LISER
LUXEMBOURG INSTITUTE OF SOCIO-ECONOMIC RESEARCH

n° 2018-01 January 2018

WORKING PAPERS

Accounting for Differences in Income Inequality across Countries: Ireland and the United Kingdom

Denisa **SOLOGON**[1]
Philippe **VAN KERM**[1, 2]
Jinjing **LI**[3]
Cathal **O'DONOGHUE**[4]

[1] LISER, Luxembourg
[2] University of Luxembourg, Luxembourg
[3] University of Canberra, Australia
[4] The National University of Ireland, Ireland

www.liser.lu

Fonds National de la Recherche Luxembourg

Accounting for Differences in Income Inequality across Countries: Ireland and the United Kingdom*

Denisa M. Sologon†, Philippe Van Kerm‡, Jinjing Li§& Cathal O'Donoghue¶

4th January 2018

Abstract

This paper proposes a framework for studying international differences in the distribution of household income. Integrating micro-econometric and micro-simulation approaches in a decomposition analysis it quantifies the role of tax-benefit systems, employment and occupational structures, labour prices and market returns, and demographic composition in accounting for differences in income inequality across countries. Building upon EUROMOD (the European tax-benefit calculator) and its harmonized datasets, the model is portable and can be implemented for any cross-country comparisons within the EU. An application to the UK and Ireland—two countries that have much in common while displaying different levels of inequality—shows that differences in tax-benefit rules between the two countries account for roughly half of the observed difference in disposable household income inequality. Demographic differences play negligible roles. The Irish tax-benefit system is more redistributive than UK's due to a higher tax progressivity and higher average transfer rates. These are largely attributable to policy parameter differences, but also to differences in pre-tax, pre-transfer income distributions.

Keyords: income inequality, decompositions, cross-national comparisons, microsimulation, tax and transfer policy

JEL Codes: D31,H23,J21,J22,J31

*This research is part of the SimDeco project (*Tax-benefit systems, employment structures and cross-country differences in income inequality in Europe: a micro-simulation approach*) supported by the National Research Fund, Luxembourg (grant C13/SC/5937475). Emilia Toczydlowska and Carina Toussaint provided invaluable research assistance. The results presented here use EUROMOD version G2.0+. EUROMOD is maintained, developed and managed by the Institute for Social and Economic Research (ISER) at the University of Essex, in collaboration with national teams from the EU member states. We are indebted to the many people who have contributed to the development of EUROMOD. The process of extending and updating EUROMOD is financially supported by the European Union Programme for Employment and Social Innovation 'Easi' (2014-2020). The results and their interpretation are our responsibility. We are grateful for comments from Francesco Andreoli and Stephen Jenkins and from participants of the International Microsimulation Association conferences 2015 and 2017, the "Inequality in the Labour Market 2015" workshop , APPAM 2016, ECINEQ 2017, EALE 2017, and the SimDeco workshop 2017 on "Understanding international differences in income inequality".

†Corresponding author: denisa.sologon@liser.lu, Luxembourg Institute of Socio-Economic Research

‡philippe.vankerm@liser.lu, Luxembourg Institute of Socio-Economic Research and University of Luxembourg

§jinjing.li@canberra.edu.au, University of Canberra

¶cathal.odonoghue@nuigalway.ie, The National University of Ireland, Galway

1 Introduction

Trends in income inequality since the 1980s have been the subject of considerable attention; see for example the comprehensive review of thirty countries experiences in Nolan et al. (2014). Examinations of cross-national differences in inequality and, especially, of the driving forces behind those differences are by comparison much less common (Förster and Tóth, 2015). Yet, variations in income inequality levels across countries tend to be more striking than changes observed within any rich country in recent years. For example, according to OECD (2011), the biggest increase in the Gini coefficient of income between 1985 and 2008 among 22 OECD countries—a change of 0.07 observed in Sweden and in New Zealand—is only half the difference of 0.13 observed between the Gini coefficients for Denmark and the USA in 2008. Among EU countries, the Gini coefficient of income currently ranges from 0.24 in the Slovak Republic to 0.37 in Bulgaria, Romania or Lithuania (Eurostat, 2017), a gap larger than any trends recently observed in the EU.

By proposing a generic framework for studying international differences in the distribution of household income and an analysis of the UK and Ireland, this paper makes a contribution to a literature that has been suprisingly small in recent years. Integrating micro-econometric and micro-simulation approaches in a decomposition analysis, the objective is to quantify the contribution of four main potential drivers to inequality differences between countries: differences in tax-benefit systems, differences in employment and occupational structures, differences in labour prices and market returns, and differences in demographic composition—four factors identified as part of the grand drivers of inequality in Förster and Tóth (2015).

Even though Brandolini and Smeeding (2010, p.97) once remarked that "… attempts to model and understand causal factors and explanations for differences in level and trends in income inequality across nations is the ultimate challenge to which researchers on inequality should all aspire", the overwhelming majority of recent research has focused on examination of the determinants of trends in inequality within countries rather than on the sources of cross-country differences in the *level* of inequality; see for example Belfield et al. (2017); Biewen and Juhasz (2012); Brewer and Wren-Lewis (2016); Daly and Valletta (2006); Herault and Azpitarte (2016); Hyslop and Mare (2005); Jenkins et al. (2013); Larrimore (2014) to mention only recent studies that examined the distribution of household income.[1] Although factors that drive changes in inequality may also explain why inequality differs across countries, countries differ considerably with respect to their tax-benefit systems, labour market and social institutions, market structures, and demographic factors, and with respect to social norms and behaviours, culture and history, etc. (Alesina and Glaeser, 2005; Haveman et al., 2011). Direct examination of the drivers behind cross-national inequality difference remain indispensible.

A body of literature has used cross-country regressions to tease out the importance of various economic and institutional variables on inequality. Aggregate indicators of inequality for a range of countries and years are regressed on macro-level economic, political or institutional variables as potential explanatory factors. Notably, empirically testing the relationship between inequality and GDP—the Kuznets hypothesis—has been central to this literature. But inequality determinants go well beyond economic growth. The comprehensive review of recent studies exploiting cross-country regressions in Förster and Tóth (2015) however leads to the disappointing conclusion that "inconclusiveness prevails for many possible drivers of inequality (...) which can often but

[1] An even larger body of literature has examined the trends in earnings and wage inequality (see, e.g., Atkinson, 2007), while the distribution of wealth has recently received growing attention especially since Piketty (2013).

not always be traced back to different country samples, time periods, data and methodological specifications" (p.1804).

Decomposition methods have been the main alternative to cross-country regressions. In their simplest form, inequality decomposition methods determine the contributions of a small number of components—particular sources of income (Lerman and Yitzhaki, 1985; Shorrocks, 1982) or particular partitions of the population (Shorrocks, 1980)—which add up to total inequality in a country. This naturally leads to comparisons of the composition of inequality across countries or over time as a way to 'explain' inequality differences (see, e.g., Brewer and Wren-Lewis, 2016). This approach is however limited to particular inequality measures and makes it difficult to examine multiple factors simultaneously. More recently, flexible decomposition approaches have modelled the full income distribution (rather than specific summary indices) and jointly examine several determinants. Typically, the contribution of a number of factors to the differences in inequality is assessed using (a sequence of) simulated counterfactual distributions of household disposable incomes that would prevail in each country (or time period), if these factors were common to different countries (or years).[2] This is the approach we follow here. It should be clear that this is not trying to generate the income distribution that would realistically be observed if one was to exogenously change the components of interest and then let households, policy-makers and the economy adjust to the change in the long run. Instead, the magnitude of the model's response to the simulated transformation is used to quantify the relative contribution of each of a number of factors of interest to the aggregate difference between two populations. While this has become a popular approach for examinations of changes in inequality, only very few studies have attempted such a decomposition in a cross-country analysis.

The present paper builds upon the approach developed in Bourguignon et al.'s (2008) study of the determinants of inequality difference between Brazil and the USA. The procedure relies on a parametric representation of the link between the components of household income (individual earnings and unearned income) and household or individual socio-demographic characteristics, complemented by a non-parametric reweighting technique to account for demographic profiles. This approach extends the ubiquitous Oaxaca-Blinder and Juhn-Murphy-Pierce decompositions in two ways: first it deals with the entire income distribution, not just mean earnings, and second it builds a parametric income-generation process based on a system of equations for multiple income sources for the household, not just a parametric earnings process for individual wages.

While our "income generation model" bears resemblance with Bourguignon et al.'s (2008), our implementation differs in several dimensions. The most important difference is the treatment of taxes and benefits. Unlike Bourguignon et al. (2008), we focus on household disposable income after taxes and transfers and explicitly study how much differences in tax-benefit systems across countries account for inequality differences. This is critical not least because policy design and parameters can be modified by government decisions, unlike demographic factors or labour market structure and returns. To obtain the most realistic contribution of taxes and benefits to household income, we incorporate tax-benefit rules by means of the pan-European tax-benefit micro-simulation engine EUROMOD (Sutherland and Figari, 2013). This allows us to represent accurately the relationship between household characteristics, market incomes (from labour and capital) and taxes and benefits. As we show in the comparison of the UK and Ireland, the tax-benefit policy differences turn out to be the main force behind the higher income inequality observed in the UK. The second main development over Bourguignon et al.'s (2008) framework is the introduction of endogenous

[2]See Fortin et al. (2011) for a review of methods.

labour supply adjustments in the generation of simulated counterfactual distributions. While the approach remains descriptive by nature, the apparatus offers sufficient sophistication to allow detailed analysis of the way tax-benefit systems can interact with labour market structures, labour prices and returns to capital, and demographics in determining the distribution of household disposable income.

Such microeconometric decomposition approaches are sometimes dismissed as being too difficult to construct to be of general practical use (Cowell and Fiorio, 2011). We address this concern by developing a framework that is portable across all European countries. The model is constructed on the basis of household survey data that are available in harmonized form in all EU countries. Using cross-country comparable data, the income distribution model has a common specification for each country so as to permit the simulation of counterfactual distributions holding components constant across countries. Also, by exploiting EUROMOD, the heterogeneity of tax-benefit rules is easily incorporated. This means that the model can be used to examine inequality differences across any pairs of European countries using a common analytic framework at a minimal development cost.

For the purpose of demonstrating the potential of the new framework, we undertake a comparison of Ireland and the UK. Ireland and UK have always formed a common travel area and labour market, they share a language and border and their Welfare States have evolved contemporaneously, influenced by similar drivers and political philosophy principles. Nevertheless there are also sufficient differences to call for examination of the factors that have resulted in different levels of inequality. We examine the year 2007, which is the latest year before the financial and economic crisis hit both countries. The Gini coefficient was 0.28 in Ireland—a relatively low figure by international standards—while it had reached 0.32 in the UK. Our results show that the direct effect of the differences in tax-benefit rules between the two countries accounts for roughly half of the observed difference in income inequality. The Irish tax-benefit system is more redistributive than the UK system due to a higher tax progressivity and more generous average transfer rates. These are largely attributable to differences in policy parameters, but also to differences in pre-tax, pre-transfer (market) income distributions.

The paper is organized as follows. Sections 2 and 3 present our methodology. Section 2 describes our representation of the "income generation process": we define the components of household income that we examine and we describe the parametric specifications to capture how these income sources vary with observed and unobserved individual characteristics. Section 3 proposes four generic "transformations" of income generation processes and shows how these transformations applied to estimates of income generation processes for different countries can be used to account for cross-country inequality differences. Our illustrative application to the UK–Ireland contrast is presented in Section 4. Section 5 concludes.

2 A representation of the household disposable income generation process

The core of the decomposition exercise is a representation of household incomes on the basis of (i) a set of basic observable characteristics (demographic characteristics (size, age and gender) and education level of adults), (ii) a vector of "parameters" describing how the receipt and level of income sources vary with household and individual characteristics, and (iii) a vector of household-specific 'residuals' linking predictions from model parameters to observed income sources. In a basic unidimensional example—say for looking at individual earnings—this corresponds to a Mincerian regression $y_h = x_h\beta + u_h$, and the three components would be (i) x_h the household characteristics, (ii) β the "parameter vector" and (iii) u_h the household's idiosyncratic residual. This is the set up used in Juhn et al. (1993) which we extend to a multivariate model. Each of the three factors can drive inequality of incomes between households and account for cross-national differences in inequality. Differences in demographic and education characteristics reflect basic, observable population heterogeneity; parameter vectors capture how much differences in such basic characteristics lead to differences in (various sources of) income; residuals capture both the magnitude of income differences between households with identical characteristics and the correlation between different sources of incomes (some are substitutes—benefits and replacement income tend to substitute labour incomes—while others can be complementary—think of capital and labour incomes).

A simple univariate regression model is a poor representation of total household incomes and recent research has modelled the bigger complexity of the household income generation process. Building upon Bourguignon et al. (2008), we develop and estimate a detailed income generation process with hierarchically structured, multiple equation specifications for detailed sources of income.

