Revealing Inequality Aversion from Tax Policy and the Role of Non-Discrimination*

Kristoffer Berg†

December 2023

Abstract

Governments have increasing access to individual information, but they exploit little of it when setting taxes. This paper shows how to reveal inequality aversion from observed tax policy choices of such governments. First, I map governments' priorities into concerns for vertical and horizontal equity. While vertical equity underlies inequality aversion, horizontal equity introduces a restriction against tax discrimination. This restriction affects the measurement of inequality aversion. Second, I apply the model to a hypothetical gender tax using Norwegian tax return data. The main result is that inequality aversion is overestimated when the horizontal equity restriction is ignored.

JEL: D63, H21, I38. Keywords: optimal income taxation; social preferences; tagging

*I have received helpful comments from Rolf Aaberge, Geir Asheim, Giacomo Brusco, Michael Devereux, Naomi Feldman, Emanuel Hansen, Shafik Hebous, Martin Holm, Morten Håvarstein, Etienne Lehmann, Egor Malkov, Andreas Myhre, Frikk Nesje, Andreas Peichl, Paolo G. Piacquadio, Marius Ring, Emmanuel Saez, Kjetil Storesletten, Magnus E. Stubhaug, Thor O. Thoresen, Gaute Torsvik, Andrea Weber, Matthew Weinzierl, and audiences at LAGV, CESifo, ESEWM, IIPF, NTA, Frischsenteret, IRLE, NMBU, Norges Bank, NTNU, OsloMet, Oxford University CBT, Statistics Norway, UC Berkeley, Uppsala University, University of Amsterdam, University of Bergen and University of Oslo. This project has received funding from the European Research Council under the ERC Starting Grant VALURED (Grant agreement No. 804104) and from the Research Council of Norway through Oslo Fiscal Studies.

†Centre for Business Taxation, Saïd Business School, University of Oxford and Department of Economics, University of Oslo. Email: kristoffer.berg@sbs.ox.ac.uk.

1 Introduction

Which equity concerns support actual tax policies? Standard welfare criteria—such as utilitarianism—suggest optimal policies emerge as a balance between *efficiency* (it is an improvement that someone becomes better off) and *inequality aversion* (taking an equal amount from someone better off and giving it to someone worse off is an improvement).

What is less commonly acknowledged is that standard welfare criteria imply that it is optimal to exploit all relevant information about individuals in setting taxes. For example, since females on average earn less than males, a utilitarian policy maker would, all else equal, set lower taxes for females than males earning the same income (an instance of *tagging*). Yet, in actual tax policy, there are much fewer cases of differential taxation across characteristics than utilitarianism would recommend.[1]

In this paper, I develop a theory that rationalizes both the observed levels of redistribution and the equal treatment of different characteristics in actual tax systems. To do so, I build on classic work in taxation (Musgrave 1959), and distinguish between *vertical equity*, the priority on reducing differences across income levels, and *horizontal equity*, the priority on equal treatment of individuals with similar incomes.

This theory has important implications. My first result is that by accounting for horizontal equity, implied social preferences are less averse to inequality. In other words, the government has access to lower cost redistribution than we assume when we neglect the possibility to tag. Since the true cost of redistribution that is available to the government is lower, then for a given a tax policy the implied benefit of redistribution (the inequality aversion) must be lower as well. Inequality aversion is a key parameter in many optimal policy contexts, such as in minimum wage setting and environmental policy. If preferences that are less averse to inequality more accurately reflects social preferences, this could warrant a lower concern for inequality reduction when evaluating other policies too.

My second result is that recognizing the government's choice to respect horizontal equity can significantly affect inferred inequality aversion. In an application to gender-based taxation, I estimate the relevant parameters using Norwegian register data. I find that

[1] See Mankiw and Weinzierl (2010) on the relationship between utilitarianism and tagging.

the level of inequality aversion is overestimated by 8.3 % when one ignores that the government has access to, but chooses not to use, lower-cost redistribution by tagging. This means that an 8.3 % lower average willingness to reduce inequality between the rich and poor can rationalize tax policy when horizontal equity is accounted for. In an application to environmental policy this translates into a lower social discount rate, such that earlier action against climate change is optimal.

The form of discrimination considered here is to condition taxes on characteristics that are immutable to tax policy, such as gender, height and age.[2] There is a longstanding literature on optimal tagging, beginning with Akerlof (1978).[3] However, actual tax systems display a limited use of tagging based on immutable characteristics.[4] At the same time, it is a well-established empirical fact (see also results for Norway in this paper) that income distributions and tax responses differ across characteristics, providing vertical equity and efficiency rationales for conditioning taxes on these characteristics. Since governments appear reluctant to exploit information on most characteristics in actual tax policy, one explanation is that society holds a counteracting *equity* rationale for not exploiting information on certain characteristics. Hence, I introduce a concern for horizontal equity.[5]

This paper offers a simple interpretation of horizontal equity to rationalize actual tax policy. Horizontal equity here means that tax policy is not allowed to exploit information on certain characteristics (non-discrimination). This definition differs from traditional interpretations of horizontal equity as non-rearrangement of the relative position between the pre-tax and the post-tax distribution (King 1983, Jenkins 1988 and Auerbach and Has-

[2]A taxpayer's gender is not immutable in general and may not be immutable to tax policy. Assigned sex at birth is an alternative immutable characteristic that also would imply a horizontal equity concern.

[3]Recent contributions include Cremer, Gahvari, and Lozachmeur (2010), Alesina, Ichino, and Karabarbounis (2011) on gender tags, Mankiw and Weinzierl (2010) on the optimal taxation of height, and Weinzierl (2011), Bastani, Blomquist, and Micheletto (2013) and Heathcote, Storesletten, and Violante (2020) on age-dependent taxation.

[4]Examples clearly exist. In the US, EITC payments are higher for single mothers, some countries have levied "bachelor's taxes" on unmarried males, and certain countries set lower income taxes for young workers. Still, few existing tags in tax systems are based solely on immutable characteristics, and the standard criterion suggests much wider use than what is currently observed.

[5]See Sausgruber and Tyran (2014) for experimental evidence on discriminatory taxes being unpopular.

sett 2002), and is inspired by the Atkinson (1980) view that horizontal equity is fundamentally about protecting against discrimination.[6]

To measure the relative priority on vertical and horizontal equity, I build on the *inverse optimal tax problem*. Following Bourguignon and Spadaro (2012), the framework exploits actual tax systems to reveal the *marginal welfare weights*, the priority on increasing consumption at an income level, that make the current tax system optimal. Contributions to this literature include Bargain et al. (2014) for the US and certain European countries, Spadaro, Piccoli, and Mangiavacchi (2015) for major European countries, Lockwood and Weinzierl (2016) for the US over time, Bastani and Lundberg (2017) for Sweden, Jacobs, Jongen, and Zoutman (2017) for political parties in the Netherlands, and Hendren (2020), who relates the inverse optimum approach to cost-benefit criteria.[7] A typical implicit assumption in these contributions is that marginal welfare weights are informative about society's level of inequality aversion. Making less specific assumptions about the welfare criterion, Saez and Stantcheva (2016) show that the social value of one more dollar of consumption to an income group can be interpreted as a *generalized* social marginal welfare weight on that group. Then, these weights can reflect a multitude of equity principles, including horizontal equity.[8] However, the link between horizontal equity and the inverse optimal tax problem has not been studied yet, and this is the first paper to do so.

I develop the inverse optimum framework to measure the separate contributions of vertical and horizontal equity concerns in supporting actual tax policy. In order to decompose the marginal welfare weights that support the actual tax system as an optimum, one requires estimates of marginal welfare weights both when tagging is used and not used.[9]

[6]The two definitions are related, as characteristics-based taxation implies rearrangement of relative positions. Furthermore, my main results do not depend on which particular fairness view restricts the use of tagging, and are thereby robust to different foundations for non-discrimination.

[7]For earlier contributions with similar approaches, see Christiansen and Jansen (1978) with an application to indirect taxation in Norway and the test for Pareto optimality in Ahmad and Stern (1984).

[8]The discussion of horizontal equity in Saez and Stantcheva (2016) is concerned with establishing that generalized social marginal welfare weights can reflect a concern for horizontal equity. They do not show how to distinguish different priorities from observed tax policy or how to quantify their importance.

[9]Instead of assuming that tax policy is set optimally, one can also view the approach here as deriving the plausible equity preferences to have in order to consider current tax policy as optimal.

Since the actual tax system respects horizontal equity, the standard inverse optimal tax approach reveals the marginal welfare weights in the case when tagging is not used. The sufficient statistics required for this exercise are standard: elasticities of taxable income, income distribution parameters, and actual tax rates.

