Discussion Papers

Statistics Norway
Research department

No. 765
November 2013

Statistics Norway

Statistisk sentralbyrå

Thor O. Thoresen, Zhiyang Jia and Peter J. Lambert

Distributional benchmarking in tax policy evaluations

Discussion Papers No. 765, November 2013
Statistics Norway, Research Department

Thor O. Thoresen, Zhiyang Jia and Peter J. Lambert

Distributional benchmarking in tax policy evaluations

Abstract:
Given an objective to exploit cross-sectional micro data to evaluate the distributional effects of tax policies over a time period, the practitioner of public economics will find that the relevant literature offers a wide variety of empirical approaches. For example, studies vary with respect to the definition of individual well-being and to what extent explicit benchmarking techniques are utilized to describe policy effects. The present paper shows how the concept of distributional benchmarking can be exploited to describe methodological options for the tax policy analyst. We present classifications of the various approaches which can be found in the literature for evaluations of individual taxation along these lines, and provide empirical illustrations and interpretations for the case of the personal income tax schedule of Norway for the period 2000–2010.

Keywords: Tax policy, Common base, Welfare metric

JEL classification: H23, H24, I31, J22

Acknowledgements: We thank Bård Lian who has assisted the work to obtain empirical illustrations. We are grateful to Arvid Raknerud and Andreas Peichl and other participants of the the 2nd Joint IZA/Statistics Norway/OFS Tax Policy Workshop (Oslo, May, 2013) for comments.

Address: Thor O. Thoresen, Statistics Norway, Research Department, and Oslo Fiscal Studies, Department of Economics, University of Oslo, Norway, Email: tot@ssb.no

Zhiyang Jia, Statistics Norway, Research Department, Email: zhiyang.jia@ssb.no

Peter J. Lambert, Department of Economics, University of Oregon, Eugene, USA and Statistics Norway, Oslo, Norway, email: plambert@uoregon.edu

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:
Statistics Norway
Telephone: +47 62 88 55 00
E-mail: Salg-abonnement@ssb.no

ISSN 0809-733X
Print: Statistics Norway

Sammendrag

I analyser som har som formål å evaluere effekter av skattepolitikk over tid anvendes det ulike metoder. Utgangspunktet for dette arbeidet er at det i slike studier er nyttig å vurdere skattepolitikken opp mot kontrafaktiske utfall, dvs å etablere en «benchmark» i analysene. Det vises til hvordan mikrodata-baserte evalueringer av skattepolitikk over tid anvender benchmarking-teknikker. I gjennomgangen av litteraturen på feltet vises det både til analyser som er baserte på anvendelse av simuleringsmodeller og studier som anvender observerte data. I den siste gruppen av studier er benchmarkingen mer implisitt. Videre benyttes data og simuleringsmodeller for Norge til å beskrive hvordan de ulike teknikkene kan anvendes i evalueringer av norsk skattepolitikk på 2000-tallet, herunder fordelingseffekter av skattereformen i 2006. Skatteendringene over en tidsperiode beskrives både ved endringer i inntekt og for alternative mål på velferd (som nytte).

1. Introduction

Evaluation of distributional effects of tax policies over time is an important task for the practitioner of public economics, and the literature offers numerous analyses with a wide variety of different methodological angles, see for instance Pechman and Okner (1974), Feldstein (1983), Bradford (1995), Boadway and Keen (2000), and Lambert (2001). The present paper suggests that studies which use cross-sectional micro data to evaluate the distributional effect of tax policies over time can be prolifically classified and contrasted by using the concept of "distributional benchmarking". What is meant by the term "distributional benchmarking" in this context? First of all, it means that studies which seek to isolate the distributional effects of tax changes, do so by establishing a baseline and implicitly or explicitly adopting counterfactual scenarios which describe what would have happened in the absence of the tax policy change. Distributional benchmarking is related to, but should be kept separated from, other methodological choices, such as the choice of measure of well-being (or welfare metric) and techniques to aggregate individual level measures. The aim of the present study is to provide an overview over the alternative benchmarking methodologies that the practitioner faces when setting out to evaluate the distributional effects of tax policies in retrospect by using micro data.

Standard time series of yearly measures of redistribution are basically silent about the contribution of tax policy, since they are influenced by many other factors, such as demographical developments and the business cycle. One therefore needs to activate techniques to extract the contribution of policy makers. The concepts of equivalent variation (*EV*) and compensating variation (*CV*) are essential microeconomic tools for evaluations of changes in the taxation of individuals. Based on the duality theory of microeconomics, *EV* and *CV* represent money metric utility measures of policy effects, derived from the notion of equalization of utility before and after a policy change. Distributional analysis based on money metric utility represents an attractive empirical approach, founded on an appealing notion of well-being (utility). As implementation typically relies on simulation modeling to describe individual behavior under different choice restrictions, the baseline and the counterfactual scenario are straightforwardly established, see Muellbauer (1974), Deaton (1977), King (1983) and Jorgenson and Slesnick (1984). However, simulation models can only provide simplified expositions of the actual choice process, and the distributional benchmarking in this case also relies on controversial procedures for interpersonal comparability of utility and the resulting ranking of individual well-being.

Therefore, instead of depending on simulation results derived from stylized models, practitioners of public economics often employ information on well-being in terms of observed indicators, such as income or consumption.[1] A wide variety of empirical approaches can be seen among studies which discuss distributional effects of tax policy changes in terms of changes in incomes; thus neglecting the contribution from leisure in the characterization of well-being. However, it is often less clear in income-based studies, than in studies based on more comprehensive welfare metrics (such as money metric utility), exactly what the tax policies are measured against, i.e., what the baseline is. In income-based non-behavioral and behavioral simulation studies the baseline and the counterfactual scenario are straightforwardly established and conceptualized, but not so in many other analysis based on income, for instance in income decomposition studies. Recently, however, methods have been developed to benchmark tax policy effects when addressing information on incomes directly (i.e., not applying simulation tools). In these studies tax policy effects are typically characterized by measures of redistribution, where the pre-tax income distribution is used in the benchmarking.

We review studies both based on more comprehensive representation of well-being, such as money metric utility, and on income-based frameworks of analysis. Whereas the framework based on approximations to individual utilities, which we may characterize as a welfarist approach (Kaplow and Shavell, 2001), often applies simulation models which are useful tools for the benchmarking, a main message of the present study is that the benchmarking concept can also be gainfully exploited to characterize analyses founded on effects of tax policy changes on incomes.

We go on to demonstrate and contrast the implications of these various approaches by applying them to the discussion of the distributional effects of the personal income tax for Norway in the period 2000–2010, which includes the tax reform of 2006. We use both income data and modeling tools for Norway to this end.

The reminder of the paper is organized as follows. Section 2 describes analyses which use approximations to individual utility as the key concept of individual well-being, i.e., the welfarist approach to tax policy analysis. Section 3 focuses on studies which use income as the main measure of well-being, and divides the covered studies into analyses that use simulation models and approaches which rely only on observed information or actual data. In the case of simulation studies, we consider both those which involve behavioral effects, and those which do not. Section 4 provides illustrative measures for Norway in the period 2000–2010 using some of the techniques. Section 5 concludes the paper with a summary and discussion of the way ahead.

[1] For instance, the Luxembourg Income Studies database (LIS) provides harmonized income data for many countries around the world.

2. Benchmarking in the welfarist approach

According to a Bergson-Samuelson social welfare function, society's overall well-being over a period of time is given by an aggregation over individual utilities (u), $W = \int u_{it}$, where $i = 1,2, \ldots . n$ indexes individuals and $t = 1,2, \ldots s$ represents the time period. A wide variety of empirical work has been carried out to estimate the measures u_{it} and utilize them in discussing effects of fiscal policy. One way to obtain information about u_{it} is to develop a structural decision model to describe how individuals make decisions according to preferences and constraints.[2] Given a specific cardinalization of utility, this information in turn can be utilized to describe well-being across individuals and across different policy regimes, also often employing aggregation techniques to implement social welfare judgments.

The conversion from preference heterogeneity in a positive sense to normative implications has in practical terms raised concern regarding the choice of welfare metric, and some authors avoid individual choice as basis for interpersonal comparability of utility by using an identical utility function for all individuals, see King (1983) and Aaberge and Colombino (2013).[3] The step from describing peoples' heterogeneous preferences to descriptions of the social good has generated substantial concern, which goes back to Arrow's impossibility theorem. The criticism comes from different angles, see Slesnick (1998); Sen (1977) criticizes the narrowness of the description of the social good, whereas Blackorby and Donaldson (1988) point out that social welfare functions defined over utilities may not be quasiconcave and may therefore have unattractive ethical characteristics.[4]

A standard procedure is to let welfare change be represented by a measure of change in money metric utility, see Blundell, Preston and Walker (1994) and Slesnick (1998). The effects of policy changes can then be measured by the concepts of equivalent variation (EV) and compensating variation (CV). Let the utility function be written as a function of prices and the tax and transfer system, f, and the non-wage income, y, $U = v(f, y)$. The pre-reform and post-reform situations are characterized by (f^0, y^0) and (f^1, y^1). The measurement of EV uses the current tax system, f^0, say, as the base and asks what income change given the current (pre-reform) schedule would be equivalent to the change in policy in terms of effects on utility. EV is seen as an addition to income, y, to keep utility constant at the post-reform level, in terms of the indirect utility function, $v(f^0, y^0 + EV) = u^1$. The reference for

---

[2] Recently, we have witnessed exchanges in the literature concerning the validity of structural model results, see Angrist and Pischke (2010), Deaton (2010), Heckman (2010), Heckman and Urzua (2010), Imbens (2010), and Keane (2010), with some of the authors arguing that results derived from quasi-experimental research designs are preferable.

