Joonas Ollonqvist

Accounting for the role of tax-benefit changes in shaping income inequality: A new method, with application to income inequality in Finland

Aboa Centre for Economics

Discussion paper No. 122

Turku 2018

The Aboa Centre for Economics is a joint initiative of the economics departments of the University of Turku and Åbo Akademi University.

ISSN 1796-3133

Printed in Uniprint
Turku 2018

Joonas Ollonqvist

Accounting for the role of tax-benefit changes in shaping income inequality: A new method, with application to income inequality in Finland

Aboa Centre for Economics
Discussion paper No. 122
December 2018

ABSTRACT

This paper introduces a new method of analysing how the changes in tax-benefitt-system have been reflected in income inequality. This method is a combination of microsimulation based decomposition (Bargain and Callan, 2010) and a multivariate regression based decomposition (Fields, 2003; Yun, 2006). It allows analysing how the policy changes have affected the importance of different individual characteristics in income inequality. With the variance of log of incomes, the decomposition can be made further to separate the changes directly related to policy decisions from the overall price- and residual effects. This method is applied to analyse the evolution of income inequality in Finland from 1993 to 2014.

JEL Classification: D31, H24

Keywords: income inequality, microsimulation, regression, tax-benefit-system

Contact information

Joonas Ollonqvist Department of Economics
University of Turku
FI-20014, Finland
Email: joonas.ollonqvist (at) utu.fi

Acknowledgements

I sincerely thank Kaisa Kotakorpi, Matti Viren, Markus Jäntti, Momi Dahan, Ilpo Suoniemi, Oskari Vähämaa and Erik Mäkelä for valuable comments and discussions. I thank the participants in the NORFACE WSF Final Conference (2018), D06: Inequality session in IIPF conference (2018), FDPE Public Economics and Labour Economics workshop II/2017 and ACE-workshop (2017) for their helpful comments and suggestions. I thank NORFACE for funding this research. All errors remain my own.

1 Introduction

The rising inequality is one of the main concerns in the modern world. During the past decades, income inequality has risen substantially among the developed countries (OECD, 2011; Atkinson and Bourguignon, 2015) and Finland is not an exception in this matter (see Figure 1). There are many different reasons behind this evolution, but the question is not completely solved. Crudely speaking there are four different issues which affect income distribution: 1) changes in sociodemographic characteristics, 2) changes in importance of different characteristics on individual's income, 3) changes in politics and 4) business cycles. What makes the analysis difficult is that these factors may also interact with each other. Raising labour taxes, for example, lowers the value of being employed. Especially from a policy maker's perspective it is important to be able to separate the factors driving the evolution of economic inequality, since the policy actions will vary depending on the reason. Currently there is no method for isolating the effect of policy changes (changes in income inequality that are accounted to the changes in tax-benefit-system) from the overall evolution of importance of individual characteristics in income inequality. To fill this gap in the literature, I propose a new method, which is a synthesis of two decomposition methods used in the literature.

Generally, the policy effects are analysed by using microsimulation based methods (Bargain and Callan, 2010; Bargain, 2012a,b; Herault and Azpitarte, 2016) and the effects of different characteristics can be analysed by a multivariate regression based decomposition (Fields, 2003; Yun, 2006).[1] In the literature of microsimulation the focus is on questions like '*How changes in tax-benefit-systems have affected income distribution?*'. With the microsimulation methods it is possible to isolate the policy effect on income inequality from the other effects and it can be extended to analyse the behavioural effects of the policy changes. However, these methods do not tell anything about how these changes affect the importance of different characteristics in income inequality. Whereas the multivariate regression based decomposition tells '*How individual/household characteristics affect income inequality?*'. Also, it allows to further analyse whether the evolution has been driven by the price-, quantity- or residual effect.[2] The downside of this method is, that it cannot distinguish whether or not the changes are driven by the changes in politics.

My proposed method combines the benefits of both methods and the synthesis of these two is surprisingly simple to form. This new method tells how the changes in tax-benefit-system have affected the importance of each characteristic in income inequality. With the variance of log of incomes, this method also allows isolating the policy effect from the total price- and residual effect, providing more information about the reasons behind the evolution of these two. For instance, it can be analysed how much of the change in education premium is accounted to tax-benefit changes and how it has affected the income distribution.

[1] For an extensive overview of other decomposition methods used analysing distribution of incomes see Fortin, Lemieux and Firpo (2011).

[2] Price effect is the effect caused by a change in the importance of a variable on income and the quantity effect is caused by the change in the distribution of a characteristic among the population.

The proposed method is applied to study the evolution of income inequality in Finland and there are several reasons why Finland is an excellent subject to study. Figure 1 illustrates the evolution of Gini coefficient for original and simulated datasets from 1993 onwards. During the 1990's inequality rose rapidly in Finland and since 2005 only small changes have happened. From the earlier research with Finnish data, we know that policy changes have increased income inequality (Bargain and Callan, 2010; Honkanen and Tervola, 2014), which is somewhat different from the observations found in the international context (see Hills, Paulus, Sutherland and Tasseva (2014); Figari, Paulus and Sutherland (2015)). However, policy changes are not the

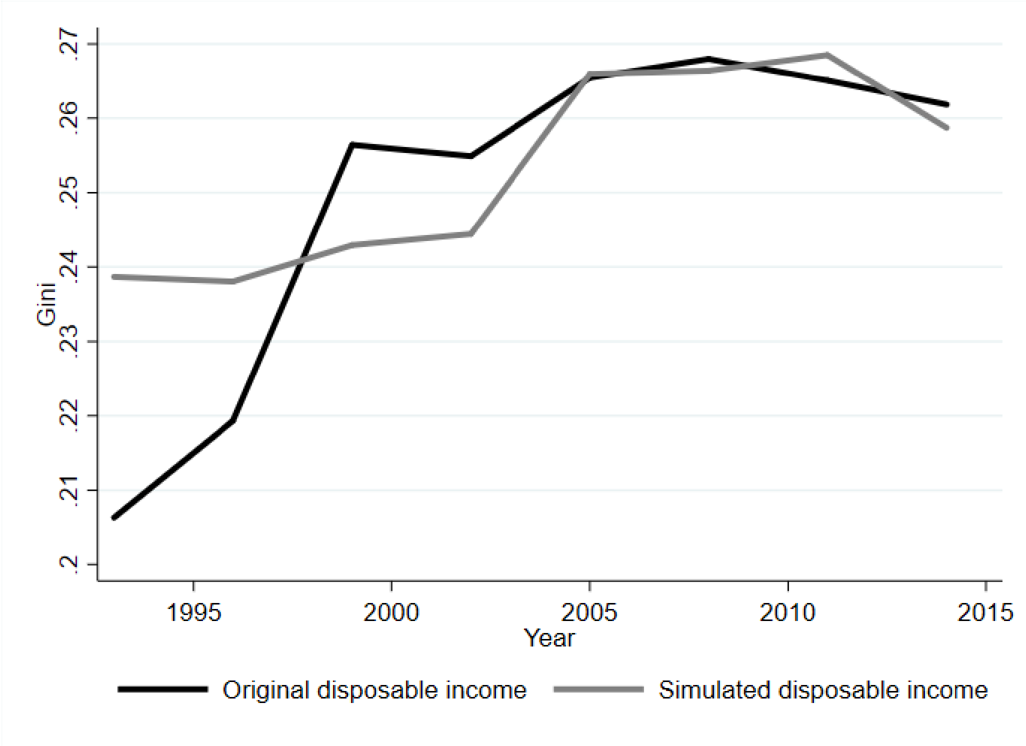

Figure 1: Gini coefficient in Finland

Note: Majority of the simulations are done using data from year 2011. Simulated year 2011 is formed by using data from year 2014

Source: Author's own calculation based on service data of income distribution and SISU-model

only reason behind the change in inequality and one other explanation is the rise in capital incomes among the top earners (Riihelä, Sullström and Tuomala, 2010). Also, the collapse of Soviet Union was a large unexpected shock to Finnish economy and it changed the composition of it (Gorodnichenko, Mendoza and Tesar, 2012). Finally, the distribution of sociodemographic characteristics in Finland have changed considerably since the early 1990's. For example, in 1993 only 4.6% of the over 15 year old had master's degree or higher and in 2014 the share was 10.9%. These changes in the distribution of the characteristics may have altered the income distribution, but also the importance of these characteristic in individual incomes may have changed as well.

For these reasons, I am using this proposed method to analyse the evolution of income inequality in Finland. My first aim is to study how much inequality is accounted to different individual/household characteristics and how policy changes have altered it. Second aim is to investigate which effects have driven the evolution. I find that price-, quantity and residual effects all explain around 1/3 of the evolution in income inequality. I also find that policy changes had a significant role in the rise in inequality, but since 2005 those have equalised the income distribution. My final aim is to analyse how policy changes have altered the income distribution. I find that policy changes have mostly changed the importance of individual characteristics, since the majority of the price effect can be traced to policy changes. Still, policy changes have also affected income distribution in a way that cannot be explained.

This paper is organised as follows: in Section 2 is presented microsimulation- and multivariate regression based decompositions and the synthesis of these two. Third section discusses the data and the empirical strategy and shows the decomposition results for Finland. Then the final, fourth, section is for conclusions.

2 Unified framework

2.1 Decomposition of policy effect with microsimulation

First, defining some notation and terminology. Household's socio-economic characteristics in year j are described by vector $\mathbf{X}_j$ and their original (gross) incomes (in year j) are denoted by a vector $\mathbf{Y}_j$. Following Bargain and Callan (2010) and Figari et al. (2015), I distinguish between the tax-benefit function (e.g. the rules of the taxation and benefits) and the monetary parameters (e.g. tax brackets). Tax-benefit system k is a function defined as $f_k(\mathbf{X}, \mathbf{Y}, m_k)$, where parameters m_k are the monetary parameters used in tax-benefit system. Household disposable income in year k with the tax-benefit system from year j is then

$$\gamma_j(\mathbf{X}_k, \mathbf{Y}_k, m_j) = \mathbf{Y}_k + f_j(\mathbf{X}_k, \mathbf{Y}_k, m_j)$$

In this paper, I restrict the attention to the static effects of policy changes, but it is possible to take into account the behavioural (indirect) effects of policy changes as is done in Bargain (2012a,b). Therefore, the direct effect of policy changes from A to B on household disposable income, while keeping the original incomes and characteristic unchanged, is

$$\Delta\gamma = \gamma_B(\mathbf{X}_A, \mathbf{Y}_A, m_B) - \gamma_A(\mathbf{X}_A, \mathbf{Y}_A, m_A)$$

This is the so called "morning after" policy effect. These kinds of calculations are usually used to form counterfactuals in "what if" kind of setup. For example, when analysing what would be household's disposable income in year B if we had the tax-benefit-system from the year A, denoted as γ_B^A (To shorten notation $\gamma_A^A = \gamma_A$). However, in these kinds of studies we need to adjust the policy parameters according to the changes in prices or mean incomes. Let α be this adjusting factor.[3] Now, when using end period B data (denoting the initial period with A) the change in disposable income is

$$\begin{aligned}\Delta\gamma =& \gamma_B(\mathbf{X}_B, \mathbf{Y}_B, m_B) - \gamma_A(\mathbf{X}_B, \mathbf{Y}_B, \alpha m_A)\\ =& \gamma_B - \gamma_B^A\end{aligned}$$

and with the initial period data A it is

$$\begin{aligned}\Delta\gamma =& \gamma_B(\mathbf{X}_A, \mathbf{Y}_A, \alpha^{-1} m_B) - \gamma_A(\mathbf{X}_A, \mathbf{Y}_A, m_A)\\ =& \gamma_A^B - \gamma_A\end{aligned}$$

Where the subscript of γ denotes the year of original incomes and characteristics (i.e. population) and superscript of γ indicates the year of the policy-parameters and the tax-benefit function.