We first detail in Section 2.1 the different components of income for which we estimate a specific parametric model. We then describe in Section 2.2 the details of those parametric specifications.

2.1 Household disposable income components

We examine five main constituents of total household disposable income y_h, namely (gross) labour incomes, capital incomes, other non-benefit pre-tax incomes (e.g., private pensions, alimonies), public transfers and (minus) direct taxes:

$$y_h = y_h^L + y_h^K + y_h^O + y_h^B - t_h. \tag{1}$$

One can broadly think of the first three sources as returns to human capital, returns to physical or financial capital, and other private (or market) incomes. The last two reflect the direct intervention of the State through transfers and taxes.

Most of these five sources are themselves aggregates of smaller components of income—notably contributions of individuals to overall household income—which we are going to model separately in order to have a parametric representation that is defined at a fine level of disaggregation. The details of our disaggregation are as follows. Labour income is the sum of employee and self-employment incomes of each household member:

$$y_h^L = \sum_{i=1}^{n_h} I_{hi}^{lab} \left(I_{hi}^{emp} y_{hi}^{emp} + I_{hi}^{se} y_{hi}^{se} \right) \tag{2}$$

where, in all expressions, I_j^S refers to a binary indicator equal to 1 if person (or household) j receives any income from source S (and 0 otherwise) whereas y_j^S refers to the actual amount of income source S received. So $I_{hi}^{lab} = 1 - (1 - I_{hi}^{emp})(1 - I_{hi}^{se})$ identifes whether person hi receives any labour income and I_{hi}^{emp} and I_{hi}^{se} identify whether she receives income from salaried employment and from self-employment. (Although the use of binary indicators is not required at this stage, their use will become clear below.) Capital income is the sum of investment income and property income received by each household member:

$$y_h^K = \sum_{i=1}^{n_h} \left(I_{hi}^{inv} y_{hi}^{inv} + I_{hi}^{prop} y_{hi}^{prop} \right).$$

The ‘other incomes’ component include private pension payments and a residual, catch-all measure that aggregates all non-benefit individual incomes that are not included in labour and capital incomes (mainly private transfers such as alimonies),

$$y_h^O = \sum_{i=1}^{n_h} \left(I_{hi}^{pripen} y_{hi}^{pripen} + I_{hi}^{other} y_{hi}^{other} \right).$$

The sum $y_h^L + y_h^K + y_h^O$ is the total pre-transfer, pre-tax income of household h (or what we will also call ‘market income’ of household h). We then add public transfers y_h^B and deduct taxes t_h to arrive at disposable income. Public transfers are composed of a range of individual replacement incomes (pension, survivor pension, disability, sickness and unemployment benefits), of household-level means-tested social assistance (including housing support) and of universal transfers to which household h is eligible (including child support). For simplicity we will refer to three broad household-level aggregates—public pensions, means-tested benefits and non-means-tested benefits:

$$y_h^B = y_h^{pens} + y_h^{mtb} + y_h^{nmtb}.$$

The fifth and final component is the level of direct taxes and social contributions paid by household h. Direct taxes and social contributions are determined by the tax schedule in place as a function of the vector of gross incomes and household characteristics and composition:

$$t_h = tax_h + \sum_{i=1}^{n_h} ssc_{hi}.$$

For reference, Table A–2 in Appendix A summarizes all separate income components included in our household income decomposition structure.

2.2 Parametric specifications

Now that we have identified the five main sources of household income and their components, we specify parametric relationships between observed household characteristics and each of the components. In order to handle the mix of individual-level income sources and household-level sources, we define the parametric relationship at the most disaggregate level described above. As Bourguignon et al. (2008) do, we give special treatment to labour income in order to take into account the role of the occupational and industrial structure of employment in determining the labour market returns to individual characteristics. We use reduced-form log-linear specifications for most pre-tax income components (expressing the probability of receipt and the level of income as separate equations). We then combine parametric specifications with a tax-benefit micro-simulation engine, EUROMOD, for the determination of taxes and benefits in order to represent accurately the role of tax-benefit parameters in the determination of disposable income. This leads to an overall strategy that combines ‘micro-econometric’ and ‘micro-simulation’ approaches.

2.2.1 Labour incomes

Labour income is the central component of household income. Total labour incomes depend on households' labour supply (participation and hours worked) as well as the average wage rates paid to household members. The wage rate varies with individual human capital characteristics but also depends on the type of occupations held, as well as the sector of employment and the industry. To capture the respective impacts of those factors, the labour income of each adult member of a household is modelled individually and we parameterise labour incomes on the basis of a set of nested equations. Other sources of (pre-tax) incomes are treated in a more conventional way.

The first equation captures the probability to be at work and to have any labour income conditional on individual characteristics; this is the indicator I_{hi}^{lab} for person i in household h appearing in equation (2). Binary outcomes are represented with logistic models (see, e.g., Agresti, 2010). Assuming a latent variable representation $I_{hi}^{s*} = x_{hi}\gamma^s + \epsilon_{hi}^s$, where $I_{hi}^s = 1$ if $I_{hi}^{s*} > 0$ and $I_{hi}^s = 0$ otherwise, the participation indicator is

$$I_{hi}^s = \mathbf{1}\left[\epsilon_{hi}^s > -x_{hi}\gamma^s\right]$$

where $\mathbf{1}\left[cond\right]$ is equal to one if *cond* is true and 0 otherwise. The logistic regression model assumes that ϵ_{hi}^s is distributed logistic so that

$$\Pr(I_{hi}^s = 1|x_{hi}) = \Pr(x_{hi}\gamma^s + \epsilon_{hi}^s > 0) = \Pr(-\epsilon_{hi} < x_{hi}\gamma^s) = \frac{\exp(x_{hi}\gamma^s)}{1+\exp(x_{hi}\gamma^s)}.$$

So, with S being the labour income component, the participation indicator for person hi is completely determined by her individual characteristics x_{hi} and her random residual ϵ_{hi}^{lab} given model parameters γ^{lab}. The characteristics that we include in x_{hi} are the person's age (and age squared), academic achievement (whether holds a university degree), marital status, number of own children in the household (separating children under 4, children between 4 and 11 and children between 12 and 15), and citizenship. Separate sets of parameters are allowed for men, single women, and women in couple.

A second equation captures the probability to earn income from salaried employment versus self-employment, conditional on being at work. For simplicity, we treat salaried employment and self-employment as mutually exclusive, so $I_{hi}^{se} = 1 - I_{hi}^{emp}$. For workers reporting both sources of income, we take the salaried employment status and treat all incomes as employee income. We parameterize the indicator with a logistic model as above and the salaried vs. self-employed indicators are thus determined by x_{hi} and a random residual ϵ_{hi}^{emp} given model parameters γ^{emp}.

The earnings of self-employed workers are described by a log-linear regression model:

$$y_{hi}^{se} = \exp(x_{hi}\beta^{se} + \upsilon_{hi}^{se})$$

where υ_{hi}^{se} is a zero-mean residual with homoscedastic variance $\sigma^{2,se}$.[3]

For people in salaried employment ($I_{hi}^{emp} = 1$), we further model occupation, sector and industry of their main job before specifying the earnings equation. Occupation is first represented by an 8-categoires 1-digit ISCO code. We use multinomial logistic models for occupation (and industry), as Bourguignon et al. (2008) do. A latent variable $I_{hi}^{k,occ*} = x_{hi}\delta^{k,occ} + \epsilon_{hi}^{k,occ}$ is associated

[3] We treat the variance of the residuals as part of the parameter vector. Note that the variance is set to unity in the logistic regression models.

to each of the $k \in m^{occ}$ alternative occupations with $\epsilon_{hi}^{k,occ}$ following an extreme value distribution. The observed occupation for person hi, say j ($I_{hi}^{j,occ} = 1$ and $I_{hi}^{k,occ} = 0$ for $k \neq j$), is such that $I_{hi}^{j,occ*} > I_{hi}^{k,occ*}$. Under extreme value distribution for the residuals, the probability of occupation j is

$$\Pr(I_{hi}^{j,occ} = 1|x_{hi}) = \frac{\exp(x_{hi}\delta^{j,occ})}{\sum_{k=1}^{m}\exp(x_{hi}\delta^{k,occ})}$$

with the parameter vector for the first alternative normalized to $\delta^{1,occ} = 0$. This is equivalent to the binary logistic probabilities when only two choices are available. So, occupation is represented by x_{hi} and a set of m^{occ} individual-specific extreme value distributed residuals $\epsilon_{hi}^{k,occ}$, given parameter vector δ^{occ} determining the probability distribution of potential occupations.

Industry of employment can be primary, secondary, or tertiary and is similarly modelled with a multinomial logistic model and $m^{ind} = 3$. Sector of employment is either public or private (public sector includes public administration jobs but also army, health and education). Public sector employment is parameterized by a binary logistic model. Note that we add occupation as a conditioning variable in the models for industry and public sector employment, which are thus determined by (x_{hi}, occ_{hi}) and residuals $\epsilon_{hi}^{k;ind}$ and ϵ_{hi}^{pub} given parameter vectors δ^{ind} and δ^{pub}.

Income from salaried employment is then given by

$$y_{hi}^{emp} = w_{hi}s_{hi}$$

where w_{hi} is the average (hourly) wage rate for person hi and s_{hi} is her total number of hours of employment. We project s_{hi} onto x_{hi} using a linear model

$$s_{hi} = x_{hi}\gamma^{hrs} + \epsilon_{hi}^{hrs}.$$

The final step in the parameterization of labour incomes is a specification for wages w_{hi}. Given the central importance of wages in the distribution of household income, we adopt a parametric specification that connects individual characteristics to the whole conditional wage distribution and not only to conditional means as in the OLS regressions used for other sources of income. To do so, we assume that wages follow a Singh-Maddala distribution F_X

$$F_{X=z}(w) = \mathrm{SM}(w; a(z), b(z), q(z)) = 1 - \left[1 + \left(\frac{w}{b(z)}\right)^{a(z)}\right]^{-q(z)}$$

where the X indicate that the distribution is conditional on a vector of characteristics z. The Singh-Maddala distribution is a flexible unimodal three-parameter distribution that has been shown to provide good fit to wage distributions (Van Kerm et al., 2016). The parameter $q(z)$ is a shape parameter for the 'upper tail', $a(z)$ is a shape parameter affecting both tails of the distribution ('spread'), and $b(z)$ is a scale parameter. Each of these parameters is allowed to vary log-linearly with individual characteristics $\theta(z) = \exp(z\beta^{\theta,emp})$, as in Biewen and Jenkins (2005) or Van Kerm (2013). Individual wage is then given by

$$w_{hi} = F_{X=z}^{-1}(v_{hi}^{emp}) = b(z)[(1 - v_{hi}^{emp})^{-\frac{1}{q(z)}} - 1]^{\frac{1}{a(z)}}$$

where v_{hi}^{emp} is a random term uniformly distributed and the conditioning variables z contain both x_{hi} and occupation, industry and sector of occupation, $z = (x_{hi}, occ_{hi}, ind_{hi}, pub_{hi})$.

The parameterisation of wages closes the model for labour incomes. To summarize: household labour income y_h^L is ultimately characterised by all adult members' characteristics x_{hi} and residual heterogeneity terms $(\epsilon_{hi}^{lab}, \epsilon_{hi}^{emp}, \epsilon_{hi}^{k,occ}, \epsilon_{hi}^{k,ind}, \epsilon_{hi}^{pub}, \epsilon_{hi}^{hrs}, \upsilon_{hi}^{se}, \upsilon_{hi}^{emp})$ for $i = 1, \ldots, n_h$ given the model parameters $(\gamma^{lab}, \gamma^{emp}, \delta^{\text{occ}}, \delta^{ind}, \delta^{pub}, (\beta^{se}, \sigma^{se}), \gamma^{hrs}, (\beta^{a,emp}, \beta^{b,emp}, \beta^{q,emp}))$. Appendix A provides details of the model components in tabular format.

2.2.2 Other market incomes

We adopt much simpler parameterizations for all other market incomes that compose capital income (y_h^K) and other pre-tax, pre-transfer incomes (y_h^O). As for labour incomes, each aggregate is the sum of the contributions of individual members hi. For any of those four sources, say S, the probability of receiving any income from the source S is modeled with a binary logistic regression described above, again conditioning on individual exogenous variables x_{hi}: $I_{hi}^s = \mathbf{1}\left[\epsilon_{hi}^s > -x_{hi}\gamma^s\right]$. The amount received is log-linearly related to x_{hi}: $y_{hi}^S = \exp(x_{hi}\beta^S + \upsilon_{hi}^S)$ with $Var(\upsilon_{hi}^S) = \sigma^{2,S}$. So both capital incomes and other incomes are determined by all household members' characteristics x_{hi} and two residual heterogeneity terms and two sets of parameters $(\gamma^S, (\beta^S, \sigma^S))$.