[3] However, this leads to inconsistencies, whereby individuals' welfare metric may differ from their utility levels determining their labor supply (Creedy and Hérault, 2012).

[4] These interpretational and methodological challenges may have resulted in reticence among practitioners to enter into utility based measures of the social good. However, Slesnick (1998) argues that "Rather than avoid the issue, a more appealing strategy is to confront it head on and make explicit the subjective judgements that are required to aggregate household welfare measures consistently into measures of social welfare" (p. 2137). Donaldson (1992) argues along the same lines.

measuring *EV* is thus the final utility level and the initial tax schedule. Correspondingly, *CV* uses the new schedule as the base for the measurement and the question is how much money is needed after the tax change to make the agent equally well off as before the tax changed: $v(f^1, y^1 - CV) = u^0$, which signifies that the reference level for measurement of utility is the initial level of utility. Thus, in the money metric utility approach, benchmarking is carried out by holding indirect utility constant before and after a policy change. The two methods vary with respect to using the pre-reform (*CV*) or the post-reform (*EV*) utility levels and pre-reform (*EV*) or the post-reform (*CV*) tax schedules as the benchmark.

Both *CV* and *EV* are invariant under increasing monotonic transformations of the underlying preference function. However, as already noted, a main challenge is to relate the tax policy effects to distributional implications. Any description of money metric utility change must be accompanied by a procedure to relate welfare changes to the baseline description of well-being. Preston and Walker (1999) and Decoster and Haan (2010) discuss alternative welfare metrics which can be used in that respect, the latter study clearly describing the sensitivity of results with respect to metric.[5] Creedy and Hérault (2011) obtain their measures for the assessment of welfare change from pre-reform and post-reform measures of money metric utility based on a measure of full income, using the maximum weekly hours of work (80 hours) in the calculations,[6] but they also show results for alternative (non-welfarist) measures of rankings of well-being. Similar to the latter, studies rank individual measures of measures of *CV* and/or *EV* by disposable income; see for example Labeaga, Oliver and Spadaro (2008) and Dagsvik, Locatelli and Strøm (2009).

The rich informational content of this line of research is clearly seen in several tax policy studies. It is however often used for discussing prospective or hypothetical policy changes and not for evaluations of already implemented policies, which is the issue discussed here. Moreover, many studies center the attention on aggregate welfare implications of changes instead of distributional aspects; see for instance Hausman (1981), Blomquist (1983), Aaberge, Dagsvik and Strøm (1995) and Blundell and Shephard (2012). Aronsson and Palme (1998), Labeaga, Oliver and Spadaro (2008), and Dagsvik, Locatelli and Strøm (2009) are examples of tax evaluation studies that discuss distributional effects in terms of money metric utility based on simulations of estimated labor supply models.[7] In Section 4, we shall use a discrete labor supply model, based on a random utility formulation, to measure changes in money metric utility in the Norwegian case as one of our empirical illustrations.

---

[5] Bargain et al. (2013a) discuss the role of choice of welfare metric for international comparison.

[6] They therefore derive measures of *EV* by addressing money metric utility before and after a reform, m_0 and m_1, for individual (after-tax) wage and non-labor income, w and I, and for identical maximum working hours, $h = 80$: $m_0(wh + I + EV) = m_1(wh + I)$.

[7] One may extend this reasoning to an intertemporal setting. Thus, Fullerton and Rogers (1996) use general equilibrium modeling and a lifetime perspective to discuss distributional effects of fundamental tax reforms. They determine *EV* for 12 different households, from poorest to richest when ranked by lifetime income.

3. Studies based on income

3.1 Conceptual issues

In most countries one can find micro data which straightforwardly can be employed by the practitioner to analyze income distributions over time. Compared to moving into structural modeling to recover underlying preferences, it is usually far less complicated to employ micro data and observed income in tax policy evaluations. Using income as the measure of well-being, welfare can now be described by an aggregation over individual income, $W = \int y_{it}$, where y_{it} is post-tax income. This form characterizes the so-called social evaluation function, pioneered by Atkinson (1970). See also Lambert (2001) on abbreviated social welfare functions and Champernowne and Cowell (1998) on reduced-form social welfare.

Kaplow and Shavell (2001) argue that departing from welfarism implies violation of the Pareto principle, and Banks and Diamond (2010) "do not see an underlying normative basis for reaching the conclusion that taxes should be related to taxable capacity without full consideration of the equilibrium consequences of following such an approach" (p. 610). However, even if descriptions of full equilibrium distributional effects are not accessible, it is important to control for the most important partial effects. Correspondingly, many studies that discuss distributional effects of tax policy changes in terms of changes in income go beyond a simple descriptive approach and use various methods to discuss effects of counterfactual policies.[8] This is sometimes characterized as "what if" analysis (Bourguignon and Ferreira, 2003; Clark and Leicester, 2004; Poterba, 2007), to signify that there is a counterfactual course, which serves as a benchmark. This type of analysis can also take an "ex ante perspective" (Bourguignon and Spadaro, 2006), meaning that one assesses effects of prospective reforms.

In Table 1 we present a classification scheme for some of the different approaches which have appeared in the literature. Studies are differentiated with respect to whether distributional effects are categorized with respect to application of simulation models, or alternatively, the use of observed incomes (actual data). The elements of this classification scheme will be discussed next.

[8] Other distributional analyses with focus on the counterfactual include studies of wage and income distributions through different transformations of the Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973) decomposition of differences in means into distributional differences; examples are Juhn, Murphy and Pierce (1993), DiNardo, Fortin and Lemieux (1996), Bourguignon, Fournier and Gurgand (2001), Hyslop and Maré (2005), Bourguignon, Ferreira and Leite (2008), and Biewen and Juhasz (2012). Of course, the counterfactual is a key concept in econometrics. Examples of studies with an econometrical leaning, which explicitly pinpoint the role of the counterfactual for studies of distributional aspects, are Cunha, Heckman and Navarro (2006), Firpo, Fortin and Lemieux (2009) and Rothe (2010).

Table 1. Tax policy evaluations based on income

Model simulation		Actual data	
Non-behavioral	Behavioral	Yearly measures and decompositions	Common base
Clark and Leicester (2004), Poterba (2007), Kasten et al. (1994), Alm et al. (2005), Bargain and Callan (2010)	Haan and Steiner (2005), Elmendorf et al. (2008), Eissa and Hoynes (2011), Bargain et al. (2013a)	Jenkins (1988, 1995), Aronson, Johnson and Lambert (1994)	Dardanoni and Lambert (2002), Lambert and Thoresen (2009), Thoresen et al. (2012)

3.2 Studies based on model simulations

We distinguish between non-behavioral and behavioral simulation studies. Non-behavioral tax-benefit models are extensively used by policy-makers in many economies; some examples are NBER's TAXSIM (Feenberg and Coutts, 1993) and Urban-Brookings Tax Policy Center Microsimulation Model (Rohaly, Carasso and Saleem, 2005) for the U.S., EUROMOD (Sutherland, 2001) for countries of the European Union, IFS's TAXBEN (Giles and McCrae, 1995) for the UK, LOTTE (Aasness, Dagsvik and Thoresen, 2007) for Norway, and MITTS (Creedy et al., 2001) for Australia.

The tax-benefit model represents a powerful tool for discussing tax policy implications. The benchmarking in studies categorized under "non-behavioral model simulation" is rather straightforward. It is done by means of various types of "fixed income" approaches, where the pre-tax income distributions are kept fixed; a base year being chosen and exposed to taxation as per the various tax schemes of the period. If we denote the tax-law by d and let the income vector be represented by X, the fixed income approach corresponds to measuring outcomes of n schedules, $d_1(X_b), d_2(X_b), \dots, d_n(X_b)$ with respect to the same base income, X_b, which, for instance, invokes the possibility of measuring the difference between $d_j(X_b)$ and $d_k(X_b)$ using some aggregation measure.