Then using similar notation, the effect of these policy changes to some inequality measure I is

$$\begin{aligned}\Delta I =& I\left[\gamma_B(\mathbf{X}_B, \mathbf{Y}_B, m_B)\right] - I\left[\gamma_A(\mathbf{X}_B, \mathbf{Y}_B, \alpha m_A)\right]\\ =& I_B - I_B^A \end{aligned}\tag{1}$$

and

$$\begin{aligned}\Delta I =& I\left[\gamma_B(\mathbf{X}_A, \mathbf{Y}_A, \alpha^{-1} m_B)\right] - I\left[\gamma_A(\mathbf{X}_A, \mathbf{Y}_A, \alpha m_A)\right]\\ =& I_A^B - I_A \end{aligned}\tag{2}$$

As Bargain and Callan (2010) points out, this decomposition gives the absolute policy effect on income distribution and it is possible to conduct the decompositions with either end or base year data. They also argue, that if there is access to both the end and base year data, the relative policy effect can be formed by using Shorrocks-Shapley decomposition.[4] The Shorrocks-Shapley decomposition is in this case just the average of these two effects:

$$\begin{aligned}\Delta_P =& \frac{1}{2}\left[I_B - I_B^A\right] + \frac{1}{2}\left[I_A^B - I_A\right]\\ =& \frac{1}{2}\left[I_B - I_B^A + I_A^B - I_A\right]\end{aligned}\tag{3}$$

In other words, the above equations capture the direct policy effect on income inequality when the other factors are kept constant.

[3]How to choose the parameter α is discussed in Hills et al. (2014)

[4]See for example Shorrocks (2013).

2.2 Multivariate regression based decomposition

Following the work of Fields (2003), the observed inequality can be decomposed to the contributions accounted to each household/individual characteristics. This decomposition can be formed with virtually any inequality measure. However, according to Yun (2006), with the variance of log of incomes the decomposition can be made further to distinct the price-, quantity- and residual effects from each other.

The procedure of the decomposition is simple. First, the income generating function is estimated by using the OLS:

$$y_i = \sum_{c=0}^{N} \beta_c X_{ci} + \varepsilon_i$$

where $y_i = \ln(\gamma_i)$ is the log of household disposable income, β:s are the regression coefficients, X:s represent the set of household/individual characteristics ($X_i \in \mathbf{X} \quad \forall i$) and ε is the error term. To ease the notation, I suppress the individual subscripts in the equations.

Then the fitted values of the estimation are used to form the relative characteristics inequality weights:

$$s_c = \frac{\text{cov}(\beta_c X_c, y)}{\sigma^2(y)}, \tag{4}$$

These weights are invariant of the choice of inequality and the share of the residual can be calculated the same way. The weights are scaled down with the variance of log of incomes and so these shares present the percentage shares accounted to each characteristic. Thus, those will sum up to one.

The absolute contribution of each characteristic is just the product of the calculated weight and the value of the inequality measure: $S_c = s_c I$.[5] Then, the absolute change caused by some characteristic c is

$$\Delta S_c = s_{cB} I_B - s_{cA} I_A$$

And the total change can be expressed as:

$$\Delta I = \sum_{c=1}^{N} (s_{cB} I_B - s_{cA} I_A) \tag{5}$$

Where subscript A denotes the base period and B denotes the end period.

Yun (2006) shows that it is possible to do the decomposition further to separate the quantity-, price- and residual effects, when using the variance of the log of incomes as the inequality measure.[6] Then the decomposition takes the form:

[5] Here I indicates an inequality measure that is calculated using disposable income.

[6] One problem with the variance of log of incomes is that it cannot be guaranteed to satisfy the Pigou-Dalton

$$\Delta\sigma_y^2 = \sum_{c=1}^{N}\left(s_{cB}\sigma_{y_B}^2 - s_c^*\sigma_{y^*}^2\right) + \sum_{c=1}^{N}\left(s_c^*\sigma_{y^*}^2 - s_{cA}\sigma_{y_A}^2\right) + \sigma_{\varepsilon_B}^2 - \sigma_{\varepsilon_A}^2$$
$$= \Delta^Q + \Delta^P + \Delta^\varepsilon \tag{6}$$

where A and B are defined as before, σ_y^2 is the variance of the log of incomes, superscript $*$ refers to values that are formed using an auxiliary income distribution, where the coefficients of characteristics are replaced while keeping the characteristics intact. Formally defined as:

$$y^* = \sum_c \beta_{cB} X_{cA} + \varepsilon_A \tag{7}$$

The first terms in equation (6) capture the *quantity effect*, the second ones are the *price effect* and the last terms present the change accounted to the changes in the residual.[7] Price effect is the effect caused by a change in the importance of a variable on income (the rise (decrease) in the price effect means that the particular variable has become the more (less) important determinant of an individual's income). Whereas the quantity effect is caused by the change in the distribution of a characteristic among the population (the rise (decrease) in the quantity effect means that the particular variable has become more (less) unequally distributed among the population).

There is also other possible way to form the auxiliary income distribution, it can be formed by replacing the characteristics while keeping the coefficients intact:

$$y^{**} = \sum_c \beta_{cA} X_{cB} + \varepsilon_B \tag{8}$$

With the auxiliary income distribution defined in equation (8) the decompositions of the price-, quantity and residual effects takes the form:

$$\Delta\sigma_y^2 = \sum_{c=1}^{N}\left(s_{cB}\sigma_{y_B}^2 - s_c^{**}\sigma_{y^{**}}^2\right) + \sum_{c=1}^{N}\left(s_c^{**}\sigma_{y^{**}}^2 - s_{cA}\sigma_{y_A}^2\right) + \sigma_{\varepsilon_B}^2 - \sigma_{\varepsilon_A}^2$$
$$= \Delta^P + \Delta^Q + \Delta^\varepsilon \tag{9}$$

Therefore price- and quantity effects presented can be calculated in three ways: 1) using equation (6), 2) using equation (9) or 3) taking average of these two. The last one corresponds to Shorrocks-Shapley decomposition and it uses more information than the other two. Also, there is no particular reason to prefer the first or the latter decomposition and therefore in this paper the price- and quantity effects are formed using the average of equations (6) and (9).[8]

The Fields' method can be extended like factor source decomposition (presented in Shorrocks

principle of transfers.

[7]Due to construction of OLS, $s_\varepsilon\sigma_y^2 = \sigma_\varepsilon^2$.

[8]In Appendix G are shown the price- and quantity effects formed according to equations (6) and (9).

(1982)) to distinguish between the pure- and interaction effects of character c in income inequality.[9] For this reason, the price- and quantity effects are possible to decompose the similar way.[10]

2.3 Synthesis

In this paper, I am only considering the static effects of policy changes and therefore, policy change has two ways to affect the income distribution: i) it may alter the importance of the characteristics (i.e. price effect) and/or ii) change the explanatory power of the characteristics (i.e. change the residual).[11] For this reason, combining these two methods is straight forward to do to form the price- and residual effects of the policy change on each characteristic. As before, there exist three possible combinations to analyse: 1) using end period data, 2) using initial period data or 3) using both.

First, the income generating function is estimated and the relative characteristic inequality weights are formed according to equation (4) with the simulated and original datasets. (Denoting $\ln(\gamma) = y$.) Then, accordingly combining equations (5) and (1) gives the decomposition with the end period data.

$$\begin{aligned}\Delta I_{PB} =& \sum_{c=1}^{N} \left[s_{cB} I_B - s_{c_B^A} I_B^A \right] + s_{\varepsilon_B} I_B - s_{\varepsilon_B^A} I_B^A \\ =& \Delta_{PB}^{P} + \Delta_{PB}^{\varepsilon},\end{aligned} \tag{10}$$

where subscript B indicates which year's population is used in the decomposition. Similarly for the initial period data the following decomposition holds:

$$\begin{aligned}\Delta I_{PA} =& \sum_{c=1}^{N} \left[s_{c_A^B} I_A^B - s_{c_A} I_A \right] + s_{\varepsilon_A^B} I_A^B - s_{\varepsilon_A} I_A \\ =& \Delta_{PA}^{P} + \Delta_{PA}^{\varepsilon}\end{aligned} \tag{11}$$

and when having access to both initial and end period data we get the following:

$$\begin{aligned}\Delta I_{PAB} =& \frac{1}{2} \left[\sum_{c=1}^{N} \left[s_{c_B} I_B - s_{c_B^A} I_B^A + s_{c_A^B} I_A^B - s_{c_A} I_A \right] \right] \\ &+ \frac{1}{2} \left[s_{\varepsilon_B} I_B - s_{\varepsilon_B^A} I_B^A + s_{\varepsilon_A^B} I_A^B - s_{\varepsilon_A} I_A \right] \\ =& \Delta_{PAB}^{P} + \Delta_{PAB}^{\varepsilon}\end{aligned} \tag{12}$$

In all of the above decompositions, the term inside the sum operator is the change in the absolute contribution of each character c caused by the policy change. This change can be

[9]Details are shown in Appendix A.
[10]Proof is shown in Appendix B
[11]In a dynamic setting, it could also have an impact on the gross incomes and distribution of characteristics, namely labour market status.

interpreted as the price effect of the policy change on income inequality, since the population and original incomes are kept intact.[12] The second terms capture the change in the contribution of the residual term accounted to the policy change.

Before this, I have not made any specific assumptions about the inequality measure. It is possible to conduct the above decompositions with virtually any well behaving inequality measure as is the case with Fields' decomposition. When using the variance of the log of incomes as the inequality measure it is possible to compare the total price effect obtained from equation (6) with the price effect caused by the change in the tax-benefit-system (equations (10), (11) and (12)). The same applies to residual effect.