2.2.3 Public transfers

The final two components of household income are the public transfers received and the income tax paid. Since many transfers and taxes are directly determined by the amount of all incomes received, they cannot reasonably be calculated and simulated independently. Instead, we derive taxes and a range of benefits on the basis of a tax-benefit calculator that can estimate the amount of benefits received and income tax paid (or, at least, due) as a function of the income sources, household characteristics and a number of variables which may influence the benefit eligibility and tax liabilities according to the rules in place.

Many such calculators exist for different countries, but for cross-country comparative analysis, it is important to rely on an engine that models taxes and benefits in a consistent way across countries. For European countries, harmonized taxes and benefit calculations can be taken from EUROMOD, a large-scale pan-European tax-benefit static micro-simulation engine (Sutherland and Figari, 2013). This large-scale income calculator incorporates the tax-benefit schemes of the majority of European countries and allows computation of predicted household disposable income, on the basis of pre-tax, pre-benefit incomes, employment and other household characteristics. It also makes it possible to implement 'policy swaps' in which particular tax or benefit policies from one reference country or year are applied to other countries or time periods (see, e.g., Bargain, 2012; Bargain and Callan, 2010; Levy et al., 2007). EUROMOD simulates direct taxes and a wide range of cash transfers to households: income and property taxes, social insurance contributions, family benefits, housing benefits, social assistance, and, where relevant other income-related benefits (Figari et al., 2015).

Not all public transfers are evaluated by EUROMOD. Two main sources of public transfers are not simulated (or are only partially simulated): contributory benefits and public pensions as well as disability benefits which generally depend on past employment histories or other information (e.g., about the severity of a disability) that is usually not observed in household survey data that input the tax-benefit simulator. For those components included in y_h^B, the benefits measured at individual level are modelled like non-labour incomes with a logistic regression for receipt of the source and a log-linear specification for the amount received while benefits measured at the household level are modelled similarly except that only one household level equation is specified for each model and the exogenous characteristics x_h are composed of household-level demographic composition and of the individual characteristics of the 'household head' (where household head is defined as the person with the highest personal income or the eldest in the case of equal income).

The remaining components of y_h^B are calculated from EUROMOD on the basis of household characteristics—typically universal transfers—and, for means-tested benefits, on the basis of pre-tax household income.[4]

2.2.4 Taxes and social security contributions

Finally, we rely entirely on EUROMOD to calculate direct taxes and social security contributions as a function of all income components and household characteristics. EUROMOD takes as input the household members' 'exogenous' characteristics, as well as all income components previously modelled. It returns social security contributions and total taxes due by the household given the tax-benefit parameters in place in the country considered.[5]

2.3 Estimation of parameters

Following Bourguignon et al. (2008), each equation of the model—logistic, log-linear or Singh-Maddala—is estimated independently using standard estimators (OLS or maximum likelihood) to derive estimates of the model parameters. We do not attempt to model selectivity in the income equations. The only exception is for the coefficients of the Singh-Maddala model for wages. The model is estimated for men and women separately. For women, we estimate a participation-corrected model as proposed in Van Kerm (2013). Given the size of the model, we do not attempt to estimate equations jointly or model correlations across equations parametrically. Of course, model parameters are not meant to capture causal relationships between the various 'endogenous' variables y_{hi}^S and I_{hi}^S and the few 'exogenous' variables x_{hi}. The parametric relationships are reduced-form projections that aim to describe the empirical associations between basic demographic variables and various components of income.

3 Counterfactual distributions and the decomposition of cross-country inequality differences

The household income generation representation is now used to describe the overall household income distribution and to create counterfactual distributions. We are interested in studying the distribution F of the random variable Y which represents household disposable income among individuals in the population. More specifically, we aim to study some summary index measure

[4]In effect some sources of benefits are of mixed type. Eligibility can be determined from household characteristics (and therefore modelled with a logistic regression) while the level of transfer is calculated by the tax-benefit parameters. Appendix B provides details on the sources and treatment of public transfers.

[5]A few additional variables that are not part of household income can influence household taxes and benefits (at least in some countries). These variables—for example mortgages, rents paid, and contributions for private pensions — are also modelled to calculate liabilities. For instance, for those who are out of the labour market, we model the propensity of being unemployed and being formally retired. The exact status for those out of the labour market may have an impact on the final household disposable income via the tax and benefit system. Each of these variables is modelled as the 'other income sources' described above: the existence of such payment is according to a logit regression model, and the amount paid is based on a log-linear regression model. The values taken by those variables do not determine household income directly, but they are fed into the tax-benefit microsimulation engine to calculate taxes and benefits for household h. Appendix B describes those auxiliary variables.

$\theta(F)$—say, the Gini coefficient—and to examine why this index differs from the index observed in another country or another time period $\theta(G)$.

At this stage, it is convenient to think of our representation of the income generation process as a generic non-separable model

$$Y = m(X, \Upsilon) \tag{3}$$

where Y is income, X is a vector of 'exogenous' characteristics and Υ is a vector of unobserved heterogeneity (residual) terms (Matzkin, 2003; Rothe, 2010). The function m describes jointly the relationship between household characteristics and income and the heterogeneity in Y that is not 'explained' by X. The derivative of m with respect to its first argument reflects variations in Y across households that can be attributed to differences in observable household characteristics while the derivative of m with respect to its second argument reflects variations in Y across households of identical observable characteristics.

The parametric functional forms adopted for the different components of our income generation model imply a particular parametric shape for m, so

$$Y = m^{\xi}(X, \Upsilon; \xi) \tag{4}$$

where m^{ξ} represents the specific parametric structure adopted for the income generation model and ξ is the vector of parameter values. Equation (4) has no 'structural' interpretation but it should be viewed as a set of reduced form equations linking household characteristics and income—a relationship that may arise from an unknown, broader structural model—through earnings functions, equations for employment and occupational and industrial structure, equations for non-labour income and replacement incomes and through tax-and-transfer rules. The distribution function F—and therefore any functional of interest $\theta(F)$—depends on the (joint) distribution function of X and Υ in the population through m^{ξ} and ξ.

In this model, the distribution of income in two countries can differ because of differences in the distribution of X, in the distribution of residual heterogeneity terms Υ and differences in m. To make progress, we assume that all countries can be represented by a common parametric model of the form m^{ξ} but that countries differ in the values taken by the parameters ξ. In order to quantify the relative contributions of these factors, we define a number of 'transformations' that, when applied to the model, allow us to capture how sensitive the income distribution is to specific dimensions of the model. The transformations are then calibrated to reflect actual differences across countries in the factors concerned and lead to a decomposition of cross-country differences in income distributions into specific factors of interest.

3.1 Four transformations of the income generation process

We focus on four types of 'transformations' that will help us capture the relative contributions of four broad factors (or subsets thereof): (i) a *demographic transformation*, (ii) a *labour market structure transformation*, (iii) a *price-and-returns transformation* and (iv) a *tax-benefit transformation*. These transformations follow from the construction of the income generation process, although these are specific choices among many other possibilities offered by the model.

The *demographic transformation* involves modification of the distribution of the random variables X:

$$m(\tilde{X}(X), \Upsilon; \xi)$$

which lead to a new, counterfactual distribution of outcome Y denoted F^d. The impact of a demographic transformation, $m^{\xi}(\tilde{X}(X), \Upsilon; \xi) - m^{\xi}(X, \Upsilon; \xi)$, on distribution functionals of interest θ is then given by $\Delta^d_\theta(F) = \theta(F^d) - \theta(F)$. This measure is called a 'partial distributional policy effect' in Rothe (2012), or simply a 'policy effect' in Firpo et al. (2009).

A *labour market structure transformation* works through the parameter vector ξ. The transformation involves modifying the parameters of the equations characterising the employment probabilities and hours worked (γ^{lab}, γ^{emp}, γ^{hrs}), and the occupational and industrial structure (δ^{occ}, δ^{ind}, δ^{pub}),

$$m^{\xi}(X, \Upsilon; \tilde{l}(\xi))$$

which, just like the demographic transformation, leads to a new counterfactual distribution of outcome Y denoted F^l. The impact of the labour market structure transformation $m^{\xi}(X, \Upsilon; \tilde{l}(\xi)) - m^{\xi}(X, \Upsilon; \xi)$ on distribution functionals θ is given by $\Delta^l_\theta(F) = \theta(F^l) - \theta(F)$.

A *price-and returns transformation* again acts through the parameter vector ξ. The transformation involves changing the parameters of the equations characterising the level of earnings ((β^{se}, σ^{se}), ($\beta^{a,emp}, \beta^{b,emp}, \beta^{q,emp}$)) and all other pre-tax incomes (($\beta^{inv}, \sigma^{inv}$), ($\beta^{prop}, \sigma^{prop}$), ($\beta^{pripen}, \sigma^{pripen}$), ($\beta^{other}, \sigma^{other}$)),

$$m^{\xi}(X, \Upsilon; \tilde{r}(\xi))$$

and the impact on θ is denoted by $\Delta^r_\theta(F) = \theta(F^r) - \theta(F)$. Observe that this transformation is analogous, albeit in a multiple equations setup, to the manipulation of the vector of coefficients of Mincerian earnings regressions in order to capture 'price' effects (as distinct from 'composition effects) in traditional Oaxaca-Blinder decomposition exercises. It closely resembles the decomposition of Juhn et al. (1993) in the way residual variances is accounted for.

The fourth and last form of transformation that we use is a *tax-benefit transformation*. The tax-benefit transformation is a particular transformation of the parameter vector ξ which modifies (i) the regression parameters determining the level of public transfers received by households and (ii) the parameters of the EUROMOD tax-benefit calculator which evaluate the tax liabilities (and the residual benefits, mostly universal benefits, which are determined directly by EUROMOD and not modelled parametrically; see Appendix B for details):

$$m^{\xi}(X, \Upsilon; \tilde{tb}(\xi)).$$

As above, we write the effect of a tax-benefit transformation on θ as $\Delta^{tb}_\theta(F) = \theta(F^{tb}) - \theta(F)$ where F^{tb} denotes the distribution function of household incomes after the tax benefit transformation is applied to the income-generation model.

3.2 Cross-country comparisons

In an analysis of cross-country differences in income distributions, a natural way to use the transformations just defined is to build the income generation model for countries A and B separately and to calibrate transformations so as to 'transplant' components of the income generation model across countries. For example, for the labour market structure transformation applied to country A, $\tilde{l}(\xi)$ is composed of the subset of parameters from ξ^A for the fixed parameters and of parameters taken from ξ^B for the transformed parameters which capture the characteristics of the labour market structure in the income generation model. The transformed income generation process for country A thereby leads to a simulated distribution for country A as if it had a labour market structure as country B and all other components of the model remained unchanged. The difference between the simulated and the observed distributions in country A (and inequality functionals defined over them) provides a quantification of the contribution of labour market structure differences to the overall difference in income distribution between the two countries. Once again this procedure is fully analogous to standard Oaxaca-Blinder decompositions—swapping regression coefficients across earnings equations for alternative groups—although it is implemented in a multiple equations model.

The transplantation of country B parameters onto country A's model is done similarly for the price-and-returns transformation by swapping the relevant subset of parameters. Unlike the labour market structure transformation, the price-and-returns transformation involves swapping variance terms, σ^2. This is achieved as in Juhn et al. (1993) by rescaling the residuals of country A by the ratio $\frac{\sigma^S_B}{\sigma^S_A}$ for each of the five components S that are affected by the transformation. This procedure scales the distribution of residual terms while preserving the rank correlation of the residuals across the different equations of the income generation model.

The calibration of the tax-benefit transformation combines both swapping model parameters as above (for the equations describing benefits) and using the EUROMOD tax-benefit calculator to apply the tax and benefit rules and parameters of country B onto the market incomes and household characteristics of country A. Such transplantation of tax-policy rules and parameters is most often done for analysis of *trends* in income distributions (see Bargain, 2012; Bargain and Callan, 2010; Herault and Azpitarte, 2016; Paulus and Tasseva, 2017), but it has also been applied to cross-country analysis (Levy et al., 2007).[6] The underpinnings of transplantation of tax-and-transfer rules across countries are discussed in Dardanoni and Lambert (2002).

Finally, the demographic transformation involves modifying the distribution of population characteristics of country A in such a way that it has the (joint) distribution of country B. The distribution of X is modified but the *conditional* distribution of Υ given X must not be affected to remain as it is in A. As shown in DiNardo et al. (1996) and Barsky et al. (2002), this can be achieved semi-parametrically by reweighting. In evaluating F or $\theta(F)$, population A households are reweighted by a factor

$$\omega(X) = \frac{\Pr(X|B)}{\Pr(X|A)} = \frac{\Pr(B|X)}{\Pr(A|X)} \frac{\Pr(A)}{\Pr(B)}. \tag{5}$$

The probabilities in (5) can be estimated by standard techniques for binary responses; see e.g., Biewen and Juhasz (2012) for a recent application of this approach.