To estimate marginal welfare weights in the counterfactual tax system, without the horizontal equity restriction, I develop an optimal tax algorithm. The idea to is to exploit the marginal welfare weights that support actual tax policy to find the optimal tax system when tagging is used. Since tagging is used in a counterfactual tax system, it requires a model of government behavior outside of the actual tax system. When exploiting characteristics in setting taxes, the government sets separate tax schedules for each characteristic and transfers between them. The marginal welfare weights that support this counterfactual tax system as an optimum must rationalize both the characteristic-specific income taxes and the between-characteristic transfers. My algorithm accounts for the fact that tagging enables less distortive redistribution, which implies a different marginal value of inequality reduction. Together, the estimates of marginal welfare weights when tagging is used and when it is not used permit estimation of the size of the bias to inequality aversion when horizontal equity is ignored.

In an empirical application, I estimate the effect of horizontal equity across gender in Norway when the government has access to information about gender-specific income distributions and taxable income elasticities. First, since females on average earn less than males, redistribution between high and low income levels can be achieved at a lower cost by imposing lump sum transfers from males to females. Second, since females adapt their pre-tax incomes more to tax changes than males do, distortions from income taxation can be reduced by lowering marginal income taxes on high-earning females while increasing them on high-earning males. In conclusion, as gender is observable to the government, correlates with income and elasticities of taxable income and is a controversial characteristic to base taxes on, it serves as a policy relevant example of the consequences of non-discrimination in optimal taxation.

The paper contributes to two main strands of literature. First, it contributes to optimal taxation in the Mirrlees (1971) tradition and in particular the growing literature on normative principles in taxation (Mankiw and Weinzierl 2010, Weinzierl 2014, Saez and Stantcheva 2016, Lockwood and Weinzierl 2016, Fleurbaey and Maniquet 2018, Weinzierl 2018, and Berg and Piacquadio 2020). This is done by introducing horizontal equity as a government-chosen constraint against tagging, solving the optimal tax problem with tagging in the local optimum framework (Saez 2001), and deriving the implications for revealed preferences using the inverse optimum approach. Second, it contributes to the broad literature on revealed social preferences, which has been achieved through surveys (Kuziemko et al. 2015 and Stantcheva 2020), experiments (Cappelen et al. 2007 and Bruhin, Fehr, and Schunk 2019), and, as in this paper, from observed policy (Bourguignon and Spadaro 2012 and Groom and Maddison 2019). The contribution to this literature is to derive social preferences that reflect equity considerations from a policy with multiple aspects (progressivity and limited use of information).

The paper proceeds as follows. Section 2 presents the model for the equity principles, before deriving the decomposition of marginal welfare weights into vertical and horizontal equity concerns. Section 3 introduces the continuous optimal taxation model and the inverse optimum tax problem. Section 4 presents the empirical application, where I provide estimates on heterogeneity in tax responses and apply the findings to the tax model. Section 5 concludes.

2 Model of vertical and horizontal equity

2.1 Description

Individuals

There is a continuum of individuals $i \in I$, with mass normalized to 1. Each individual is characterized by a wage rate, $w_i \geq 0$, and a utility function, $u_i(c_i, l_i)$, which is weakly concave in consumption, $c_i > 0$, and strictly convex in labor supply, $l_i \geq 0$, with $\partial u_i(c_i, l_i)/\partial c_i > 0$ and $\partial u_i(c_i, l_i)/\partial l_i < 0$. Individuals maximize utility subject to the budget constraint $z_i - T_i(.) \geq c_i$, where $z_i = w_i l_i \in (0, \infty)$ is pre-tax income and is distributed according to $h(z)$, and $T_i(.)$ is the tax payment of individual i. What tax payments can be a function of is explained below.

Each individual is also characterized by a *tag* that represent different types, k_i, and each type is a *characteristic*. A tag is informative if there are income distribution differences across characteristics. Denote by p_k the proportion of each characteristic in the population, $\sum_k p_k = 1$. Within each characteristic, income is distributed according to $h_k(z)$. Denote by $\hat{x}(z)$ the average of any variable $x_k(z)$ across characteristics at income level z, $\hat{x}(z) = \sum_k \left(h_k(z)/\sum_k p_k h_k(z)\right) x_k(z)$. Denote the average of variables $x(z)$ and $x_k(z)$ over the total distribution, $h(z)$, and the characteristic-specific distributions, $h_k(z)$, by $E\left(x(z)\right) = \int_0^\infty x(z)h(z)dz$ and $E_k\left(x_k(z_k)\right) = \int_0^\infty x_k(z_k)h_k(z_k)dz_k$, respectively.

Government

The government sets taxes T_i as an (for now unspecified) function of information about individuals in order to raise revenue $\sum_i T_i = R$. In the absence of a horizontal equity concern (it is introduced below), it maximizes welfare, which is the integral of individual-specific concave transformations $G_i(u_i(c_i, l_i))$ of individuals' utilities,

$$\max W = \int_i G_i\left(u_i(c_i, l_i)\right)di. \tag{1}$$

In the following, the government can set taxes based on w, z and/or k, depending on the information requirement. As long as $z_k'(w_k) \geq 0$, which I assume from now on, the government's objective can be rewritten in terms of the characteristic-specific income distributions, as any heterogeneity within a characteristic-specific income level does not have any impact on welfare judgements and policy choice. Subsequently the subscript i is dropped for readability. Since the social welfare function is additively separable in individual welfare, social welfare can be written as the average welfare across characteristics,

$$W = \sum_k p_k E_k\left[G_k\left(u_k(c_k(z_k), l_k(z_k))\right)\right]. \tag{2}$$

Characteristic-specific marginal welfare weights are defined by

$$g_k(z_k) \equiv \frac{\partial W}{\partial c_k(z_k)} \in (0, \infty), \tag{3}$$

and are the government's valuation of increased consumption, at each income level for each characteristic. These are normalized such that $\sum_k p_k E_k\left[g_k(z_k)\right] = 1$. A government that exploits all information takes account of the different characteristic-specific redistribution to set taxes, the average marginal welfare weights across unconditional income for this government is given by $\hat{g}(z)$. When a government does not account for differences across characteristics, it sets taxes across the unconditional income distribution and is associated with marginal welfare weights $g(z)$.

The next assumption is that marginal welfare weights are equal across characteristics for a given utility level, such that the government respects a form of anonymity (at the same utility, individuals of different characteristics are given the same weight). This assumes that the only reason the government may use information on characteristics is to redistribute more efficiently, and not because it favours one of the groups. If marginal welfare weights are (weakly) falling in utility, then the government is characterized as *redistributive*.[10]

Sorting means that the ordering of incomes (before tax) is the same as the order of consumption levels (after tax) over the income distribution, which emerges if there is a monotonically increasing relation between c and z, $c_k'(z_k) \geq 0$. In the following, I assume sorting within the income distributions exploited by the government to set taxes, such that if the government exploits the characteristic-specific income distributions, sorting is assumed within each of these distributions. In addition, I assume that within the income distributions exploited by the government to set taxes, individuals with higher income have (weakly) higher utility.

Next, I discuss the relationship between marginal welfare weights and certain cases of information availability and preferences of the government. Instead of explicitly solving the optimal tax problem (which is presented in Section 3), the cases rely on features of the first- and second-best optimal policy problems.

[10]In an optimum for a government with a standard utilitarian social welfare function: $W = E\left(G(u(c, z/w))\right)$, then $g(u(c,l)) = G'\left(u(c,l)\right)u_c(c,l)$, such that concavity of G implies that marginal welfare weights are (weakly) falling in utility (as $u_c(c,l)$ is assumed to be weakly concave in c).

2.2 Cases

Both w and l are unobserable to the government, while it observes z and k. The government chooses whether to exploit information on characteristics or not.

Case 1: No tagging

Information about k is not exploited, since the government chooses not to. This is the case where the government does not act simply according to Equation 2, but also constrains itself to no tagging. Taxes must therefore be set only according to individuals' pre-tax income, z. If the government is redistributive, the information problem introduces a *cost of redistribution*. The cost emerges from individuals' responses to income taxes. A redistributive government that lacks information about w and chooses not to exploit information about k, places (weakly) higher marginal welfare weights on lower incomes, $g'(z) \leq 0$. This follows directly from the definition of a redistributive government and that utility is sorted over the income distribution.

Case 2: Tagging

Taxes are now allowed to be characteristic-specific, $T_k(z_k)$, which entails that taxes and consumption may differ for individuals at the same income level.[11]

Proposition 1. *For a given economy, a redistributive government that exploits an informative tag chooses a different optimal allocation and given sorting, it is associated with a flatter marginal welfare weight schedule over income on average compared to the government that does not exploit tagging,* $E(|\hat{g}'(z)|) < E(|g'(z)|)$.

Proof. Initially the government does not impose tagging, such that marginal welfare weights are given by $g(z)$. Assume without loss of generality that there are two groups, a and b such that $u_a(c_a, c_a) < u_b(c_b, l_b)$. The government chooses to impose tagging by increasing consumption for a by m_a, such that $\Delta u_a(c_a, l_a) > 0$, and reducing consumption for b by m_b, such that, $\Delta u_b(c_b, l_b) < 0$. All other individuals are unaffected. The government imposes the transfer whenever $g\left(u_a(c_a, l_a) + \Delta u_a(c_a, l_a)\right) - g\left(u_b(c_b, l_b) + \Delta u_b(c_b, l_b)\right) >$

[11] Sorting is now assumed within each characteristic, $c_k'(z) \geq 0$ and $u_k'(z) \geq 0$.