The price effect calculated by the equation (6) or (9) is the total price effect between periods A and B conditional to base or end year data. Those can be expressed as a sum of the change caused by policy changes and the change caused by other factors and the same holds for their average as well. Therefore, with the variance of the log of incomes the following holds:

$$\Delta^P = \Delta^P_P + \Delta^P_O, \tag{13}$$

Where Δ^P is the total price effect, Δ^P_P is the part caused by policy change and Δ^P_O is the price effect caused by other factors than the changes in the tax-benefit-system. Δ^P and Δ^P_P are obtained from the earlier decompositions and, thus, it is possible to calculate the Δ^P_O. The same can be done with the residual effect:

$$\Delta^\varepsilon = \Delta^\varepsilon_P + \Delta^\varepsilon_O, \tag{14}$$

As before, Δ^P_P and Δ^ε_P can be calculated either with initial or end period data or using both. However, both initial and end period data is required to form the total price- and residual effects. Also, the price effect can be formed in three possible ways (equation (6), (9) or both). Therefore, there is 9 possible combinations to form Δ^P_O and three possible combinations to form Δ^ε_O.

Even though, here I am only studying the total effects of policy changes, the same equations are applicative to study how some of the changes in the tax-benefit system have affected. For example, it could be in our interest to study how the changes in the benefit side or taxation have separately affected.

3 Application to income inequality in Finland

3.1 Data and empirical strategy

The analysis is performed with triennial cross-sectional data from 1993 to 2014. The data used is the service data of income distribution collected by Statistics Finland, which can be

[12]This is shown formally with the variance of logs in the appendix C.

(partly) used with Finnish microsimulation model SISU (Official Statistics of Finland (OSF), 2014). The yearly sample size is around 25 000 individuals in approx. 10 000 households and it includes a large amount of information about the individual/household characteristics and their incomes.

The analysis is conducted in two steps. First, counterfactuals are formed by using SISU microsimulation model. The majority of the Finnish tax-benefit system is encoded in SISU model from year the 1993 onwards. However, the SISU model is not compatible with the data before the year 2011 and therefore the simulations are carried out by using data from the years 2011 and 2014. 2011 data is used to form counterfactuals for years 1993, 1996, 1999, 2002, 2005, 2008 and 2014 whereas counterfactual for year 2011 is done by using the data of 2014.

Doing the simulations this way means I can not calculate the policy effect of each sub period as is presented in equation (1) or (2).[13] Therefore I have to approximate the policy effect between these years. So, keeping the same notation as earlier, but looking three different years: A, B and C, where C is the only year applicative to microsimulation. Then, the policy changes $C \rightarrow B$ and $C \rightarrow A$ are:

$$\begin{aligned} \Delta_{P1} =& I_C - I_C^B \\ \Delta_{P2} =& I_C - I_C^A \end{aligned}$$

The difference between these two then gives the approximation of the policy change $B \rightarrow A$ conditional to the data from year C:

$$\Delta_P = I_C^B - I_C^A$$

and, in this paper, this approximation is used to form the policy effect of each subsequent period before the year 2008.

As mentioned earlier, the value of money changes over time and for this reason the monetary parameters are not straight applicative from different years with data from other year. There are at least three different possibilities how to choose the adjusting parameter: 1) the Consumer price index (CPI), 2) the Market income index (MII) or 3) no indexation. Differences between these three and how those affect counterfactuals are discussed more details in Hills et al. (2014).

I use the CPI because of a couple of reasons. First, usually benefits in Finland, if those are tied, are tied to CPI. Second, using CPI means that every family can afford to buy the same basket of goods over time. So, the policy changes reflects better the welfare changes associated with the changes in legislation. However, as with any index, there are problems associated with the CPI. First, not every benefit is tied to index. Meaning that, adjusted simulated benefits may be higher than they really should be. Second problem is related to growth in incomes. If incomes grow the higher rate than than the CPI, households on benefits lose relatively compared

[13]For example, the policy effect from the year 1993 to 1996 cannot be done that way since the earliest data year which can be simulated is 2011.

with families with earnings. Indicating that the tax revenues grow faster than the expenditures of benefits. Also, families with earnings are better off in real terms since tax-brackets do not grow as fast as incomes.

In the second step for both actual data years and counterfactual dataset are analysed according to Fields' and Yun's methods. I follow the example of Brewer and Wren-Lewis (2016), all the characteristics are transformed to a set of indicator variables and these variables are used in the analysis. At the end relative shares are summed up back together to form the total effect of each characteristic. This is possible to do, because of the additive nature of the method.

The decompositions are conducted with the log of household's disposable income scaled by the modified OECD equivalence scale. The decompositions are done at the individual level by using household level weights multiplied with the number of people in the household. Only household heads are used in the analysis, but the household head is given the information about any spouses.

There have been changes in the variables during the time span of the analysis. There are new variables and for some of the variables the definitions have changed. In the analysis I only use variables that are found in every yearly sample[14] and I try to make them comparable between each sample year. Therefore some variables are recoded.[15] The most crucial change in the variables happened in 1997 when the definition of the level of education changed. Before 1997, secondary schooling was divided into two categories: lower- and upper secondary schooling. Since the 1997 there has been only one category in secondary schooling and at the same time some of the educations were moved from the upper secondary to the lowest- or lower tertiary levels. Engineers were moved from the upper secondary school to the lower tertiary level and nurses were moved from upper secondary school to the lowest tertiary level, for example. It is impossible to reliably identify the person's field of education before 1997 and therefore the results for secondary schooling and lowest/lower tertiary level education are not entirely comparable before and after 1997.

3.2 Results

To better illustrate the evolution of income inequality, in Figure 2 is presented the indexed evolutions of the variance of logs and the Gini coefficient for both the actual and simulated data. It can be seen that the changes of the simulated data are clearly smaller than the actual evolution, indicating that both policy changes and other factors have had an important role in the evolution of inequality. This holds with both the Gini coefficient and the variance. The evolutions of the Gini coefficient and the variance differ in levels, but the overall trends are similar between these two. Therefore, using the variance of log of incomes should not distort

[14]For this reason some of the relevant variables are not included in the analysis, for example the field of the education is excluded.

[15]Full information about the variables is listed in Appendix E.

the results.

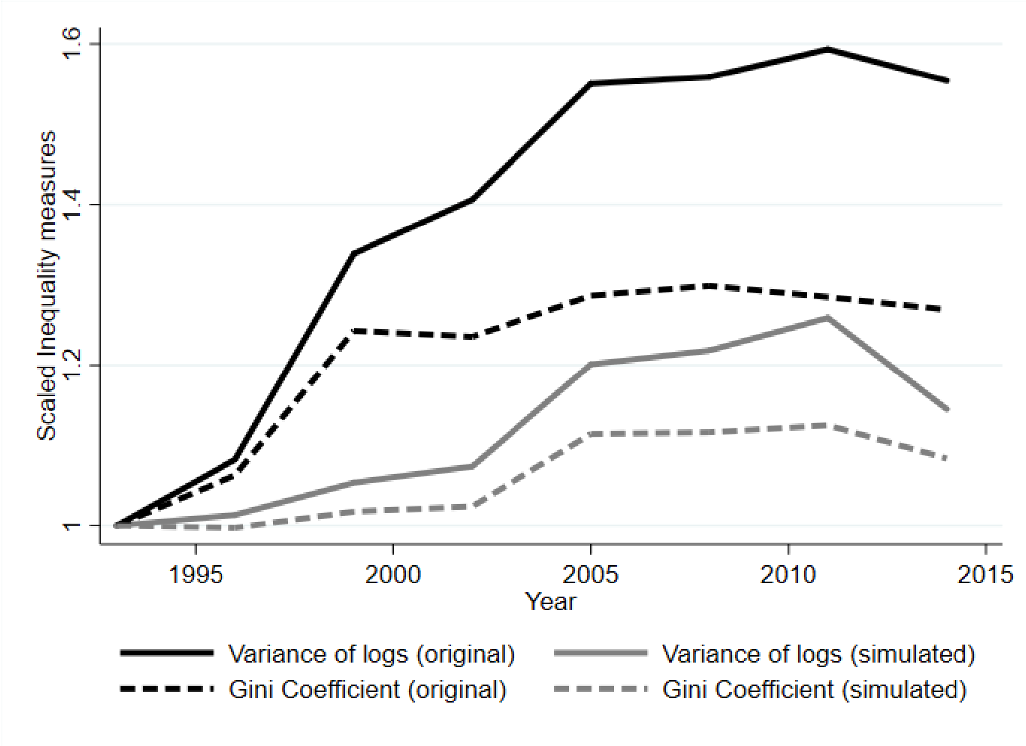

Figure 2: Scaled inequality measures for actual and simuated data in Finland
Note: Majority of the simulations are done using data from year 2011. Simulated year 2011 is formed by using data from year 2014. All the values have been scaled by the base year values, so in year 1993 every measure gets the value 1.
Source: Author's own calculation based on service data of income distribution and SISU-model

The largest difference between original and counterfactual datasets is related to the timings of the rapid changes. As shown earlier, inequality rose rapidly in Finland during the 1990's but with the simulated data only a small increase in inequality can be seen during that time. The majority of the changes in inequality with the simulated data can be traced after the year 2002. This indicates, that before 2002 the majority of the change is associated with the factors other than policy changes. After 2002 it seems that this has changed and the policy changes have become the dominant driver of the evolution of income income inequality.

Now inequality is decomposed by six individual (age, employment status and education by both genders) and two household characteristics (region and household type). The decomposition is done by using the characteristics of the household heads and the characteristics of any partner of the household head. For each of the characteristics, indicator variables are created according to which subgroup the individual belongs to. Then, the total contribution by each group is formed as a sum of the shares of the indicator variables belonging to that particular group.[16]

Table 1[17] presents the shares of each characteristic and the share of the residual with the actual and simulated data (formed according to equation (4)). According to Table 1 the largest

[16]See Appendix E.
[17]The shares and absolute contributions for each year are shown in Appendix F.

share with both datasets is accounted to the residual meaning that there is a lot of variation that cannot be explained with these characteristics. However, the share of the residual has substantially decreased so the explanatory power of the variables used has increased. In 1993 only 40.0% could be explained, but in the year 2014 the same variables explained 50.2%. With the simulated data, the percentages are 45.2% and 48.1%.

Table 1: Shares of characteristics in income inequality (%)

Actual data									
			Household	Age		Employment status		Education	
Years	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	60.0	3.1	4.5	5.6	3.7	5.2	6.9	6.2	4.8
2002	55.2	1.9	3.8	6.6	3.7	8.1	7.1	8.3	5.2
2014	49.8	2.1	2.2	9.0	3.7	9.3	6.9	8.7	8.3
Simulated data									
			Household	Age		Employment status		Education	
Years	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	54.8	1.9	3.1	7.5	2.4	7.6	7.6	7.9	7.1
2002	51.3	1.9	3.1	7.5	2.4	8.9	8.9	8.4	7.6
2014	51.9	2.0	3.1	7.5	2.6	8.5	8.0	9.0	7.4

Note: Simulated years are formed using data from the year 2011. 95% confidence intervals are shown in Appendix G.

