 37(2), pages 407-412, February.
- Central Statistics Office (2012), *Household Budget Survey 2009/10 – Volume 1*, Dublin, Stationery Office.
- Collins, Michael (2014) "Total Direct and Indirect Tax Contributions of Households in Ireland: Estimates and Policy Simulations" Working Paper 18, Nevin Economic Research Institute
- Decoster, A., Rock, B. D., Swerdt, K. D., Loughrey, J., O'Donoghue, C. and Verwerft, D. (2007). "Techniques to impute expenditures into an income data set" EUROMOD AIM-AP deliverable 3.4, University of Leuven
- Decoster, A., De Swerdt, K., Loughrey, J., O'Donoghue, C. and Verwerft, D (2009) Imputation of expenditures into the income datasets of five European countries. EUROMOD AIM-AP Deliverable 3.5, University of Leuven.
- Decoster, A., Loughrey, J., O'Donoghue, C. and Verwerft, D. (2010). How regressive are indirect taxes? A microsimulation analysis for five European Countries, *Journal of Policy Analysis and Management* **29**(2): 326 – 350.
- Decoster, André, Jason Loughrey, Cathal O'Donoghue and Dirk Verwerft, (2011) "Microsimulation of indirect taxes" *International Journal of Microsimulation*, International Microsimulation Association, vol. 4(2), pages 41-56.
- Duan, Naihua, et al, (1983) "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business & Economic Statistics*, American Statistical Association, vol. 1(2), pages 115-26, April.
- Figari, Francesco and Alari Paulus (2012) "GINI DP 28: The impact of indirect taxes and imputed rent on inequality: A comparison with cash transfers and direct taxes in five EU countries," GINI Discussion Papers 28, AIAS, Amsterdam Institute for Advanced Labour Studies.
- Keane, Claire, Tim Callan, Michael Savage, John Walsh and Kevin Timoney (2014) "Identifying Policy Impacts in the Crisis: Microsimulation Evidence on Tax and Welfare" *Journal of the Statistical and Social Inquiry Society of Ireland*, Vol. XLII, 2012-13, pp. 1-14

- Leahy, Eimear, Seán Lyons and Richard S. J. Tol (2011) "The Distributional Effects of Value Added Tax in Ireland," *The Economic and Social Review*, Economic and Social Studies, vol. 42(2), pages 213-235.
- Manning, Willard G. (1998) "The logged dependent variable, heteroscedasticity, and the retransformation problem," *Journal of Health Economics*, Elsevier, vol. 17(3), pages 283-295, June.
- Meyer, B. and J. Sullivan (2011) "Further results on measuring the well-being of the poor using income and consumption" *Canadian Journal of Economics* 44(1), 52-87.
- Mirrlees, James, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond and Robert Chote (2011) "Tax by Design" Oxford University Press, Oxford.
- Minh le, Tuan (2003) "Value Added Taxation: Mechanism, Design, and Policy Issues" Paper prepared for the World Bank course on Practical Issues of Tax Policy in Developing Countries, Washington D.C., April 28-May 1, 2003.
- O'Donoghue, Cathal and Baldini, Massimo and Mantovani, Daniela (2004) "Modelling the redistributive impact of indirect taxes in Europe: an application of EUROMOD" EUROMOD Working Papers EM7/01, EUROMOD at the Institute for Social and Economic Research.
- Office for National Statistics (2013) "The effects of taxes and benefits on household income, 2009/10: Further analysis and methodology" Office of National Statistics Publications, Newport, South Wales.
- Pestel, Nico and Eric Sommer (2013). "Shifting Taxes from Labor to Consumption: Efficient, but Regressive?" IZA Discussion Papers 7804, Institute for the Study of Labor (IZA).
- Picos-Sánchez, F. (2011), "Consumption Taxation as an Additional Burden on Labour Income", *OECD Taxation Working Papers*, No. 7, OECD Publishing.
- Revenue Commissioners of Ireland *Statistical Report* (2012) [accessed at <http://www.revenue.ie/en/about/publications/statistical/2012/index.html> on 1/10/14]
- Sah, R. K. (1983). "How much redistribution is possible through commodity taxes?" *Journal of Public Economics* 20(1), 89-101.
- SEAI (2013) *Domestic Fuel Cost Archive*, [accessed at [http://www.seai.ie/Publications/Statistics\\_Publications/Fuel\\_Cost\\_Comparison/Domestic\\_Fuel\\_Cost\\_Archives.pdf](http://www.seai.ie/Publications/Statistics_Publications/Fuel_Cost_Comparison/Domestic_Fuel_Cost_Archives.pdf) on 1/10/14]
- Sutherland, H., Taylor, R. and Gomulka, J. (2002). "Combining household income and expenditure data in policy simulations" *Review of Income and Wealth* 48(4): 517 – 536.
- Weber, Richard and Dominic Weber (2013) "Statistical matching of EU-SILC and the Household Budget Survey to compare poverty estimates using income, expenditures and material deprivation" *Eurostat Working Papers and Methodologies* 2013 Edition ISSN 1977-0375.