3.3 Decomposition

A decomposition procedure aims to (additively) decompose the total difference $\Delta_\theta(F^A, F^B) = \theta(G) - \theta(F)$ into a number of factors that capture the contribution of different components of the model:

$$\Delta_\theta(F^A, F^B) = \sum_{k=1}^{K} \Delta^k_\theta(F^A, F^B).$$

[6] In an intertemporal context, studies usually attempt to disentangle the contribution of structural changes in policies from the mere uprating of policy parameters defined in nominal monetary units (Bargain, 2012; Figari et al., 2015). The issue is less relevant in the present context and we swap both structural differences and nominal parameters at once. For the conversion of nominal parameters across countries, all monetary units expressed in different currencies are converted on the basis of exchange rates in the year concerned.

A common way to build such a decomposition is to build a sequence of counterfactual distributions by applying each of the four transformations one after the other from the original distribution, say F^A, to the target distribution F^B and to define the components of the decomposition as

$$\Delta_\theta^k(F^A, F^B) = \theta(F^{A,B(k)}) - \theta(F^{A,B(k-1)})$$

where $F^{A,B(k)}$ is a counterfactual distribution obtained by composing k transformations of the income generation model for country A calibrated to the structure of country B (and we define $F^{A,B(0)} = F^A$ and $F^{A,B(K)} = F^B$). Note that the last factor K is a 'residual' (or 'unexplained') factor that is not modelled and transplanted explicitly but that collects all residual difference between the target distribution F^B and the counterfactual distribution obtained after all frou transformations have been composed and applied to the income generation model for country A.[7] The drawback of such a sequential decomposition is its path-dependence, the dependence on the sequence of composition of transformations in quantifying the contribution of each factor.[8]

To reduce issues of path-dependence we prefer to examine 'direct effects' to assess the impact of each factor from the same initial benchmark distribution : $D_\theta^k(F^A, F^B) = \theta(F_A^k) - \theta(F^A)$ where F_A^k is the counterfactual distribution obtained by applying one particular transformation k and avoid composing transformations. As Biewen and Juhasz (2012) argue, comparing 'direct effects' is a natural way to compare the effects of alternative transformations. However the sum of direct effects does not add up to the overall inequality change. A decomposition can be expressed as

$$\begin{aligned}\Delta_\theta(F^A, F^B) &= D_\theta^d(F^A, F^B) + D_\theta^l(F^A, F^B) + D_\theta^r(F^A, F^B) + D_\theta^{tb}(F^A, F^B) \\ &\quad + I_\theta(F^A, F^B) + R_\theta^\Upsilon(F^A, F^B)\end{aligned}$$

where (i) the term $I_\theta(F^A, F^B) = \left(\theta(F_A^{tb,r,l,d}) - \theta(F^A)\right) - \left(\sum_{k\in\{d,r,l,tb\}} D_\theta^k(F^A, F^B)\right)$ is equal to the difference between the combined effect of the four transformations composed and the sum of direct effects which captures all two-way and three-way interactions between the four components in the model (Biewen, 2014), and (ii) the residual difference $R_\theta^\Upsilon(F^A, F^B) = \theta(F^B) - \theta(F_A^{tb,r,l,d})$ captures factors that are not transplanted across countries by any of the transformations, namely the distribution of residual heterogeneity terms Υ.

3.4 Incorporating labour supply responses

The optimal tax literature has long recognized the importance of accounting for labour supply responses to tax changes (Mirrlees, 1971) and recent analysis of tax-benefit policy changes

[7] Concretely, we would ascribe $\Delta_\theta^1(F^A, F^B) \equiv \Delta_\theta^d(F^A, F^B) = \theta(F_A^d) - \theta(F^A)$ to the demographic differences between countries—F_A^d is the counterfactual distribution obtained after applying the demographic transformation to country A. The contribution of differences in labour market structure is $\Delta_\theta^2(F^A, F^B) \equiv \Delta_\theta^{l|d}(F^A, F^B) = \theta(F_A^{l,d}) - \theta(F_A^d)$ where $F_A^{l,d}$ is obtained by composing the demographic and labour market structure transformations, that is $m^\xi(\tilde{X}(X), \Upsilon; \tilde{l}(\xi))$. Similarly, the contributions of prices-and-returns and of tax-benefits are respectively defined as $\Delta_\theta^3(F^A, F^B) \equiv \Delta_\theta^{r|l,d}(F^A, F^B) = \theta(F_A^{r,l,d}) - \theta(F_A^{l,d})$ and $\Delta_\theta^4(F^A, F^B) \equiv \Delta_\theta^{tb|r,l,d}(F^A, F^B) = \theta(F_A^{tb,r,l,d}) - \theta(F_A^{r,l,d})$ with composition of multiple transformations. The residual difference $\Delta_\theta^5(F^A, F^B) \equiv \Delta_\theta^\Upsilon(F^A, F^B) = \theta(F^B) - \theta(F_A^{tb,r,l,d})$ captures factors that are not transplanted across countries by any of the transformations, namely the distribution of residual heterogeneity terms Υ. So, we have $\Delta_\theta(F^A, F^B) = \left[\Delta_\theta^d(F^A, F^B) + \Delta_\theta^{l|d}(F^A, F^B) + \Delta_\theta^{r|l,d}(F^A, F^B) + \Delta_\theta^{tb|r,l,d}(F^A, F^B)\right] + \Delta_\theta^\Upsilon(F^A, F^B)$.

[8] Some authors have proposed to calculate the contribution of each factor in all possible sequence of introduction of factors and average across sequences (Chantreuil and Trannoy, 2013; Devicienti, 2010; Shorrocks, 2013). This approach can however be computationally prohibitive for complex models and does not necessarily improve the economic interpretation of the estimated components.

incorporate analysis of labour supply responses to estimate the full effect of a reform (see, e.g., Aaberge et al., 1995; Bargain, 2012). While the counterfactual distributions that we construct are meant to quantify the contribution of various factors to inequality differences across countries in a static perspective—we are not trying to describe accurately what would actually happen in country A if, say, its labour market structure was to morph suddenly into the structure of B—we incorporate a structural labour supply model which evaluates employment probabilities (including part-time employment) by household members as a function of the household demographic characteristics and disposable income where disposable income is itself a function of tax-benefit parameters, wages and non-labour incomes and individual characteristics. The labour supply model identifies household labour supply elasticities that allow us to adjust employment probabilities as a result of either changes in tax-benefit parameters (after a tax-benefit transformation) or changes in market incomes following from a price-and-return or labour market structure transformation.[9]

The labour supply response adjustment to a transformation of the parameters affects both individual employment I_{hi}^{lab} and hours worked s_{hi} of all household members in the income generation model. Appendix C describes the labour supply model implemented in our empirical application and how predictions are incorporated in the representation of the income generation process. Comparing the decomposition results with and without allowing for labour supply adjustments to our cross-country transformations informs us of the potential importance of understanding behavioural responses (at least in the short run). We isolate below the contribution of labour supply responses by taking the difference between the counterfactual distribution when labour supply responses are considered and the completely "static" counterfactual. The total effect of the "swap" is decomposed into a direct effect and indirect effect from the labour supply response.

4 Application to Ireland and the United Kingdom

We now provide an illustrative application of the methods to Ireland and the United Kingdom. The two countries share a common language and border and have much in common, historically and with respect to labour market and Welfare State policies which were influenced by similar political philosophy principles. Nonetheless, the contemporaneous income distributions in the two countries differ quite substantially. In 2007, the last year before the financial and economic crisis hit both countries, the Gini coefficient was 0.28 in Ireland—a relatively low figure by international standards—while it was 0.32 in the UK—among the highest EU figure. To put the gap in perspective, a difference of four Gini points corresponds to the increase in disposable income inequality reported by OECD (2011) for the USA between the mid-1980s and the late 2000's—a period during which inequality is thought to have increased dramatically.

4.1 Data

We exploit two nationally representative household surveys: the European Union Statistics of Income and Living Conditions (EU-SILC) for Ireland and the Family Resources Survey (FRS) for the United Kingdom. These surveys contain detailed information about household incomes as well as a wide range of variables about the characteristics of households and their members. They have been the key sources of official statistics about the distribution of income in both countries.

[9]By construction, the demographic transformation does not lead to any change in our household labour supply predictions since it does not lead to any change in relevant individual or household-level variables.

A central component of our model is the tax-benefit microsimulation engine EUROMOD, so we use the "EUROMOD input data" versions of the FRS and EU-SILC datasets which have been standardized to common definitions for (market) income variables and household characteristics by the EUROMOD team (Sutherland and Figari, 2013). The definition of disposable household income in EUROMOD includes the sum across all household members of market incomes and public pensions plus cash benefit minus taxes and social insurance contributions. Note that cash benefits and taxes are not reported by survey respondents but are calculated by the tax-benefit calculator. This assumes away any tax evasion and assumes full take-up of benefits. However, in some countries with high non-take-up rates—including both Ireland and the UK—EUROMOD applies a correction to the data so as to match external statistics on take-up proportions (Sutherland and Figari, 2013). [10]

We study the distribution of income in 2007 in both countries and incomes are expressed in 'single adult equivalent' by dividing total household income by the square root of household size. Currency values for the UK are converted in euros using the exchange rate of 1.484 British pounds per euro. Samples sizes are 12,516 individuals (5,247 households) in the Irish data and 57,276 individuals (25,088 households) in the UK sample.

4.2 Inequality in Ireland and the UK compared

The distributions of income in the two countries are shown in Figure 1 in the form of Pen's parades (quantile functions divided by mean income). (The vertical axis is on a logarithmic scale: a vertical shift of the curves corresponds to a proportional increase of all incomes—a transformation that leaves most commonly-used inequality measures unchanged.) Overall, the UK displays a steeper profile, especially above the 50th percentile. Until the 80th percentile, incomes (relative to the mean) are higher in Ireland than in the UK whereas incomes of the richest 20% are higher in the UK relative to the country mean—see the bottom panel. These differences translate into a higher inequality in the UK than in Ireland as measured by the Gini index (see Table 1). Figure 2 shows Lorenz dominance of the Irish distribution over the UK distribution so the country ranking in inequality is robust to the choice of inequality index.

[10] The correction consists in randomly imputing 'non take-up' and therefore assigning zero benefits to a fraction of the sample households. This is done separately for different sources of benefits.

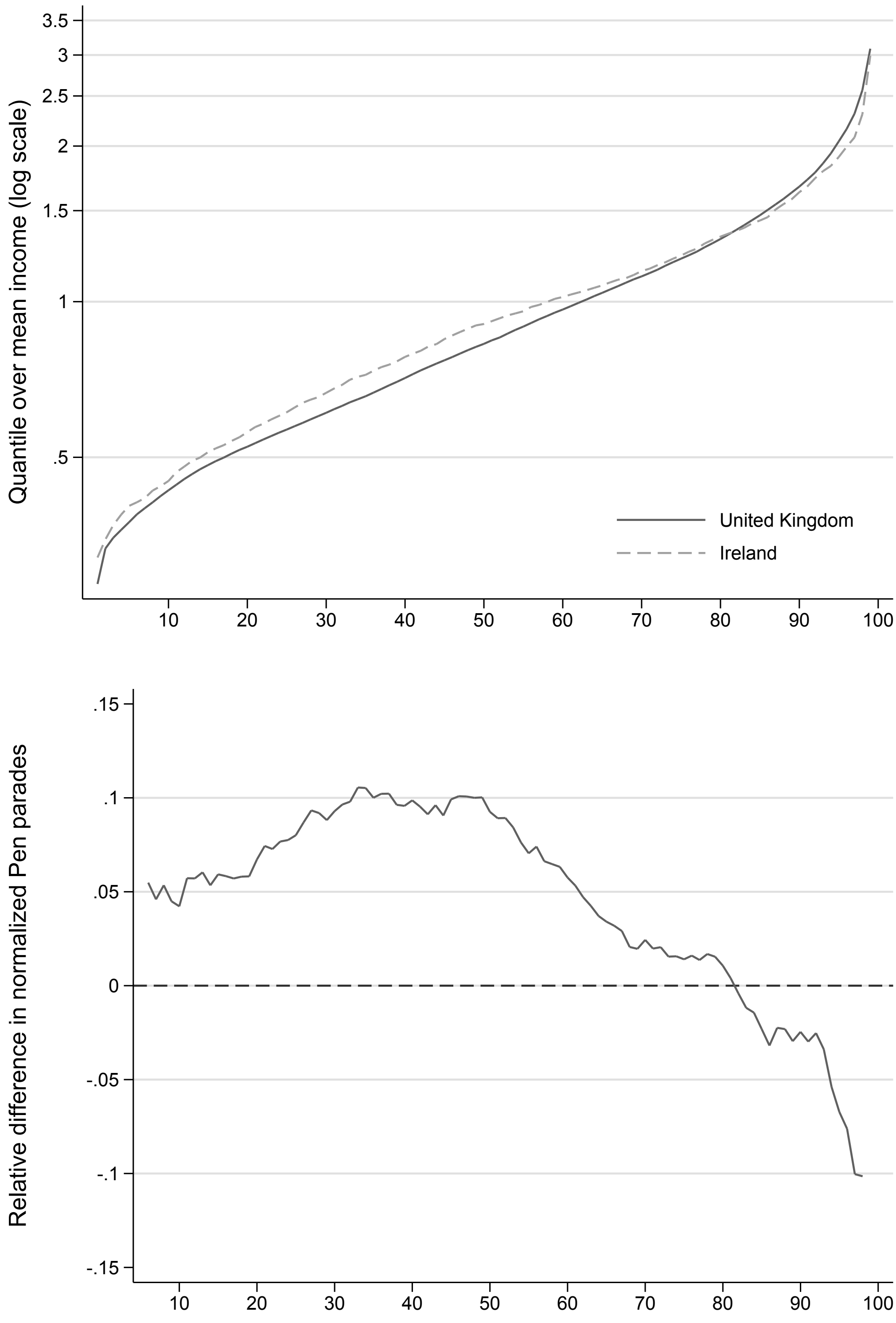


Figure 1: Distribution of equivalised household disposable income in Ireland and the UK: Normalized quantile functions (Pen's parades) (top) and relative differences (IE/UK-1) (bottom)