The choice of indexation of the tax-rule becomes crucial under this procedure, and alternatives include indexation by development in prices, wages and incomes. For instance, Kasten, Sammartino and Toder (1994) and Bargain and Callan (2010) use income growth rates to inflate or deflate tax rules, whereas wage developments are employed for indexation in Thoresen (2004). Clark and Leicester (2004) report that results are highly dependent on the choice of uprating methods, in their case whether price or income is used as the deflator. Further, the choice of base year is critical, as emphasized by Lambert and Thoresen (2009). Kasten, Sammartino and Toder present results for different choices of base income, X_b, namely they apply tax laws for 1980, 1985, 1989 and 1993 on incomes both for 1985 and 1989 (with data for the US), and report that "the income-fixed simulations show substantially different results, dependent on the year to which the tax changes are applied" (page 11).

The base dependence problem is further explored by studies categorized under "Common base" in Table 1 and studies which are influenced by Shorrocks-Shapley decomposition methods (Shorrocks, 2013). We will return to the former type of studies in the next sub-section, whereas studies inspired by the Shorrocks-Shapley method benefit from applying simulation tools (and thereby fit into the "Model simulation" category of Table 1), as for example seen in Bargain and Callan (2010).[9] Bargain and Callan identify three effects: change in the tax-benefit structure, change in nominal levels of tax parameters relative to the change in income levels, and non-tax changes. The Shapley reasoning comes into play by acknowledging that there is no obvious choice between using the final or initial state as the point of departure. In Bargain and Callan (2010) the reference to the Shapley technology is used to justify taking the average of results according to the initial and final configurations. Thus, tax policy effects are generated by $\frac{1}{2}\{g[d_a(X_a)] - g[d_b(X_a)]\} + \frac{1}{2}\{g[d_a(X_b)] - g[d_b(X_b)]\}$, where g symbolizes an aggregation function (such as the Gini coefficient) and subscripts a and b denote initial and final periods. The procedure opens up possibilities for identifying effects from other (non-tax) factors too.[10] This technique is further elaborated upon in Bargain et al. (2013b), where identification of tax policies is obtained by addressing information of yearly fixed income simulations over the time 1979 to 2007 for the U.S. (and again averaging over results for the initial and final income base). By comparing results of simulations with the developments in actual measures over time, they distinguish between tax policy effects and other factors.

Different approaches are followed to account for behavioral effects in studies using simulation approaches. Both Poterba (2007) and Thoresen (2004) use simulation tools for describing direct effects of tax changes, but not when describing behavioral adjustments. Poterba adds theoretical discussions and some main findings from the literature, whereas Thoresen provides estimates of the elasticity of taxable income from panel data regressions, and discusses behavioral response estimates in relation to the results of non-behavioral simulations.[11] Estimates of the elasticity of taxable income are also used by both by Elmendorf et al. (2008) and Bargain et al. (2013b) as input to a (non-behavioral) micro simulation model to control for distributional effects of responses (when discussing effects of US tax policy).[12]

[9] Shorrocks (2013) refers to the Shapley value (after Lloyd Shapley) of a cooperative game, where the solution assigns each player the marginal contribution averaged over all possible coalitions, as the method suggests to take the average over all possible elimination sequences in decompositions. The reference to the Shapley value in the base dependence context (as in Bargain and Callan, 2010) mainly stems from applying a technique based on averaging.

[10] Identified by $\frac{1}{2}\{g[d_a(X_b)] - g[d_a(X_a)]\} + \frac{1}{2}\{g[d_b(X_b)] - g[d_b(X_a)]\}$. See also Bargain (2012a) for application of this method.

[11] See details about this method to obtain measures of tax responsiveness in Feldstein (1995) and Saez, Slemrod and Giertz (2012).

[12] Using measures of income instead of simulation tools based on decision models may be seen as more practical. A realistic, structural decision model would normally be constructed for a subgroup of the population, which means that one needs different modeling tools for various groups, such as wage earners and self-employed.

Turning to studies which are based on (explicit) individual decision models in income distribution studies, Bargain (2012b) uses a discrete choice labor supply model to account for the contribution of tax induced behavioral changes (UK tax policies 1998–2001), given a Shapley decomposition procedure. As identification of the behavioral effects can be carried out with different combinations of income distributions and policies as the benchmark (initial income and initial policy, initial income and final policy, and so on), the Shapley method is used to provide average effects. Discrete choice models are also used by Haan and Steiner (2005) when evaluating German tax reforms. We will return to an application of the discrete choice labor supply model when presenting empirical illustrations for Norway in Section 4.

3.3 Studies based on actual data

Among studies based on observed data rather than simulation models, we find studies that are descriptive in nature, with no methodological efforts to isolate the effects of tax changes within the observed levels of inequality or income redistribution, as in Aronson, Johnson and Lambert (1994), Bishop et al. (1997), and Lambert and Ramos (1997). Further, one finds several studies which employ different types of decomposition methods in order to discuss effects of taxes. In Table 1 we refer to analyses which apply decomposition by subgroups (Jenkins, 1988, 1995); a recent example is Wang and Caminada (2011).

There are implicit or explicit benchmarks invoked in all of the~~se~~ studies categorized under "Yearly measures and decompositions".[13] For Aronson, Johnson and Lambert (1994), which measures income redistribution by decomposing post-tax inequality across pre-tax equals or close equals groups (see Urban and Lambert, 2008, on the latter), benchmarking refers to the setting up of the baseline as the counterfactual disposable income distribution under an equal-yield proportional income tax. The redistribution measure is Reynolds and Smolensky's (1977), $G_X - G_{X-T}$, where $G_X$is the pre-tax Gini coefficient and G_{X-T} is the post-tax Gini coefficient. Other Gini-based indices which may be used are essentially normalizations of this (Musgrave and Thin's (1948) $\frac{1-G_{X-T}}{1-G_X}$ and Pechman and Okner's (1974) $\frac{G_X-G_{X-T}}{G_X}$ normalize by pre-tax equality and pre-tax inequality respectively, but involve the same implicit benchmark of the yield-neutral proportional tax). Authors such as Duclos (1993) have used extended Gini coefficients analogously, involving a distributional judgment parameter *v*, set by the analyst. Blackorby and Donaldson (1984) use Atkinson indices (I) rather than Gini coefficients, their index is $\frac{I_X(e)-I_{X-T}(e)}{1-I_X(e)}$, where e is an inequality aversion parameter chosen by the analyst, whilst Kiefer's (1985) measure, $I_X(e) - I_{X-T}(e)$, is an unnormalized version of this. The choices of

[13] Also, one may remark that even though there is less controversy concerning the use of income as a measure of individual well-being, there are substantial measurement challenges in interpersonal comparisons here too, for instance due to people living in households of different sizes.

distributional judgement parameter, *v*, and inequality aversion parameter, *e*, determine, for the respective inequality indices, the income weighting structure. The higher is *v* or *e*, the greater is the weight placed on lower incomes relative to higher ones. The values of *v* and *e* that are chosen by the analyst are subjective judgements, which are required to aggregate household measures into welfare measures. In the case of the Blackorby-Donaldson and Kiefer indices, the benchmark is of a welfare-neutral (rather than revenue-neutral) proportional income tax, the welfare function being the one that is conditioned by the chosen degree of inequality aversion, *e*, as in Atkinson (1970).

A measurement exercise of a rather different nature, which is due to Fellman, Jäntti and Lambert (1999), while still using the Gini or an extended Gini coefficient, sets its benchmark tax scheme as the one which top-slices all high incomes to the same value while maintaining the same total revenue (notionally). This has been characterized as an "optimal yardstick": actual inequality reduction is measured relative to the maximal that would notionally be possible given the revenue constraint.[14]

The decomposability properties of the Gini coefficient, particularly, have led to demographic subgroup decomposition studies of post-tax income inequality, see e.g. Jenkins (1988, 1995). The counterfactual for post-tax inequality analysis is, simply, the pre-reform level of post-tax inequality or some target value for post-reform post-tax income inequality, whilst, as we have remarked, for income redistributional studies it is an equal yield proportional tax, relative to which a tax of the same size has a pure inequality effect.

The Gini coefficient permits other decompositions too. As Pfähler (1990) demonstrates, the elements of a tax scheme can be introduced one by one (the allowances, deductions, marginal rate structure and any tax credits) to permit an assessment of the roles of each of these elements in income redistribution. The overarching benchmark scenario is still that of an equal yield proportional tax, but Pfähler's sequencing procedure means that the redistributive effect of each constituent is measured against a "local" benchmark provided by the previous calculation.[15] Lambert's (2001, pp. 214-216) reformulation of Pfähler's decomposition uses gross income as a fixed benchmark for every constituent.

The Gini index can also be decomposed across income sources (using concentration coefficients), as shown by Shorrocks (1982). This too can be used in the study of income redistribution. See Fuest, Niehues and Peichl (2010) for a study of tax benefit redistributional effects in EU countries, Iyer and

[14] Empirical properties of all of these approaches are compared and contrasted for Scandinavia and the U.S. in Lambert, Nesbakken and Thoresen (2012), and are found to have somewhat differing characteristics (as one would expect).