For both datasets employment status, education and age of males explain large fractions of the observed income inequality, whereas region, household type and age of females do not. Difference between the shares calculated with the actual and simulated data is how those have evolved. Actual changes in the contributions of residual, household type, education, age of males and male employment status are statistically significant. Whereas, with the simulated data, none of the change in the relative contribution of the characteristics is statistically significant.

Employment status has several interesting features. First of all, its' role, especially among males, has increased over time and this change is statistically significant. Also, policy changes have some effect on its' contribution to income inequality and a little surprising is that the female employment status was more important factor than males' employment status in income inequality, but no more in the 2000s.

The most interesting finding in Table 1 is the role of education. For both genders the contribution of education in income inequality has increased, but those have evolved differently. The majority of the change in the contribution of male education happened before 2002 and since that nearly no changes. For females', it is the other way around.

Age is the only factor that behaves completely different between the genders. The males' age makes a clear (both relative and absolute) contribution to income inequality and it has become a more important factor since 1993. Whereas the role of females' age is clearly smaller and its' importance have remained fairly constant since 1993. These findings can not be traced to policy changes, since the role of age for both genders does not vary between the simulated years.

Household type is generating less inequality in 2014 than in 1993. Only at the beginning of the 2008 crisis its contribution peaked. This indicates that the recession hit differently according to the type of the household. However, the relative contribution has decreased and not even in 2008 the relative contribution was higher than in 1993. Surprisingly with the simulated data the role of the household type has not changed since 1993. This is surprising since the majority of the benefits in Finland are related at some level to household type.[18] One explanation for this might be that the amounts of household related benefits are small at the aggregate level and therefore those have only small effect on income inequality.

Next the absolute contributions of each characteristic are decomposed to the price-, quantity and residual effects according to the average of equations (6) and (9). These results along the side with the total policy effect are shown in Table 2.[19] After that, the price- and residual effects are further decomposed to separate the evolution accounted to policy changes (see equations (13) and (14)) from the other effects. These findings are reported in Table 3. Approximately 36.6% of the change in income inequality is accounted to price effects, 31.9% to quantity effects and 31.5% to residual effect. Policy effects explain 46.9% of the change in income inequality, but as mentioned, these are already included in the price- and residual effects.

Before moving forward, some cautionary notes are in place. The results of the policy effect are highly sensitive and probably not robust for, at least, three reasons. First, as Bargain and Callan (2010) mention, the longer the time span between the data and legislation year the more inaccurate the results will be. This inaccuracy can be reduced by using more data to smaller the distance between the legislation and data year. Also, using both initial and period data will reduce inaccuracy, but, unfortunately, neither one is possible to do with the microsimulation model in hand. Second, the choice of the index will affect the results especially with the long time span. Solution to this problem is the same as previously: reducing the distance between the legislation and data year. Third, Pigou-Dalton principle of transfers cannot be guaranteed to hold with the variance of log of incomes. Using more frequent data cannot solve this problem and it is an undesired feature of the method proposed here. Luckily, it seems that the calculated policy effects are fairly similar between the Gini coefficient and variance of log of incomes (see Figure 2 Appendices F and G) and this should not change drastically the results.

The price-, quantity-, residual- and policy effects have not evolved steadily over time and some of them have turned their direction. For instance, residual effect accounts the largest share in the evolution of income inequality before the year 2002, but after that it has decreased inequality. The price and quantity effects have increased inequality during the whole time span, but their magnitudes have become smaller in the second half. Especially price effect has been substantially smaller since 2002 than before that. Policy effect, on the other hand,

[18]For instance, also the incomes of your spouse determines if you are eligible to receive some benefits and how much you are eligible to get. Also, some benefits are directly targeted to certain types of households, for example the child allowance.

[19]One thing to be noted is that, price- and quantity effects are not time additive, whereas the other effects calculated are. Therefore, for these two effects, the yearly decompositions presented in Appendix F will not sum up to the results shown in Table 2.

Table 2: Changes in income inequality, price-, quantity and residual effects (1000x variance of log of incomes)

Years	Total Change	Price Effect	Quantity Effect	Residual Effect	Policy Effect	Region		Household type		Age				Employment status				Education			
										Male		Female		Male		Female		Male		Female	
						P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q
1993 - 2002	59.8	19.3	14.5	26	13.5	-0.7	0.1	1.1	0.1	2.7	2.8	0.6	1.7	10.2	-1.2	2.3	2.3	3.7	4.5	-0.5	4.1
2002 - 2014	21.8	10.1	12	-0.3	24.7	0.9	-0.2	-2.4	-0.2	5.1	1.7	-0.7	1.5	3.4	1.2	1.6	-0.6	-1.5	4.1	3.9	4.5
1993 - 2014	81.6	29.9	26	25.7	38.3	0	0.1	-1	-0.4	7.7	4.6	0	3.1	13.5	0.1	3.7	1.9	3.2	7.5	2.8	9.2

Note: Negative values indicate that on average the effect has negative contribution to income inequality. Annual decompositions are presented in appendix F.

Source: Author's own calculations based on the service data of income distribution.

Table 3: Price and residual effects decomposed to policy and other effects (1000x variance of log of incomes)

Years	Price effect		Residual effect		Region		Household type		Age				Employment status				Education			
	Total								Male		Female		Male		Female		Male		Female	
	Δ_O^P	Δ_P^P	Δ_O^ε	Δ_P^ε	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P
1993 - 2002	6.2	13.1	25.6	0.4	-0.9	0.1	0.6	0.5	1.7	1	0.1	0.5	6.7	3.5	-1.3	3.6	1.7	2	-2.4	1.9
2002 - 2014	3.7	6.4	-18.6	18.3	0.5	0.4	-3.7	1.3	3.4	1.8	-2.1	1.4	3.1	0.3	2.7	-1.1	-3.5	2	3.5	0.4
1993 - 2014	10.4	19.6	7	18.7	-0.6	0.6	-2.8	1.8	5	2.8	-1.9	1.9	9.8	3.7	1.2	2.5	-0.9	4	0.5	2.3

Note: Negative values indicate that on average the effect has negative contribution to income inequality. Annual decompositions are presented in appendix F. Change from 2011 to 2014 is calculated as an average of these two.

Source: Author's own calculations based on the service data of income distribution and SISU-model.

is significantly higher during the second half than in the first half. Since 2002, policy effect has been even larger than the total rise in inequality. However 2/3 of the total policy effect is accounted to the year 2005 and whether to include the year to the first or second half will change the results. Since 2005 policy effects have had an equalising effect on income inequality and the same holds also with the Gini coefficient.[20]

When, looking the channels of how policy changes have affected the inequality interesting aspects come up. First of all, from 1993 to 2002 policy changes were solely affecting the importance of different characters on individual's income. During that period around 2/3 of the price effect was accounted to policy changes. During the second half, the channel turned upside down. Still, policy changes accounted about 2/3 of the total price effect, but the majority effect and the residual effect accounted to policy changes cancels each other out and therefore the residual effect has been nearly zero since 2002. In other words, since 2002 the policy changes have affected the income distribution mostly in a way that can not be explained.

Employment status, especially employment status of men and education are affecting the most to the change. From 1993 to 2002 changes in education accounts for around 19.7% of the total change in inequality and from 2002 to 2014 it accounts for about 50.5%. For employment status, the numbers are 22.7% and 25.7%. The most striking finding is the differences in quantity and price effects between these two. The changes in the employment status of men are solely due to the price effect and it single hands accounts 45% of the total price effect. Whereas the quantity effect accounts for the majority of the change in education. This indicates that the changes in the distribution of education have increased income inequality. Education has also become more important determinant of individual's income and therefore it is generating more inequality.[21] Another interesting difference between these two is that share of the price effect accounted to policy changes vary. Only 1/4 of the price effect of the employment status of men is accounted to policy changes. Whereas for the education and employment status of women the price effect is almost solely driven by policy changes.

Here also, the role of age has interesting features in explaining the changes. The age of males has clearly affected the evolution. In the first half, it accounts 9.2% of the change in inequality and during the second period the share increased to 31.2%. The changes in the age distribution of both genders are generating more inequality. However, only for males age has become more important factor in individuals' income. The part of price effect accounted to policy changes is almost the same between genders, but other factors have driven the evolutions to opposite directions. For males other effects have increased the importance of age in income, but for females it has decreasing effect. This may be due to the fact that in Finland females have educated themselves on a large scale much shorter time than males and that men receive more self-employment and capital incomes, which usually tend to increase over age.

The changes accounted for the household type are unique, but not statistically significant.

[20] Results for the Gini coefficient are shown in Appendix G.

[21] The change in the definition of the level of education in 1997 might have affected the results obtained from 1993 to 2002. Especially the result for quantity effect might be larger than it in reality should be.

It is the only variable that has equalised the income distribution, but this is not due to policy changes. On the contrary, policy changes have made the household type more important factor in income. Here also, the changes in the distribution have increased the inequality. The last variable used was the region and it has nearly no effect on changes in inequality.

4 Conclusions

In this paper, a new method of analysing how the policy changes have affected the income distribution was introduced. This new method combines the benefits of the microsimulation- and regression based decompositions. It allows analysing how policy changes have affected the importance of sociodemographic characters in income inequality. With the variance of log of incomes decomposition can be made even further to isolate which part of the price- and residual effect is accounted to policy changes and which part is not. This new method was applied to study how different characteristics have affected the income distribution in Finland since 1993. First aim was to quantify how much policy changes, price effect, quantity effect and residual effect have affected the evolution of income inequality in a static setting. Second aim was to investigate how much policy changes have altered the price effect of individual/household characteristics.

The price-, quantity- and residual effect all account around 1/3 of the total change in inequality from 1993 to 2014. This indicates that both the changes in the distribution of the sociodemographic characteristics and the changes in the importance of these characteristics explain significant part of the rise in inequality. Results with policy changes were in line with the earlier findings of Honkanen and Tervola (2014). Policy changes account the largest share in the evolution of income inequality in Finland (about 47% since 1993). However, this result is very sensitive and the policy changes have not increased inequality during the whole time span. Since 2005 policy changes have equalised the income distribution. About half of policy effect could be traced to the price effect of different characteristics and other half was affecting the residual term. The most interesting finding was that the majority of the change (about 2/3) in inequality accounted to the change in importance of the characteristics in individuals' income were caused by the policy changes. The same does not apply to residual effect, where the majority of the changes remained unexplained.

I also find interesting features in the changes of the contributions of different characteristics on inequality. For instance, the most important character was the employment status of men and it was solely driven by the price effect. Whereas the changes in the distribution of level of education made substantial increase in income inequality.

This new method does not allow, at least yet, studying how policy changes have affected the quantity effect. It can be the case that policy changes have had indirect effects that may have altered the distribution of the characteristics and it would be interesting to analyse these effects. For now this is left for future research.