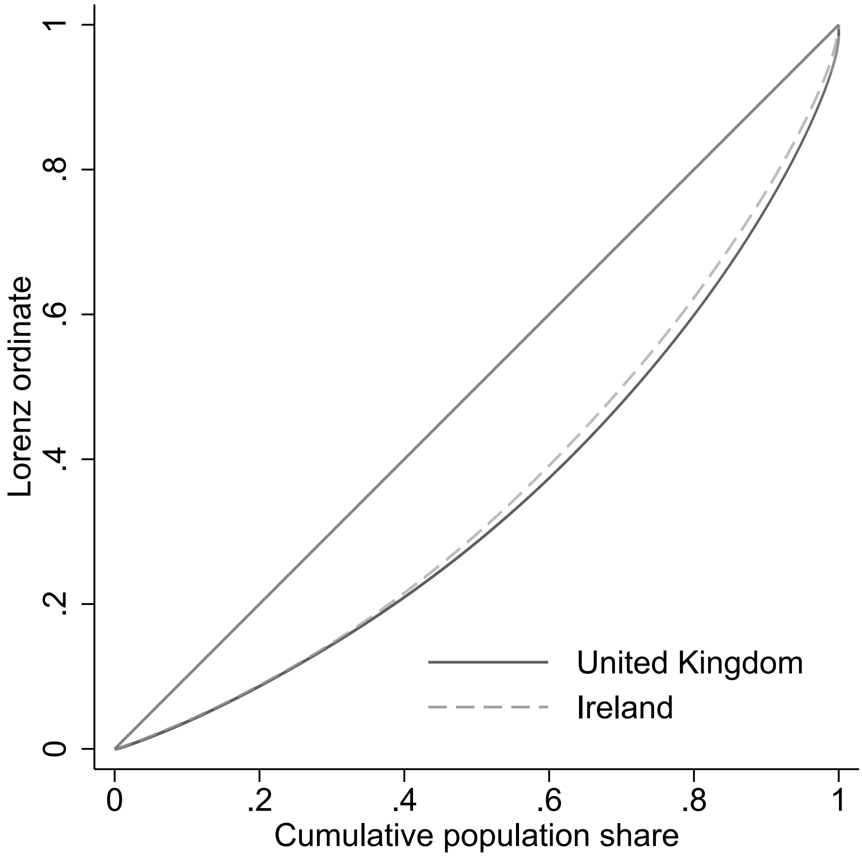


Figure 2: Lorenz curves for the United Kingdom and Ireland

Table 1: Equivalized household disposable income in the UK and Ireland, 2007 (monthly, in euros)

	Mean	Median	Gini
UK	2341	2243	0.319
Ireland	2541	2491	0.277

Table 2 shows differences across the two countries in a number of population charateristics and labour market structures. The two countries have similar demographic profiles; with some notable exceptions however. According to our samples, the population aged 25–64 is generally more educated in Ireland than in the UK, and there are larger shares of people aged 65+ and smaller shares of children aged 4+ in the UK than in Ireland. The share of people aged 16+ at work is 6 percentage points higher in Ireland than in the UK but their distribution across occupations are similar (we suspect that differences with respect to "professionals" and "associate professionals" reflect differences in data definitions). The distribution of workers across sectors of activity is however remarkably different: the share of workers in agriculture is almost four times larger in Ireland than in the UK, the share in industy is less than half, and the services sector employs 10 percentage points more people in Ireland. Public sector employment is also larger in the UK. Table 2 also points to differences in the household income composition. The prevalence of non-labour income sources differs between the two countries, with the UK exhibiting larger shares of people with private pensions, capital and other sources of income.

Table 2: Population and labour market structures (shares of total population)

	UK	Ireland
Demographic		
Tertiary Education	0.263	0.314
People 16-65	0.667	0.676
People >65	0.146	0.102
Child 0-3	0.052	0.051
Child 4-11	0.087	0.108
Child 12-15	0.048	0.063
Married	0.504	0.485
Citizen	0.906	0.927
Male	0.486	0.496
Labour market		
In-work	0.557	0.619
Employee/Self-Employed	0.883	0.864
Occupation		
Managers	0.158	0.174
Professionals	0.136	0.182
Associate Prof.	0.149	0.049
Clerks	0.119	0.115
Service	0.153	0.180
Craft	0.113	0.162
Plant	0.072	0.045
Unskilled	0.101	0.092
Industry		
Agriculture	0.014	0.048
Industry	0.218	0.087
Services	0.769	0.865
Public/Private	0.283	0.234
Other market factors		
With private pensions	0.342	0.146
With capital income	0.576	0.214
With other income	0.099	0.052

Notes: The estimates are weighted. The shares for education refer to age-group 25-64; for married, sex to age >= 16; for in-work to ages 16 to 80; for employees, occupation, industry and sector to those in work aged [16, 80); for citizen to the entire sample. The shares for private pensions refer to ages >= 45, for capital age>= 16.

Both the Irish and UK tax-benefit systems are part of the Anglo-Liberal system of Welfare States. With mainly flat rate or means-tested benefit instruments, the primary objective of the transfer system is poverty alleviation. While there are differences, many of the historical developments in the Irish benefit system have derived from reforms in the UK system. Both are characterized by (i) flat rate and means-tested income replacement benefits (the main difference rests in the presence of a previous earnings related component in the UK system, that although contemplated at various stages in Ireland was never introduced); (ii) in-work transfers (both countries have transfers targeted at low-income families with children, where payments are made once a particular number of hours have been made; there is an additional child care component in the UK system); (iii) flat rate universal child benefits; and (iv) housing benefits (coverage has been lower historically in Ireland, but has been increasing).

Both countries have progressive income taxation systems and earnings-related social insurance contributions that vary by employee and self-employed. There are however some notable differences. The Irish income taxation system is joint, while the UK system is an individualized system.

Table 3 documents the redistributive effects of the tax-benefit system in both countries. Note first that inequality in market income is more similar in the two countries than inequality of disposable income. The benefit schedule increases the difference in inequality between the two countries by dropping inequality to a larger extent in Ireland compared with the UK, effect driven by a higher degree of redistribution in Ireland. The benefit schedule is more regressive in the UK (more low incomes receive benefits), but the average benefit rate is also much lower. The tax system increases further the percentage point difference in inequality between Ireland and the UK. This is due to a more progressive and a more redistributive tax system in Ireland. As taxes are progressive and benefits are regressive, the net schedule is equalizing in both countries. The Reynolds-Smolensky index of net redistributive effect shows the Irish tax-benefit system is more redistributive than the UK system (Lambert, 2001). Whether these differences are due to policy design or to differences in the market income distribution is revealed in the decomposition analysis.

Table 3: Progressivity and redistribution of taxes and benefits on household equivalized disposable income

	UK	Ireland	Ratio: IRL/UK
Gini Gross Income	0.497	0.483	0.972
Gini Gross Income (incl. benefits)	0.377	0.341	0.905
Averate transfer rate	0.155	0.242	1.558
Benefit Regressivity (K)	0.936	0.769	0.822
Benefit Redistribution (RS)	0.120	0.142	1.183
Gini (gross + benefits - income taxes)	0.332	0.289	0.871
Averate tax rate	0.159	0.132	0.826
Tax Progressivity (K)	0.242	0.354	1.460
Tax Redistribution (RS)	0.045	0.053	1.156
Gini Disposable Income	0.319	0.277	0.869
Net Redistributive Effect	0.178	0.206	1.157

Notes: K = Kakwani; RS = Reynolds-Smolensky.

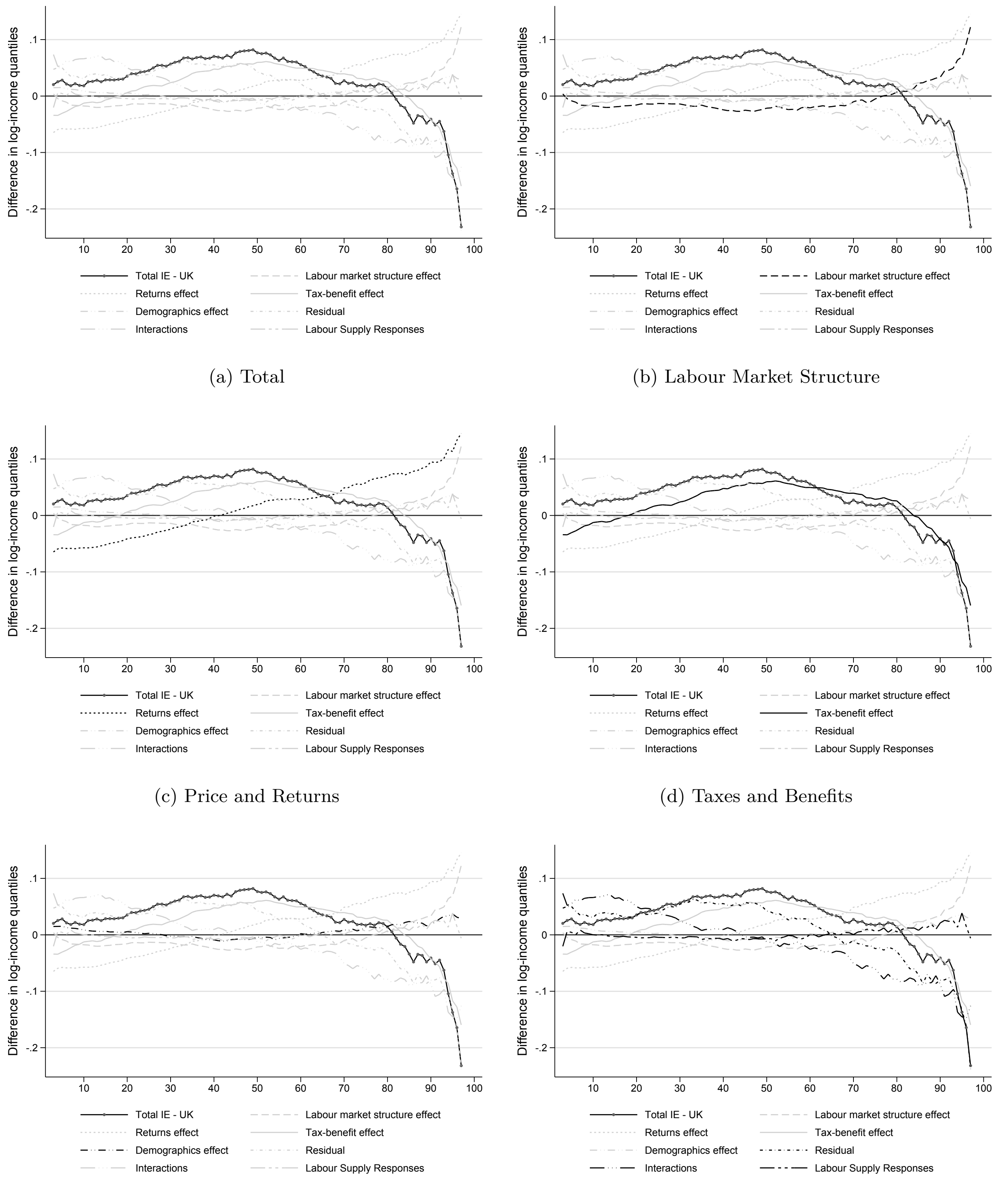
4.3 Accounting for differences in income inequality

Preliminary inspection of the characteristics of the population and of the income distribution points towards a few explanatory factors of the difference in income inequality, mainly the tax-benefit system, but also differences in the industrial structure of the two countries and in the distribution of non-income sources. Applying the counterfactual decomposition exercise quantifies the respective roles played by such factors.