[15] In Fuest, Niehues and Peichl (2010), a slightly different manifestation of the sequencing problem arises. These authors analyze EU tax benefit systems, instrument by instrument, in each case comparing the status quo with a counterfactual distribution which lacks the instrument in question.

Reckers (2012) for a U.S. study, and Kristjánsson's (2013) theoretical analysis for a dual income tax (DIT) system. The counterfactual for Kristjánsson's study is a comprehensive proportional tax at the same overall rate as that of the combined labor income and capital income components of the DIT.

A line of research initiated by Dardanoni and Lambert (2002) involves a different type of benchmarking to assess the redistributional effect of a tax reform. The procedure, which we shall characterize as the "transplant and compare" method shortly (see on), is as follows. First, observed differences in pre-tax inequality before and after the reform are eliminated (or minimized) by controlling for location and spread differences in logged distributions. This means identifying an appropriate linear "deformation function", whose parameters are given by the intercept and slope estimates in an OLS regression. The R^2 statistic of this regression provides the measure of goodness-of-fit for this deformation. If the fit is sufficiently good, then the pre-reform and deformed post-reform distributions of income before tax differ in logarithms essentially only by location and scale, that is, by an isoelastic transformation (in levels rather than logs). Second, the post-reform income distribution after tax is correspondingly deformed; this means, that a conjugate of the reformed income tax is now at work between deformed distributions. If the slope parameter in the regression just referred to is less than (more than) unity, this conjugate is more (less) progressive than the actual schedule. Finally, the redistributional properties of the reform may be assessed in terms of this conjugate, now having corrected for pre-tax inequality differences occurring concomitantly with the reform, using conventional (e.g. Gini-based) methodology. Thus, this method is closely connected to the fixed income approach, as the benchmarking according to both methods rests upon establishing a baseline pre-tax income distribution.[16] Practical issues concerning implementation of this approach are discussed in Lambert and Thoresen's (2009) application of the method to Norway for the period 1992–2004.

One practical advantage of the transplant and compare method, contrasted to for example the fixed income approach (described in Section 3.2), is that simulation tools are not required for this approach. Provided only that the shape dissimilarities between the chosen baseline distribution and the actual pre-tax distributions involved are confined to be linear in logarithms, as confirmed by high R^2 statistics, redistributional findings will not vary (Dardanoni and Lambert, 2002, pp. 105–106). Thus, instead of applying a detailed tax schedule representation, which for example under the fixed income procedure has shown to be susceptible to base dependence problems, the transplant and compare approach involves a more general approximation procedure, letting the pre-tax income distributions be

[16] Even though the measurement of redistributional effects may provide a suitable point of departure according to both methods, in effect, the benchmarking procedures turn the underlying welfare reasoning into a standard approach, with the (normalized) distribution of post-tax income as the objective. In this perspective, the pre-tax income distribution plays a role as tool in the benchmarking procedure, implying that the pre-tax income distribution is assigned a normative weight which may be disputed (Kaplow and Shavell, 2001).

described by parametric distribution functions. Detailed tax policy analysis, as for the fixed income approach, is traded for a more general approximation based on overall descriptions of income distributions.

Further, as the transplant and compare procedure is based on repeated observations of actual pre-tax income distributions, each year's pre-tax income distribution incorporates the behavioral effects due to the adjoining tax schedules. Thoresen et al. (2012) describe how these indirect effects can be separated out by using estimates of the elasticity of taxable income.

4. Illustrations of methods

In this section we demonstrate and contrast the implications of some of the approaches we have discussed, by applying them to the discussion of the distributional effects of the personal income tax for Norway in the period 2000–2010, a period which includes the tax reform of 2006. Before providing the empirical illustrations, we present some details of the data and modeling tools we use. First, we consider the non-behavioral simulation tool, to be used in evaluation procedures based on actual data and non-behavioral simulations.[17] We then briefly describe the discrete choice random utility labor supply model, which we employ to simulate behavioral effects of the tax reform of 2006, leading to more comprehensive measures of well-being.

4.1 Data and tax-benefit model

As already denoted, non-behavioral tax-benefit models are extensively used by policy-makers; some examples of models are referred to in Section 3.[18] These models share a rather simple structure. The model used here, the Norwegian tax-benefit model LOTTE-Skatt (Aasness, Dagsvik and Thoresen, 2007) consists of four parts: individual income data, a program that projects data to the year of interest, a set of tax rules, and a simulation routine that applies the tax rules to individual records. Standard model results include estimates of the aggregate revenue effects and distributional effects of various policies.

LOTTE's main source of data is individual income tax returns. In Norway, income tax return data are now electronically available for all citizens, and the national register of ground addresses and buildings can be used to obtain information on the composition of households. This allows us (in principle) to simulate tax burdens for the whole population. However, in order to reduce computing

[17] The data source of the tax-benefit model is the Income statistics for household. The difference between the post-tax income distribution observed in data and generated by the tax-benefit model is negligible.

[18] Terminology used to describe models varies. Some authors (such as Bourguignon and Spadaro, 2006) refer to non-behavioral simulation models as arithmetical models, whereas for instance Creedy and Duncan (2005) distinguish between first round (non-behavioral) and second round (behavioral) effects (with third round effects denoting endogenous wage changes).

time, one typically runs the model for a sub-sample, as we do here. We use macroeconomic predictions for income growth, income composition, etc., to project the data forward. Detailed tax rules for different years are compiled in the tax law module of the model.

4.2 A discrete choice labor supply model

LOTTE-Skatt is available to Norwegian decision-makers through the model system LOTTE, a collection of simulation tools which also includes the labor supply model LOTTE-Arbeid. See the survey by Blundell and MaCurdy (1999) on issues arising in the modeling of labor supply responses to tax policy change. Since van Soest (1995), the adaptation of random utility discrete choice models has become increasingly popular, because this approach simplifies the implementation of complicated nonlinear budget constraints; see survey in Creedy and Kalb (2005).

We now give a summary description of the discrete choice labor supply model that is applied in the analysis to follow. The discrete choice labor supply model departs from the theory of random utility models (see McFadden, 1984). Let $U(C,h)$ denote the agent's utility function of real disposable income C and hours of work h, and assume that

(4.1) $$U(C_j,h_j) = v(C_j,h_j) + \eta(C_j,h_j),$$

where $v(C_j,h_j)$ is a positive deterministic term that represents the mean utility across observationally identical agents and $\eta(C_j,h_j)$ is a random term that is uncorrelated with the structural term $v(C_j,h_j)$, with the c.d.f. $\exp(-\exp(-z))$ defined for any real z.[19] The budget constraint is given by

(4.2) $$C_j = f(hw_j, I),$$

where w and I are the wage and non-labor income respectively and $f(\cdot)$ is the function that transforms gross income into after-tax household income. The probability $p(h_j)$ that the agent will supply h_j hours of work, given a set D of feasible hours of work, the budget constraint in (4.2), the wage rate w and non-labor income I, is given by

(4.3) $$p(h_j) = P(\widetilde{U}(h_j) = \max_{z\in D} \widetilde{U}(z)) = \frac{exp(\psi(C_j,h_j))}{exp(\psi(0)) + \sum_{z\in D} exp(\psi(z))}.$$

[19] This c.d.f. is often called the type III (standard) extreme value distribution, or Gumbel distribution. Other authors call this the type I extreme value distribution.

The structural part of the utility function, $v\left(C_j, h_j\right)$, is assumed to have a convenient functional form and is allowed to depend on individual covariates. The version used in the following, see Dagsvik et al. (2012), takes into account that job offers vary across individuals and across the distribution of working hours. The resulting choice model is analogous to a multinomial logit model with representative utility terms, $\psi\left(h_j\right)$, weighted by the frequencies of available jobs, $m\left(h_j\right)$, which implies that the choice probabilities can be seen as: $p\left(h_j\right) = \frac{exp\left(\psi(C_j,h_j)\right)m(h_j)}{exp(\psi(0))+\sum_{z\in D} exp(\psi(z))m(z)}$.

4.3 Measures of money metric utility

Given the stochastic character of the discrete choice labor supply model, measures of the social good also become stochastic, see Dagsvik and Karlstrøm (2005). For analytical and computational challenges generated by the use of microsimulation models based on specifications that are nonlinear in income, see for instance McFadden (1999) and Herriges and Kling (1999).

As before, let the pair (f^0, y^0) and (f^1, y^1) represent pre-reform and post-reform structures. Then, when using derivation of the measurement of *EV* for illustration, *EV* is implicitly the value that solves (for *k* alternatives with working hours h_j),

(4.4) $$\max_k(\, v(f^0, y^0 + EV) + \varepsilon_k) = \max_k(\, v(f^1, y^1) + \varepsilon_k).$$

Hence *EV* is a random variable that may depend on all stochastic terms $\{\varepsilon_j\}$. This is because the maximum of the left hand side of (4.4) is not necessarily attained at the same discrete alternative as the maximum of the right hand side, except in special cases. Consequently, the random terms on each side will not cancel. A key assumption, followed in most of the literature, is that there is no change in the error terms between the pre-reform and the post-reform situations.