References

Atkinson, A. B. & Bourguignon, F. (2015) Introduction: Income distribution today. In *Handbook of Income Distribution*, eds. A. B. Atkinson & F. Bourguignon, vol. 2 of *Handbook of Income Distribution*, xvii – lxiv, Elsevier, URL `http://www.sciencedirect.com/science/article/pii/B9780444594280099896`.

Bargain, O. (2012a) Decomposition analysis of distributive policies using behavioural simulations. *International Tax and Public Finance*, vol. 19 (5), 708–731.

Bargain, O. (2012b) The Distributional Effects of Tax-benefit Policies under New Labour: A Decomposition Approach. *Oxford Bulletin of Economics and Statistics*, vol. 74 (6), 856–874.

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

Brewer, M. & Wren-Lewis, L. (2016) Accounting for changes in income inequality: Decomposition analyses for the uk, 1978–2008. *Oxford Bulletin of Economics and Statistics*, vol. 78 (3), 289–322, URL `http://dx.doi.org/10.1111/obes.12113`.

Fields, G. S. (2003) *Worker Well-Being and Public Policy (Research in Labor Economics, Volume 22)*, chap. ACCOUNTING FOR INCOME INEQUALITY AND ITS CHANGE: A NEW METHOD, WITH APPLICATION TO THE DISTRIBUTION OF EARNINGS IN THE UNITED STATES, 1–38. Emerald Group Publishing Limited.

Figari, F., Paulus, A. & Sutherland, H. (2015) Chapter 24 - microsimulation and policy analysis. In *Handbook of Income Distribution*, eds. A. B. Atkinson & F. Bourguignon, vol. 2 of *Handbook of Income Distribution*, 2141 – 2221, Elsevier, URL `http://www.sciencedirect.com/science/article/pii/B978044459429700025X`.

Fortin, N., Lemieux, T. & Firpo, S. (2011) Decomposition methods in economics. In *Handbook of Labor Economics*, eds. O. Ashenfelter & D. Card, vol. 4A, chap. 01, 1–102, Elsevier, 1 edn., URL `https://EconPapers.repec.org/RePEc:eee:labchp:4-01`.

Gorodnichenko, Y., Mendoza, E. G. & Tesar, L. L. (2012) The finnish great depression: From russia with love. *American Economic Review*, vol. 102 (4), 1619–44.

Herault, N. & Azpitarte, F. (2016) Understanding changes in the distribution and redistribution of income: A unifying decomposition framework. *Review of Income and Wealth*, vol. 62 (2), 266–282, URL `http://dx.doi.org/10.1111/roiw.12160`.

Hills, J., Paulus, A., Sutherland, H. & Tasseva, I. (2014) A lost decade? Decomposing the effect of 2001-11 tax-benefit policy changes on the income distribution in EU countries. ImPRovE Working Papers 14/03, Herman Deleeck Centre for Social Policy, University of Antwerp.

Honkanen, P. & Tervola, J. (2014) Vero- ja tulonsiirtojräjestelmän vaikutus tulonjakoon suomessa 1995-2013. *Yhteiskuntapolitiikka*, vol. 79, 306–317, in Finnish.

OECD (2011) Growing income inequality in oecd countries: What drives it and how can policy tackle?

Official Statistics of Finland (OSF) (2014) *SISU-malli - Käyttöopas tulonsiirtojen ja verotuksen mikrosimulointiin*. Official Statistics of Finland (OSF), in Finnish.

Riihelä, M., Sullström, R. & Tuomala, M. (2010) Trends in top income shares in finland 1966–2007. Research Reports 157, Government Institute for Economic Research Finland (VATT).

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

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

Yun, M.-S. (2006) Earnings inequality in usa, 1969–99: Comparing inequality using earnings equations. *Review of Income and Wealth*, vol. 52 (1), 127–144, URL http://dx.doi.org/10.1111/j.1475-4991.2006.00179.x.

Appendices

A Pure and interaction effects

Shorrocks (1982) showed that the contribution of income source k can be decomposed to two parts: A) the pure contribution of income source k on inequality and B) the interaction effect of income source k on inequality. Fields' (2003) decomposition method is very close to the factor source decomposition and thus Shorrocks example is easy to apply in this set up.

Following Shorrocks' example, the contribution of character c can be regarded in two ways: A) the inequality which would be observed if character c was the only character affecting income and B) the amount by which inequality would fall if differences in characteristics c were eliminated. Formally, these can be expressed as:

$$\begin{aligned} C_c^A =& I\left(\beta_c X_c + (\mu - \overline{\beta_c X_c})\right) \\ C_c^B =& I(y) - I\left(y - \beta_c X_c + \overline{\beta_c X_c}\right) \end{aligned}$$

Where I is some inequality measure, y log of income, μ is the mean of log of income and over line presents the mean as well.

Shorrocks (1982) showed that this can consistently be done with the variance of incomes and square of the coefficient of variation as an inequality measure. With the variance of log of incomes C_c^A becomes:

$$C_c^A = \sigma^2\left(\beta_c X_c + (\mu - \overline{\beta_c X_c})\right) = \sigma^2\left(\beta_c X_c\right)$$

And similarly C_c^B is

$$\begin{aligned} C_c^B =& \sigma^2(y) - \sigma^2\left(y - \beta_c X_c + \overline{\beta_c X_c}\right) \\ =& \sigma^2(y) - \sigma^2\left(y - \beta_c X_c\right) = \sigma^2(y) - \sigma^2(y) - \sigma^2(\beta_c X_c) + 2\mathrm{cov}(y, \beta_c X_c) \\ =& -\sigma^2(\beta_c X_c) + 2\mathrm{cov}(\beta_c X_c, \beta_c X_c) + 2\mathrm{cov}(\beta_c X_c, y - \beta_c X_c) \\ =& \sigma^2(\beta_c X_c) + 2\mathrm{cov}(\beta_c X_c, y - \beta_c X_c) \end{aligned}$$

And the contributions derived from the equation (4) are then:

$$S_c = s_c \sigma_y^2 = \mathrm{cov}(\beta_c X_c, y) = \frac{1}{2}\left(C_c^A + C_c^B\right)$$

B Price- and quantity effects decomposed to pure- and interaction effects

The income generating functions are defined as:

$$\begin{aligned} y_A &= \sum_c \beta_{cA} X_{cA} + \varepsilon_A \\ y_B &= \sum_c \beta_{cB} X_{cB} + \varepsilon_B \\ y^* &= \sum_c \beta_{cB} X_{cA} + \varepsilon_A \end{aligned}$$

And the pure- and interaction effects are defined as:

$$\begin{aligned} C_c^A =& \sigma^2 \left(\beta_c X_c\right) = \beta_c^2 \sigma^2 \left(X_c\right) \\ C_c^B =& \sigma^2 (\beta_c X_c) + 2\text{cov}(\beta_c X_c, y - \beta_c X_c) \end{aligned}$$

Given C_c^A and C_c^B, the contributions derived from the equation (4) are then:

$$S_c = s_c \sigma_y^2 = \text{cov}(\beta_c X_c, y) = \frac{1}{2}\left(C_c^A + C_c^B\right)$$

The share of C_c^A and C_c^B in the total change of inequality is then:

$$\Delta S_c = \frac{1}{2}\left[\Delta C_c^A + \Delta C_c^B\right]$$

From the equation (6) we know that:

$$\Delta I = \Delta^Q + \Delta^P + \Delta^\varepsilon = \sum_c \left[\Delta_c^Q + \Delta_c^P + \Delta_c^\varepsilon\right]$$

Where Δ_c^Q is

$$\begin{aligned} \Delta_c^Q =& s_{cB} \sigma_{y_B}^2 - s_c^* \sigma_{y^*}^2 \\ =& \text{cov}(\beta_{cB} X_{cB}, y_B) - \text{cov}(\beta_{cB} X_{cA}, y^*) \\ =& \frac{1}{2}\left(C_{cB}^A + C_{cB}^B\right) - \frac{1}{2}\left(C_{c^*}^A + C_{c^*}^B\right) \\ =& \frac{1}{2}\left(C_{cB}^A - C_{c^*}^A\right) + \frac{1}{2}\left(C_{cB}^B - C_{c^*}^B\right) \\ =& \frac{1}{2}\left[\beta_{cB}^2 \text{Var}(X_{cB}) - \beta_{cB}^2 \text{Var}(X_{cA})\right] \\ &+ \frac{1}{2}\left[\beta_{cB}^2 \text{Var}(X_{cB}) - \beta_{cB}^2 \text{Var}(X_{cA}) + 2\text{cov}(\beta_{cB} X_{cB}, y_B - \beta_{cB} X_{cB}) - 2\text{cov}(\beta_{cB} X_{cA}, y^* - \beta_{cB} X_{cA})\right] \\ =& \frac{1}{2}\left[\beta_{cB}^2 \Delta \text{Var}(X_c)\right] + \frac{1}{2}\left[\beta_{cB}^2 \Delta \text{Var}(X_c) + 2\sum_{i \neq c} \beta_{cB} \beta_{iB} [\Delta \text{cov}(X_c, X_i)]\right] \end{aligned}$$

And similarly Δ_c^P is

$$\Delta_c^P = \frac{1}{2}\left[\mathrm{Var}(X_c)\Delta\beta_c^2\right] + \frac{1}{2}\left[\mathrm{Var}(X_c)\Delta\beta_c^2 + 2\sum_{i\neq c}\mathrm{cov}(X_{cA}, X_{iA})\Delta\beta_c\beta_i\right]$$

C Effect of the policy change in price- and residual effects

To show that policy effect really is either price- or/and residual effect, the counterfactuals need to be formed first. Then the change between counterfactual and actual data is decomposed according equation (6). With the end, period prices it takes the form:

$$\begin{aligned}\Delta\gamma =& \gamma_B(\mathbf{X}_B, \mathbf{Y}_B, m_B) - \gamma_A(\mathbf{X}_B, \mathbf{Y}_B, \alpha m_A) \Rightarrow \\ \Delta I(\gamma) =& I[\gamma_B(\mathbf{X}_B, \mathbf{Y}_B, m_B)] - I[\gamma_A(\mathbf{X}_B, \mathbf{Y}_B, \alpha m_A)]\end{aligned}$$

Now the change can be decomposed like before to the price-, quantity-, and residual effects:

$$\begin{aligned}\Delta\sigma_y^2 =& \sum_{c=1}^{N}\left(s_{cB}\sigma_{y_B}^2 - s_c^*\sigma_{y^*}^2\right) + \sum_{c=1}^{N}\left(s_c^*\sigma_{y^*}^2 - s_{c_B^A}\sigma_{y_B^A}^2\right) + \sigma_{\varepsilon_B}^2 - \sigma_{\varepsilon_B^A}^2 \\ =& \Delta^Q + \Delta^P + \Delta^\varepsilon\end{aligned}$$