$g\left(u_a(c_a,l_a)\right)-g\left(u_b(c_b,l_b)\right)$. Since the government is redistributive, marginal welfare weights are lower for higher utility levels. This means that when tagging is implemented, inequality in utility has gone down and marginal redistribution becomes less valuable. Hence, the difference in marginal welfare weights becomes smaller between group a and b. Due to the sorting property, group a has lower income than group b, and the marginal welfare weight schedule becomes flatter over income $E(|g'(z)|) - E(|\hat{g}'(z)|) > 0$. □

However, it cannot be guaranteed that the marginal welfare weight schedule shifts in a specific way everywhere when the government introduces tagging when there is within-group income variation, only that it becomes flatter on average. For example, consider the case of females and males. If all the high-income earners are male while the middle-income earners consist of mostly females, the government may use gender-specific taxation to increase taxes on high-income earners while decreasing them on middle-income earners. Then, the marginal welfare weight on the high-income earners increases, since they receive lower consumption, while the weight on middle-income earners decreases, which increases the steepness of the marginal welfare weight schedule from low-income to middle-income earners. Still, since the redistributive government chose to introduce the lump sum transfer between the groups, it must increase welfare and reduce inequality, such that the average steepness of its marginal welfare weight schedule goes down.

Next, I use these cases to derive the relation between vertical and horizontal equity.

2.3 Equity principles

Vertical equity

Vertical equity is society's priority on reducing inequality across consumption levels. The vertical equity principle may be provided with further foundation from various theories of justice, such as prioritarianism (Parfit 1991) and egalitarianism (Temkin 1993). Here, one should think of it as the resulting priority on reducing inequality, irrespective of its moral foundation.

Consider again the government that maximizes welfare and exploits all information and label their welfare by W^{VE},

$$W^{VE} = \sum_k p_k E_k \left[G_k \left(u_k(c_k(z_k), l_k(z_k)) \right) \right].$$

where $g_k(z_k) = \partial G_k(z_k)/\partial c_k$. The extent of the redistributive motive for this government can be measured by how much priority it places on increasing consumption for high versus low income levels. Define the weight from vertical equity at income level z averaged over characteristics by $VE(z) = \hat{g}(z) - 1$. If $VE(z) > 0$, then income level z is given extra priority due to the vertical equity concern. Vertical equity represents the government's priority in Case 2. It follows that $VE'(z) = \hat{g}'(z)$.

As long as $VE'(z) < 0$, the vertical equity concern means that the government accepts a higher cost of giving an extra dollar to low income individuals relative to high income individuals on the margin. For example, $VE(z) = 1.5$ means that the vertical equity concern imposes that the government accepts a 50 percent larger cost on increased consumption at income level z compared to distributing the transfer equally to everyone. In other words, if there are 15 individuals, 1 dollar to each is as desirable as 10 dollars to the individual with income z. The government is more redistributive the higher is the average marginal cost of vertical equity for a given level of redistribution, as it is willing to pay a higher price in terms of total consumption to redistribute from the rich to the poor.

Importantly, a (weakly) decreasing vertical equity schedule cannot alone characterize the government in Case 1. As the government in Case 1 could have reduced inequality at the same efficiency cost by exploiting more information, it cannot be represented by a standard inequality averse government. I turn to horizontal equity to characterize this government.

Horizontal equity

Horizontal equity reflects an aversion to treating individuals with the same circumstance unequally. Inspired by Atkinson (1980), I account for horizontal equity by introducing a constraint that prohibits tagging based on certain characteristics. Such a constraint violates Pareto efficiency (Kaplow 1989). Alternative representations of horizontal equity are possible, see Feldstein (1976) for a tax-reform based measure, Auerbach and Hassett (2002) for a horizontal inequality index that respects Pareto efficiency, and Saez and Stantcheva (2016) for a representation based on marginal welfare weights that only allows Pareto-improving tagging. Non-utilitarian welfare criteria, such as those based on equal sacrifice (Weinzierl 2014 and Berg and Piacquadio 2020), may be less redistributive and thereby imply that tagging would produce less efficiency gains and smaller between-characteristic transfers. However, less inequality aversion is in itself insufficient to rationalize the non-use of a tag (as long as the criterion is Paretian, the tax is distortive and tagging can reduce inequality), as tagging relaxes self-selection constraints also for such criteria.

Introducing horizontal equity as a constraint is less restrictive when revealing government preferences from observed policy. If there are no Pareto improvements to be made by violating the constraint, then one cannot tell whether the government would be willing to violate Pareto efficiency or not. If the government does violate Pareto efficiency for the sake of horizontal equity, then a constraint rationalizes a feature of actual tax policy that other representations could not.

The constraint introduced is

$$T_k(z) = T(z)\ \forall k, \tag{4}$$

at each income level z. This imposes that all individuals at the same income level face the same tax. If it binds, the horizontal equity constraint makes reaching the government's other objectives more costly. Define $HE(z)$ as the Lagrange multiplier associated with the constraint at each income level in the government's maximization problem, and it measures the shadow price of horizontal equity at each income level. The next proposition states the relationship between social preferences of redistributive governments that have information on k and may exploit or not exploit the information depending on whether they have a concern for horizontal equity.

Proposition 2. *The shadow cost of horizontal equity, $HE(z)$, represents the difference in marginal welfare weights between not exploiting information on k (Case 1) and exploiting the information (Case 2), $HE = g(z) - \hat{g}(z)$. The shape of the cost of horizontal equity, $HE'(z)$, is negative on average for a redistributive government that does not exploit information on k.*

Proof. With horizontal equity as a constraint, the government maximizes welfare subject to $T_k(z) = T(z)$ at each z. These constraints are added to construct a new social welfare function labeled W^{HE},

$$W^{HE} = \sum_k p_k \left(E_k \left[G_k \left(u_k(c_k(z_k), l_k(z_k)) \right) \right] - E_k \left[HE(z) \left(T_k(z) - T(z) \right) \right] \right),$$

where E_k is the integral of the constraint for each characteristic. Remember that $T_k(z_k) = z_k - c_k(z_k)$. Consider how this government values a marginal increase in $c_k(z_k)$ at income level z_k,

$$\frac{\partial WE^{HE}}{\partial c_k(z_k)} = p_k \left(g_k(z_k) + HE(z) \right).$$

By averaging this over characteristics one considers the welfare increase from increased consumption at income level z across all characteristics,

$$\sum_k \frac{\partial WE^{HE}}{\partial c_k(z_k)} = \hat{g}(z) + HE(z).$$

The optimum according to W^{HE} is the same as when information is not available (Case 1), as the problem is the same (maximize integrals of welfare for each individual using only income taxation). At such an optimum, marginal welfare weights over income are equivalent to when information on k is not exploited, $g(z)$, such that

$$\hat{g}(z) + HE(z) = g(z).$$

Hence, $HE(z)$ is the change in the cost of redistribution at income z imposed by the horizontal equity concern.

To determine the shape of the horizontal equity concern, consider $HE'(z) = g'(z) - \hat{g}'(z)$. By Proposition 1, $|\hat{g}'(z)| < |g'(z)|$ on average, and since $g'(z) < 0$ on average, $HE'(z) < 0$ on average. $\square$

When the horizontal equity concern limits a redistributive government from exploiting information, the concern imposes costs in terms of achieved vertical equity.

Proposition 3. *For a government concerned with efficiency, vertical equity and horizontal equity, marginal welfare weights at each income level can be decomposed into*

$$g(z) = VE(z) + HE(z) + 1.$$

If horizontal equity is not accounted for and the information that remains unused is informative of w, total willingness to pay for vertical equity is overestimated.

Proof. The decomposition follows from $VE(z) = \hat{g}(z) - 1$ and $g(z) = \hat{g}(z) + HE(z)$. Together with $HE'(z) < 0$ on average from Proposition 2, this implies that if horizontal equity is not accounted for, the average $-VE'(z)$, a measure of willingness to pay for inequality reduction, will be overestimated. □

Inequality aversion

While vertical equity establishes the local priority on inequality reduction, *inequality aversion* is the measure for the total priority on reducing inequality. Inequality aversion can be defined in multiple ways. One possible measure is the average value of the steepness of the marginal welfare weights over consumption in the case when tagging is exploited,

$$IA = E\left(|\hat{g}'(z)|\right). \tag{5}$$

This measure directly aggregates the vertical equity measures. The simple inequality measure when tagging is not used and horizontal equity is ignored, $E\left(|g'(z)|\right)$, is biased. One bias measure, b, to IA from not accounting for horizontal equity is

$$b = E\left(|g'(z)|\right) - E\left(|\hat{g}'(z)|\right).$$

Ignoring horizontal equity implies a positive bias, b, in the measurement of inequality aversion, IA. By Proposition 2, $-E\left(|\hat{g}'(z)\right)| < E\left(|g'(z)|\right)$, such that $b > 0$. Hence, the level of inequality aversion is overestimated when horizontal equity is ignored.