There exist different procedures to derive welfare measures; Dagsvik and Karlstrøm (2005) derive explicit formulae for the expected value of the Hicksian choice probabilities and for the expenditure function, whereas McFadden (1999) and Creedy, Hérault and Kalb (2011) propose simulation procedures that establish probability distributions for welfare measures based on drawing sets of random utility components.[20]

[20] See discussion of different alternatives in Creedy et al. (2013).

4.4 Evaluation of Norwegian tax policy 2000–2010

4.4.1 Income based measures

We start the illustrations of tax policy effect by simply using actual data to evaluate tax redistribution over time. In Figure 1, two indices of redistribution presented in Section 3.3, the Reynolds-Smolensky index and the Blackorby-Donaldson index, are used to describe the redistribution by the Norwegian personal income tax schedule from 2000 to 2010.[21] We see that the two measures provide rather similar descriptions of the trend in tax redistribution. As there is close correspondence between taxation of dividends and redistribution in the Norwegian case,[22] the pattern of Figure 1 is clearly influenced by a temporary tax on dividends in 2001 as well as the tax reform of 2006. The tax reform of 2006 implied a substantial realignment of dividend income and wage income taxation under the Norwegian dual income tax system (introduced in 1992),[23] as the top marginal tax rates were reduced combined with the introduction of a tax on dividends (above a normal rate of return); we will return to the effect of the lower marginal tax rates on earned income.

As the taxation of dividends was pre-announced prior to both the temporary tax in 2001 and the more permanent change in 2006, one faces measurement problems, such as timing effects resulting from the shifting of income between personal and corporate tax bases. The third time series in Figure 1 (the one that stops in 2008, taken from Thoresen et al., 2012) is based on an alternative income concept which seeks to account for the measurement problems due to timing effects by adding shares of corporate profits to individuals each year, regardless of whether these profits were transferred to the individuals or not. This latter curve basically lies below the other curve based on the Reynolds-Smolensky index, but most importantly shows less variation, as it is less sensitive to the fluctuation in income due to individual timing effects. This example shows clearly that in general the choice of income definition is important. However, as made clear in the Introduction, we prefer to keep benchmarking as separate from other methodological issues, such as choice of income concept.

Next, we describe tax policy contributions for the same time period by applying the fixed income approach and the transplant and compare method (for the latter method showing results both for the Reynolds-Smolensky index and the Blackorby-Donaldson index). As seen in Figure 2, the two

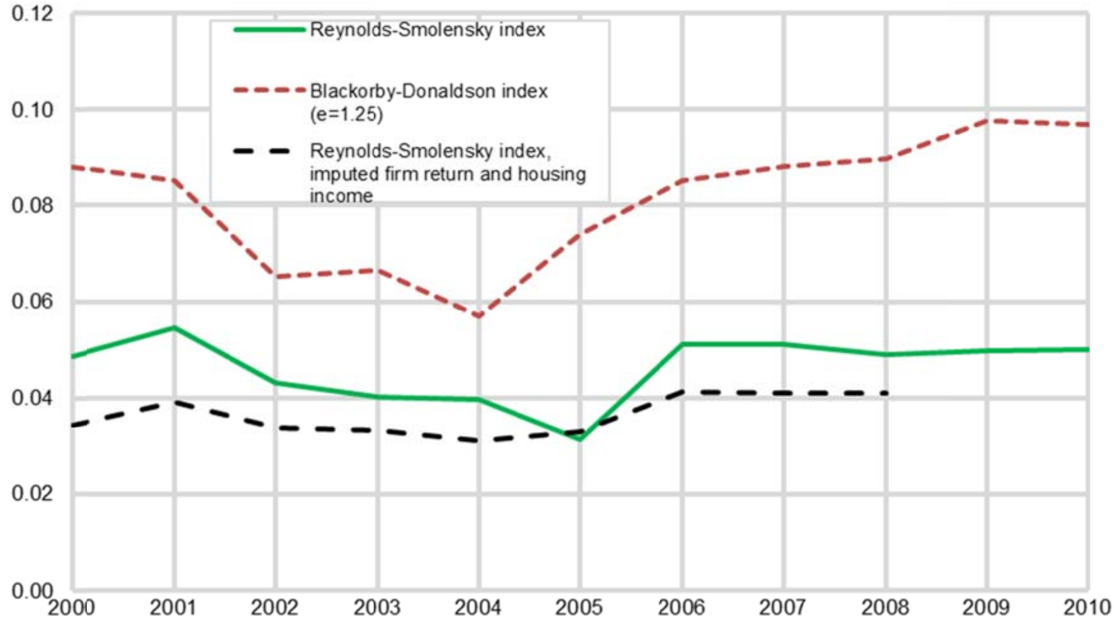
[21] The choice of equivalence scale and unit of analysis is also a key methodological issue in analysis of income distributions, which we so far have not brought up. Note that in all presentations of results incomes are measured in “equivalent values”, which means that the nominal values of aggregate income of the family have been weighted by an equivalence scale (the square root of the number of family members). The representation of each family when obtaining summary measures of redistributional effects depends on the number of family members. This is often characterized as employing the individual as the unit of analysis. Thus, incomes have been readjusted for interpersonal comparison similarly to what Ebert (1997) denotes as Method 3.

[22] The reason is that dividend income in Norway is an income component that almost exclusively benefits people at the high end of the income distribution. For example, 95 percent of dividends were received by individuals in decile 10 in 2004.

[23] A dual income tax system is characterized by separate tax schedules for capital and wage income; prior to the 2006 reform capital income and wage income were taxed by a (basic) flat rate of 28 percent, whereas a two-tier surtax supplemented the basic rate with respect to wage income.

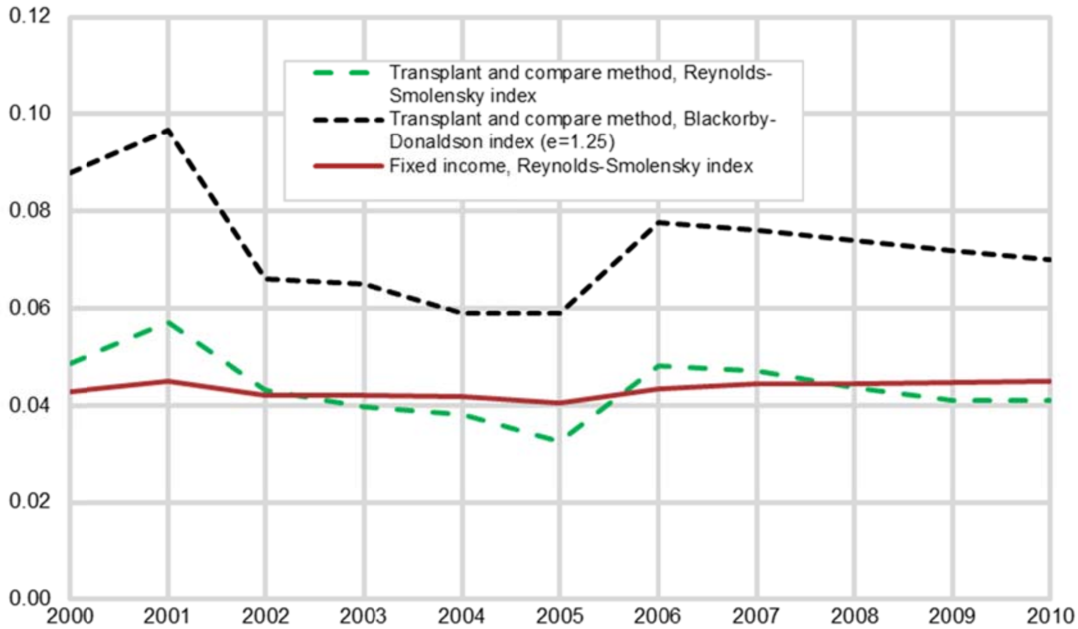
evaluation techniques provide very different descriptions of the distributional effect of the tax policies. The fixed income approach is not always able to pick up tax policy changes. For instance, we used 2010 as the base year for application of all years' tax schedules, and find that the taxation of dividends from 2006 and onward gives only a small increase in redistributional effect. This is easy to understand given that relatively low levels of dividends transferred to households in 2010. If 2005 had been used as the base, the effect of the dividend tax (from 2006 and onward) would have been more pronounced. The likelihood of base-dependent findings is considerable, if not inevitable, using the fixed income approach.