Where the income generating function is estimated with OLS as:

$$\begin{aligned}y_B &= \sum_c \beta_{cB}X_{cB} + \varepsilon_B \\ y_B^A &= \sum_c \beta_{cB}^A X_{cB}^A + \varepsilon_B^A\end{aligned}$$

And y^* is defined as:

$$y^* = \sum_c \beta_{cB}X_{cB}^A + \varepsilon_B^A$$

Since X_{ciB}^A is equal to X_{cB} (the same data set used in the analysis), the above equations take the forms:

$$y_B = \sum_c \beta_{cB}X_{cB} + \varepsilon_B \tag{15}$$

$$y_B^A = \sum_c \beta_{cB}^A X_{cB} + \varepsilon_B^A \tag{16}$$

$$y^* = \sum_c \beta_{cB}X_B + \varepsilon_B^A \tag{17}$$

Therefore,

$$
\begin{aligned}
s_{cB}\sigma^2_{y_B} =& \operatorname{cov}(\beta_{cB}X_{cB}, y_B) = \operatorname{cov}(\beta_{cB}X_{cB}, \sum_c \beta_{cB}X_{cB} + \varepsilon_B) \\
=& \operatorname{cov}(\beta_{cB}X_{cB}, \sum_c \beta_{cB}X_{cB}) + \beta_{cB}\overbrace{\operatorname{cov}(X_{cB}, \varepsilon_B)}^{=0 \text{ eq. (15)}} \\
=& \operatorname{cov}(\beta_{cB}X_{cB}, \sum_c \beta_{cB}X_{cB}) + \beta_{cB}\overbrace{\operatorname{cov}(X_{cB}, \varepsilon_B^A)}^{=0 \text{ eq. (16)}} \\
=& s_c^*\sigma^2_{y^*}, \quad \forall c \Rightarrow \\
\Delta^Q =& \sum_{c=1}^{N}\left(s_{cB}\sigma^2_{y_B} - s_c^*\sigma^2_{y^*}\right) = 0
\end{aligned}
$$

Whereas the price- and residual effects won't (necessarily) vanish since:

$$
\begin{aligned}
s_c^*\sigma^2_{y^*} =& \operatorname{cov}(\beta_{cB}X_{cB}, y_B) = \operatorname{cov}(\beta_{cB}X_{cB}, \sum_c \beta_{cB}X_{cB} + \varepsilon_A) \\
=& \beta^2_{cB}\operatorname{cov}(X_{cB}, X_{cB}) + \sum_{i\neq c}\beta_{cB}\beta_{iB}\operatorname{cov}(X_{cB}, X_{iB}) + \beta_{cB}\overbrace{\operatorname{cov}(X_{cB}, \varepsilon_B^A)}^{=0 \text{ eq. (16)}} \\
\neq& \beta^2_{c_B^A}\operatorname{cov}(X_{cB}, X_{cB}) + \sum_{i\neq c}\beta_{c_B^A}\beta_{i_B^A}\operatorname{cov}(X_{cB}, X_B) + \beta_{c_B^A}\overbrace{\operatorname{cov}(X_{cB}, \varepsilon_B^A)}^{=0 \text{ eq. (16)}} \\
=& s_{c_B^A}\sigma^2_{y_B^A} \Rightarrow \\
\Delta^P \neq& 0
\end{aligned}
$$

and

$$
\varepsilon_B^A \neq \varepsilon_B \quad \Rightarrow \quad \sigma^2_{\varepsilon_B^A} \neq \sigma^2_{\varepsilon_B} \quad \Rightarrow \quad \Delta^\varepsilon \neq 0
$$

The same can be applied to the case where initial period data is used in the decomposition.

D Data and microsimulation

Data used in this analysis is the service data of income distribution collected by Statistics Finland. From this data set is created the Finnish data of EU-SILC. Years covered in the analyses are 1993, 1996, 1999, 2002, 2005, 2008, 2011 and 2014.

Simulations are conducted using Statistics Finland's microsimulation model SISU. The majority of the Finnish tax-benefit-system is encoded in SISU model from the year 1993 onwards. However, the SISU model is not compatible with the data before the year 2011. Therefore the simulations are carried out by using data from the years 2011 and 2014. 2011 data is used to form counterfactuals for years 1993, 1996, 1999, 2002, 2005, 2008 and 2014 whereas counterfactual for year 2011 is done by using the data of 2014.

The SISU model has 12 different sub-models: Health care insurances; Unemployment benefits; Home care allowance; National pensions; Student allowance; Taxation; Estate taxation; Child allowance; Pensioner's housing benefit; Housing benefit; Day care fees and Social assistance. All of these sub-models are used in the simulations except estate taxation.

All the tax-benefit changes cannot be simulated. Some of the changes for eligibility or changes in the duration of benefit payments cannot be simulated. For example, the duration of unemployment insurance was decreased from 500 days to 400 days, which effect can not be simulated with SISU model.

To make different years' tax-benefit system more comparative to other years data, the monetary parameters are indexed by the price index.

Disposable incomes are at the household level and thus those are scaled by the modified OECD-equivalence scale. The analyses are conducted by using only the households' heads, which are given the information about the characteristics of any spouses. Weight used is the household level weight multiplied with the number of members in the household. This way the analyses are done at the individual level.

E Definition of population subgroups

Full details of the characteristics are presented here. *Education:* Elementary school or no education or education unknown; secondary school; lower/lowest tertiary level; upper tertiary level or higher. *Region:* Uusimaa; Varsinais-Suomi; Satakunta; Kanta-Häme; Pirkanmaa; Päijät-Häme; Kymenlaakso; Etelä-Karjala; Etelä-Savo; Pohjois-Savo; Pohjois-Karjala; Keski-Suomi; Etelä-Pohjanmaa; Pohjanmaa; Keski-Pohjanmaa; Pohjois-Pohjanmaa; Kainuu; Lappi; Ahvenanmaa. *Household type:* 1 adult, no children; 2 adults, no children; 3 or more adults, no children; single parents, youngest child under 7 years old; two adults, youngest children under 7 years old; 3 or more adults, youngest children under 7 years old; single parents, youngest child over 6 years old; two adults, youngest child over 6 years old; 3 or more adults, youngest child over 6 years old; 1 adult, household head over 64 years old; other families with household head over 64 years old. *Age:* under 25 years old; 25–34; 35–44; 45–54; 55–64; 65–74; over 74 years old. *Employment status:* employed; unemployed and others; self-employed; pensioner; student; schoolchild; unknown.

F Annual decompositions

In this section are presented the annual decompositions of the results.

Table F.1: Shares of characteristics in income inequality (%)

			Household	Age		Employment status		Education	
Years	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	60	3.1	4.5	5.6	3.7	5.2	6.9	6.2	4.8
1996	57.2	1.4	3.9	6.5	3.6	7.4	8.6	6.6	4.9
1999	55.7	2.3	2.6	7.5	3.1	6.6	7.4	8.7	6.1
2002	55.2	1.9	3.8	6.6	3.7	8.1	7.1	8.3	5.2
2005	55.2	1.6	3.5	5.2	2.6	7.3	7.4	10	7.1
2008	53.9	2.2	4.4	6.4	2.6	8.4	7.9	7.5	6.7
2011	53.3	1.8	3.1	7.4	2.8	8.2	7.9	8.5	7
2014	49.8	2.1	2.2	9	3.7	9.3	6.9	8.7	8.3

Table F.2: Shares of characteristics in income inequality (Simulated data, %)

			Household	Age		Employment status		Education	
Years	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	54.8	1.9	3.1	7.5	2.4	7.6	7.6	7.9	7.1
1996	54.1	1.9	2.8	7.4	2.3	8	8.1	8.1	7.3
1999	51.8	1.9	2.8	7.5	2.3	9	8.8	8.4	7.5
2002	51.3	1.9	3.1	7.5	2.4	8.9	8.9	8.4	7.6
2005	52.0	1.9	2.5	7.3	2.2	9.1	8.2	9.2	7.6
2008	51.6	1.9	2.7	7.3	2.3	9.1	8.3	9.1	7.6
2011	47.5	2.1	1.8	9.0	3.5	10.0	7.9	9.3	8.8
2014	51.9	2.0	3.1	7.5	2.6	8.5	8.0	9.0	7.4

Note: Simulated data for year 2011 is formed by using data from year 2014. For the rest of simulated years data used is from year 2011.

Table F.3: Absolute contribution of characteristics in income inequality (1000 $\times$ variance of logs)

				Household	Age		Employment status		Education	
Years	Total	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	147.2	88.3	4.6	6.6	8.3	5.4	7.7	10.2	9.1	7.1
1996	159.4	91.2	2.3	6.2	10.3	5.8	11.7	13.7	10.5	7.8
1999	197.1	109.9	4.5	5.2	14.8	6.0	13	14.5	17.1	12.1
2002	207.0	114.3	4.0	7.8	13.8	7.8	16.7	14.8	17.3	10.7
2005	228.3	126.1	3.7	7.9	11.8	6.0	16.7	17	22.9	16.1
2008	229.5	123.6	5.1	10.0	14.7	5.9	19.2	18.2	17.3	15.4
2011	234.6	125.0	4.3	7.3	17.4	6.7	19.2	18.5	19.8	16.4
2014	228.8	114.0	4.7	5.1	20.6	8.5	21.3	15.8	19.8	19.0

Table F.4: Absolute contribution of characteristics in income inequality (Simulated data, 1000 $\times$ variance of logs)

				Household	Age		Employment status		Education	
Years	Total	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	183.0	100.4	3.6	5.6	13.8	4.3	14	13.9	14.5	13.1
1996	185.5	100.5	3.5	5.2	13.8	4.3	14.8	15.0	15.0	13.5
1999	192.9	99.9	3.6	5.5	14.5	4.5	17.3	16.9	16.1	14.5
2002	196.6	100.8	3.7	6.1	14.8	4.8	17.4	17.5	16.5	14.9
2005	219.8	114.3	4.2	5.6	15.9	4.9	20.0	18	20.3	16.6
2008	223.0	115.1	4.3	6.0	16.3	5.1	20.4	18.6	20.3	16.9
2011	230.4	109.6	4.8	4.2	20.7	8.1	23.0	18.2	21.5	20.4
2014	209.6	108.8	4.1	6.4	15.7	5.3	17.9	16.8	18.9	15.6

Note: Simulated data for year 2011 is formed by using data from year 2014. For the rest of simulated years data used is from year 2011.