It follows that marginal welfare weights derived from actual tax policy reflect both vertical and horizontal equity. Typically, $g(z)$ is interpreted both as the cost of redistribution (fiscal externality), as in Hendren (2020), and as the willingness to pay for reduced inequality, as in Bourguignon and Spadaro (2012). Proposition 3 establishes that horizontal equity drives a wedge between the cost measure and the willingness to pay interpretation. This is because the horizontal equity constraint increases the cost of redistribution, which is reflected in marginal welfare weights. Hence, part of the value of marginal welfare weights reflects horizontal rather than vertical equity.

To illustrate the bias to inequality aversion from ignoring horizontal equity, assume quasi-linear utility, $u_i = c_i - v(l_i)$ and that the social welfare function exhibits constant relative inequality aversion in equivalent income, $e(c,l)$, such that $u(e(c,l),\bar{l}) = u(c,l)$, where l is some fixed level of labor supply. Then, $SWF = E\left(\left(e(c,l)\right)^{1-\gamma} / \left(1-\gamma\right)\right)$, where γ is the inequality aversion parameter. The inequality aversion parameter is given by

$$\gamma = -E\left(log(g(z))/log(e(z))\right). \tag{6}$$

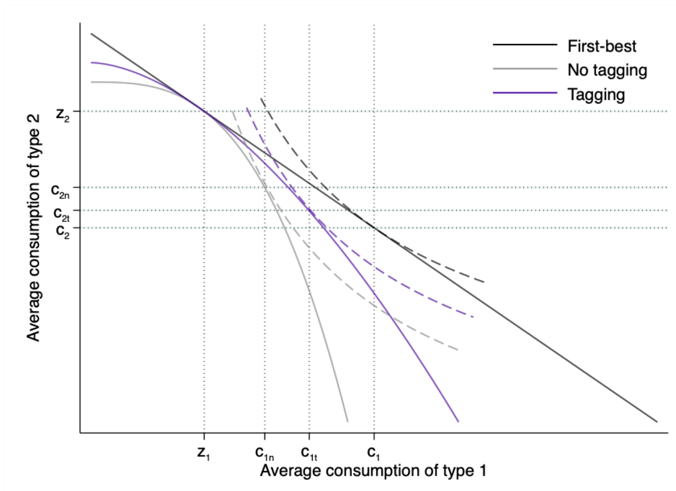
The inequality aversion measure depends on the optimal allocation. When tagging is not used, inequality aversion is measured in a different optimum than when tagging is not used. This means that each optimum also reflects the priority on horizontal equity. Interestingly, for a fixed level of inequality aversion, a government that respects horizontal equity reduces inequality by less than a government that does not respect horizontal equity, as it faces higher costs of inequality reduction. The government that respects horizontal equity is thereby trading off both some inequality reduction and total income in order to respect the horizontal equity constraint.

2.4 Two-wage type illustration

To illustrate the relationship between inequality aversion and horizontal equity Figure 1 presents a two-wage type model. There are two wage types and two characteristics. An inequality averse social welfare function implies social indifference curves that rank different allocations, but does not account for a horizontal equity concern. The cases presented above are associated with different consumption possibility frontiers, reflecting different costs of redistribution away from the "laissez-faire" ($c_1 = z_1, c_2 = z_2$). If the government places no value on vertical equity, it chooses the laissez-faire (z_1, z_2) independently of its information set. In the first-best, a government that only values vertical equity may choose consumption levels (c_1, c_2). When the problem is second-best and the government exploits tagging (Case 2) it chooses consumption levels such that the average consumption across characteristics are (c_{1t}, c_{2t}), while when the government values both vertical and horizontal equity (Case 1) it chooses allocation (c_{1n}, c_{2n}).

Vertical equity induces the move from the laissez-faire to allocation the tagging allocation, while horizontal equity induces the move from tagging to no tagging. The average consumption difference across types and the steepness of social indifference curve is lower at the allocation when tagging is used. Hence, the presence of the horizontal equity constraint increases redistributive costs, which lowers redistribution and increases the cost the government incurs to reduce inequality. In conclusion, part of the cost the government incurs to reduce inequality when it does not tag is due to its priority of horizontal equity.

Figure 1: The effects of vertical and horizontal equity concerns



3 Optimal taxation with and without tagging

3.1 Optimal tax model

This section provides the theory to quantify the importance of horizontal equity for inequality aversion. To do this, I require estimates of marginal welfare weights in the cases when tagging is used and not used, $\hat{g}(z)$ and $g(z)$, respectively. I adopt the tax perturbation approach to optimal taxation (Saez 2001), and extend it to a setting with tagging. The model is the same as in Section 3, but I further specify the optimal taxation problem here.

For simplicity, assume everyone works (excluding extensive margin responses), quasilinear utility (no income effects) and no exogenous revenue requirement, $R = 0$. The behavioral response to taxes may differ across characterstics, but for simplicity I assume that it is constant within each characteristic, $\varepsilon_k(z) = \varepsilon_k$ for all k. The government have preferences described by Equation (2) and faces the budget requirement

$$R = \sum_k p_k E_k \left(T_k(z)\right) = 0, \tag{7}$$

and the structure of the tax system is

$$T_k(z) = t_k(z) + R_k, \tag{8}$$

where $T_k(z)$ is the nonlinear tax for each characteristic, separated into lump sum transfers R_k and income-dependent taxes $t_k(z)$. It appears that the government has $2k$ instruments, $t_k(z)$ and R_k for each k, but these are related through $\sum_k p_k E_k \left(t_k(z)\right) = \sum_k p_k R_k$, such that the government has $2k-1$ independent instruments.

As in Mankiw and Weinzierl (2010), the problem can be separated, which means that one can solve for the optimal within-characteristic tax rates for a given transfer and then solve for the optimal between-characteristic transfer. Consider a small perturbation of one characteristic's tax schedule, keeping the other schedule (and the transfer) constant. The perturbation is an increase in the tax rate τ_k by $d\tau_k$ at the income level z for the characteristic k, which has the revenue effect

$$dR_k = d\tau_k dz \left(1 - H_k(z) - h_k(z)\varepsilon_k \frac{T_k'(z)}{1 - T_k'(z)}\right), \tag{9}$$

where dR is the change in revenue. It depends on how many individuals pay the new tax, $1 - H_k(z)$, and how individuals respond to the tax, $h_k(z)\varepsilon_k T_k'(z) / \left(1 - T_k'(z)\right)$. This tax change has a welfare effect which is a combination of the welfare gain for everyone from increased revenue and the welfare loss of lower consumption for those with income above z. In the (local) optimum, the welfare change must be zero

$$dW_k = dR_k \sum_k p_k E_k \left(g_k(z)\right) - d\tau_k dz \int_{z>z_i}^{\infty} g_k(z) h_k(z) dz = 0. \tag{10}$$

Combining Equation 9 and 10 (Saez (2001) without income effects), the within-characteristic optimal tax rate is

$$T_k'(z) = \frac{1 - \bar{G}_k(z)}{1 - \bar{G}_k(z) + \alpha_k(z)\varepsilon_k} \tag{11}$$

where $\alpha_k(z) = z h_k(z) / \left(1 - H_k(z)\right)$ is the characteristic-specific local hazard rate and $\bar{G}_k(z) = \int_{z \geq z_i}^{\bar{z}} g_k(z) h_k(z) dz / \left(1 - H_k(z)\right)$ is the characteristic-specific average marginal welfare weight above income level z.

Following the inverse optimum approach (Bourguignon and Spadaro 2012) one can infer marginal welfare weights at each income level, $g(z)$, from the actual tax schedule. *The inverse problem* is to find the marginal welfare weights $g_k(z)$ for which the *current* tax system is a solution to the optimal tax problem. This is achieved by solving Equation 11 for $g_k(z)$. The marginal welfare weights from the inverse optimal problem are given by

$$g_k(z) = -\frac{1}{h_k(z)}\frac{d}{dz}\left[(1 - H_k(z))\left(1 - \frac{T_k'(z)}{1 - T_k'(z)}\rho_k(z)\varepsilon_k\right)\right]. \tag{12}$$

Assuming (for simplicity) that $T(z)$ can be approximated by a piece-wise linear tax system (Bastani and Lundberg 2017), the marginal welfare weights from the inverse optimal problem are given by

$$g_k(z) = 1 - \frac{T_k'(z)}{1 - T_k'(z)}\rho_k(z)\varepsilon_k, \tag{13}$$

where $\rho_k(z) = -\left(1 + zh_k'(z)/h_k(z)\right)$ is the characteristic-specific *elasticity of the income distribution* (Hendren 2020).[12] It measures how the characteristic-specific income distribution changes with income.