Figure 1. Redistributional effects of personal income tax in Norway 2000–2010, Reynolds-Smolensky index and Blackorby-Donaldson index



For the transplant and compare method, we chose 2000 as the base year. Note that for both the Reynolds-Smolensky and Blackorby-Donaldson indices, Figure 2 shows a constantly-changing assessment of redistributional effect as time progresses. It is easy to understand why. Consider 2001, for example. As relative pre-tax income differentials in 2001 were smaller than in 2000, the slope estimate of the OLS regression is above unity for that year (we estimate b to 1.06). One is using a conjugate of the 2001 tax schedule which is less progressive than the actual 2001 tax schedule to compute redistributional effect for 2001; using the deformed 2001 pre-tax distribution, which is, in essence, the same as the 2000 pre-tax income distribution. Now consider 2005, as another example. Pre-tax inequality is higher in 2005 than in 2000, whence the slope parameter is less than unity in this regression (b equal to 0.88): this means that the conjugate of the 2005 tax schedule used in the transplant and compare procedure is more progressive than the actual one in 2005 (and again, is applied in essence to the 2000 pre-tax income distribution). In this way, the progressivities of the tax schedules after 2000 are adjusted year by year – and all are applied, in essence, to the base year pre-tax distribution, that of 2000 – to take into account the changing pre-tax inequality profile across the years. This results in a graphical depiction over the whole period in which “a lot is happening”, and this would also have been true if we had made a different choice of base year.

Figure 2. Common-base (fixed income and transplant and compare) redistributional effects of personal income tax in Norway 2000–2010, Reynolds-Smolensky index and Blackorby-Donaldson index



The main drawback of this method is that income distributions are described by only two parameters, to comply with the isoelastic deformation function. If income distributions differ in shapes, the fit may not sufficiently close and the reproduction of the base year becomes less trustable. For instance, given the example of Norway 2000–2010, we find that 2005 stands out, and is hard to reshape into the base year by only two parameters.[24]

Incorporating indirect (behavioral) effects to the tax policy effects can be accounted for in the evaluation by applying the labor supply model that was presented in Section 4.2. As already discussed, controlling for behavioral effects can also be done by applying measures of income responses derived from reduced form data analysis, such as the elasticity of taxable income.

4.4.2 Evaluation by other measures of well-being

Next, we illustrate the use of other metrics of well-being than income alone to evaluate the redistributional effect of tax policies. Due to the fact that our behavioral model covers only the labor

[24] The time period 1992–2004, analyzed in Lambert and Thoresen (2009), appears to provide better fits.

supply dimension in individual/household's response to tax policies, we base our illustration on a sample of married couples where both are wage earners. Results are shown for tax policy changes induced by the 2006 tax reform, which means that the 2004 tax system is used to describe the "no-reform" counterfactual. Note that, in our setting, the utility function is stochastic, which means the outcome variables such as labor supply hours, wage income and measures of well-being will be stochastic as well.

Equation 4.4 describes the principle of how a money metric measure of utility change (*EV*) can be derived. As noted in Section 4.3, there exist different practical procedures to derive measures of changes in money metric utility; here we employ the procedure of Creedy, Hérault and Kalb (2011). One way of looking at the redistribution effect of a tax policy change using *EV* is to describe the distribution of *EV* over the population with respect to observable variables, such as (pre-reform) pre-tax income, post-tax income, household type, etc. In Table 2, we rank married couples by pre-tax income under the 2004 tax system and report the mean *EV* for each decile. From the table, we see that households benefit in general from the tax rule change, as the 2006 tax reform implied a tax reduction for most tax-payers (Thoresen, Aasness and Jia, 2010). Moreover, *EV* increases with the gross income level.

Table 2. Effect of the 2006 tax reform, mean *EV* for levels of pre-tax income in deciles, NOK

Decile	Pre-reform pre-tax income	EV
1	330,263	12,488
2	396,843	16,440
3	431,639	17,600
4	456,801	18,329
5	477,823	19,050
6	496,467	20,078
7	516,927	20,929
8	541,723	22,766
9	576,137	25,274
10	644,065	29,031

In Table 2 two rather different concepts to describe distributional effects are combined, employing pre-tax income to rank individuals and *EV* to describe effects of the reform. Fixing interpersonal comparison by employing welfare criteria not based on preferences is similar to approaches seen in Kornstad and Thoresen (2006), Labeaga, Oliver and Spadaro (2008) and Dagsvik, Locatelli and Strøm (2009), but they use post-tax (instead of pre-tax) income for the ranking. To emphasize that the literature offers different procedures for descriptions of results, we therefore show results for alternative individual welfare metrics: a full income measure obtained from Creedy and Hérault (2011), and the so-called "rente criterion" of Decoster and Haan (2010). We also compare estimates according to these concepts to (summary) measures based on disposable income (as in Table 2).

Creedy and Hérault (2011) propose to assign the pre-reform money metric utility value for each individual as the net full income under the pre-reform tax rule, i.e. the net income when individuals works the maximum hours available. Once the pre-reform money metric m_0 is assigned, the post-reform money metric is straightforward to calculate: $m_1 = m_0 + EV$. One can then calculate standard inequality measures, such as the Gini and the Atkinson indices, based on the pre- and post-reform distribution of money metric measures of full income. The problem with this approach is that choice of pre-reform money metric utility value is somewhat arbitrary; an issue which is further discussed by Decoster and Haan (2010), where different measures based on the concept of "subset dominance" are presented. The "rente criterion" is defined (in a net income and hours worked diagram) as the intercept of the indifference curve for net wage equal to zero, which is similar to "intercept income" in the list of welfare metrics of Preston and Walker (1999).

Table 3 presents summary statistics and inequality measures based on disposable income, full income and the rente criterion.[25] The redistributive effects of the two tax systems are compared directly using inequality measures for the definition of well-being, as the differences in the pre- and post-distributions of the same metric can be attributed to the tax changes. When using the full income metric, inequality increased as the result of the 2006 reform, suggesting a stronger redistributive effect from changing from the 2004 tax rules to the 2006 tax rules. The inequality measures are almost unchanged when individual utility levels are measured using the rente criterion, whereas if we use post-tax income to define individual well-being, the 2006 tax rules become more redistributive. Thus, these results highlight the importance of characterizing the initial distribution of well-being when employing benchmarking techniques in tax policy evaluations.

Table 3. Summary statistics and measures of inequality for different characterizations of individual well-being, for 2004 (pre-reform) and 2006 (post-reform)

		Mean (NOK)	Standard deviation	Gini index
Full income	Pre-reform	458,816	40,338	0.049
	Post-reform	479,010	43,719	0.051
Rente criterion	Pre-reform	284,335	43,450	0.086
	Post-reform	302,956	46,360	0.086
Disposable income	Pre-reform	343,685	55,075	0.090
	Post-reform	368,778	58,026	0.088

[25] Note that the effect of the 2006-reform on income earners is small, and pre-reform and post-reform differences are not obviously statistically significant.

5. Summary

As policy makers often are judged by their efforts to redistribute, evaluations of the distributional effects of tax policies are a key issue of empirical public finance. Correspondingly, there exist a wide variety of empirical approaches for the practitioner of public finance to employ when pursuing this objective. This survey of the literature has identified two main avenues of studies, separated by the concept used to describe individual well-being: income and utility. The utility based studies rely on simulation models to approximate measures of well-being before and after a policy change. Whereas a counterfactual scenario is easily established when applying simulation models, and thus the benchmarking is straightforward, the dependence on characteristics and qualities of modeling tools may be seen as a disadvantage. In particular, to establish decision models for each relevant subgroup, such as wage earners, self-employed, families with preschool children, etc., may be seen as an obstacle. Moreover, issues have been raised concerning the comparability of individual well-being when utility based concepts are used.

In this perspective, studies of observed income may be preferred. However, the identification of the contribution of tax policy changes is demanding, and none of the techniques put forward are without weaknesses. A base dependence problem is often inescapable in the fixed income simulation studies, and the transplant and compare procedure relies on income distributions to be sufficiently close to each other.

To summarize, there is certainly need for more efforts to develop new methods and improve current techniques in this field. As more and better data become available, this remains to be followed by new advances in methods to identify tax policy contributions in a distributional context. The present study suggests that the concept of distributional benchmarking serves as the guiding notion in these endeavors.

References

Aaberge, R. and U. Colombino (2013). Using a Microeconometric Model of Household Labour Supply to Design Optimal Income Taxes. *Scandinavian Journal of Economics*, 115, 449–475.

Aaberge, R., J.K. Dagsvik and S. Strøm (1995). Labor Supply Responses and Welfare Effects of Tax Reforms. *Scandinavian Journal of Economics*, 97, 635–659.

Aasness, J., J. K. Dagsvik and T. O. Thoresen (2007). "The Norwegian Tax-benefit Model System LOTTE". In A. Gupta and A. Harding (eds): *Modelling Our Future: Population Ageing, Health and Aged Care*, *International Symposia in Economic Theory and Econometrics*, Amsterdam: Elsevier Science, North-Holland, 513–518.