Table F.5: Change in the absolute contribution of characteristics in income inequality (1000 $\times$ variance of logs)

				Household	Age		Employment status		Education	
Years	Total	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993 to 1996	12.2	2.9	-2.3	-0.4	2	0.3	4	3.5	1.4	0.7
1996 to 1999	37.7	18.7	2.2	-1	4.5	0.3	1.3	0.8	6.6	4.3
1999 to 2002	9.9	4.4	-0.5	2.6	-1	1.7	3.7	0.3	0.2	-1.4
2002 to 2005	21.2	11.8	-0.3	0.1	-1.9	-1.7	0	2.2	5.7	5.5
2005 to 2008	1.2	-2.5	1.4	2.1	2.9	-0.1	2.5	1.2	-5.7	-0.7
2008 to 2011	5.1	1.3	-0.8	-2.8	2.7	0.7	0	0.3	2.6	1
2011 to 2014	-5.8	-11	0.4	-2.2	3.1	1.8	2.1	-2.7	0	2.6

Table F.6: Absolute policy effect of characteristics in income inequality (1000 $\times$ variance of logs)

				Household	Age		Employment status		Education	
Years	Total	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993 to 1996	2.5	0.1	-0.1	-0.4	0	0	0.8	1.1	0.6	0.5
1996 to 1999	7.3	-0.6	0.1	0.2	0.7	0.2	2.6	1.9	1.1	1
1999 to 2002	3.7	0.9	0.1	0.6	0.3	0.3	0.1	0.6	0.4	0.4
2002 to 2005	23.2	13.5	0.5	-0.5	1.2	0.1	2.6	0.5	3.7	1.7
2005 to 2008	3.2	0.8	0.1	0.4	0.3	0.3	0.3	0.6	0	0.3
2008 to 2011	11.6	9.8	0	1.3	1.2	1.5	-1.2	-0.1	-0.5	-0.5
2011 to 2014	-13.3	-5.9	-0.2	0.1	-0.9	-0.5	-1.5	-2.1	-1.3	-1.1

Note: Last change calculated as an average of 2011 and 2014

Table F.7: Changes in income inequality, price-, quantity and residual effects (1000x variance of log of incomes)

Years	Total Change	Price Effect	Quantity Effect	Residual Effect	Region		Household type		Age				Employment status				Education			
									Male		Female		Male		Female		Male		Female	
					P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q
1993 to 1996	12.2	7.4	1.9	2.9	-2.1	-0.2	-0.3	-0.1	1.6	0.4	0.4	0	4.7	-0.7	3.3	0.2	0.1	1.3	-0.2	0.9
1996 to 1999	37.7	10.4	8.7	18.7	2.1	0	-1	0	2.9	1.6	-0.2	0.5	1.7	-0.4	-1	1.9	4.4	2.2	1.4	2.9
1999 to 2002	9.9	-0.8	6.3	4.4	-1	0.5	2.3	0.3	-2.1	1.1	0.3	1.5	2.7	1	0.1	0.2	-0.9	1.1	-2.1	0.7
2002 to 2005	21.2	10	-0.5	11.8	0	-0.3	0.5	-0.4	-2.2	0.2	-2	0.3	2.2	-2.2	3.7	-1.5	3.6	2	4.1	1.4
2005 to 2008	1.2	0.1	3.7	-2.5	1.3	0.1	1.1	1	2.6	0.3	-0.6	0.5	2.7	-0.1	0.2	1	-5.7	0.1	-1.5	0.7
2008 to 2011	5.1	-0.1	3.8	1.3	-0.9	0.2	-2.4	-0.3	2.9	-0.1	0.5	0.2	-2.2	2.2	0	0.3	1.8	0.8	0.5	0.6
2011 to 2014	-5.8	-1.9	7.1	-11	0.4	0	-2	-0.2	1.5	1.6	0.9	1	0.5	1.6	-2.7	0	-1.2	1.1	0.6	2

Note: Price- and quantity effects are formed as an average of equations (6) and (9).

Table F.8: Price and residual effects decomposed to policy and other effects (1000x variance of log of incomes)

Years	Price Effect		Residual Effect		Region		Household type		Age				Employment status				Education			
									Male		Female		Male		Female		Male		Female	
	Δ_O^P	Δ_P^P	Δ_O^ε	Δ_P^ε	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P	Δ_O^P	Δ_P^P
1993 to 1996	5	2.4	2.8	0.1	-2	-0.1	0.1	-0.4	1.6	0	0.3	0	3.9	0.8	2.2	1.1	-0.5	0.6	-0.7	0.5
1996 to 1999	2.5	7.9	19.2	-0.6	2	0.1	-1.3	0.2	2.2	0.7	-0.4	0.2	-0.9	2.6	-2.9	1.9	3.4	1.1	0.4	1
1999 to 2002	-3.6	2.8	3.5	0.9	-1.1	0.1	1.7	0.6	-2.4	0.3	0	0.3	2.6	0.1	-0.5	0.6	-1.3	0.4	-2.5	0.4
2002 to 2005	0.2	9.7	-1.7	13.5	-0.5	0.5	1	-0.5	-3.3	1.2	-2.1	0.1	-0.4	2.6	3.3	0.5	-0.1	3.7	2.4	1.7
2005 to 2008	-2.3	2.3	-3.3	0.8	1.2	0.1	0.7	0.4	2.3	0.3	-0.8	0.3	2.3	0.3	-0.5	0.6	-5.7	0	-1.8	0.3
2008 to 2011	-1.8	1.8	-8.5	9.8	-1	0	-3.7	1.3	1.7	1.2	-1	1.5	-1.1	-1.2	0.1	-0.1	2.2	-0.5	0.9	-0.5
2011 to 2014	5.5	-7.4	-5.1	-5.9	0.6	-0.2	-2	0.1	2.4	-0.9	1.3	-0.5	2	-1.5	-0.6	-2.1	0.2	-1.3	1.7	-1.1

Note: Simulated data for year 2011 is formed by using data from year 2014. For the rest of simulated years data used is from year 2011.

G Robustness checks

In this section is presented the robustness checks of the results obtained earlier. In the first table is shown the absolute policy effects (price-effect accounted to policy changes) by using 100 × Gini coefficient as an inequality measure. In the next tables are shown the price- and quantity effects formed using equations (6) and (9). Then the final tables presents the bootstrapped 95% confidence intervals of the results. Confidence intervals are formed using 1 000 replications.

Table G.1: Policy effect of characteristics in income inequality (100 × Gini coefficient)

				Household	Age		Employment status		Education	
Years	Total	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993 to 1996	-0.063	0.138	-0.019	-0.06	-0.031	-0.007	0.072	0.11	0.041	0.032
1996 to 1999	0.489	0.794	0.008	0.019	0.062	0.013	0.29	0.207	0.102	0.092
1999 to 2002	0.151	0.2	0.008	0.066	0.01	0.029	-0.017	0.049	0.024	0.031
2002 to 2005	2.152	0.856	0.048	-0.08	0.091	-0.006	0.255	-0.002	0.397	0.153
2005 to 2008	0.044	0.119	0.003	0.043	0.013	0.023	0.01	0.047	-0.027	0.007
2008 to 2011	-0.133	-0.496	-0.025	0.106	0.029	0.138	-0.265	-0.133	-0.183	-0.163
2011 to 2014	-0.652	-0.444	-0.002	0.037	-0.043	-0.033	-0.097	-0.168	-0.076	-0.063
1993 to 2002	0.576	1.132	-0.003	0.025	0.042	0.035	0.345	0.366	0.167	0.155
2002 to 2014	1.411	0.035	0.024	0.106	0.09	0.121	-0.097	-0.257	0.112	-0.066
1993 to 2014	1.988	1.167	0.021	0.132	0.132	0.156	0.248	0.109	0.279	0.089

Note: Last change calculated as an average of 2011 and 2014

Table G.2: Changes in income inequality, price-, quantity and residual effects (1000x variance of log of incomes)

Years	Total Change	Price Effect	Quantity Effect	Residual Effect	Region		Household type		Age				Employment status				Education			
									Male		Female		Male		Female		Male		Female	
					P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q
1993 to 1996	12.2	7.9	1.4	2.9	-2.1	-0.3	-0.3	-0.1	1.6	0.4	0.4	-0.1	5.1	-1.1	3.3	0.2	0	1.3	-0.2	0.9
1996 to 1999	37.7	11.5	7.5	18.7	2.2	-0.1	-0.7	-0.3	2.7	1.8	-0.1	0.4	2	-0.7	-0.7	1.6	4.4	2.2	1.8	2.6
1999 to 2002	9.9	-0.5	6	4.4	-0.9	0.4	2.2	0.4	-2.2	1.2	0.4	1.3	2.8	0.9	0.2	0.1	-0.9	1.1	-2	0.6
2002 to 2005	21.2	9.8	-0.4	11.8	0	-0.3	0.4	-0.3	-2.2	0.2	-1.9	0.1	2.1	-2.2	3.9	-1.7	3.5	2.1	3.9	1.6
2005 to 2008	1.2	0.2	3.5	-2.5	1.3	0.1	1.2	0.9	2.4	0.5	-0.6	0.5	2.9	-0.3	0.1	1.1	-5.7	0	-1.4	0.6
2008 to 2011	5.1	-0.1	3.8	1.3	-1	0.2	-2.5	-0.3	2.7	0	0.5	0.2	-2	1.9	-0.1	0.4	1.7	0.9	0.5	0.6
2011 to 2014	-5.8	-2	7.2	-11	0.4	0	-1.9	-0.2	1.4	1.7	0.9	1	0.4	1.7	-2.6	-0.1	-1.2	1.2	0.7	1.9

Note: Price- and quantity effects are formed according to equation (6).

Table G.3: Changes in income inequality, price-, quantity and residual effects (1000x variance of log of incomes)

Years	Total Change	Price Effect	Quantity Effect	Residual Effect	Region		Household type		Age				Employment status				Education			
									Male		Female		Male		Female		Male		Female	
					P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q
1993 to 1996	12.2	6.9	2.3	2.9	-2.1	-0.2	-0.2	-0.2	1.6	0.4	0.3	0.1	4.3	-0.3	3.2	0.3	0.1	1.3	-0.3	0.9
1996 to 1999	37.7	9.2	9.9	18.7	2	0.1	-1.4	0.4	3.1	1.4	-0.3	0.6	1.5	-0.2	-1.3	2.1	4.5	2.1	1	3.3
1999 to 2002	9.9	-1	6.5	4.4	-1.1	0.6	2.4	0.2	-2	1	0.1	1.6	2.7	1.1	0	0.3	-0.9	1.1	-2.1	0.7
2002 to 2005	21.2	10.1	-0.6	11.8	0.1	-0.4	0.6	-0.4	-2.2	0.2	-2.1	0.4	2.2	-2.2	3.6	-1.4	3.8	1.9	4.2	1.3
2005 to 2008	1.2	-0.1	3.8	-2.5	1.3	0.1	1.1	1	2.8	0.1	-0.6	0.5	2.5	0.1	0.2	1	-5.8	0.1	-1.6	0.8
2008 to 2011	5.1	0	3.8	1.3	-0.9	0.1	-2.4	-0.4	3	-0.3	0.5	0.2	-2.5	2.5	0	0.3	1.8	0.8	0.5	0.6
2011 to 2014	-5.8	-1.8	7	-11	0.5	-0.1	-2	-0.1	1.6	1.5	0.9	1	0.6	1.5	-2.7	0	-1.2	1.1	0.6	2

Note: Price- and quantity effects are formed according to equation (9).