4.3.1 Counterfactual distributions

A decomposition of the differences between the mean-normalized quantile functions of the two distributions of equivalised disposable income is displayed in Figure 3.[11] Remember that mean-normalized quantiles are higher in Ireland than in the UK up to the 80th quantile beyond which the difference turns negative; the observed difference is marked by dots on the plot. The counterfactual differences obtained by applying each of the four transformations defined in Section 3 onto the UK data are shown in Figure 3. Figure 3(a) shows the total, observed difference; Figures 3(b,c,d,e) show how the 'first order', direct effect of applying each of the four transformations compares to the total difference; Figure 3(f) highlights the residual and interaction terms and the impact of labour supply responses.

[11]The fit of our simulation model is described in Appendix D. The whole set of model parameter estimates is not reported for the sake of brevity, but they are available on request.



(a) Total

(b) Labour Market Structure

(c) Price and Returns

(d) Taxes and Benefits

(e) Demographics

(f) Interactions, residuals, labour supply responses

Figure 3: Distributional differences across quantiles of equivalised household disposable income and counterfactuals after labour market structure transplant, prices and returns transplant, taxes and benefits transplant, demographic transplant, and resulting interactions, residual and labour supply response components

The key observation comes from the tax-benefit transformation (highlighted in panel (d)). Applying the Irish tax-benefit rules onto the UK data generates a counterfactual difference to the UK distribution that is remarkably close to the actual difference between Ireland and the UK. The match is particularly close for incomes above the median. For the bottom half of the distribution, tax-benefit differences do not appear to fully explain the higher values observed in Ireland. These results clearly points to differences in social and fiscal policies as the most important determinant of income distribution differences between the two countries. As the tax-benefit effect is basically the policy effect controlling for differences in the market income distributions between the two countries, the larger deviation of the tax-benefit effect from the total difference observed for the bottom 50% suggests larger market income differences between the two countries at the bottom rather than at the top. The negative tax-benefit effect for the bottom and top 20% reveals that the UK tax-benefit rules are more advantageous to the poorest and the richest households than the Irish rules. Overall, this is consistent with a higher benefit regressivity of the UK system and a higher tax progressivity of the Irish system.

The transformation having the second strongest impact is the price-and-returns transformation. However its impact mostly works *against* the observed difference. Applying the Irish structure of returns to demographic and labour market characteristics onto the UK data would tend to reduce the incomes of the bottom 40 percent and increase incomes for the top 60 percent—overall a "disequalizing effect". The counterfactual quantile function is nowhere close to the actual difference observed between the two countries (except to the point where the two curves cross around the 65th percentile). A similar lesson is drawn from the labour market structure transformation: applying the employment rate, the occupational and industrial characteristics of Irish employment, and the prevalence of non-market incomes in Ireland onto the UK population also works against the observed difference. It leads to slightly lower incomes throughout the bottom 75 percent of the population and larger incomes for the top 25 percent. This pattern can be tracked to the higher employment rate and the larger share of workers in the service sector in Ireland, although it does not seem to fit the higher share of non-labour incomes found in the UK. The demographic transformation has the smallest impact of all. Applying the Irish demographic characteristics onto the UK population only leads to a modest increase of income for the top 20 percent. This pattern is consistent with the higher share of tertiary educated and "prime age" individuals in Ireland documented in Table 2, but it also discards demographic differences as the source of the greater inequality observed in the UK.

The picture that emerges is one where the Irish tax-benefit system "undoes"—actually "more than undoes"—what otherwise appear to be disequalizing factors in the Irish labour market structure and in the returns to characteristics in terms of market income, compared to the UK.

These four univariate factors are not however additive. Adding up the four counterfactuals separately applied to the UK leads us relatively far from the observed difference to the Irish distribution: incomes of the bottom 60 percent remain underestimated compared to the observed Irish incomes and incomes at the top 40 tend to be overestimated. The magnitude of correction factors—consisting mostly in (i) the correlation of income sources in the households (including partners earnings) which is not modelled and "swapped" in our counterfactuals and (ii) multiplicative interactions between the transformations—are reflected in the last panel of Figure 3. The correlation of incomes is difficult to model parametrically and is therefore left as a "residual" explanation. Interactions are inherent to the non-linear transformations applied to the income distributions (Biewen, 2014). For example we apply the Irish tax-benefit transformation directly on the UK distribution of market incomes and population structure. But if we had applied the tax-benefit transformation on a "modified" UK distribution reflecting the "disequalizing" adjustments due to the population structure or the different returns to characteristics, we can expect that its impact would have been stronger than shown in Figure 3. These interactions beyond the "first order" effect of the transformations are bundled into the interaction term. As we show below in the decomposition of Gini indices, the two factors—residual and interaction—seem to be of approximately equal sizes.

Panel (f) of Figure 3 also identifies the contribution of labour supply responses. Our transformations so far ignore the potential behavioural shifts. For instance, if the transplanted taxation system favours the mid-high income earners with lower taxes, disregarding the behavioural shift might underestimate the size of the high-income population as there is now a greater incentive for a worker with mid-high earning capacity to work. Additionally, the labour supply model captures the labour supply preference of the population, allowing us to swap the incentive structure of the economic system while retaining both the observed population characteristics and unobserved preferences to some extent. The labour supply preference interacts with all other components as the hours of labour supply is a function of the population structure, tax benefit system, the market composition and returns. Overall, the behavioural response, as an indirect factor, appears to have a much smaller impact than the direct transformations.

4.3.2 Gini coefficients of disposable and market incomes and the net redistributive impact of taxes and transfers

We now move to examination of Gini indices. The first column of Table 4 presents decomposition results for the difference in the Gini coefficients of disposable income—a 0.042 difference between 0.319 in the UK and 0.277 in Ireland (rows 1–2). Rows 3–6 show the direct effects of applying each of our four transformations onto the UK distribution: the difference between the Gini coefficient in the counterfactual distribution that would prevail in the UK if we transplant each factor in isolation from Ireland (assuming no labour supply responses) and the original distribution. Rows 7-9 report indirect effects, namely the additional change once labour supply adjustments are taken into account. Rows 10 and 11 capture respectively the interaction effects (the difference between the sum of direct and indirect effects and the Gini observed when all four transformations are jointly applied) and the residual difference (that conflates all factors not explicitly modelled in the income generation model, most notably the cross-equation correlation of unobserved heterogeneity terms). The direct effect of differences in the two tax-benefit systems (–0.015) is just under half the observed difference in disposable income inequality (–0.042). As was already apparent from Figure 3, the cross-country difference in disposable income inequality is largely attributable to differences in tax-benefit systems which counterbalance the disequalizing effect of the differences in market composition and returns between Ireland and the UK. The effect of labour supply responses are relatively small.

The difference in gross income inequality between the two countries (column 2) is smaller than in disposable income, yet inequality remains smaller in Ireland. Gross income includes only labour market income, private pensions, capital and "other" pre-tax incomes. Gross income inequality would be larger in the UK had the UK the market composition of Ireland but it would be smaller if the Irish returns to assets and human capital where transplanted in the UK (in this case any difference in gross income inequality between the UK and Ireland would disappear). Again, the contribution of demographic differences is comparatively small.[12]

Table 5 refines this picture by showing the contribution of more disaggregated transformations of the income distributions from our model. For the sake of brevity, we do not detail the construction of all those sub-transformations but they should be self-explanatory from the table labels and notes: they involve swapping only a *subset* of model parameters from 'labour market structure' and 'price and returns' transformations across countries (either parameters reflecting the returns to some specific characteristics or reflecting the relative prevalence of such characteristics). The effect of the labour market structure transformation is mostly driven by differences in the prevalence of non-labour incomes in household portfolios. The price and returns transformation impact on disposable income mostly arises from the parameters of the non-labour incomes equations again (concerning mostly capital incomes) while equations for labour income and private pensions are driving the impact of the transformation on gross incomes.

[12]The small effect of the taxes and benefits transformation on gross income inequality is due to adjustments to minimum wages which are included in the taxes and benefit transformation.

Table 4: Decomposition of differences in Gini coefficients and in redistribution measures between Ireland and the UK

2007	Gini Disposable (1)	Gini Gross Income (2)	Net Redistr. (3)	Benefit Regressivity (4)	Avg. Benefit Rate (5)	Tax Progressivity (6)	Avg. Tax rate (7)
UK	0.319	0.497	0.178	0.936	0.155	0.242	0.159
Country Differences							
IE-UK	-0.042	-0.014	0.028	-0.167	0.087	0.112	-0.028
Contribution of direct effects (UK*-UK) to cross-national differences							
LMS	0.015	0.022	0.006	0.009	0.007	0.009	0.002
Returns	0.015	-0.012	-0.027	-0.062	-0.022	-0.032	0.006
TB	-0.015	0.002	0.017	-0.170	0.058	0.113	-0.023
Demographics	-0.004	0.006	0.010	-0.012	0.016	0.003	-0.002
Labour supply responses to component swaps							
LMS	0.007	0.020	0.014	0.010	0.017	0.008	-0.002
Returns	-0.004	-0.002	0.002	-0.003	0.004	0.000	-0.003
TB	0.000	0.001	0.000	0.001	0.000	0.000	-0.000
Interactions	-0.023	-0.018	0.004	0.080	-0.017	-0.002	0.007
Residuals	-0.034	-0.032	0.002	-0.020	0.023	0.012	-0.013

Notes: LMS: labour market structure; TB: tax-benefit system.

Table 5: Decomposition of differences in Gini coefficients: disaggregation of the and transformations

2007	Gini Disposable	Gini Gross Income
UK	0.319	0.497
Country Differences		
IE-UK	-0.042	-0.014
Contributions to cross-national differences (UK*-UK)		
LMS	0.015	0.022
LMS Components		
In-work	-0.001	-0.016
Employed/Self-employed	0.002	0.003
Occupation/Industry/Sector	-0.001	-0.000
Has NL income	0.013	0.037
Other	0.004	0.000
Interactions	-0.001	-0.003
Returns	0.015	-0.012
Returns Components		
Labour Income	0.001	-0.009
Private Pensions	-0.001	-0.010
Other	0.016	0.005
Interactions	-0.000	0.002

Notes: LMS: labour market structure; NL: non-labour.

The difference between the Gini coefficient for gross incomes and for net disposable incomes is a standard measure of the net redistributive effect of taxes and transfers (which convert gross to net incomes). Accordingly, the net redistribution through taxes and transfers appears larger in Ireland than in the UK (see column 3 in Table 4). However such measures of net redistribution are the result of the combination of a country's tax-benefit system applied to its own distribution of gross incomes and are therefore difficult to compare across countries (see, e.g., Verbist and Figari, 2014). We can exploit our decomposition framework to examine how much of the different model components account for the overall redistributive effect of taxes and transfers. It turns out that only about half of the total difference in net redistributive effect is accounted for by differences in tax and benefit parameters (as captured by application of the taxes-and-benefits transformation). Columns 4–7 of Table 4 detail the channels through which this happens by examining the progressivity/regressivity of benefits (col. 4) and taxes (col. 6) and the average benefit-to-income ratio (col. 5) and average tax rates (col. 7). The Irish tax-benefit system appears to be more redistributive than the UK system because of a higher tax progressivity and larger average benefits rates. The UK system has a higher benefits regressivity and lower average tax rates. Even when measured on identical market income distributions (that is, after transplanting the Irish TB system on the UK data), tax progressivity and average benefit rates, and overall net redistribution are larger under the Irish tax-benefit system. By design, the Irish tax schedule is more progressive than the UK tax schedule, but the Irish lead in tax progressivity is eroded partially by the differences in gross income distributions. The difference in average benefit rates due solely to policy differences is lower (5.8 percentage points) than the observed difference (8.7 percentage points), suggesting that market income difference leads to a greater difference in the average benefit rates between the two countries.

5 Concluding Remarks

This paper develops a methodological framework for exploring the drivers of cross-national differences in the distribution of household disposable income, focusing on the role of tax-benefit systems, labour market structures, prices and returns, and demographic structures. Our framework draws upon the Bourguignon et al. (2008) methodology by developing a household income distribution which incorporates a flexible parametric approach of modelling wage differentials across the entire distribution. We incorporate the complexity of tax-benefit rules through micro-simulation (EUROMOD) and integrate potential labour supply responses in the generation of counterfactual distributions. The result is an integrated framework that is portable across EU countries for generating and simulating the distribution of household disposable income under alternative scenarios, thereby enabling the study of the various drivers of the cross-national distributional differences in household disposable income. The framework proposed is descriptive by nature, but the apparatus offers sufficient sophistication to allow detailed analysis of the way tax-benefit systems can interact with labour market structures, income structures and demographics in determining the distribution of household disposable income and in explaining the cross-national differences in disposable income inequality. Compared to alternative decomposition approaches such as Oaxaca-Blinder, and some variance decomposition techniques for inequality index, the method can better capture the effect of policy through simulated counterfactuals instead of the observed data which contains a mixture of policy and interaction effects. Additionally, this method allows better control of the socio-economic characteristics including the entire market income distribution.

The paper illustrates use of the method through the analysis of two European neighbouring English-speaking countries–the UK and Ireland–that share many similarities while displaying at the same time sufficient differences to merit understanding more clearly of the factors that have resulted in different levels of inequality. We explored the drivers of distributional differences between these countries in 2007, the latest year before the economic crisis in both countries. Whereas market income distribution is roughly 1.4 Gini point less unequal in Ireland than in the UK, the difference in inequality in disposable income is almost three times larger (4.2 Gini points). Our decomposition analysis reveals that differences across countries are largely due to policy parameter differences which more than offset otherwise disequalizing contributions of labour market structures and returns. Differences in demographic characteristics of the two populations (including differences in educational attainment) play a negligible role. Comparison of disposable and market incomes suggest that the Irish tax-benefit system is more redistributive than the UK system due to a higher tax progressivity and more generous average transfer rates. Market income distributional differences reinforce the net redistributive policy effect via both market composition and demographic differences. The effect of the differences in market composition (mainly via the assignment of non-labour income sources and the occupational structure) and in demographics stems primarily from the positive effect on both average transfer rates and tax progressivity. Our model also allows labour supply behavioural shifts once each main component is swapped. Given the relatively inelastic labour supply for the general population, the behavioural impact for the whole population income distribution is limited, with the static first round effect dominating the results.

References

Aaberge, R., Dagsvik, J. K., and Strøm, S. (1995). Labor supply responses and welfare effects of tax reforms. *Scandinavian Journal of Economics*, pages 635–659.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Wiley Series in Probability and Statistics. Wiley.

Alesina, A. and Glaeser, E. (2005). *Fighting Poverty in the U.S. and Europe: A World of Difference*. Oxford University Press.

Atkinson, A. B. (2007). The distribution of earnings in OECD countries. *International Labour Review*, 146(1-2):41–60.

Bargain, O. (2012). Back to the future: decomposition analysis of distributive policies using behavioural simulations. *International Tax and Public Finance*, 19(5):708–731.

Bargain, O. and Callan, T. (2010). Analysing the effects of tax-benefit reforms on income distribution: a decomposition approach. *Journal of Economic Inequality*, 8(1):1–21.

Barsky, R., Bound, J., Charles, K., and Lupton, J. (2002). Accounting for the black-white wealth gap: A nonparametric approach. *Journal of the American Statistical Association*, 97:663–673.

Belfield, C., Blundell, R., Cribb, J., Hood, A., and Joyce, R. (2017). Two decades of income inequality in Britain: The role of wages, household earnings and redistribution. *Economica*, 84(334):157–179.

Biewen, M. (2014). Additive decompositions and interaction effects. *Applied Economics Letters*, 21(9):636–642.

Biewen, M. and Jenkins, S. P. (2005). A framework for the decomposition of poverty differences with an application to poverty differences between countries. *Empirical Economics*, 30(2):331–358.

Biewen, M. and Juhasz, A. (2012). Understanding rising income inequality in Germany, 1999/2000-2005/2006. *Review of Income and Wealth*, 58(4):622–647.

Bourguignon, F., Ferreira, F. H., and Leite, P. G. (2008). Beyond Oaxaca-Blinder: Accounting for differences in household income distributions. *Journal of Economic Inequality*, 6(2):117–148.

Brandolini, A. and Smeeding, T. (2010). Income inequality in richer and OECD countries. In Salverda, W., Nolan, B., and Smeeding, T., editors, *The Oxford Handbook of Economic Inequality*, pages 71–100. Oxford University Press.

Brewer, M. and Wren-Lewis, L. (2016). Accounting for changes in income inequality: Decomposition analyses for the UK, 1978–2008. *Oxford Bulletin of Economics and Statistics*, 78(3):289–322.

Chantreuil, F. and Trannoy, A. (2013). Inequality decomposition values: the trade-off between marginality and efficiency. *Journal of Economic Inequality*, 11(1):83–98.

Colombino, U., Locatelli, M., Narazani, E., and O'Donoghue, C. (2010). Alternative basic income mechanisms: An evaluation exercise with a microeconometric model. *Basic Income Studies*, 5(1):1932–0183.

Cowell, F. and Fiorio, C. (2011). Inequality decompositions—a reconciliation. *Journal of Economic Inequality*, 9(4):509–528.

Creedy, J. and Duncan, A. (2002). Behavioural microsimulation with labour supply responses. *Journal of Economic Surveys*, 16(1):1–39.

Daly, M. C. and Valletta, R. G. (2006). Inequality and poverty in United States: The effects of rising dispersion of men's earnings and changing family behaviour. *Economica*, 73(289):75–98.

Dardanoni, V. and Lambert, P. J. (2002). Progressivity comparisons. *Journal of Public Economics*, 86(1):99–122.

Devicienti, F. (2010). Shapley-value decompositions of changes in wage distributions: a note. *Journal of Economic Inequality*, 8(1):35–45.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

Eurostat (2017). Gini coefficient of equivalised disposable income - EU-SILC survey. European Commission, http://ec.europa.eu/eurostat/web/products-datasets/-/tessi190 (accessed June 12 2017).

Figari, F., Paulus, A., and Sutherland, H. (2015). Microsimulation and policy analysis. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 2B, chapter 24, pages 2141–2221. North-Holland, Oxford, UK.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Förster, M. and Tóth, I. G. (2015). Cross-country evidence of the multiple causes of inequality changes in the OECD area. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 2B, chapter 19, pages 1729–1843. North Holland.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4A, pages 1–102. North-Holland, Amsterdam.

Haveman, R., Heinrich, C., and Smeeding, T. (2011). Policy Responses to the Recent Poor Performance of the U.S. Labor Market. *Journal of Policy Analysis and Management*, 31(1):153–195.

Herault, N. and Azpitarte, F. (2016). Understanding changes in the distribution and redistribution of income: A unifying decomposition framework. *Review of Income and Wealth*, 62(2):266282.

Hyslop, D. R. and Mare, D. C. (2005). Understanding New Zealand's Changing Income Distribution, 1983 - 1998: A Semi-parametric Analysis. *Economica*, 72(287):469 – 495.

Jenkins, S., Brandolini, A., Micklewright, J., and Nolan, B., editors (2013). *The Great Recession and the Distribution of Household Income*. Oxford University Press, Oxford, UK.

Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3):410–442.

Keane, C., Kelly, E., Callan, T., and Savage, M. (2012). Ireland 2007–2010. Euromod country report, EUROMOD, University of Essex.

Keane, M. and Moffitt, R. (1998). A structural model of multiple welfare program participation and labor supply. *International Economic Review*, pages 553–589.

Lambert, P. J. (2001). *The Distribution and Redistribution of Income.* Manchester University Press, Manchester.

Larrimore, J. (2014). Accounting for United States household income inequality trends: The changing importance of household structure and male and female labor earnings inequality. *Review of Income and Wealth*, 60(4):683–701.

Lerman, R. I. and Yitzhaki, S. (1985). Income inequality effects by income. *Review of Economics and Statistics*, 67(1):151–156.

Levy, H., Lietz, C., and Sutherland, H. (2007). Swapping policies: Alternative tax-benefit strategies to support children in Austria, Spain and the UK. *Journal of Social Policy*, 36(4):625–647.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, 71(5):1339–1375.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2):175–208.

Nolan, B., Salverda, W., Checchi, D., Marx, I., McKnight, A., Tóth, I. G., and van de Werfhorst, H. G., editors (2014). *Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences.* Oxford University Press.

OECD (2011). *Divided We Stand: Why Inequality Keeps Rising.* OECD, Paris.

Paulus, A. and Tasseva, I. (2017). Decomposition of changes in the EU income distribution in 2007–2011. EUROMOD Working Paper EM 9/17, Institute for Social and Economic Research, University of Essex, UK.

Piketty, T. (2013). *Capital in the Twenty-First Century.* Harvard University Press.

Rothe, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, 155(1):56–70.

Rothe, C. (2012). Partial distributional policy effects. *Econometrica*, 80(5):2269–2301.

Shorrocks, A. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48:613–625.

Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica*, 50(1):193–211.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *Journal of Economic Inequality*, 11(1):99–126.

Sutherland, H. and Figari, F. (2013). EUROMOD: the European Union tax-benefit microsimulation model. *International Journal of Microsimulation*, 6(1):4–26.

Sutherland, H., Tumino, A., and Zantomio, F. (2012). United Kingdom 2006–2011. Euromod country report, EUROMOD, University of Essex.

Van Kerm, P. (2013). Generalized measures of wage differentials. *Empirical Economics*, 45(1):465–482.

Van Kerm, P., Choe, C., and Yu, S. (2016). Decomposing quantile wage gaps: a conditional likelihood approach. *Journal of the Royal Statistical Society (Series C)*, 65(4):507–527.

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

van Soest, A., Das, M., and Gong, X. (2002). A structural labour supply model with flexible preferences. *Journal of Econometrics*, 107(1):345–374.

Verbist, G. and Figari, F. (2014). The redistributive effect and progressivity of taxes revisited: An international comparison across the European Union. *FinanzArchiv: Public Finance Analysis*, 70(3):405–429.

Appendices

A Summary of income generation process components

Tables A–2 and A–3 summarize the components of the income generation process. Table A–2 lists all income sources examined and identifies whether the source is an aggregate of sub-components or whether it is modelled directly, it identifies the type of model applied and the conditioning variables, and it also identifies to what family of transformations the model corresponds to. Table A–3 describes the demographic and labour market variables in the same way.

Variable	Definition	Level	Treatment	Transformation	Model	Conditioning variables
y_h	total household disposable income	household	aggregate		–	–
y_h^L	gross labour income	household	aggregate		–	–
I_{hi}^{emp}, y_{hi}^{emp}	employee income (wage*hours)	individual	aggregate	Returns (wage rates) and /LM struc (hours)	–	–
I_{hi}^{se}, y_{hi}^{se}	self-employment income (receipt, amount)	individual	modelled	Returns	logit, log-linear	x_{hi}
y_h^K	capital income (investment, property)	household	aggregate	Returns	–	
I_h^{inv}, y_h^{inv}	investment income (receipt, amount)	individual	modelled	Returns	logit,log-linear	x_{hi}
I_h^{prop}, y_h^{prop}	property income (receipt, amount)	individual	modelled	Returns	logit,log-linear	x_{hi}
I_h^{pripen}, y_h^{pripen}	private pensions (receipt, amount)	individual	modelled	Returns	logit,log-linear	x_{hi}
y_h^O	other non-benefit incomes (receipt, amount)	individual	aggregate, modelled	Returns	logit, log-linear	x_{hi}
y_h^B	public transfers	household	aggregate	TB	–	–
y_{hi}^{repl}	replacement income (pensions, unemployment)	individual	aggregate	TB	–	–
I_{hi}^{unemp}, y_{hi}^{unemp}	unemployment benefits (receipt, amount)	individual	aggregate, modelled	TB	logit,log-linear, EUROMOD	x_{hi}
I_{hi}^{pens}, y_{hi}^{pens}	public (state, survival, occupational pensions) (receipt, amount)	individual	aggregate, modelled	TB	logit,log-linear, EUROMOD	x_{hi}
$I_{hi}^{disability}$, $y_{hi}^{disability}$	disability (receipt and amount) sickness (receipt and amount)	individual	aggregate, modelled	TB	logit, log-linear, EUROMOD	x_{hi}
$I_{hi}^{sickness}$, $y_{hi}^{sickness}$	sickness (receipt, amount)	individual	modelled	TB	logit, log-linear, EUROMOD	x_{hi}
$I_h^{housing}$, $y_h^{housing}$	housing benefits (receipt, amount)	household	modelled	TB	logit, log-linear, EUROMOD	x_h
y_h^{sa}	social assistance	household	modelled	TB	EUROMOD	x_h
y_h^{osw}	other social welfare	household	modelled	TB	logit, log-linear, EUROMOD	x_h
y_h^{fb}	family benefits	household	modelled	TB	EUROMOD	x_h
y_h^{mb}	maternity benefit	household	modelled	TB	logit,log-linear, EUROMOD	x_h
y_h^{cb}	child benefit	household	modelled	TB	EUROMOD	x_h
t_h	taxes and social security contributions	household	aggregate, modelled	TB	EUROMOD	y_h^L, y_h^K, y_h^O ,y_h^B ,x_h

Table A–2: Definition of income components and summary modelling information

Variable	Definition	Level	Treatment	Factor	Model	Conditioning variables
n_h	household size	household	observed	Demo	–	–
x_h	household-level demographic characteristics (number of children aged 0–3, 4–11 and 12–15) and individual characteristics of the household head (marital status, gender, age and age squared, university education)	household	observed	Demo	–	–
x_{hi}	individual-level characteristics: gender, age and age squared, university education, marital status, number of children in the household (aged 0–3, 4–11 and 12–15)	individual	observed	Demo	–	–
occ_{hi}	Occupation (1-digit ISCO); for employees only	individual	modelled	LM Struct	multinomial logit	x_{hi}
ind_{hi}	Sector (primary, secondary or tertiary); for employees only	individual	modelled	LM Struct	multinomial logit	x_{hi} occ_{hi}
pub_{hi}	Public or private sector job; for employees only	individual	modelled	LM Struct	logit	x_{hi} occ_{hi}
s_{hi}	Number of hours worked	individual	modelled	LM Struct	linear	x_{hi}
w_{hi}	Average wage rate; for employees only	individual	modelled	Returns	Singh-Maddala	x_{hi} occ_{hi} ind_{hi} pub_{hi}

Table A–3: Demographic and labour market variables

B The simulation of taxes and benefits in system swaps between countries using EUROMOD: detailed components and construction

EUROMOD calculates three main output groups of benefits (public pensions, means-tested benefits and non means-tested benefits), taxes and social security contributions.

A challenge in swapping the tax-benefit rules between two countries is to standardize the two input datasets as to permit each system to be run on either input data. This implies that each input data needs to have the core variables required to run either system.

This process is made cumbersome by several factors. For one, not all benefits and taxes are simulated in EUROMOD. EUROMOD simulates those instruments for which there are available data, whereas the rest are included from the data as input variable and are chosen as components of output variables. The EUROMOD country reports (Keane et al. (2012); Sutherland et al. (2012))discuss in detail which instruments are taken from the data and which are simulated or partially simulated. Partially simulated instruments involve an eligibility condition that is based on actual receipt plus other relevant conditions being satisfied. Our income generation model deals with non-simulated or partially simulated instruments as the other income components discussed above. We estimate parametric models for the presence of the income source and the level received. These variables serve as input variables in the tax-benefit simulator. EUROMOD "simulate" them as components of output variables.

Furthermore, each country system has specific rules. Some instruments could be (partially) simulated in one country, while used as input in the other. For example, state pension (basic contributory) in the UK is directly included from the data because part of the eligibility informaiton is unavailable (e.g. contribution history or retirement date). In Ireland, the contributory state pension is partially simulated: the output variable depends on the value of the input variable (taken from the data) and a set of other eligibility conditions. So in both countries, the input data needs to indicate who gets the transfer and the level. For these two conditions, we estimate parametric models for the presence of the income source and the level. When swapping the two tax-benefit systems, the two distributions are swapped as well in the input data. In the case of the non-contributory state pension in Ireland these steps are not required, as the transfer is simulated.

To start with, we mapped the instruments simulated in each system and their dependency on the input data. We cross-checked the overlapping and non-overlapping input variables. ("Input variables" are variables present in the input data sets provided by the EUROMOD team.) The resulting matrix can be summarized as follows:

Variables in input data needed for simulations	Country tax-benefit system
W, Y, D	A
Z, Y, D	B

D are overlapping variables that do not need parametric modelling (e.g. demographics) and Y are overlapping variables that need modelling (e.g. employment, wage, occupation, contributory unemployment benefit). For the overlapping variables that would be swapped parametrically (e.g. contributory unemployment benefit) we estimate parametric models for the presence of the income source and the level. These will allow swapping the two distributions between countries when the two tax-benefit systems are swapped. W and Z are non-overlapping variables that require

Category	Definition	Level	Factor	Structure	Model
Old Age	Public pensions (contributory)	Individual	TB	Presence/ Level	logit, log-linear, EUROMOD
Survivor	Survivor pensions (contributory)	Individual	TB	Presence/ Level	logit, log-linear, EUROMOD
Disability	Benefits (contributory, non-contributory, invalidity)	Individual	TB	Presence/ Type/ Level	logit, log-linear, EUROMOD
Sickness	Benefits	Individual	TB	Presence/ Level	logit, log-linear, EUROMOD
Unemployment	Benefits contributory/ non-contributory	Individual	TB	Presence/ Type/ Level	logit, log-linear, EUROMOD
Maternity	Benefits	Household	TB	Presence/ Level	logit, log-linear, EUROMOD
Housing	Benefits	Household	TB	Presence/ Level	logit, log-linear, EUROMOD
Other Social Welfare	Benefits	Household	TB	Presence/ Level	logit, log-linear, EUROMOD

Table B–1: Modelling of non-simulated/partially simulated benefit instruments and supporting variables for EUROMOD system swaps

modelling, with W needed for system A and Z needed for system B. We estimate parametric models for W and for Z in the data where they are present. When system A (B) is imported in country B (A), we simulate the statistical distribution of variables W (Z) using the parametric relationship estimated in country A (B) and data of country B (A). For example, the number of months receiving unemployment benefit is required by the Irish system and is present in the Irish data, but it is not required by the UK systems and it is not present in the UK data. We estimate parametric models for receiving unemployment benefit and the number of months in the Irish data and when importing the Irish system in the UK, we preserve the conditional distribution of months receiving unemployment benefits from Ireland by simulating the reduced-form projections.

In order to fit all pieces of the puzzle, we built a standardized structure of equations for the non-simulated/partially simulated instruments and supporting variables (e.g. number of months receiving unemployment benefits) required in simulating taxes and benefits. We classified them in eight categories, as shown in Table B–1. Further information for country-specific details can be obtained by request.

Several variables are required to calculate benefits and tax liabilities. Although they are not part of household income *per se*, they are modelled as part of the model and are used in constructing counterfactual distributions by swapping their distributions depending on which component they belong to. Table B–2 describes the variables concerned.

Further information for country-specific details can be obtained upon request.

Variable	Definition	Level	Factor	Model	Conditioning variables
$IsHomeOwn_h$	Whether homeowner	household	Market	logit	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$(demographics, in work, employee)
$Mort_h$	Whether mortgage being paid and how much interest paid	household	Market	logit, log-linear	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, in work, employee)
$IsPubRent_h$	Whether benefits from public rent	household	Market	logit	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, in work, employee)
$PaysRent_h$	Whether it pays rent	household	Market	logit	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, in work, employee)
$Rent_h$	Value of rent	household	Market	log-linear	$x_{hi} I_{hi}^{lab}, I_{hi}^{emp}, IsPubRent_h$(demographics, in work, employee, is renter)
$Rooms_h$	Number of rooms	household	Market	log-linear	x_{hi} (demographics)
PPC_{hi}	Voluntary contributions to private pension plans	individual	Market	logit, log-linear	x_{hi} (demographics)
ChC_{hi}	Childcare cost	household	TB	logit, log-linear	$x_h, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, employee, with disability)
$Insurance_{hi}$	Insurance scheme	individual	TB	logit	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, labour market characteristics)
$Months_{w/em}$	Months in work/employed	individual	TB	logit/ log-linear	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, labour market characteristics)
$Months_{ub}$	Months receiving unemployment benefits	individual	TB	logit/ log-linear	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, labour market experience)
$Months_{unem}$	Months in unemployment	individual	TB	logit/ log-linear	$x_{hi}, I_{hi}^{lab}, I_{hi}^{emp}$ (demographics, labour market experience)

Table B–2: Non-income components entering tax calculations

C A model for labour supply responses to tax-benefit parameters

As labour supply behaviours respond to the taxation incentives, we also estimate a structural labour supply model to allow us to simulate the behaviour patterns in both countries labour market. The model endogenises individual labour supply and simulates behavioural changes induced by a change in the tax benefit system. We opted for discrete choice labour supply models, similar to van Soest (1995), Creedy and Duncan (2002) etc. Note that the labour supply response is only applied to those who are between age 18 and 60 years old who are neither in education nor retirement. Individuals outside of the age range, students, retirees are assumed to be inelastic in their labour supply.

Empirically, we use the direct quadratic utility specification as in Keane and Moffitt (1998)

$$u = \beta_c c + \beta_{cc} c^2 + \beta_{ch} c \cdot h + \beta_h h + \beta_{hh} h^2 + \beta_{ue} d_u + \beta_{iw} d_w$$

where c is the household disposable income, and h is the number of hours worked per week. The coefficients β are heterogeneous, varying linearly with several taste-shifters such as age, presence of children, marriage status etc. We also incorporate two 'state' dummies, namely one work dummy, which captures the cost of working, and one unemployment dummy d, which captures the cost of claiming unemployment benefit in the case of inactive. The fixed cost dummy (d_w) is a flexible way of specifying the utility, which may explain the skewed distribution in the number of working hours (van Soest et al., 2002), and to some extent captures the demand side constraints (Aaberge et al., 1995). The unemployment dummy (d_u) is expected to have a negative coefficient, reducing the overall utility should the individual be involuntarily unemployed. Similar to Colombino et al. (2010), there are five possible choices for each individuals, namely

$$y = \begin{cases} h = 0,\ d_u = 0,\ d_w = 0 & \text{inactive} \\ h = 0,\ d_u = 1,\ d_w = 0 & \text{unemployed} \\ h = 20,\ d_u = 0,\ d_u = 1 & \text{working part-time (20 hours per week)} \\ h = 40,\ d_u = 0,\ d_w = 1 & \text{working full time (40 hours per week)} \\ h = 50,\ d_u = 0,\ d_w = 1 & \text{working extended full time (50 hours per week)} \end{cases}$$

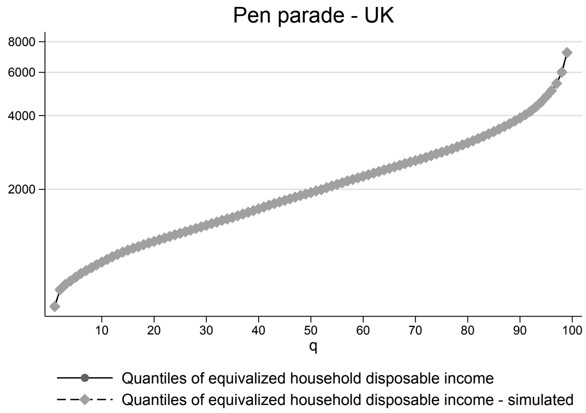
Our estimation is based on the tax unit which can be composed of one or two decision makers. Households with multiple families are split into multiple tax units based on the relationship status. A tax unit contains one adult and his or her dependent children in the case of single person household or single-parent, or two partnered adults and their children in the case of a married or de facto couple. We estimated singles and couples separately with a unitary utility function applied to couples. The model can be estimated under the conditional logit framework with a Type I extreme value distributed error term. As the residuals are latent, we draw the residuals for all non-chosen choices following the Type I extreme value distribution, and a truncated Type I extreme value distribution where $\epsilon_j > max(x_i\beta + \epsilon_i) - x_j\beta$ for all $i \neq j$ for the observed choices j. Such approach allows us to preserve the individual heterogeneities of their labour supply preferences.

To estimate the response to a tax-benefit change, we simulate the predicted labour supply choices of all individuals with the disposable income resulting from the new tax-benefit parameters.

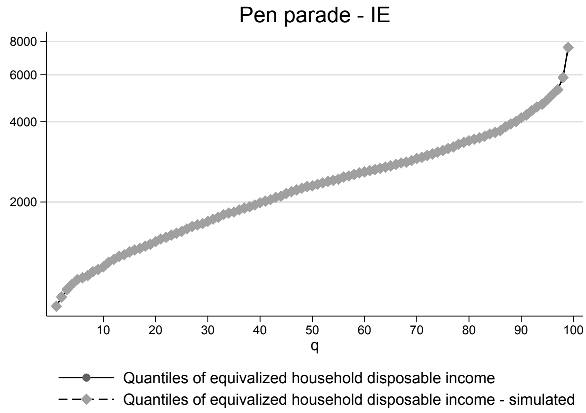
If the predicted outcome differs from the observed outcome under the original regime, we update the employment characteristics and the associated auxiliary variables, such as working hours s_{hi}, employment dummies and other characteristics in the income generation model (I_{hi}^{lab}), to match the predicted labour supply choice. The addition of the labour supply behavioural model alters the sequence of the simulation compared with what was described in Section 2. Unemployment, non-participation, and working hours are now jointly simulated instead of sequentially determined. The logic for the rest of the simulation remains unchanged. The simulated disposable income is used for further analyses on the inequality.

D Validation results

To validate our simulations we compare the original, observed distribution of equivalised household disposable income with the one resulting from a full simulation: fully simulated labour market, income and tax-benefit structure. The two distributions are compared in Figure D–1. For both countries, the simulation closely matches the original, observed distributions.



(a) UK



(b) IE

Figure D–1: Actual vs. Simulated Quantiles of Household Disposable Income (EURO)

LUXEMBOURG INSTITUTE OF
SOCIO-ECONOMIC RESEARCH

11, Porte des Sciences
Campus Belval
L-4366 Esch-sur-Alzette

T +352 58 58 55-1
F +352 58 58 55-700

www.liser.lu