Alm, J., F. Lee and S. Wallace (2005). How fair? Changes in Federal Income Taxation and the Distribution of Income, 1978 to 1998. *Journal of Policy Analysis and Management*, 24, 5–22.

Angrist, J.D. and J.-S. Pischke (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24, 3–30.

Aronson, J.R., P. Johnson and P.J. Lambert (1994). Redistributive Effect and Unequal Income Tax Treatment. *Economic Journal*, 104, 262–270.

Aronsson, T. and M. Palme (1998). A Decade of Tax and Benefit Reforms in Sweden: Effects on Labour Supply, Welfare and Inequality. *Economica*, 65, 39–67.

Atkinson, A.B. (1970). On the Measurement of Inequality. *Journal of Economic Theory*, 2, 244–263.

Banks, J. and P. Diamond (2010). "The Base for Direct Taxation". In J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba (eds): *Dimensions of Tax Design: the Mirrlees Review*, Oxford: Oxford University Press, 548–648.

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

Bargain, O. (2012b). Decomposition Analysis of Distributive Policies Using Behavioural Simulations. *International Tax and Public Finance*, 19, 708–731.

Bargain, O. and T. Callan (2010). Analysing the Effects of Tax-Benefit Reforms on Income Distribution: a Decomposition Approach. *Journal of Economic Inequality*, 8, 1–21.

Bargain, O., A. Decoster, M. Dolls, D. Neumann, A. Peichl, and S. Siegloch (2013a). Welfare, Labor Supply and Heterogeneous Preferences: Evidence for Europe and the US. *Social Choice and Welfare*, forthcoming.

Bargain, O., M. Dolls, H. Immervoll, D. Neumann, A. Peichl, N. Pestel and S. Siegloch (2013b). Partisan Tax Policy and Income Inequality in the U.S., 1979–2007. IZA Discussion Paper No. 7190, Institute for the Study of Labor.

Biewen, M. and A. Juhasz (2012): Understanding Rising Income Inequality in Germany, 1999/2000–2005/2006. *Review of Income and Wealth*, 58, 1475–4991.

Bishop, J.A., K.W. Chow, J.P. Formby and C.-C. Ho (1997). Did Tax Reform Reduce Actual US Progressivity? Evidence from the Taxpayer Compliance Measurement Program. *International Tax and Public Finance*, 4, 177–197.

Blackorby, C. and D. Donaldson (1984). Ethical Social Index Numbers and the Measurement of Effective Tax/Benefit Progressivity. *Canadian Journal of Economics*, 17, 683–694.

Blackorby, C. and D. Donaldson (1988). Money Metric Utility: A Harmless Normalization. *Journal of Economic Theory*, 46, 120–29.

Blinder, A.S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8, 436–55.

Blomquist, N. S. (1983). The effect of income taxation on the labour supply of married men in Sweden.*Journal of Public Economics*, 22, 169–197.

Blundell, R. and T. MaCurdy (1999). "Labor Supply: A Review of Alternative Approaches". In O.C. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, Vol. 3A, Amsterdam: Elsevier, 1559–1695.

Blundell, R., I. Preston and I. Walker (eds.) (1994). *The Measurement of Household Welfare*, Cambridge: Cambridge University Press.

Blundell, R. and A. Shephard (2012). Employment, Hours of Work and the Optimal Taxation of Low-Income Families. *Review of Economic Studies*, 79, 481–510.

Boadway, R. and M. Keen (2000). "Redistribution". In A.B. Atkinson and F. Bourguignon (eds.): *Handbook of Income Distribution*, Vol. 1, Amsterdam: North-Holland, 677–789.

Bourguignon, F. and F.H.G. Ferreira (2003). "Ex-ante evaluation of policy reforms using behavioral models". In F. Bourguignon and L. Pereira da Silva, (eds): *The Impact of Economic Policies on Poverty and Income Distribution*, World Bank-Oxford University Press, 123–141.

Bourguignon, F., F.H.G. Ferreira and P.G. Leite (2008): Beyond Oaxaca–Blinder: Accounting for Differences in Household Income Distributions. *Journal of Economic Inequality,* 6, 117–148.

Bourguignon, F., M. Fournier and M. Gurgand (2001). Fast Development with a Stable Income Distribution: Taiwan, 1979–94. *Review of Income and Wealth,* 47, 139–163.

Bourguignon, F. and A. Spadaro (2006). Microsimulation as a Tool for Evaluating Redistribution Policies, *Journal of Economic Inequality*, 4, 77–106.

Bradford, D.F. (1995): *Tax Policy*. Edited by D.F. Bradford, Washington DC: The AEI Press.

Champernowne, D.G. and F.A. Cowell (1998). *Inequality and Income Distribution*, Cambridge: Cambridge University Press.

Clark, T., and A. Leicester (2004). Inequality and Two Decades of British Tax and Benefit Reforms. *Fiscal Studies*, 25, 129–158.

Creedy, J., J.K. Dagsvik, N. Hérault, Z. Jia, G. Kalb, and T.O. Thoresen (2013). On the Measurment of Welfare Changes in Random Utility Labour Supply Models, Manuscript, Statistics Norway/University of Melbourne.

Creedy, J. and A. Duncan (2005). Aggregating Labour Supply and Feedback Effects in Microsimulation. *Australian Journal of Labour Economics*, 8, 277–290.

Creedy, J., A. Duncan, G. Kalb, and H. Kew (2001). The Melbourne Institute Tax and Transfer Simulator (MITTS). Melbourne Institute Working Paper No. 16/01, The University of Melbourne.

Creedy, J. and N. Hérault (2011). Welfare-Improving Income Tax Reforms: A Microsimulation Analysis. *Oxford Economic Papers*, 64, 128–150.

Creedy, J. and N. Hérault (2012). Decomposing Inequality and Social Welfare Changes: The Use of Alternative Welfare Metrics, Working Paper 06/2012, Working Papers in Public Finance, University of Wellington.

Creedy, J., N. Hérault and G. Kalb (2011). Measuring Welfare Changes in Behavioural Microsimulation Modelling: Accounting for the Random Utility Component. *Journal of Applied Economics*, 14, 5–34.

Creedy, J. and G. Kalb (2005). Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation. *Journal of Economic Surveys*, 19, 697–734.

Cunha, F., J. Heckman and S. Navarro (2006). "Counterfactual Analysis of Inequality and Social Mobility". In G. Fields, D. Grusky and S. Morgan (eds): *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, Stanford: Stanford University Press, 290–346.

Dagsvik, J. and A. Karlström (2005). Compensating Variation and Hicksian Choice Probabilities in Random Utility Models that are Nonlinear in Income. *Review of Economic Studies*, 72, 57–76.

Dagsvik, J.K., Z. Jia, T. Kornstad and T.O. Thoresen (2012). Theoretical and Practical Arguments for Modeling Labor Supply as a Choice among Latent Jobs, forthcoming in *Journal of Economic Surveys*.

Dagsvik, J.K., M. Locatelli and S. Strøm (2009). Tax Reform, Sector-Specific Labor Supply and Welfare Effects. *Scandinavian Journal of Economics*, 111, 299–321.

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

Deaton, A. (1977). Equity, Efficiency, and the Structure of Indirect Taxation. *Journal of Public Economics*, 8, 299–312.

Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*,48, 424–455.

Decoster, A. and P. Haan (2010). Empirical Welfare Analysis in Random Utility Models of Labour Supply. Center for Economic Studies Discussion Paper Series 10.30, Katholieke Universiteit Leuven.

DiNardo, J., N.M. Fortin and T. Lemieux (1996). Labour Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach. *Econometrica*, 64, 1001–1044.

Donaldson, D. (1992) On the Aggregation of Money Measures of Wellbeing in Applied Welfare Economics. *Journal of Agricultural and Resource Economics*, 17, 88–102.

Duclos, J.-Y. (1993). Progressivity, Redistribution, and Equity, with Application to the British Tax and Benefit System. *Public Finance/Finances Publiques*, 48, 350–365.

Ebert, U. (1997). Social Welfare when Needs Differ: An Axiomatic Approach. *Economica*, 64, 233–244.

Elmendorf, D.W., J. Furman, W.G. Gale and B.H. Harris (2008). Distributional Effects of the 2001 and 2003 Tax Cuts: How Do Financing and Behavioral Responses Matter?" *National Tax Journal*, 61, 365-380.

Eissa, N. and H. Hoynes (2011). Redistribution and Tax Expenditures: The Earned Income Tax Credit. *National Tax Journal*, 64, 689–730.

Feenberg, D.R. and E. Coutts (1993). An Introduction to the TAXSIM Model. *Journal of Policy Analysis and Management*, 12, 189–194.

Feldstein. M. (1983). *Behavioral Simulation Methods in Tax Policy Analysis*, edited by M. Feldstein, Chicago: University of Chicago Press.