Table G.4: Shares of characteristics, 95% confidence intervals for actual data

Years	Residual	Region	Household type	Age		Employment status		Education	
				Male	Female	Male	Female	Male	Female
1993	(57.6% – 62.4%)	(2.5% – 3.8%)	(3.4% – 5.6%)	(4.5% – 6.8%)	(2.8% – 4.6%)	(4.3% – 6.2%)	(5.8% – 8%)	(5.1% – 7.3%)	(3.9% – 5.7%)
1996	(55.1% – 59.3%)	(0.9% – 1.9%)	(2.9% – 4.9%)	(5.3% – 7.6%)	(2.6% – 4.6%)	(6.2% – 8.6%)	(7.5% – 9.7%)	(5.4% – 7.7%)	(3.9% – 5.8%)
1999	(53.5% – 58%)	(1.7% – 2.9%)	(1.5% – 3.7%)	(6.3% – 8.8%)	(2.1% – 4%)	(5.4% – 7.8%)	(6.1% – 8.6%)	(7.4% – 9.9%)	(5.1% – 7.2%)
2002	(52.7% – 57.7%)	(1.3% – 2.6%)	(2.7% – 4.8%)	(5.4% – 7.8%)	(2.8% – 4.7%)	(6.8% – 9.3%)	(6% – 8.3%)	(7.2% – 9.5%)	(4.1% – 6.2%)
2005	(52.3% – 58.1%)	(1.1% – 2.1%)	(2.4% – 4.5%)	(4% – 6.4%)	(1.7% – 3.6%)	(6.1% – 8.6%)	(6.3% – 8.6%)	(8.7% – 11.4%)	(5.9% – 8.2%)
2008	(51.7% – 56.1%)	(1.7% – 2.8%)	(3.3% – 5.5%)	(5.2% – 7.7%)	(1.7% – 3.5%)	(7.2% – 9.6%)	(6.7% – 9.1%)	(6.2% – 8.8%)	(5.6% – 7.8%)
2011	(50.3% – 56.2%)	(1.3% – 2.3%)	(2.1% – 4.2%)	(6.1% – 8.8%)	(1.9% – 3.8%)	(6.7% – 9.7%)	(6.7% – 9.1%)	(7.2% – 9.7%)	(5.8% – 8.2%)
2014	(47.9% – 51.7%)	(1.5% – 2.6%)	(1.2% – 3.3%)	(7.5% – 10.5%)	(2.7% – 4.8%)	(8% – 10.7%)	(5.8% – 8%)	(7.4% – 9.9%)	(7.2% – 9.4%)

Table G.5: Shares of characteristics, 95% confidence intervals for simulated data

Years	Residual	Region	Household type	Age		Employment status		Education	
				Male	Female	Male	Female	Male	Female
1993	(52% – 57.7%)	(1.4% – 2.5%)	(2% – 4.1%)	(6.1% – 9%)	(1.4% – 3.3%)	(6.3% – 9%)	(6.3% – 8.8%)	(6.6% – 9.2%)	(6% – 8.2%)
1996	(50.9% – 57.4%)	(1.4% – 2.4%)	(1.8% – 3.9%)	(5.9% – 8.9%)	(1.4% – 3.3%)	(6.6% – 9.3%)	(6.8% – 9.4%)	(6.7% – 9.4%)	(6.2% – 8.4%)
1999	(49.3% – 54.3%)	(1.4% – 2.4%)	(1.8% – 3.9%)	(6.1% – 8.9%)	(1.4% – 3.3%)	(7.6% – 10.4%)	(7.5% – 10%)	(7.1% – 9.6%)	(6.4% – 8.6%)
2002	(48.8% – 53.7%)	(1.4% – 2.4%)	(2% – 4.2%)	(6.1% – 8.9%)	(1.5% – 3.4%)	(7.5% – 10.2%)	(7.6% – 10.2%)	(7.2% – 9.6%)	(6.5% – 8.7%)
2005	(49.8% – 54.3%)	(1.4% – 2.4%)	(1.5% – 3.6%)	(5.9% – 8.6%)	(1.3% – 3.2%)	(7.8% – 10.5%)	(6.9% – 9.4%)	(8% – 10.4%)	(6.4% – 8.7%)
2008	(49.4% – 53.9%)	(1.4% – 2.4%)	(1.7% – 3.7%)	(6% – 8.6%)	(1.4% – 3.2%)	(7.8% – 10.5%)	(7.1% – 9.6%)	(7.9% – 10.3%)	(6.5% – 8.7%)
2011	(45.6% – 49.5%)	(1.6% – 2.7%)	(0.8% – 2.8%)	(7.5% – 10.4%)	(2.5% – 4.5%)	(8.6% – 11.3%)	(6.8% – 9%)	(8.1% – 10.6%)	(7.7% – 10%)
2014	(49.6% – 54.2%)	(1.5% – 2.5%)	(2% – 4.1%)	(6.1% – 8.8%)	(1.6% – 3.5%)	(7.2% – 9.9%)	(6.8% – 9.2%)	(7.8% – 10.2%)	(6.3% – 8.6%)

Table G.6: Absolute contribution of characteristics in income inequality, 95% confidence intervals for actual data (1000 × variance of logs)

				Household	Age		Employment status		Education	
Years	Total	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	(137.8 – 156.7)	(80 – 96.6)	(3.6 – 5.7)	(4.9 – 8.3)	(6.5 – 10)	(4.1 – 6.7)	(6.4 – 9)	(8.6 – 11.8)	(7.5 – 10.7)	(5.8 – 8.4)
1996	(150.8 – 168)	(85 – 97.4)	(1.4 – 3.1)	(4.6 – 7.9)	(8.5 – 12.1)	(4.1 – 7.4)	(9.7 – 13.7)	(11.7 – 15.6)	(8.6 – 12.3)	(6.2 – 9.4)
1999	(185.8 – 208.4)	(101 – 118.8)	(3.3 – 5.7)	(3.1 – 7.3)	(12.2 – 17.4)	(4.1 – 8)	(10.6 – 15.4)	(12.2 – 16.9)	(14.2 – 19.9)	(9.9 – 14.3)
2002	(195.2 – 218.9)	(104.3 – 124.3)	(2.7 – 5.3)	(5.7 – 9.9)	(11.2 – 16.3)	(5.7 – 9.8)	(14.1 – 19.3)	(12.5 – 17.1)	(14.8 – 19.8)	(8.5 – 12.8)
2005	(209.2 – 247.3)	(110.6 – 141.6)	(2.5 – 4.9)	(5.4 – 10.5)	(9.4 – 14.3)	(3.9 – 8.2)	(13.9 – 19.5)	(14.5 – 19.5)	(19.5 – 26.3)	(13.5 – 18.8)
2008	(217.6 – 241.4)	(114 – 133.2)	(3.8 – 6.4)	(7.6 – 12.5)	(11.8 – 17.7)	(3.8 – 8.1)	(16.4 – 22)	(15.4 – 21)	(14.2 – 20.3)	(12.7 – 18.1)
2011	(214.2 – 255)	(108.6 – 141.4)	(3.1 – 5.6)	(4.8 – 9.7)	(14.3 – 20.5)	(4.2 – 9.1)	(16.2 – 22.1)	(15.2 – 21.8)	(16.9 – 22.8)	(13.8 – 19)
2014	(218.3 – 239.4)	(107.1 – 120.8)	(3.5 – 5.9)	(2.6 – 7.7)	(16.9 – 24.2)	(6 – 11)	(18.2 – 24.5)	(13.2 – 18.3)	(16.8 – 22.8)	(16.3 – 21.7)

Table G.7: Absolute contribution of characteristics in income inequality, 95% confidence intervals for simulated data (1000 × variance of logs)

				Household	Age		Employment status		Education	
Years	Total	Residual	Region	type	Male	Female	Male	Female	Male	Female
1993	(170.9 – 195.2)	(89.8 – 111)	(2.6 – 4.6)	(3.6 – 7.6)	(11.4 – 16.1)	(2.6 – 6)	(11.7 – 16.2)	(11.4 – 16.4)	(12.2 – 16.7)	(11 – 15.1)
1996	(171.8 – 199.2)	(88.2 – 112.8)	(2.5 – 4.5)	(3.2 – 7.3)	(11.4 – 16.2)	(2.5 – 6.1)	(12.5 – 17)	(12.4 – 17.5)	(12.7 – 17.3)	(11.4 – 15.6)
1999	(182.1 – 203.6)	(91.2 – 108.6)	(2.6 – 4.6)	(3.3 – 7.6)	(12 – 17)	(2.7 – 6.4)	(14.8 – 19.9)	(14.1 – 19.7)	(13.8 – 18.5)	(12.3 – 16.7)
2002	(185.7 – 207.5)	(92 – 109.6)	(2.7 – 4.8)	(3.9 – 8.2)	(12.2 – 17.3)	(2.9 – 6.7)	(14.9 – 20)	(14.7 – 20.4)	(14.1 – 18.9)	(12.7 – 17.2)
2005	(208.4 – 231.3)	(105.5 – 123.1)	(3 – 5.4)	(3.3 – 7.9)	(13.1 – 18.7)	(2.8 – 6.9)	(17.1 – 22.9)	(14.9 – 21)	(17.4 – 23.1)	(14.1 – 19.1)
2008	(211.4 – 234.5)	(106.3 – 124)	(3.1 – 5.5)	(3.7 – 8.3)	(13.4 – 19.1)	(3.1 – 7.2)	(17.4 – 23.4)	(15.5 – 21.7)	(17.4 – 23.2)	(14.4 – 19.4)
2011	(221.6 – 239.2)	(103.9 – 115.2)	(3.5 – 6.1)	(1.9 – 6.5)	(17.1 – 24.2)	(5.8 – 10.4)	(19.8 – 26.2)	(15.6 – 20.8)	(18.6 – 24.4)	(17.6 – 23.1)
2014	(198.3 – 220.9)	(100 – 117.5)	(3 – 5.3)	(4.2 – 8.7)	(13 – 18.4)	(3.4 – 7.3)	(15.2 – 20.7)	(13.8 – 19.7)	(16.1 – 21.6)	(13.3 – 18)

The **Aboa Centre for Economics (ACE)** is a joint initiative of the economics departments of the Turku School of Economics at the University of Turku and the School of Business and Economics at Åbo Akademi University. ACE was founded in 1998. The aim of the Centre is to coordinate research and education related to economics.

Contact information: Aboa Centre for Economics, Department of Economics, Rehtorinpellonkatu 3, FI-20500 Turku, Finland.

www.ace-economics.fi

ISSN 1796-3133