3.2 Marginal welfare weights when tagging is not used

Now, the government also respects horizontal equity. Hence there is no tagging, $T_k(z) = T(z)$, and inverse optimum marginal welfare weights are given by

$$g(z) = 1 - \frac{T'(z)}{1 - T'(z)}\rho(z)\varepsilon(z), \tag{14}$$

where $g(z)$, $T'(z)$, $\rho(z)$ and $\varepsilon(z)$ are defined over the joint income distribution. Differences in the composition of characteristics across the distribution may create variation in the behavioral response $\varepsilon(z)$ over the joint income distribution (Jacquet and Lehmann 2020). For example, if females and males respond differently to tax changes, the varying composition of females and males over the income distribution implies heterogeneous responses over the joint income distribution.

[12] Equation (13) assumes that despite the piece-wise nature of the system, there is no bunching at kinks in the tax schedule.

3.3 Marginal welfare weights when tagging is used

I provide a new method to reveal marginal welfare weights for the counterfactual tax system. The point is that marginal welfare weights reflect the allocation in question. If a specific relation between the allocation and weights can be inferred from the shape of inverse optimum marginal welfare weights for actual tax policy, one can arrive at a new set of weights for the allocation when tagging is used.

A government that exploits tagging can set lump sum transfers between characteristics. These transfers must be accounted for to obtain an estimate of $\hat{g}(z)$. The idea is that we can learn about the counterfactual tax system when tagging is used from the inferred priorities of the actual tax system. Then, the difference between tax systems with and without tagging determine the contribution of vertical and horizontal equity in supporting the actual tax schedule. While the standard inverse optimum approach relies on local marginal welfare weights, the trick here is to exploit the broader shape of the marginal welfare weight schedule. However, my method requires a specific utility function.

Proposition 4. *Ceteris paribus (fixed allocations and marginal welfare weight schedules for income levels that do not receive transfers), a redistributive government's new marginal welfare weight schedule with a transfer m to income level z can be obtained from the original marginal welfare weight schedule by the relation*[13]

$$\tilde{g}(z) = g(z_y),$$

where z_y is the income of the average individual with the same pre-transfer utility as the average individual at income z post-transfer.

Proof. The income level z_y is such that $u_y(c_y, l_y) = u_z(c_z + m, l_z)$. Without loss of generality, assume three individuals, i, j and n, with $c_n < c_i < c_j$ and $g_n(u_n(c_n, l_n)) > g_i(u_i(c_i, l_i)) > g_j(u_j(c_j, l_j))$. Initially there is no difference in the relation between consumption and income across individuals. Now, n receives a transfer m, such that n obtains the same utility as i. The after-transfer marginal welfare weight is denoted by $\tilde{g}(z)$. By separability, the

[13] It does not account for that marginal welfare weights must rationalize both within-characteristic tax rates and between-characteristic transfers, and that tax changes induce behavioral responses (a first-order approach). The algorithm below accounts for these factors.

transfer leaves the relative marginal welfare weight of i and j unchanged. □

The condition relates the current marginal welfare weights over income to new marginal welfare weights with transfers. It exploits that individuals are weighted equally given their utility, such that the weight attached to an individual who receives a transfer is the same as an individual who receives the same consumption by earning higher income. The relation relies on the local stability of marginal welfare weights, which will not hold for non-marginal policy changes, such as the introduction of tagging. The algorithm I present next addresses this issue.

Between-characteristics transfers

Between-characteristics transfers are lump sum, such that they are the same for all income levels for each characteristic. The characteristic-specific marginal income tax, $T'_k(z)$, affects within-characteristic income distributions through behavioral responses. As I have assumed no income effects, transfers do not directly affect the pre-tax income distribution, but they still affect the marginal welfare weights over the income distribution by changing consumption levels across characteristics. To measure the effect of tagging on the marginal welfare weight schedule, exploiting current marginal welfare weights, assume that there are no transfers that differ across characteristics prior to tagging.[14]

Now, the optimal between-characteristic transfer, m_k, is found when a change in the transfer keeps welfare unchanged, where dm is defined as the transfer from characteristic k to characteristic $\not{k}$

$$dW = dmE_k\left(g(c_k(z))\right) - dmE_{\not{k}}\left(g(c_{\not{k}}(z))\right) = 0 \; \forall k. \tag{15}$$

This implies setting transfers such that the average marginal welfare weight on individuals of each characteristic is equal, because if not, the government could increase total (weighted) welfare by changing transfers such that $E_k\left(g(c_k(z))\right) = \bar{g}$ for all k. An updating of marginal welfare weights is necessary to satisfy the requirement that the transfer

[14] Any transfer that does not affect pre-tax income and is equal across characteristics in the actual tax system will have no effect on the relation between weights.

from tagging should equalize average marginal welfare weights, since if the transfer did not affect marginal welfare weights, the condition could never be satisfied.

Since transfers only increase or reduce individuals' consumption, there is no direct effect on the income distributions, $h_k(z)$. The initial estimate is obtained by Proposition 4, $g_k(z) = g\left(z + c^{-1}(m_k)\right)$. Depending on the transfer, average consumption among individuals of one characteristic increases while individuals of the other decreases. Marginal welfare weights are still equal across characteristics given the same consumption level, while they now differ for the same income level. The algorithm that solves the optimal tagging problem is:

1. Transfers m_k are set by
$$E_k\left(g\left(c_k(z)\right)\right) = \bar{g}\ \forall k,$$
which depends on $h_k(z)$. This determines c_k, which implies a new $g_k(z)$.

2. Tax rates $T_k'(z)$ are set by
$$T_k'(z) = \frac{1 - \bar{G}_k(z)}{1 - \bar{G}_k(z) + \alpha_k(z)\varepsilon_k}\ \forall k,$$
which depends on marginal welfare weights weights $g_k(z)$. A tax change $dT_k'(z)$ induces a behavioral response $dz_k(z)$ which implies a new $h_k(z)$.

3. Repeat step 1 and 2 by replacing weights and income distributions until marginal welfare weights rationalize both m_k and $T_k'(z)$.

4. Calculate the resulting joint marginal welfare weights $\hat{g}(z)$ as averages of the characteristic-specific marginal welfare weights.

The process can be seen as follows:

$$g(z) \rightarrow m_k \rightarrow g_k(z) \rightarrow T_k'(z) \rightarrow h_k(z) \rightarrow m_k \rightarrow \ldots \rightarrow \hat{g}(z).$$

The endogenous variables are $h_k^t(z) = h_k(z + \Delta_k z)$ with $\Delta_k z \approx \epsilon_k\left(z/(1 - T_k'(z)\right)\Delta T_k'(z)$ and $g_k^t(z) = g(z + c_k^{-1}\left(\Delta m_k\right))$, where t denotes the number in the cycle of the algorithm. The behavioral response to the tax change creates the endogeneity, such that if there was no behavioral response to the new tax rates, the algorithm would be redundant. The algorithm may not converge if the effect on marginal welfare weights from the transfer is too large or if the behavioral response to taxes are too large. It turns out to converge in the application presented next.

4 Application: Hypothetical gender tag in Norway

4.1 Description of application

The application is a hypothetical introduction of a gender tag in the Norwegian tax system. The application relies on empirical estimates of three types of sufficient statistics: income distribution parameters, actual tax rates, and the elasticity of taxable income.[15]

4.2 Norwegian income data

My analysis focuses on the labor income tax for wage earners. I use Norwegian income register data for the period 2001 to 2010 (Statistics Norway 2005). The main analysis is for wage earners in the year 2010. I exclude individuals who are under 25 and above 62 years old, who do not have wage earnings as their primary income source, and those with earnings below two times the government basic amount (NOK 75,641 in 2010, ≈ USD 12,500) for all years 2001-2010. The resulting balanced panel consists of about 800,000 individuals. Main variables include wage income, gender, age, county of residence, educational level and educational field. See Table 1 for summary statistics for 2010.

[15] I assume that marginal welfare weights do not depend on the household composition, see Bargain et al. (2006) more on the household, welfare and tax.

Table 1: Summary statistics for main variables in 2010

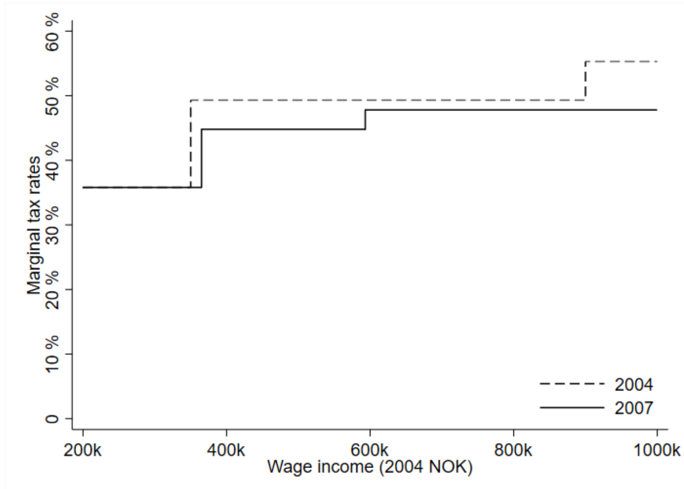
	Mean	Standard deviation
Wage income	541,432	329,576
Age	46.8	7.3
Share of males	57.4 %	
Share born in Norway	94.0 %	
Share with children	67.3 %	
Share married	61.0 %	
N	787,722	

4.3 Tax system

The Norwegian tax system applies different tax rates to different types of incomes (a ”dual income tax”). Specifically, it combines a flat tax on ”ordinary” income and a two-step top income tax applied to ”personal income”, where deductions are applied to ordinary income. The 2006 tax reform introduced a new dividend tax and partly aligned the tax treatment of different income types. As part of the reform, marginal tax rates on wage income were reduced, shown in Figure 2. To calculate total individual marginal tax rates, I employ the LOTTE tax-benefit calculator (Hansen et al. 2008).[16] It includes the standard tax rate and the two-bracket top income tax rates, the lower tax rates applied to certain areas in Northern Norway, certain income-dependent transfers (mainly social assistance and housing support), and I add a flat 20 percent VAT rate (roughly the average rate across goods) for all individuals. Panel a in Figure A2 in the Appendix shows the resulting average marginal tax schedule over the income distribution.

[16] I thank Bård Lian for assistance with the tax-benefit simulator.

Figure 2: The 2006 tax reform



Notes: Marginal tax rates on total wage earnings (ordinary + personal income) in 2004 and 2007.

4.4 Elasticity of taxable income

The optimal tax rate depends on how individuals respond to tax changes. After Feldstein (1995), the response is typically summarized by the elasticity of taxable income (ETI). The ETI is the percentage change in taxable income when the net-of-tax rate changes by one percent

$$\varepsilon(z) = \frac{(1-\tau)}{z}\frac{\partial z}{\partial(1-\tau)}. \tag{16}$$

In my setup, z is labor income for individuals who primarily obtain income from labor income. There is a large literature estimating ETIs and estimates differ widely across countries (see the survey by Saez, Slemrod, and Giertz (2012)). Most comparable to the setting here, Kleven and Schultz (2014) estimate ETIs in Denmark and obtain a response for wage earnings around 0.05, which is similar to what Thoresen and Vattø (2015) find for Norway, exploiting the same tax reform as here.

The difference is that I account for heterogeneity in tax responses across immutable observable characteristics. Here, this is exemplified by estimating separate ETIs for females and males. I estimate the ETIs using using a standard first-difference panel data approach with a Weber (2014) style instrument and a Kopczuk (2005) type mean-reversion control. See panel a in Figure A1 in the Appendix for the evolution of average income for those with tax reductions and for those with no tax change. See Table A1 for summary statistics of the same groups by gender. Specifically, the approach is a three-year first-difference panel data approach including a spline function in base-year income and the lag of base-year income to control for mean reversion and exogenous trends in income. The identifying variation in tax rates comes from the Norwegian 2006 tax reform. The estimating equation is

$$\begin{aligned} \Delta_3 \log(z_{i,t}) = &\alpha_t + \beta D_k \Delta_3 \log(1-\tau_{i,t}) + \theta \log(z_{i,t}) + \pi \Delta_1 \log(z_{i,t-1}) \\ &+ \eta M'_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{17}$$

where Δ_y is a y-year difference $x_{i,t+j} - x_{i,t}$, $z_{i,t}$ is taxable income for individual i in year t, $1-\tau_{i,t}$ is the corresponding net-of-tax-rate, D_k is a dummy for each characteristic, α_t is the year-specific effect, and $M_{i,t}$ is a vector of other observable features about the individuals. The tax rate change $\Delta_3 \log(1-\tau_{i,t})$ is instrumented by the tax rate change that would have occurred had income stayed constant $\log(1-\tau_{i,t+3}) - \log\left(1-\tau^I_{i,t}\right)$, where $\tau^I_{i,t}$ is the marginal tax rate in year $t+3$ applied to income in year $t-1$. Mean reversion and exogenous income trends create bias, such that $\log(z_{i,t})$ and $\Delta_1 \log(z_{i,t-1})$ are introduced as bias corrections (Kopczuk 2005).

The resulting estimates are shown in Table 2. Although estimates are small compared to US estimates, the main point here is that females respond about twice as much to the reform than males.

Table 2: ETI estimates

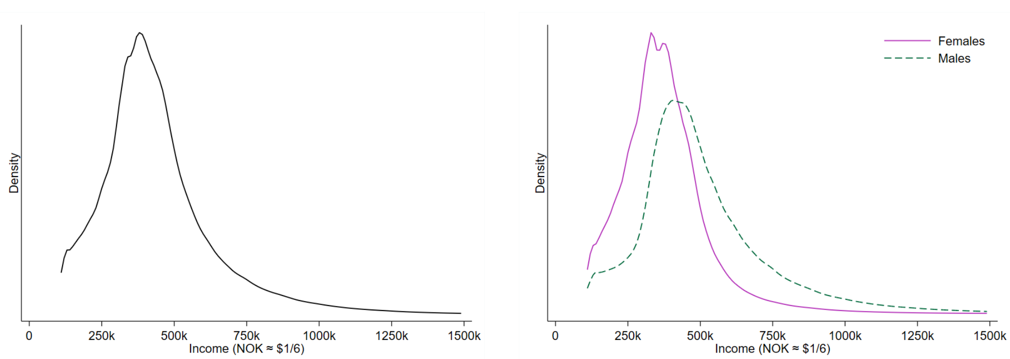
	All	Males	Females
ETI	0.081	0.054	0.101
se	0.002	0.003	0.004
N	4,723,512	2,710,870	2,012,870

Notes: ETI estimates, average and separated by gender for wage earners. The estimation is a first-difference equation where the tax rate change is instrumented by the reform-induced tax rate change. Other controls are year-dummies, the log of base year income, the first-difference of the log of income in the year prior to the base year, educational field and level, family status, county of residence, age and gender. See Table A2 for detailed results.

4.5 Income distributions

The next main determinant of marginal welfare weights is the shape of the income distribution. I follow the approach in Hendren (2020) to estimate the elasticity of the income distribution $\rho(z)$, which is done by applying an (adaptive) kernel to estimate the distribution before regressing the log of the density estimates on a fifth degree polynomial of the log of taxable income. Then, I predict the estimates of the elasticity of the income distribution at different points in the income distribution. Since the distribution is very thin at the top, I replace the kernel-based measure with a simple Pareto calculation above 1.1 million NOK (95th percentile) for the joint income distribution. Figure 4a presents the kernel estimate of the joint income distribution, while 4b shows the estimates for the female and male income distributions. Panel b in Figure A1 in the Appendix displays the estimate of the elasticity of the income distributions.

Figure 3: Norwegian wage income distributions, 2010



(a) Total distribution

(b) Distributions by gender

4.6 The effect of horizontal equity on inequality aversion

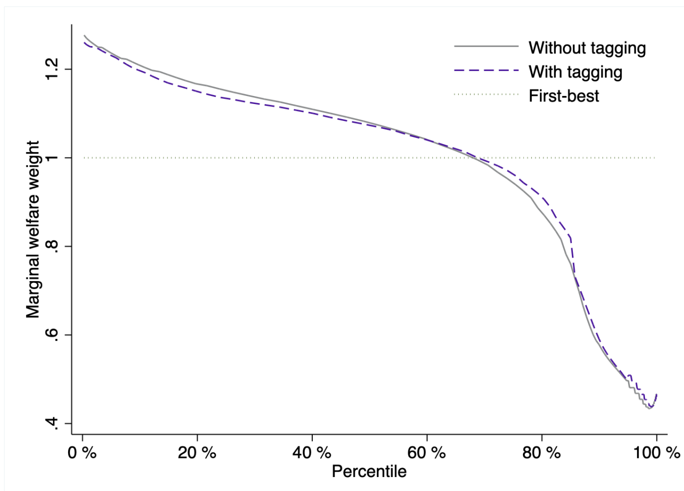
Introducing gender-based taxation reduces inequality by imposing a lump a sum transfer from males to females and by increasing marginal tax rates for males while reducing them for females. Panel a in Figure A2 presents optimal taxes by gender. Optimal gender-specific transfer from males to females is about NOK 70,000 (≈ USD 10,000). With tagging, males also face higher marginal tax rates than females, mainly due to the large (relative) difference in taxable income elasticities. Panel b in Figure A2 shows the effect of the gender-based lump sum transfer on inequality.

Using these statistics, Figure 4 presents estimates of marginal welfare weights when tagging is used averaged over characteristics at each income level, $\hat{g}(z)$, and when tagging is not used, $g(z)$. In line with Proposition 2, tagging decreases the average steepness of the marginal welfare weight schedule. The relaxation of the horizontal equity constraint implies less redistributive pressure, as redistribution has become cheaper and therefore more of it is implemented. The figure shows that a gender tag can have a visible effect on the marginal welfare weight schedule. The steepness of the inverse optimum marginal welfare weights from the actual tax system reflects both the contribution from the vertical equity concern and horizontal equity. By increasing the steepness, horizontal equity contributes in the same direction as vertical equity, which implies that if horizontal equity is ignored, the contribution from vertical equity is overestimated.

The total difference in steepness between $g(z)$ and $\hat{g}(z)$, $-E\left(g'(z) - \hat{g}'(z)\right)$, is 8.3 percent of the average steepness in the actual tax system. Hence, if inequality aversion is measured by the average steepness, it is overestimated by 8.3 percent by ignoring the horizontal equity concern. If more tags are accounted for, the effect will be larger, as redistributive costs are further reduced. Compared to the case when horizontal equity is not accounted for, less inequality averse preferences are required to rationalize current tax policy. This means that a lower concern for inequality reduction may be warranted in other policies with redistributive consequences, such as environmental or monetary policy. The policy implication is therefore that to be consistent with the social preferences for reducing inequality in tax policy, less inequality reduction is required in other policy domains. Since inequality aversion is measured here by the distortions the government is willing to accept for a redistributive tax system, the efficiency loss from not being able to tag based on gender also amounts to 8.3 percent of the total distortions imposed by the tax system.

Next, I compare the relative marginal welfare weight at specific income levels. This is similar to Okun's leaky bucket (Okun 1975), in that inequality aversion is measured by the extent of leakage society is willing to accept. In the actual tax system, society is indifferent between taking $100 from an individual with income at the 80th percentile and taking $75 from an individual with income at the 20th percentile. In the tax system when tagging is used, society is indifferent between taking $100 from an individual with income at the 80th percentile and taking $81 from an individual with income at the 20th percentile. Hence, the priority on vertical equity implies a weight of the 20th relative to the 80th percentile of 1.24, while including the horizontal equity priority increases it to 1.33. The higher relative weight on income at the 20th percentile thereby partly reflects that the government is not allowing tagging.

Figure 4: Marginal welfare weights when tagging is used and not used



Notes: Inverse optimum marginal welfare weights without tagging, $g(z)$, and when tagging is used, $\hat{g}(z)$, over the income distribution for wage earners in 2010.

4.7 Inequality aversion and climate policy

I now demonstrate the implications of overestimating inequality aversion by an application to the social discount rate used in evaluating climate change policy. In doing so, I also show how the inverse optimum tax approach can be applied to such models.

There has been a recent interest in accounting for income inequality in climate change policy evaluation (for example Anthoff and Emmerling (2019) and Groom and Maddison (2019)). One common method is to measure social preferences for redistribution from current tax policy, as is done in this paper. However, the approach in the climate policy literature is typically based on a distinct welfare criterion and ignores the second-best nature of the optimal tax problem.

The key normative parameter in evaluating climate change policy is the social discount rate, which determines how much lower priority to put on the future compared to the present. One derivation of the social discount rate is the Ramsey rule, such that

$$r = \psi + \gamma\nu, \tag{18}$$

where r is the social discount rate, ψ is the pure time preference (in percent), γ is inequality aversion (as an elasticity) and ν is the consumption growth (in percent).

I focus on the inequality aversion parameter. Estimates of marginal welfare weights must be translated into a relevant inequality aversion measure. For a government weighting individuals according to their consumption with constant relative risk aversion, inequality aversion is given by $\gamma = -E\left(log(g(z))/log(e(z))\right)$ (from Equation 6). To relate this to the social discount rate, I normalize average equivalent consumption to 1. The result is that when horizontal equity is ignored, the inequality aversion parameter is 0.41, while when horizontal equity across gender is accounted for, it drops to 0.36. Interestingly, both of these are lower than what is typically used in the climate policy literature. For illustration, set $\psi = 1$ and $\nu = 3$. With $\gamma = 0.41$, $r = 2.23\%$, and with $\gamma = 0.36$, $r = 2.08\%$. Hence, by accounting for horizontal equity, the estimated social discount rate is lower, such that earlier climate change policies are optimal.

5 Conclusion

Governments do not exploit all the relevant available information when setting taxes. This cannot be explained by standard criteria, which focus exclusively on vertical equity (and efficiency). By combining vertical equity with horizontal equity, I show that one can rationalize both the high cost the government is willing to incur to redistribute and the restriction on the type of information used in setting taxes. To measure the importance of accounting for horizontal equity, I decompose inverse optimum marginal welfare weights into the contribution from each form of equity. From the decomposition, I demonstrate that accounting for horizontal equity affects the inferred priority on vertical equity and inequality aversion.

The point of distinguishing between vertical and horizontal equity is, first, to reveal equity principles that are consistent with observed tax policy. This allows policy makers and voters to evaluate for themselves whether they find these equity principles appealing. The second point is to estimate and correct the bias in the measurement of vertical equity. Since horizontal equity increases the cost of redistribution, directly using standard inverse optimum marginal welfare weights would lead to an overestimation of the role of vertical equity in supporting the current tax system. As a consequence, I find in the empirical application to gender neutral taxation in Norway that inequality aversion is overestimated by 8.3 percent when horizontal equity is ignored. In an application to climate change policy, this translates into a lower social discount rate and earlier climate change action being optimal.

More generally, the results show that the instruments governments choose to employ to reduce inequality matter for the interpretation of their inequality aversion. Imagine a country that chooses to exploit more information in setting taxes. The extra information is used to reduce inequality while also reducing the cost of redistribution. Measuring inequality aversion directly from inverse optimum marginal welfare weights would suggest that the country has become less inequality averse, as it accepts a lower cost of redistribution than it did before. This paper has shown that this conclusion is misleading, and how to adjust inequality aversion measures for constraints that governments impose on themselves. This means that accounting for the trade-off between inequality reduction and non-discrimination is key when evaluating redistributive tax policy.

References

Ahmad, Ehtisham, and Nicholas Stern. 1984. “The theory of reform and Indian indirect taxes”. *Journal of Public Economics* 25 (3): 259–298.

Akerlof, George A. 1978. “The economics of ”tagging” as applied to the optimal income tax, welfare programs, and manpower planning”. *The American Economic Review* 68 (1): 8–19.

Alesina, Alberto, Andrea Ichino, and Loukas Karabarbounis. 2011. “Gender-based taxation and the division of family chores”. *American Economic Journal: Economic Policy* 3 (2): 1–40.

Anthoff, David, and Johannes Emmerling. 2019. “Inequality and the social cost of carbon”. *Journal of the Association of Environmental and Resource Economists* 6 (2): 243–273.

Atkinson, Anthony B. 1980. “Horizontal equity and the distribution of the tax burden”. In *The Economics of Taxation*, ed. by Henry Aaron and Michael J. Boskin. Brookings Institution Press.

Auerbach, Alan J, and Kevin A Hassett. 2002. “A new measure of horizontal equity”. *American Economic Review* 92 (4): 1116–1125.

Bargain, Olivier, et al. 2006. “Does the representation of household behavior matter for welfare analysis of tax-benefit policies? An introduction”. *Review of Economics of the Household* 4:99–111.

Bargain, Olivier, et al. 2014. “Tax-benefit revealed social preferences in Europe and the US”. *Annals of Economics and Statistics*: 257–289.

Bastani, Spencer, Sören Blomquist, and Luca Micheletto. 2013. “The welfare gains of age-related optimal income taxation”. *International Economic Review* 54 (4): 1219–1249.

Bastani, Spencer, and Jacob Lundberg. 2017. “Political preferences for redistribution in Sweden”. *The Journal of Economic Inequality*: 1–23.

Berg, Kristoffer, and Paolo Giovanni Piacquadio. 2020. “The equal-sacrifice social welfare function with an application to optimal income taxation”. *CESifo Working Paper*.

Bourguignon, François, and Amedeo Spadaro. 2012. “Tax–benefit revealed social preferences”. *The Journal of Economic Inequality* 10 (1): 75–108.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. 2019. “The many faces of human sociality: Uncovering the distribution and stability of social preferences”. *Journal of the European Economic Association* 17 (4): 1025–1069.

Cappelen, Alexander W, et al. 2007. “The pluralism of fairness ideals: An experimental approach”. *American Economic Review* 97 (3): 818–827.

Christiansen, Vidar, and Eilev S Jansen. 1978. "Implicit social preferences in the Norwegian system of indirect taxation". *Journal of Public Economics* 10 (2): 217–245.

Cremer, Helmuth, Firouz Gahvari, and Jean-Marie Lozachmeur. 2010. "Tagging and income taxation: Theory and an application". *American Economic Journal: Economic Policy* 2 (1): 31–50.

Feldstein, Martin. 1976. "On the theory of tax reform". *Journal of Public Economics* 6 (1-2): 77–104.

— . 1995. "The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act". *Journal of Political Economy* 103 (3): 551–572.

Fleurbaey, Marc, and Francois Maniquet. 2018. "Optimal income taxation theory and principles of fairness". *Journal of Economic Literature* 56 (3): 1029–79.

Groom, Ben, and David Pr. Maddison. 2019. "New estimates of the elasticity of marginal utility for the UK". *Environmental and Resource Economics* 72 (4): 1155–1182.

Hansen, Kirsten, et al. 2008. "LOTTE-Skatt-en mikrosimuleringsmodell for beregning av direkte skatter for personer". *Statistisk sentralbyrå*.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante. 2020. "Optimal progressivity with age-dependent taxation". *Journal of Public Economics* 189:104074.

Hendren, Nathaniel. 2020. "Measuring economic efficiency using inverse-optimum weights". *Journal of Public Economics* 187:104–198.

Jacobs, Bas, Egbert LW Jongen, and Floris T Zoutman. 2017. "Revealed social preferences of Dutch political parties". *Journal of Public Economics* 156:81–100.

Jacquet, Laurence, and Etienne Lehmann. 2020. "Optimal income taxation with composition effects". *Journal of the European Economic Association* 19 (2): 1299–1341.

Jenkins, Stephen P. 1988. "Empirical measurement of horizontal inequity". *Journal of Public Economics* 37 (3): 305–329.

Kaplow, Louis. 1989. "Horizontal equity: Measures in search of a principle". *National Tax Journal* 42 (2): 139.

King, Mervyn A. 1983. "An index of inequality: With applications to horizontal equity and social mobility". *Econometrica*: 99–115.

Kleven, Henrik Jacobsen, and Esben Anton Schultz. 2014. "Estimating taxable income responses using Danish tax reforms". *American Economic Journal: Economic Policy* 6 (4): 271–301.

Kopczuk, Wojciech. 2005. "Tax bases, tax rates and the elasticity of reported income". *Journal of Public Economics* 89 (11-12): 2093–2119.

Kuziemko, Ilyana, et al. 2015. "How elastic are preferences for redistribution? Evidence from randomized survey experiments". *American Economic Review* 105 (4): 1478–1508.

Lockwood, Benjamin B, and Matthew Weinzierl. 2016. "Positive and normative judgments implicit in US tax policy, and the costs of unequal growth and recessions". *Journal of Monetary Economics* 77:30–47.

Mankiw, N Gregory, and Matthew Weinzierl. 2010. "The optimal taxation of height: A case study of utilitarian income redistribution". *American Economic Journal: Economic Policy* 2 (1): 155–76.

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

Musgrave, Richard Abel. 1959. *Theory of Public Finance: A Study in Public Economy*. McGraw-Hill.

Okun, Arthur M. 1975. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press.

Parfit, Derek. 1991. *Equality or Priority*. University of Kansas.

Saez, Emmanuel. 2001. "Using elasticities to derive optimal income tax rates". *The Review of Economic Studies* 68 (1): 205–229.

Saez, Emmanuel, Joel Slemrod, and Seth H Giertz. 2012. "The elasticity of taxable income with respect to marginal tax rates: A critical review". *Journal of Economic Literature* 50 (1): 3–50.

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized social marginal welfare weights for optimal tax theory". *American Economic Review* 106 (1): 24–45.

Sausgruber, Rupert, and Jean-Robert Tyran. 2014. "Discriminatory taxes are unpopular—Even when they are efficient and distributionally fair". *Journal of Economic Behavior & Organization* 108:463–476.

Spadaro, Amedeo, Luca Piccoli, and Lucia Mangiavacchi. 2015. "Optimal taxation, social preferences and the four worlds of welfare capitalism in Europe". *Economica* 82 (327): 448–485.

Stantcheva, Stefanie. 2020. "Understanding tax policy: How do people reason?" *National Bureau of Economic Research Working Paper*.

Temkin, Larry S. 1993. *Inequality*. Oxford University Press.

Thoresen, Thor O, and Trine E Vattø. 2015. "Validation of the discrete choice labor supply model by methods of the new tax responsiveness literature". *Labour Economics* 37:38–53.

Weber, Caroline E. 2014. "Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences". *Journal of Public Economics* 117:90–103.

Weinzierl, Matthew. 2018. "Revisiting the classical view of benefit-based taxation". *The Economic Journal* 128 (612): F37–F64.

— . 2014. "The promise of positive optimal taxation: normative diversity and a role for equal sacrifice". *Journal of Public Economics* 118:128–142.

— . 2011. "The surprising power of age-dependent taxes". *The Review of Economic Studies* 78 (4): 1490–1518.

A Summary statistics and detailed results

For visualization, the treated are defined as individuals with earnings below NOK 1 Mill. whose tax rates falls by more than 3 percentage points due to the reform, while the control group consists of individuals with earnings above NOK 250,000 whose tax rates do not change. All variation in tax and income are exploited in the regressions.

Table A1: Summary of treatment and control groups

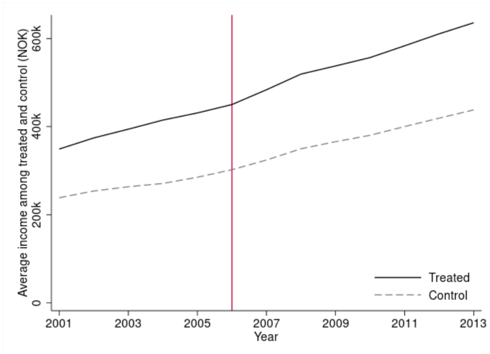
	All		Males		Females	
	Treated	Control	Treated	Control	Treated	Control
Age	40.9	40.4	40.4	39.5	41.8	40.9
Born in Norway	94.5 %	93.7 %	94.7%	92.5 %	93.9%	94.4 %
Children	69.7 %	71.9 %	70.0 %	63.8 %	69.2 %	76.6 %
Married	56.5 %	56.7 %	57.7 %	49.7 %	54.0 %	60.6 %
Completed higher education	42.2 %	40.1 %	41.6 %	38.7 %	44.1 %	42.3 %
Business or technology education	35.4 %	30.3 %	39.5 %	37.1 %	31.4 %	30.2 %
N	21,988	101,094	14,887	62,387	7,101	38,707

Table A2: Detailed ETI estimates

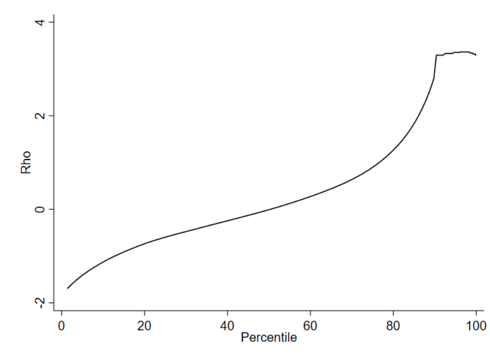
	Full	Males	Females
Tax treatment	0.081	0.054	0.101
	(0.002)	(0.003)	(0.004)
Age	0.008	0.001	0.025
	(0.000)	(0.000)	(0.000)
Birth country	0.006	0.015	0.002
	(0.001)	(0.001)	(0.001)
Children	0.005	0.008	0.002
	(0.000)	(0.000)	(0.000)
Married	0.003	0.010	-0.007
	(0.000)	(0.000)	(0.000)
Male	0.051		
	(0.000)		
N	4,723,512	2,710,226	2,012,870

Notes: Standard errors in parentheses. Other controls are year-dummies, the log of base year income, the first-difference of the log of income in the year prior to the base year, educational field and level, and residence county.

Figure A1: Tax response and income distribution

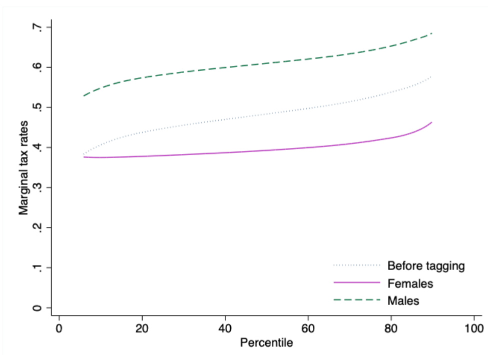


(a) Tax reform treatment

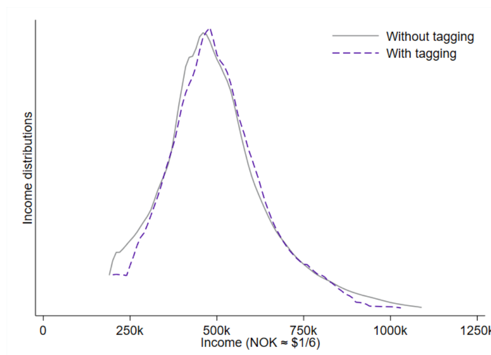


(b) Elasticity of the income distribution

Figure A2: Marginal taxes and income distributions with and without tagging



(a) Marginal tax rates



(b) Income