Feldstein, M. (1995). The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy*, 103, 551–572.

Fellman, J., M. Jäntti and P.J. Lambert (1999). Optimal Tax-Transfer Systems and Redistributive Policy. *Scandinavian Journal of Economics*, 101, 1–9.

Firpo, S., N.M. Fortin and T. Lemieux (2009) .Unconditional Quantile Regressions. Econometrica 77, 953–973.

Fuest, C., J. Niehues and A. Peichl (2010). The Redistributive Effects of Tax Benefit Systems in the Enlarged EU. *Public Finance Review*, 38, 473–500.

Fullerton, D. and D.L. Rogers (1996). ‘‘Lifetime Effects of Fundamental Tax Reform’’. In H.J. Aaron and W.G. Gale (eds): *Economic Effects of Fundamental Tax Reform*, Washington, DC: Brookings Institution Press, 321–354.

Giles, C. and J. McCrae (1995). TAXBEN: The IFS Microsimulation Tax and Benefit Model. IFS Working Papers No W95/19, Institute for Fiscal Studies, United Kingdom.

Haan, P. and V. Steiner (2005). Distributional Effects of the German Tax Reform 2000 – a Behavioral Microsimulation Analysis. *Journal of Applied Social Science Studies*, 125, 39–49.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. *American Economic Review*, 71, 662–676.

Heckman, J.J. (2010). Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature*, 48, 356–398.

Heckman, J.J. and S. Urzua (2010). Comparing IV with Structural Models: What Simple IV Can and Cannot Identify. *Journal of Econometrics*, 156, 27–37.

Herriges, J.A. and C.L. Kling (1999). Nonlinear Income Effects in Random Utility Models. *Review of Economics and Statistics*, 81, 62–72.

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

Imbens, G.W. (2010). Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, 48, 399–423.

Iyer, G.S. and P.M.J. Reckers (2012). Decomposition of Progressivity and Inequality Indices: Inferences from the US Federal Income Tax System. *Journal of Accounting and Public Policy*, 31, 258–276.

Jenkins, S.P. (1988). Empirical Measurement of Horizontal Inequity. *Journal of Public Economics*, 37, 305-329.

Jenkins, S.P. (1995). Accounting for Inequality Trends: Decomposition Analyses for the UK. *Economica*, 62, 29–64.

Jorgensen, D.W. and D.T Slesnick (1984). Aggregate Consumer Behavior and the Measurement of Inequality. *Review of Economic Studies*, 51, 369–392.

Juhn, C., K. Murphy and B. Pierce (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy*, 101, 410–442.

Kaplow, L. and S. Shavell (2001). Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle. *Journal of Political Economy*, 109, 281–286.

Kasten, R., F. Sammartino and E. Toder (1994). “Trends in Federal Tax Progressivity”. In J. Slemrod (ed.): *Tax Progressivity and Income Inequality*, Cambridge (MA): Cambridge University Press, 9–50.

Keane, M.P. (2010). Structural vs Atheoretic Approaches to Econometrics. *Journal of Econometrics*, 156, 3–20.

Kiefer, D.W. (1985) Distributional Tax Progressivity Measurement. *National Tax Journal*, 37, 497–513.

King, M.A. (1983). Welfare Analysis of Tax Reforms Using Household Data. *Journal of Public Economics*, 21, 183–214.

Kornstad, T. and T.O. Thoresen (2006). Effects of Family Policy Reforms in Norway: Results from a Joint Labour Supply and Childcare Choice Microsimulation Analysis. *Fiscal Studies*, 27, 339–371.

Kristjánsson, A.S. (2013). Redistributive Effects of Taxation in a Dual Income Tax System. *Finanzarchiv/Public Finance Analysis*, 69, 148–166.

Labeaga J.M., X. Oliver and A. Spadaro (2008). Discrete Choice Models of Labour Supply, Behavioural Microsimulation and the Spanish Tax Reforms. *Journal of Economic Inequality*, 6, 247–273.

Lambert, P.J. (2001). *The Distribution and Redistribution of Income*. Third edition. Manchester and New York: Manchester University Press.

Lambert, P.J., R. Nesbakken and T.O. Thoresen (2012). Is There More Redistribution in Scandinavia than in the U.S.? *Economics Bulletin*, 32(3), 2146-2154.

Lambert, P.J. and X. Ramos (1997). Vertical Redistribution and Horizontal Inequity. *International Tax and Public Finance*,4, 25–37.

Lambert, P.J. and T.O. Thoresen (2009). Base Independence in the Analysis of Tax Policy Effects: With an Application to Norway 1992–2004. *International Tax and Public Finance*, 16, 219–252.

McFadden, D. (1984). “Econometric Analysis of Qualitative Response Models”. In Z. Griliches and M.D. Intriligator (eds.): *Handbook of Econometrics*, Vol. II, New York: Elsevier Science Publishers, 1395–1457.

McFadden, D. (1999). “Computing Willingness-to-Pay in Random Utility Models”. In J. Moore, R. Riezman and J. Weeks (eds.): *Trade, Theory and Econometrics: essays in Honour of John S. Chipman. Routledge Studies in the Modern World*, Vol. 15, London: Routledge, 253–274.

Muellbauer, J. (1974). Household Composition, Engel Curves and Welfare Comparisons Between Households: A Duality Approach. *European Economic Review*, 5, 103–122.

Musgrave, R.A. and T. Thin (1948). Income Tax Progression, 1929-1948. *Journal of Political Economy,* 56, 498-514.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14, 693–709.

Pechman, J.A. and B.A. Okner (1974). *Who Bears the Tax Burden*, Washington, DC: Brookings Institution Press.

Pfähler, W. (1990). Redistributive Effect of Income Taxation: Decomposing Tax Base and Tax Rates Effects. *Bulletin of Economic Rersearch*, 42, 121-129.

Poterba, J. (2007). Income Inequality and Income Taxation, *Journal of Policy Modeling*, 29, 623–633.

Preston, I. and I. Walker (1999). Welfare Measurement in Labour Supply Models with Nonlinear Budget Constraints, *Journal of Population* Economics, 12, 343–361.

Reynolds, M. and E. Smolensky (1977). *Public Expenditures, Taxes, and the Distribution of Income: The United States, 1950, 1961, 1970*. New York: Academic Press.

Rohaly, J., A. Carasso and M.A. Saleem (2005): *The Urban-Brookings Tax Policy Center Microsimulation Model: Documentation and Methodology for Version 0304*, Washington DC: The Urban Institute.

Rothe, C. (2010). Nonparametric Estimation of Distributional Policy Effects. *Journal of Econometrics,* 155, 56–70.

Saez, E., J.B. Slemrod and S.H. Giertz (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: a Critical Review. *Journal of Economic Literature*, 50, 3–50.

Sen, A. (1977): On Weights and Measures: Informational Constraints in Social Welfare Analysis. *Econometrica*, 45, 1539–71.

Shorrocks, A.F. (1982). Inequality Decomposition by Factor Components. *Econometrica*, 50, 193–211.

Shorrocks, A.F. (2013). Decomposition Procedures for Distributional Analysis: a Unified Framework Based on the Shapley Value. *Journal of Economic Inequality*, 11, 99–126.

Slesnick D.T. (1998). Empirical Approaches to the Measurement of Welfare. *Journal of Economic Literature*, 36, 2108–2165.

Sutherland, H. (2001). Final Report EUROMOD: An Integrated European Benefit-Tax Model. EUROMOD Working Paper No. EM9/01.

Thoresen, T.O. (2004). Reduced Tax Progressivity in Norway in the Nineties. The Effect from Tax Changes. *International Tax and Public Finance*, 11, 487–506.

Thoresen, T.O., J. Aasness and Z. Jia (2010). The Short-Term Ratio of Self-Financing of Tax Cuts: an Estimate for Norway's 2006 Tax Reform. *National Tax Journal*, 63, 93–120.

Thoresen, T.O., E.E. Bø, E. Fjærli and E. Halvorsen (2012). A Suggestion for Evaluating the Redistributional Effects of Tax Changes: With an Application to the 2006 Norwegian Tax Reform. *Public Finance Review*, 40, 303–338.

Urban, I. and P.J. Lambert (2008). Redistribution, Horizontal Inequity, and Reranking: How to Measure Them Properly. *Public Finance Review,* 36, 563–587.

Van Soest, A. (1995). Structural Models of Family Labor Supply. A Discrete Choice Approach. *Journal of Human Resources*, 30, 63–88.

Wang, C. and K. Caminada (2011). Disentangling Income Inequality and the Redistributive Effect of Social Transfers and Taxes in 36 LIS Countries. MPRA Paper No. 32861, http://mpra.ub.uni-muenchen.de/32861/

B

Return to:
Statistisk sentralbyrå
NO-2225 Kongsvinger

From:
Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Kongens gate 6, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISSN 0809-733X

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist