Revealing Inequality Aversion from Tax Policy*

Kristoffer Berg[†]

November 19, 2020

JOB MARKET PAPER (latest version here)

Abstract

Governments have increasing access to information about individuals, but they exploit little of it in setting taxes. This paper shows how to reveal inequality aversion from observed tax policy when governments restrict the information they exploit. The first contribution is to map social marginal welfare weights into the concerns for vertical and horizontal equity. While vertical equity provides the standard inequality aversion rationale for redistributive taxation, horizontal equity introduces a restriction against tax discrimination based on certain characteristics. Building on the inverse optimal tax problem, I develop a theory and optimal tax algorithm to reveal the priority on each concern. The second contribution is to apply the model to gender taxation in Norway and estimate the necessary statistics. The main result is that inequality aversion is overestimated if horizontal equity is ignored, and by as much as 30 percent in the application.

JEL: D31, D63, H20, H21, H23, I38. Keywords: horizontal equity; optimal income taxation; social preferences; tagging

*I am grateful to my supervisor Paolo G. Piacquadio. I have benefited from discussions with Rolf Aaberge, Alberto Alesina, Annette Alstadsæter, Geir Asheim, Marcus Berliant, Giacomo Brusco, Monica Costa Dias, Johannes Fleck, Bård Harstad, Shafik Hebous, Martin Holm, Harald Høyem, Kasper Kragh-Sørensen, Claus Kreiner, Etienne Lehmann, Ben Lockwood, Andreas Myhre, Frikk Nesje, Alex Rees-Jones, Marius Ring, Emmanuel Saez, Ragnhild Schreiner, Kjetil Storesletten, Magnus Stubhaug, Dmitry Taubinsky, Thor Thoresen, Gaute Torsvik, Ragnar Torvik, Danny Yagan, Matt Weinzierl, audiences at IIPF, NTA, UC Berkeley, IRLE, UiO, Uppsala University, NMBU, NTNU, Statistics Norway and OsloMet. This project has received funding from the Research Council of Norway through Oslo Fiscal Studies.

[†]Department of Economics, University of Oslo. Email: kristoffer.berg@econ.uio.no.

1 Introduction

Which equity concerns could support actual tax policies? Most economic models derive optimal policy with utilitarianism as the welfare criterion. This would suggest that the concerns that support current policies are 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 appreciated is that utilitarianism also implies that it is optimal to exploit all relevant information about individuals in setting taxes (*tagging*). 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. Yet, in actual tax policy, there are few cases of differential taxation across characteristics such as gender, and much fewer than utilitarianism would recommend (see Mankiw and Weinzierl 2010 on the relationship between utilitarianism and tagging).

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 different characteristics with similar incomes.

The theory has important implications. My first result is that by accounting for horizontal equity, the implied level of inequality aversion is lower. The reason is that the priority on horizontal equity increases both inequality and the cost of redistribution. While the government could tag based on observable characteristics, the concern for horizontal equity prohibits the use of certain tags. This limits the government's redistributive instruments and increases the cost of redistribution. Inequality aversion is a key parameter in many economic models, such as in optimal macroeconomic and environment policy. Hence, if policy choices should reflect societies' redistributive preferences, correctly measuring revealed inequality aversion (from observed tax policy) is crucial to decide which policies are optimal. My second result is that this effect can be large. In an application to gender-neutral taxation, I estimate the relevant parameters using Norwegian register data. I find that the level of inequality aversion is overestimated by up to 30% when attributing the cost of not exploiting gender information to vertical equity.

Interestingly, both vertical and horizontal equity used to be main principles in taxation, while after Mirrlees (1971), economists have instead predominantly studied optimal taxation with a utilitarian social welfare function.[1] Such a policy maker would exploit available tags when setting tax policy. The form of tagging considered here is to condition taxes on immutable characteristics such as gender, height and age.[2] There is a longstanding literature on tagging, starting with Akerlof (1978), and recent contributions include Cremer, Gahvari, and Lozachmeur (2010), Alesina, Ichino, and Karabarbounis (2011) and Bastani (2013) 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.

However, there is limited use tagging based on immutable characteristics in actual tax systems, and most of a person's disposable income is determined by her pre-tax income.[3] 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 there is little conditioning on characteristics in actual tax systems, one natural explanation is that society holds a counteracting *equity* rationale for not exploiting information on certain characteristics. Here, the concern is horizontal equity.

While the utilitarian criterion restricts equity principles, its generalized version allows the researcher to vary the level of inequality aversion. Building on this flexi-

[1] However, there are recent non-utilitarian contributions, see Fleurbaey and Maniquet (2006; 2018), Weinzierl (2014; 2018) and Berg and Piacquadio (2020).

[2] A taxpayer's gender may not be strictly immutable to tax policy, but assigned sex at birth could be an alternative immutable characteristic that would also give rise to a horizontal equity concern.

[3] Of course, counterexamples do exist. In the US, EITC payments are higher for single mothers, some countries, including Italy, levied a "bachelor's tax" on unmarried men, and a number of countries effectively set lower taxes for the youngest workers. However, 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.

bility, a literature following Bourguignon and Spadaro (2012), under the name *inverse optimal taxation*, exploits actual tax-transfer systems to reveal the marginal welfare weights that make the current tax system the optimal one. Contributions 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, and Jacobs, Jongen, and Zoutman (2017) for political parties in the Netherlands, while Hendren (2020) relates the inverse optimum approach to cost-benefit criteria.[4] A key point in some of these contributions is that implicit marginal welfare weights is 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. However, the link between horizontal equity and the inverse optimal tax problem has not been studied yet.

The theoretical framework in this paper provides a mapping from the government's valuation of increasing consumption at each income level, the *marginal welfare weight*, into the concerns for vertical and horizontal equity. Since tagging can be exploited to increase vertical equity at the same efficiency cost, the higher observed cost of redistribution cannot be explained by vertical equity or efficiency. Hence, other equity principles such as horizontal equity are necessary to rationalize policy. This implies that if one infers the vertical equity priority by the costs governments are willing to incur for redistribution, without accounting for horizontal equity, one will overestimate the priority on vertical equity and thereby the level of inequality aversion.

Next, I develop a method to measure the separate contributions of 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 in cases with and without tagging. Since the actual tax system respects horizontal equity, the standard inverse optimal

[4]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).

tax approach reveals the marginal welfare weights in this case. To estimate marginal welfare weights in the counterfactual tax system, without this restriction, I develop an algorithm that exploits the current marginal welfare schedule in an optimal tax problem with tagging. Under certain assumptions about sufficient statistics for welfare weights, this permits estimation of the size of the bias to inequality aversion when horizontal equity is ignored. Then, in an 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.

The paper contributes to three main strands of literature. First, it contributes to optimal taxation in the Mirrlees (1971) tradition. This is done by introducing horizontal equity as a constraint and solving the optimal tax problem with tagging in the local optimum framework (Saez 2001), highlighting the implications for the inverse optimum problem and inequality aversion. Second, it contributes to a growing literature expanding normative principles in economics (see Feldstein (1976) and Atkinson (1980) for classic contributions). Here, the contribution is similar in spirit to Mankiw and Weinzierl (2010), Weinzierl (2014), Saez and Stantcheva (2016) and Lockwood and Weinzierl (2016), who argue that the traditional principles in optimal taxation do not fit well with principles people state in surveys or with actual tax policy in the US. A difference is that I combine a revealed preference approach with tagging and horizontal equity.[5] Third, it contributes to the literature on revealed social preferences, which is achieved by surveys (Kuziemko et al. 2015, Alesina, Stantcheva, and Teso 2018 and Stantcheva 2020), experiments (Cappelen et al. 2007 and Bruhin, Fehr, and Schunk 2019), and, as in this paper, revealing preferences from observed policy (McFadden 1975, Basu 1980 and Bourguignon and Spadaro 2012), by showing how social preferences for vertical and horizontal equity jointly rationalize current tax policies.

The paper proceeds as follows. Section 2 presents a simple two-type model to highlight the relation between vertical and horizontal equity. Section 3 develops the general model and equity principles, before presenting the decomposition of marginal

[5]Another related contribution is Hermle and Peichl (2018), which exploits revealed marginal welfare weights in an optimal tax problem with multiple types of income. They are however not concerned with equity and assume that marginal welfare weights stay constant under different tax systems.

welfare weights into vertical and horizontal equity contributions. Section 4 introduces the continuous optimal taxation model with tagging and the inverse optimum tax problem. Section 5 presents the empirical application, where I provide estimates on heterogeneity in tax responses and apply the findings to the tax model. Section 6 concludes.

2 A simple illustration of horizontal equity in optimal tax

To illustrate the role of vertical and horizontal equity, I present a simple two-type model. The full model is presented in Section 3. There are two types i of individuals, $i = 1$ (low type) and $i = 2$ (high type), with corresponding wage rates w_i such that $w_1 < w_2$. Each individual is associated with the observable characteristic gender, $k = m, f$, which I assume is fixed. Importantly, gender may be informative of individuals' productivites. Let the proportion of each gender k with type i be denoted by p_i^k, such that $\sum_k \sum_i p_i^k = 1$. Type- and gender-specific variables are denoted by x_i^k and averages across gender for a given type are given by $\hat{x}_i = \sum_k p_i^k x_i^k / \sum_i p_i^k$ for $i = 1, 2$. Assume also a homogeneous quasi-linear utility function for each individual, $u(c_i^k, l_i^k) = c_i^k - v(l_i^k)$, which depends on consumption, c_i^k, and labor supply, l_i^k. Individuals maximize utility subject to their budget constraint

$$\max u(c_i^k, l_i^k) \text{ s.t. } c_i^k = w_i l_i^k - T_i^k, i = 1, 2 \text{ and } k = f, m, \tag{1}$$

where $z_i^k = w_i l_i^k$ is the type- and characteristic-specific pre-tax income and T_i^k is their tax payment, such that each individual obtain the type-specific indirect utility $V_i^k = V(c_i^k, z_i^k / w_i)$.

The government sets taxes, T_i, in order to raise revenue, $\sum_i T_i = R$. It maximizes welfare, W:

$$\max W = \sum_k \sum_i p_i^k G(V_i^k), \tag{2}$$

where $G(V_i^k)$ is an equal concave transformations of individual indirect utilities. This assumes that the government respects anonymity, in that it evaluates the same amount of utility for different types and genders equally. The marginal welfare weight, the value the government attaches to increasing consumption for type i with gender k, is

$g_i^k = G'(V_i^k)$ (since $\partial V_i^k(c_i^k, l_i^k)/\partial c_i^k = 1$). The "steepness" of the welfare weight schedule is measured by the absolute value of the difference in marginal welfare weights between the less and the more productive, $|\Delta \hat{g}| = |\hat{g}_2 - \hat{g}_1|$.

Vertical equity is the local priority on reducing consumption differences. *Inequality aversion* is the average absolute change in the marginal welfare weight when consumption increases, $-G''(\hat{V}_i)$, in the case where vertical equity is the only priority (the relationship between vertical equity and inequality aversion is explored at the end of the section). *Tagging* is to exploit information on gender when setting taxes, $T^k(z) \neq T(z)$. *Horizontal equity* introduces a constraint on policy such that a gender tag is impermissible, $T^f(z) = T^m(z) = T(z)$ for all z.

2.1 Optimal taxation in the two-type model

Using the model features presented above, the optimal tax model builds on the classical Mirrlees two-type model, such as the one presented in Stiglitz (1982). The key feature is the self-selection constraints for each type, such that the allocation is incentive-compatible (the utility of each type must be weakly higher in the bundle intended for each type than the bundle intended for the other type).[6] Since the social welfare function is concave, only mimicking by the high type can emerge (Stiglitz 1982). Assume for ease of notation that the proportion of each gender and each type are equal, while the relative number of types within each gender may differ. Then the government maximizes welfare, such that the government raises enough revenue

$$\sum_k \sum_i p_i^k (z_i^k - c_i^k) \geq R \text{ with multiplier } \gamma, \tag{3}$$

and the four self-selection constraints hold

$$V_2^k\left(c_2^k, z_2^k\right) \geq V_2^s\left(c_2^s, z_2^s\right) \; \forall k = f, m \text{ and } s = f, m \text{ with multipliers } \lambda^{ks}. \tag{4}$$

The government may face three different information scenarios and a choice about whether to exploit information on gender or not. Since the choice is irrelevant when information is complete or when there is no information on gender, it leaves us with four interesting cases.

[6] I am assuming that the optimum imposes separation, which is standard.

1. The government has complete information. The first-best is obtained and tagging is irrelevant.
2. The government lacks information about w_i, l_i and k. This is the standard problem, as tagging is impossible.
3. The government lacks information about w_i and l_i. Tagging is optimal, and used in combination with an income tax.
4. The government lacks information about w_i and l_i. In addition to the constraints above, the government also imposes that it should treat individuals independently of their gender. No tagging is optimal, as the government respects horizontal equity.

Define $\Delta_n x$ as $x_2 - x_1$ for case $n = 1, 2, 3, 4$. The four cases are now discussed in order.

Case 1: First-best

This case prevails if the government has information on w_i or if gender is a perfect predictor of w_i. In the latter case, all low-type individuals have one gender and all high-type individuals have the other gender. This information is known to and exploited by the government. Since the government can distinguish abilities perfectly, it can impose the first-best allocation, and we obtain

$$\Delta_1 \hat{c} = 0 \text{ and } \Delta_1 \hat{g} = 0. \tag{5}$$

Then, taxes do not vary with income and there is no possibility to mimic, such that self-selection constraints do not bind. The result is that when information is complete or gender perfectly predicts ability (and with no limits on the government's tax instruments), consumption is equalized and the welfare weight schedule is flat.

Case 2: Standard second-best

In the standard Mirrleesian case there is no distinction across gender (so all k's are dropped from the notation), such that the optimization problem simplifies to

$$\max_{c,z} W = \sum_i p_i G(V_i(c_i, z_i)), \tag{6}$$

s.t.

$$\sum_i z_i - c_i \geq 0 \text{ with multiplier } \gamma, \tag{7}$$

and

$$V_2(c_2, z_2) \geq V_2(c_1, z_1) \text{ with multiplier } \lambda. \tag{8}$$

The solution is

$$\Delta_2 c > 0 \text{ and } \Delta_2 g = -\lambda < 0, \tag{9}$$

which means the high type has higher consumption and is assigned a lower marginal welfare weight than the low type in the standard case.

Case 3: Tagging

Now, consider the case where the government has and exploits information on gender, but gender is not a perfect predictor of ability. Since the government exploits information on k, it sets separate tax systems for each gender and there are no cross-gender incentive compatibility constraints ($\lambda^{mf} = \lambda^{fm} = 0$.) The result is

$$\Delta_3 \hat{c} > 0 \text{ and } \Delta_3 \hat{g} = -\frac{1}{2}\left(\lambda^{ff} + \lambda^{mm}\right) < 0. \tag{10}$$

As in Case 2, consumption is higher and the marginal welfare weight lower for the high type than the low type.

Case 4: No tagging

This is the case when the government has information on gender, but it respects horizontal equity, such that it does not exploit information on gender in the design of the tax system. It subjects itself to the constraint $T^f(z) = T^m(z)$ for all z. Since the government does not exploit information on gender, the between-gender self-selection constraints also enter the solution

$$\Delta_4 \hat{c} > 0 \text{ and } \Delta_4 \hat{g} = -\frac{1}{2}\left(\lambda^{ff} + \lambda^{mm} + \lambda^{fm} + \lambda^{mf}\right) < 0. \tag{11}$$

Since the government faces the same problem as in Case 2 (it has restricted itself to exploit only the information available in Case 2), we also know that $\Delta_4 \hat{g} = \Delta_2 g$.

2.2 Implications for equity principles

From the four cases, we observe that

$$|\Delta_1 \hat{g}| < |\Delta_3 \hat{g}| < |\Delta_4 \hat{g}| = |\Delta_2 g|, \tag{12}$$

which implies that the marginal welfare weight schedule is steeper when horizontal equity is a constraint (Case 4) compared to when it is not (Case 3). The ordering of differences between marginal welfare weights in the different cases is associated with a consistent ranking of consumption differences

$$\Delta_1 \hat{c} < \Delta_3 \hat{c} < \Delta_4 \hat{c} = \Delta_2 c. \tag{13}$$

Assume in the following that the government has access to information on taxpayers' gender, which is the case for most governments.[7] The priority on vertical equity (VE) is measured by the weight in Case 3, when the government is concerned with only vertical equity

$$\Delta VE = \Delta_3 \hat{g} < 0. \tag{14}$$

The priority on horizontal equity (HE) is the shadow price of being restricted to Case 4 rather than Case 3, which is measured by

$$\Delta HE = \Delta_4 \hat{g} - \Delta_3 \hat{g} < 0. \tag{15}$$

Hence, marginal welfare weights can be decomposed

$$\Delta_4 \hat{g} = \Delta VE + \Delta HE. \tag{16}$$

The main message is that vertical equity (and thereby inequality aversion) cannot be measured simply by considering steepness of the welfare weight schedule in a system where information on gender is not exploited. Because the government has access to this information but chooses not to exploit it, horizontal equity is also a priority in that system. Most governments do have information on gender and choose not to exploit it when setting income taxes (Case 4), and then marginal welfare weights do not reflect only vertical equity. Hence, if horizontal equity is not accounted for, one will overestimate the absolute value of the priority on vertical equity.

[7] All proofs on the relation between the concepts and how I measure them are presented in Section 3.

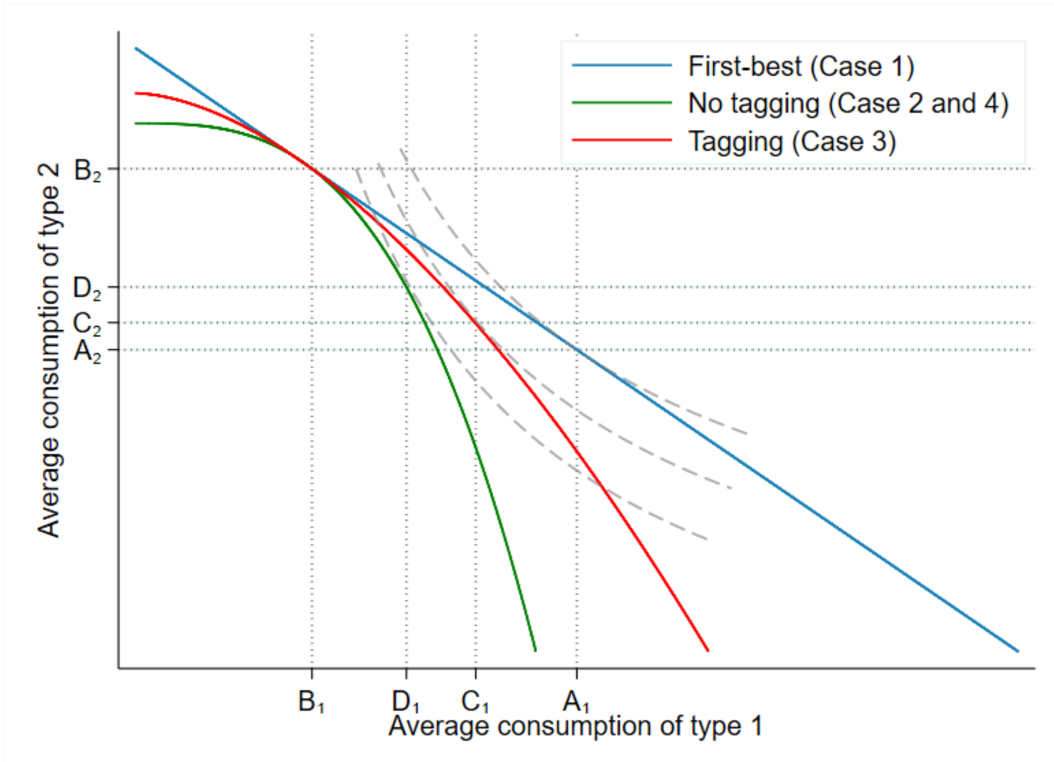
Average inequality aversion, the absolute value of the concavity of the social welfare function over types, averaged over genders, is measured by

$$-G''(\hat{V}_i) = -\frac{\partial \hat{g}_i}{\partial \hat{c}_i} \approx -\frac{\Delta \hat{g}}{\Delta \hat{c}}. \tag{17}$$

The difference in inequality aversion between cases reflects both the difference in marginal welfare weights and the difference in the allocation between the cases. $VE(z)$ is therefore not sufficient to determine the change in inequality aversion, since it does not account for the difference in the allocations of z and c. Inequality aversion with information on gender is measured by Case 3, $-\Delta_3\hat{g}/\Delta_3\hat{c}$. If $-\Delta_3\hat{g}/\Delta_3\hat{c} < -\Delta_4\hat{g}/\Delta_4\hat{c}$, then inequality aversion is overestimated for a government in Case 4 when horizontal equity is ignored. This is always the case, as there is more redistribution in Case 3 than in Case 4.

Figure 1 illustrates the point. Assume that the government need not raise any revenue, and only set taxes for redistributive purposes. Denote by $X = (X_1, X_2)$ an allocation such that $X_1 = c_1, X_2 = c_2$. The social welfare function specified in Equation (2) implies social indifference curves that rank different allocations, but does account for horizontal equity. The four cases presented earlier are associated with different consumption possibility frontiers, reflecting different costs of redistribution away from the ”laissez-faire” ($c_1 = z_1, c_2 = z_2$). When the government places no value on vertical equity, the government chooses the laissez-faire (allocation B) independently of the information available. In the first-best (Case 1) a government that only values vertical equity chooses allocation A and consumption levels are equalized. When the problem is second-best and the government exploits tagging (Case 3) it chooses allocation C, while when the government values both vertical and horizontal equity (Case 4) it chooses allocation D. Hence, vertical equity induces the move from allocation B to allocation C, while horizontal equity induces the move from allocation C to allocation D. We observe that the average consumption difference across types and the steepness of social indifference curve (reflecting the steepness of the welfare weights) is lower at the allocation when there is tagging, meaning that horizontal equity increases inequality across types and the steepness of the indifference curve at the allocation. The presence of the horizontal equity constraint increases redistributive costs, which lowers redistribution and increases the cost the government is willing to incur to reduce inequality. As the government deliberately detracts from using additional information, the cost of actual redistribution enforced increases. This is reflected by a steeper tangent for the social indifference curve.

Figure 1: The effects of vertical and horizontal equity



3 General model of vertical and horizontal equity

Here I present the general model of the concepts and the relationship between vertical and horizontal equity, including the proofs.

3.1 Description

Individuals

There is a continuum of individuals $i \in I$, with mass normalized to 1. Each individual is characterized by a wage rate $w_i \in (0, \infty)$, and a utility function $u_i(c_i, l_i)$, which depends on consumption $c_i > 0$ and 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 $h(z)$. The tax payment of individual i is T_i.

Each individual is also characterized by a *tag* that takes different values, k, and each value is a *characteristic*. 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 the $h_k(z)$. The function $c_k(z)$ translates income z into consumption $c_k(z) = z - T_k(z)$, where $T_k(z)$ is the characteristic-specific tax, such that the relation between z and c may vary across characteristics. 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 (and total) of variables $x(z)$ and $x_k(z)$ over the income distributions $h(z)$ and $h_k(z)$ by $E\left(x(z)\right) = \int_{-\infty}^{\infty} x(z)h(z)dz$ and $E_k\left(x_k(z)\right) = \int_{-\infty}^{\infty} x_k(z)h_k(z)dz$, respectively.

Government

The government sets taxes T_i in order to raise revenue $\sum_i T_i = R$. It maximizes total weighted consumption[8]

$$\max W = \sum_k p_k E_k \left[g\left(c_k(z)\right) c_k(z)\right], \tag{18}$$

where $0 < g(c_k(z)) < \infty$ for each z and is the government's valuation of increased consumption c, marginal welfare weights, at each income level z for characteristic k. These are normalized such that $\sum_k p_k E_k \left(g(c_k(z))\right) = 1$. With an explicit social welfare function, the approach appears structural, but if weights are allowed to vary freely over the income distribution, it can represent the local approach in Saez and Stantcheva (2016). The new assumption here is that marginal welfare weights are equal across characteristics for a given consumption level c. See Appendix A for an

[8]This formulation assumes Pareto efficiency, continuity, separability and anonymity. In the case of homogeneous and quasi-linear preferences $u = c - v(l)$ and $c'(z) \geq 0$, this government is equivalent to a utilitarian government that maximizes the sum of weighted utility, with Pareto weights π such that $g(c(z)) = \pi\left(1 - \frac{v(z/w)}{c(z)}\right)$.

alternative formulation based on equivalent consumption levels.[9]

If the marginal welfare weight schedule is (weakly) falling in consumption, $g'(c) \leq 0$ for all c and $g'(c) < 0$ for some c, then the government is *redistributive*.[10] The government is more redistributive the higher is the average steepness of the welfare weight over consumption, $-\sum_k p_k E_k \left(g'(c_k(z))\right)$.

Define $\hat{g}(z) = \sum_k \left(h_k(z) / \sum_k p_k h_k(z)\right) g(c_k(z))$ as the *average marginal welfare weight* across characteristics at income level z, the *average local steepness* of the welfare weight schedule over income z as $-\hat{g}'(z)$ and the *total average steepness* of the marginal welfare weights schedule over income as $-E(\hat{g}'(z))$.

The *local amount of redistribution* is the marginal tax rate averaged over characteristics at income level z. *Total redistribution* is the sum of local redistribution over all characteristics and any lump sum grant m_k, $\sum_k p_k \left(E_k \left(1 - c_k'(z)\right) + m_k\right)$.

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, $c_k'(z) \geq 0$. I assume sorting within the relevant income distributions exploited by the government to set taxes, such that if the government exploits the joint income distribution, sorting is assumed over this distribution, while if the government exploits the marginal (characteristic-specific) income distributions, sorting is assumed within each of these distributions.[11]

3.2 Cases

Depending on the government's information set and preferences, the same four cases as in Section 2 may emerge. These cases are now discussed in order.

[9] All the corresponding propositions hold for the equivalent consumption formulation.

[10] In an optimum for a government with a standard utilitarian social welfare function: $W = E\left(G(u(c, z/w))\right)$, then $g(c) = G'\left(u(c,l)\right) u_c'(c,l)$, such that strict concavity of G in u and u in c implies $g'(c) < 0$.

[11] It corresponds to the role of separation in the two-type model. As is well-known, this property does not always hold in optimal taxation. However, is also the sufficient condition for optimum in the optimal tax problem and can be verified to hold for the specific problem.

Case 1: First-best

The wage rate w_i is observable to the government. This is the first-best case (where the second welfare theorem holds), such that the government can obtain any distribution it prefers. Because the government respects anonymity (the marginal welfare weight does not depend on characteristics for a given level of consumption), information on k is redundant.

Proposition 1. *Marginal welfare weights at the first-best optimum are constant and equal to* 1.

Proof. By contradiction, assume there exists two individuals h and j with consumption levels c_h and c_j such that marginal welfare weights are different $g_h \neq g_j$. Assume, without loss of generality, that $g(c_h) > g(c_j)$. Then, contrary to the proposition, this produces the maximum level of welfare $W^* = g(c_h)c_h + g(c_j)c_j + \int_{i \neq h,j} g(c_i)c_i h(z_i)dz > W \neq W^*$. Since consumption can be allocated freely, imagine increasing the consumption for h and reducing consumption for j with the same amount, $\Delta c_h = -\Delta c_j$. By separability between individuals in the social welfare function, the weights for all other individuals stay constant under this transfer, such that the change in welfare is $\Delta W = g(c_h))\Delta c_h - g(c_j)\Delta c_h > 0$, an increase in welfare, which is a contradiction. By anonymity this generalizes to any individual's weight deviating from equality. By the normalization, the sum and average of weights is 1, such that $g(z) = 1$. □

Case 2: Standard second-best

Now, w_i, l_i and k are unobservable to the government. Taxes must therefore be set according to individuals' pre-tax income z_i. This is the standard case in the optimal tax literature. If the government is redistributive, the information problem introduces a *cost of redistribution*. The cost emerges from individuals' responses to income taxes.

Proposition 2. *A redistributive government that lacks information about w_i, l_i and k places (weakly) higher marginal welfare weights on lower incomes, $g'(z) \leq 0$.*

Proof. Now, to equalize everyone's consumption, $c_i = c_j$ for all i, j, the government must set taxes at 100 percent and redistribute lump sum. This is not optimal, since

$\partial u_i(c_i, l_i)/\partial l_i < 0$ and no individual will work, such that $c_i = 0$ for all i. Hence, $c_i < c_j$ and $g(c_i) < g(c_j)$ for some $i \neq j$. Remember that $g'(c) \leq 0$ for a redistributive government. It can observe z about individuals and there is sorting, such that it sets taxes according to its valuation $g(c(z))$. By the sorting property, marginal welfare weights are thereby also (weakly) decreasing in income z, $\partial g(c(z))/\partial z = g'(c)c'(z) \leq 0$. □

Case 3: Tagging

The wage rate w_i and labor supply l_i are unobservable to the government, while k (and z) is observable. Now, taxes can be characteristic-specific, $T_k(z)$.[12] This entails that taxes and consumption may now differ at the same income level.

Proposition 3. *A redistributive government that lacks information about w_i and l_i, but has information on k, exploits this information and increases total redistribution. This corresponds to a (weakly) flatter marginal welfare weight schedule over consumption and income on average, $-E\left(\hat{g}'(z)\right) < -E\left(g'(z)\right)$.*

Proof. By Proposition 2, the information problem introduces different consumption levels and marginal welfare weights. Assume that at least two characteristics have different income distributions, $h_a(z) \neq h_b(z)$ for some $a \neq b$ (such that information on k is useful). Then, the government can increase welfare by introducing a transfer from one characteristic to another, dm. For a marginal transfer, the increase in welfare is the average difference in marginal welfare weights across characteristics, $\Delta W = dmE_a\left(g_a(z)\right) - dmE_b\left(g_b(z)\right)$, which follows from the welfare function and that the transfer is marginal. The increase in welfare is positive whenever the average welfare weight in the group that receives the transfer is higher than the group that pays the transfer. The government therefore exploits the information to transfer income from

[12] Sorting is now assumed within each characteristic, $c_k'(z) \geq 0$. Because the government has more instruments, an income tax for each characteristic, it is optimal to violate the standard sorting property over income, while sorting within each characteristic-specific income distribution may hold in the optimum. This also implies that a redistributive government no longer by itself implies monotonically decreasing marginal welfare weights over the (joint) income distribution, $\hat{g}'(z) \leq 0$, as there could be lower costs of redistributing from income groups with a disproportionate number of the one characteristic.

characteristics with lower average marginal welfare weights to characteristics with higher average marginal welfare weights.

Since $g'(c) \leq 0$, characteristics with higher average marginal welfare weights have on average lower consumption levels. This means that aggregate redistribution increases, because the government increases the average consumption of individuals with lower consumption levels more than individuals with higher consumption levels.

To consider the effect on marginal welfare weights, assume without loss of generality that $c_i < c_j$. The government increases consumption for i by m and reduces consumption for j by n. It imposes the transfer whenever $g(c_i+m)(c_i+m)-g(c_j-n)(c_j-n) > g(c_i)c_i - g(c_j)c_j$. Since $g'(c) \leq 0$, redistribution is less valuable when consumption is more equal, $-\left(g'(c_i+m) - g'(c_j-n)\right) \leq \left(g'(c_i) - g'(c_j)\right)$, and the marginal welfare weight schedule becomes flatter over consumption on average.

This has implications for marginal welfare weights over income. By the definition of $\hat{g}(z)$, we observe that $E\left(\hat{g}'(z)\right) = \sum_k p_k E_k\left(g'(c_k(z))c_k'(z)\right)$. Hence, there are two components of marginal welfare weights over income, $\hat{g}'(z)$: the extent of the redistributive motive, $g'(c_k(z))$, and the extent of redistribution, $1 - c_k'(z)$. When the redistributive motive falls, $\Delta\left(-g_k'(c)\right) \leq 0$, and total redistribution increases compared to Case 2, it corresponds to a lower steepness over income, $-\hat{g}'(z)$, on average. $\square$

Hence, it cannot be guaranteed that the welfare weight schedule shifts in a specific way everywhere when the government obtains more information which it exploits in setting taxes. For example, the government may increase redistribution from the high earners to the middle earners, while leaving redistribution from the middle earners to the low earners unchanged. To see this, consider the case of females and males. If all the high earners are male while the middle earners consist of mostly females, the government may increase taxes on high earners while decreasing them on middle earners. Then, the marginal welfare weight on the high earners increase, since they receive lower consumption, while the weight on the middle earners decrease, which may increase the steepness of the welfare weight schedule from the low to the middle earners.

Case 4: No tagging

Again, w_i and l_i are unobservable to the government, while k (and z) is observable. However, information on k is not exploited, since the government chooses not to.

Proposition 4. *If the government is redistributive, not exploiting available information on characteristics (weakly) increases the average steepness of the welfare weight schedule.*

Proof. The government in Case 4 faces the same optimization problem as in Case 2. Since marginal welfare weights in Case 2 are on average steeper than in Case 3, the average steepness is higher also in Case 4. □

I now use these four cases to derive the relation between vertical and horizontal equity.

3.3 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 different theories of justice, such as quasi-utilitarianism (Parfit 1991 and Temkin 1993) or luck-egalitarianism (Arneson 1989, Dworkin 2002 and Roemer 2009). Here one should think of it as the resulting priority on reducing inequality across consumption levels, irrespective of its moral foundation.

$VE(z)$ measures the relative vertical priority at each income level z when the government exploits all information and instruments to reduce inequality. The absolute value of its steepness, $-VE'(z)$, reflects the marginal cost and society's local willingness to pay for VE at a particular allocation. Define also the average marginal cost of vertical equity by $-E\left(VE'(z)\right)$.

Proposition 5. *Vertical equity represents the redistributive motive in Case 2 and 3,* $VE(z) = \hat{g}(z)$*. If* $VE'(c) \leq 0$ *for all* c *and* $VE'(c) < 0$ *for some* c*, the government is redistributive. For a fixed amount of redistribution, a higher average marginal cost of vertical equity means that the welfare weight schedule is on average steeper over income.*

Proof. Consider a government that maximizes welfare (weighted consumption) and exploits all information

$$W^{VE} = \sum_k p_k E\left(g(c_k(z))c_k(z)\right) \tag{19}$$

This government values a marginal increase in consumption by $\partial WE^{VE}/\partial c = g(c) + g'(c)c$. Define

$$VE(c) = g(c) + g'(c)c. \tag{20}$$

To determine the shape of the curve, observe that $VE'(c) = g''(c)c + g'(c)c + g'(c)$. If the steepness nowhere changes too fast $(-g''(c)c \leq -\left(g'(c)c + g'(c)\right))$ and $VE'(c) < 0$, then $g'(c) < 0$, such that when the priority on vertical equity is falling in consumption, then the marginal welfare weights schedule is falling in consumption, and the government is *redistributive*.

Now, redefine marginal welfare weights such that they vary directly over income, $g_k(z)$, then

$$W^{VE} = \sum_k p_k E_k\left(g_k(z)c_k(z)\right). \tag{21}$$

Locally, on average, this government values a marginal increase in consumption at income level z (a first order approximation) by $\sum_k p_k \partial WE^{VE}/\partial c_k = \hat{g}(z)$, such that

$$VE(z) = \hat{g}(z). \tag{22}$$

To measure the redistributive properties of the government from the marginal welfare weight schedule over z, consider that $VE'(z) = \hat{g}'(z)$. Then, the average marginal cost of VE, is $-E\left(\hat{g}'(z)\right)$. There are two components of $\hat{g}'(z)$: the extent of the redistributive motive, $g'(c_k)$, and the extent of redistribution, $1 - c_k'(z)$. If the redistributive motive strengthens, $\Delta\left(-g_k'(c)\right) > 0$, it corresponds to a higher $-\hat{g}'(z)$. If we observe a higher $-\hat{g}'(z)$ and the level of redistribution is not lower, then $-g'(c)$ has increased as well. Hence, the average steepness of the marginal welfare weight schedule over consumption increases with the average cost of VE for a fixed amount of redistribution. □

The vertical equity concern makes it more expensive to give 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, or 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.

Remark. *A (weakly) decreasing vertical equity schedule over consumption cannot alone represent the government in Case 4.*

Proof. When the equity concern is $VE'(c) \leq 0$ for all z and $VE'(c) < 0$ for some z, it implies the level of redistribution in Case 3 with the information structure in Case 3 and 4. In Case 4, by exploiting the information available, consumption inequality could have been reduced further without increasing redistributive costs, which would increase vertical equity. □

The total value of vertical equity is

$$V_{VE} = E\left(\hat{g}(z)\hat{c}(z)\right) - E\left(\bar{c}(z)\bar{c}(z)\right), \tag{23}$$

where $\bar{c}(z) = z - R$ is the no-redistribution consumption function. V_{VE} measures the increase in welfare-weighted consumption from redistribution. The problem is that no redistribution induces negative consumption for many individuals in the presence of an exogenous revenue requirement. To to make the comparison achievable, consider the counterfactual economy where $R = 0$ and compare welfare-weighted consumption with redistribution and without redistribution, such that all taxation is used for redistribution and when there is no redistribution $\bar{c}(z) = z$ (the laissez-faire).

Horizontal equity

Horizontal equity reflects society's aversion to treating individuals with the same circumstance unequally. While discussions on horizontal equity often have centered around who to consider as equals and how to create an aggregate index (see among others Lambert and Ramos 1995 and Auerbach and Hassett 2002), Atkinson (1980) suggests *non-discrimination* as the normative basis for horizontal equity. Inspired by Atkinson, I account for horizontal equity by introducing a constraint that prohibits tagging based on certain characteristics. A constraint on policy does not necessarily respect Pareto efficiency, see Kaplow (1989) on the problems with this. Alternative representations of horizontal equity are possible, see Feldstein (1976) for a tax reformed 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. For the sake of the revealed preference approach presented here, it does not matter much. If the government does not violate Pareto efficiency (there are no Pareto improvements to be made by violating the constraint), then strictly speaking one still 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 the horizontal equity constraint, then arguably the constraint rationalizes a feature of actual tax policy that other representations could not.

The constraint is

$$T_k(z) = T(z)\ \forall k, \tag{24}$$

which imposes that each income level faces the same tax level. If it binds, the horizontal equity constraint makes reaching the government's objective more costly on the margin. Define $HE(z)$ as the Lagrange multiplier associated with the constraint, which measures the shadow price of horizontal equity at each income level z. Its steepness, $-HE'(z)$, reflects the marginal cost of horizontal equity. Define the average marginal cost of horizontal equity as $-E\left(HE'(z)\right)$.

Proposition 6. *The shadow cost of horizontal equity, $HE(z)$, represents the difference in marginal welfare weights between not exploiting information on k (Case 4) and exploiting the information (Case 3), $HE = g(z) - \hat{g}(z)$. Since $HE'(z) \leq 0$ on average, the average cost of horizontal equity is positive for a redistributive government in Case 4.*

Proof. Remember the government that maximizes welfare weighted consumption and exploits all information: $W^{VE} = \sum_k p_k E_k\left(g_k(z)c_k(z)\right)$. Now, with horizontal equity as a constraint, the government maximizes welfare subject to the constraint $T(z,k) =$

$T(z)$. The constraint can be added to a new (Lagrangian) social welfare function with a loss function that accounts for the constraint

$$W^{HE} = \sum_k p_k \left[E_k \left(g_k(z) c_k(z) \right) - E \left(HE(z) \left(T_k(z) - T(z) \right) \right) \right], \tag{25}$$

and this function will be associated with a new set of marginal welfare weights, $g(z)$, in the optimum

$$W^{HE} = E \left(g(z) c(z) \right). \tag{26}$$

Now, consider how this government values a marginal increase in consumption at income level z

$$\frac{\partial WE^{HE}}{\partial c} = \hat{g}(z) + HE(z) = g(z), \tag{27}$$

such that

$$HE(z) = g(z) - \hat{g}(z). \tag{28}$$

This is the extra cost of redistribution at income z imposed by the horizontal equity concern. To determine the shape of the curve, consider $HE'(z) = g'(z) - \hat{g}'(z)$, the difference in steepness between the two curves. By Proposition 3, $\hat{g}'(z) \leq g'(z)$ on average, such that $HE'(z) \leq 0$ on average, and the average marginal cost is positive, $-E(HE'(z)) \geq 0$.[13] □

The total cost and minimal valuation of horizontal equity is

$$C_{HE} = E \left(\hat{g}(z) \hat{c}(z) \right) - E \left(g(c(z)) c(z) \right), \tag{29}$$

which is the loss in weighted consumption from choosing not to tag. It measures the weighted loss due to horizontal equity, corresponding to the minimal value the government has to place on horizontal equity to rationalize no tagging, as the valuation may be higher than the current cost. This is not the case for vertical equity, as the government always could have redistributed more, while the government cannot impose more than perfect horizontal equity with respect to the tag considered.

[13] The approach resembles Negishi (1960), which supports different Pareto optimal allocations as equilibria. The difference here is that the redistributive preferences of the government adapt to the allocation, such that status quo redistribution is not imposed.

Relationship between vertical and horizontal equity

Proposition 7. *For a government that is concerned with efficiency, vertical equity and horizontal equity, marginal welfare weights at each income level z can be decomposed as*

$$g(z) = VE(z) + HE(z). \tag{30}$$

If one does not account for horizontal equity, the average willingness to pay for vertical equity is overestimated.

Proof. The decomposition follows immediately from Proposition 5 and 6. The local willingness to pay for vertical equity is $-VE'(z) = -g'(z) + HE'(z)$. By Proposition 7, $HE'(z) \leq 0$ on average, such that $-VE'(z)$ is lower on average than when horizontal equity is ignored. □

This shows 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). However, Proposition 7 establishes that horizontal equity drives a wedge between the cost measure and the willingness to pay interpretation. The reason is that part of the cost of redistribution reflects the willingness to pay for horizontal equity rather than for vertical equity.

Inequality aversion

Inequality aversion is intimately linked to vertical equity, but there is no one-to-one relationship. There are many ways in which to measure inequality aversion. Here, it is measured by the average value of the steepness of the marginal welfare weights over consumption

$$IA = -\sum_k p_k E_k \left(g'(c_k(z)) \right) = -\sum_k p_k E_k \left(\frac{g'_k(z)}{c'_k(z)} \right). \tag{31}$$

This corresponds to the definition of inequality aversion in Section 2, but with continuous types.

Remark: *The sufficient statistic for the bias to inequality aversion from not accounting for horizontal equity, b, is*

$$b = E\left(g'(c(z))\right) - \sum_k p_k E_k\left(g'(c_k(z))\right). \tag{32}$$

Ignoring horizontal equity implies $b > 0$ for an inequality averse government.

Proof. The sufficient statistic follows immediately from the definition of inequality aversion. With constant level of redistribution, $c'(z) = c'_k(z)$, such that $b = \left(E\left(g'(z)\right) - E\left(\hat{g}'(z)\right)\right)/c'(z)$. By Proposition 3, $-E\left(\hat{g}'(z)\right) > -E\left(g'(z)\right)$ and $c'(z) > 0$, such that $b > 0$. Hence, the level of inequality aversion is overestimated when horizontal equity is ignored and redistribution is constant. By Proposition 3, tagging increases total redistribution such that c is more evenly distributed on average. This means that the average steepness of welfare weights over consumption is lower on average with tagging, and that $b > 0$ when horizontal equity is ignored. □

To illustrate the bias to inequality aversion from ignoring horizontal equity, assume quasi-linear utility, $u_i = c_i - v(l_i)$ and that social welfare function exhibits constant relative inequality aversion in consumption $SWF = E\left(W(c(z))\right)$ with $W(c(z)) = c(z)^{1-\gamma}/\left(1-\gamma\right)$, where γ is the inequality aversion parameter (or, equivalently, that $W(c(z)) = u\left(c(z), l(z)\right)$ and $u = c^{1-\gamma}/\left(1-\gamma\right)$). Then, from

$$\gamma = -\frac{log(g(z))}{log(c(z))}\ \forall z, \tag{33}$$

one obtains the inequality aversion parameter. However, without tagging, inequality aversion is measured in a different optimum, and the optimum reflects the priority on horizontal equity. One can measure the bias to the inequality aversion parameter as the difference between γ, in the case without tagging, and $\hat{\gamma}$, in the case with tagging,

$$b = \hat{\gamma} - \gamma = \frac{log\left(g(z)\right)}{log(c(z))} - \frac{log\left(\hat{g}(z)\right)}{log(\hat{c}(z))}. \tag{34}$$

The intuition can be illustrated in the simpler case where redistribution stays constant. Consider a hypothetical tag that increases average consumption by the same amount at all income levels,[14] but at the same time, reduces the cost of redistribution,

[14] Specific individuals may still lose in terms of consumption, but the tag is designed such that each income level on average neither gains nor loses compared to other income levels.

such that the welfare weight schedule is flatter. Not all tags can achieve this, but the point is valid as long as such tags are feasible in principle, which they are, for example in the case of a Pareto improving tag (see more on the relation between Pareto improvements and tagging in Ziesemer (2019)). Then, $\hat{g}(z)$ changes while c increases equally for all, and the level of absolute inequality stays the same. $VE'(z)$ measures the local willingness to pay for vertical equity, and since redistribution is cheaper and inequality stays the same, the local willingness to reduce inequality must fall on average, such that inequality aversion also decreases.

More generally, vertical equity and inequality aversion are overestimated also when redistribution changes if horizontal equity is ignored. How to estimate the extent of the bias is addressed in Section 4.

3.4 Types of governments

To demonstrate the relation between marginal welfare weights and different types of governments, I connect to the discussion in Saez and Stantcheva (2016) for my decomposition of the marginal welfare weights into vertical and horizontal equity components.

A *libertarian* government does not value reductions in inequality across income levels, $VE(z) = 0$. If it must raise revenue, taxes are the same for all, $T(z) = R$. Then, information on tags is redundant, and the government obtains horizontal equity at no cost, such that $HE(z) = 0$ and $g(z) = 1$.

A *utilitarian government* is assumed in the traditional optimal tax literature. It sets taxes T_i in order to maximize the sum of equal concave transformations of individual (homogeneous) utility:

$$\max_{T_i} W = \int_i G\left(u(c_i, l_i)\right) di, \tag{35}$$

where $G(u(c_i, l_i)$ is a concave transformation of individual utility $u(c_i, l_i)$. This government respects Pareto-efficiency and can be inequality averse in consumption through the concavity of G in u or u in c. The marginal welfare weight is

$$W'(c) = G'(u(c_i, l_i)u'_c(c_i, l_i) = g(c). \tag{36}$$

When a constrained utilitarian government sets taxes, it corresponds to $VE(z) \neq 0$ for some values of z, due to concave utility functions $u_c''(c, l) < 0$ and/or concave transformations of utilities $G''(u(c, l)) < 0$. The utilitarian government fully exploits tags, such that the government in Case 4 cannot be utilitarian. It corresponds to $HE(z) = 0$, and the marginal welfare weights, $\hat{g}(z) = VE(z)$, therefore reflect only vertical equity. This government is represented by Case 3.

A constrained *inequality averse and horizontal equity-respecting* government also sets taxes that correspond to $VE(z) \neq 0$ for some values of z. However, this government does not exploit tags, such that $HE(z) \neq 0$ for some values of z, and inverse optimum marginal welfare weights reflect both vertical and horizontal equity: $g(z) = VE(z) + HE(z)$. This government is represented by Case 4 and arguably represents the preferences of actual governments.[15]

4 Optimal taxation with and without tagging

Section 4 provides the theory to quantify the importance of horizontal equity for inequality aversion. This quantification requires estimates of marginal welfare weights in the cases with and without tagging, $g(z)$ and $\hat{g}(z)$, respectively. I now provide the theory and methods to reveal marginal welfare weights for the actual and counterfactual tax system. The innovation is to develop a method to consider non-local policy changes by adding structure to how marginal welfare weights adapt to changes in allocations. 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 new allocation with tagging.

I initially adopt the tax reform approach to optimal taxation (Saez 2001). The government is fundamentally the same as the one introduced in Section 3, but I further specify the optimal taxation problem here. Assume everyone works (excluding extensive margin responses), no income effects and no exogenous revenue requirement,

[15]Political concerns (such as for re-election) may also affect government policy, but are not accounted for in this framework. Alternatively, the horizontal equity constraint can be interpreted as a political constraint on the tax system, but for this interpretation it is important that the constraint is not a fundamental feature of the economy, as the equity constraint cannot be unavoidable for the government.

$R = 0$. The behavioral response to taxes may differ across characterstics, but I assume that it is constant within each characteristic $\varepsilon_k(z) = \varepsilon_k$ for all k. The government faces the budget requirement

$$R = \sum_k p_k E_k \left(T_k(z) \right) = 0, \tag{37}$$

and the structure of the tax system is

$$T_k(z) = t_k(z) + R_k, \tag{38}$$

where $T_k(z)$ is the total nonlinear tax for each characteristic, separated into lump sum transfers R_k and income-dependent taxes $t_k(z)$. It appears like 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. This is achieved by deriving the non-linear within-characteristic tax schedule and then the optimal transfers.

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

$$dR_k = d\tau_k dz \left(1 - H_k(z) - h_k(z) \varepsilon_k \frac{T_k'(z)}{1 - T_k'(z)} \right), \tag{39}$$

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 that 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

$$dW_k = dR_k \sum_k p_k E_k \left(g_k(z) \right) - d\tau_k dz \int_{z > z_i}^{\infty} g_k(z) h_k(z) dz = 0. \tag{40}$$

Combining Equation 39 and 49 (applying Saez (2001) without income effects), the within-characteristic optimal tax rate is

$$T_k'(z) = \frac{1 - \bar{G}_k(z)}{1 - \bar{G}_k(z) + \alpha_k(z) \varepsilon_k} \tag{41}$$

where $\alpha_k(z) = zh_k(z)/(1 - H_k(z))$ is the characteristic-specific local Pareto parameter and $\bar{G}_k(z) = \int_{z \geq z_i}^{\bar{z}} g_k(z) h_k(z) dz / (1 - H_k(z))$ 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. It is simply to solve Equation 41 for $g_k(z)$. Assuming that $T(z)$ can be approximated by a piece-wise linear tax system, the marginal welfare weights from the inverse optimal problem are given by

$$g_k(z) = 1 - \frac{T_k'(z)}{1 - T_k'(z)} \rho_k(z) \varepsilon_k, \tag{42}$$

where $\rho_k(z) = -(1 + zh_k'(z)/h_k(z))$ is the characteristic-specific "elasticity of the income distribution" (Bastani and Lundberg 2017). It measures how the characteristic-specific income distribution locally is changing with income.

4.1 Marginal welfare weights with no tagging

For the case without tagging, $T_k(z) = T(z)$, inverse optimum marginal welfare weights are simply given by

$$g(z) = 1 - \frac{T'(z)}{1 - T'(z)} \rho(z) \varepsilon(z), \tag{43}$$

where $g(z)$, $T'(z)$, $\rho(z)$ and $\varepsilon(z)$ now are defined over the joint income distribution. The behavioral response $\varepsilon(z)$ may vary over the joint income distribution due to differences in composition of characteristics across the 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.

4.2 Marginal welfare weights with tagging

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 with tagging 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 welfare weight schedule.[16] Consider a transfer m to individuals at income level z.

Proposition 8. *Ceteris paribus, a redistributive government's new welfare weight schedule with a transfer m to income level z can be obtained from the original welfare weight schedule by the relation*[17]

$$\tilde{g}(z) = g\left(z + c^{-1}(m)\right). \tag{44}$$

Proof. I have already assumed that marginal welfare weights only depend on consumption levels and not on particular individuals or characteristics (anonymity), $g_i(z_i)$ $= g(c_i(z_i))$. By *separability* between individuals in the underlying social welfare function, the difference between the consumption of individual i and consumption of individual j decides the relative weight on i compared to j, and is independent of individual h's consumption. By no income effects, transfers do not directly affect income. There is initially no difference in the relation between consumption and income across individuals, $c(z)$ for all z.

Without loss of generality, assume $c_h < c_i < c_j$ with weights $g(c_h) > g(c_i) > g(c_j)$. Now, individual h receives a transfer $m = c_i - c_h$, such that h obtains the same consumption as i. The after-transfer welfare weight on income level z is $\tilde{g}(z)$. The transfer leaves the relative marginal welfare weight of i and j unchanged (by separability). To consider welfare weights over the income distribution, observe that m corresponds to the same consumption increase as an increase in income equal to $c^{-1}(m)$. Now, by $g_i(z) = g(c_i(z))$, h's new marginal welfare weight must be equal to i's, which results in the welfare weight $\tilde{g}(z_h) = g(z_h + c^{-1}(m))$ for a transfer m to individual h earning income z_h. □

[16] It resembles the distinction in Basu (1980) between the local and global social welfare function, such that my approach is "less local" than the standard inverse optimum approach and the local social welfare function.

[17] When ignoring 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). I later present an algorithm that accounts for these factors.

The condition relates the current welfare weights over income to new welfare weights with transfers. It exploits that marginal welfare weights only depend on consumption and that individuals are weighted equally given their consumption, such that the weight attached to an individual that receives a transfer is the same as an individual who receives the same consumption by earning higher income.[18] For example, when the income tax is flat, $T(z) = tz$, the inverse consumption relation simplifies to $c^{-1}(m) = m/tz$. Then, if income is taxed at 50 percent, the new welfare weight for an individual at income level z that receives a transfer equal to 10 percent of income is the same as the welfare weight of an individual with 20 percent higher income before transfers were introduced. The relation relies on the local stability of marginal welfare weights, which will not hold for non-local policy changes such as the introduction of tagging. The algorithm I present now addresses this issue.

Between-characteristics transfers

The characteristic-specific marginal income tax, $T_k'(z)$, affects within-characteristic income distributions through behavioral responses. Even though transfers do not directly affect the pre-tax income distribution, they still affect the marginal welfare weights over the income distribution by changing each characteristic's consumption level. To measure the effect of tagging on the welfare weight schedule, exploiting current marginal welfare weights, assume that there are no transfers that differ across characteristics prior to tagging.[19]

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 $\not{k}$

$$dW = dmE_k\left(g(c_k(z))\right) - dmE_{\not{k}}\left(g(c_{\not{k}}(z))\right) = 0\ \forall k. \tag{45}$$

This implies setting transfers such that the average welfare weight on each characteristic is equal, because if not, the government could increase total (weighted) welfare

[18]This updating of the welfare weight schedule is a natural way to account for other transfers and taxes too.

[19]Any transfer that does not affect income and is equal across characteristics in the actual tax system will have no effect on the relation between $g(c_k(z))$ and $g_k(z)$ and is therefore irrelevant here.

by changing transfers such that $E_k\left(g(c_k(z))\right) = \bar{g}$ for all k. We observe that an updating relation for welfare weights is necessary to make sense of the requirement that the transfer from tagging should equalize average marginal welfare weights, since if the transfer did not affect marginal welfare weights the condition could never be satisfied (which implies that the first-order approach in the standard local approach to optimal taxation is not applicable to this problem).

Since the sole impact of the transfer is to increase or reduce individuals' consumption, there is no direct effect on (pre-tax) income distributions, $h_k(z)$. The key relation is stated in Proposition 8, such that I obtain the initial estimate $g_k(z) = g\left(z + c^{-1}(m_k)\right)$. Depending on the transfer, some characteristics' average consumption increase and others' decrease. Marginal welfare weights are still equal for all characteristics given the same consumption level (by assumption), while they now differ for the same income level. The algorithm that solves the problem is then:

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:

$$g(z) \rightarrow m_k \rightarrow g_k(z) \rightarrow T'_k(z) \rightarrow h_k(z) \rightarrow m_k \rightarrow ... \rightarrow \hat{g}(z).$$

The key endogenous variables are $h_k^t(z) = h_k(z + \Delta_k z)$ with $\Delta_k z \approx \epsilon_k (z/(1 - T_k'(z)) \Delta T_k'(z)$ and $g_k^t(z) = g(z + c_k^{-1}(\Delta m_k))$, 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, and any weights implied by the optimal transfer would imply within-characteristic optimal tax rates. Unfortunately, as is often the case for optimal tax algorithms, the algorithm may not converge if the effect on welfare weights from the transfer is too large or if the behavioral response to taxes are too large. It turns out to work in the applications presented here.

5 Application: Gender tag in Norway

The main application is an hypothetical experiment of introducing a gender tag in the Norwegian tax system. I also apply the model to immigration status and age group tags, see the results Appendix C.

5.1 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 2015 (Statistics Norway 2005). The main analysis is for wage earners in the year 2010. I exclude individuals that 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, $\approx$ 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.

Table 1: Summary statistics for main variables in year 2010

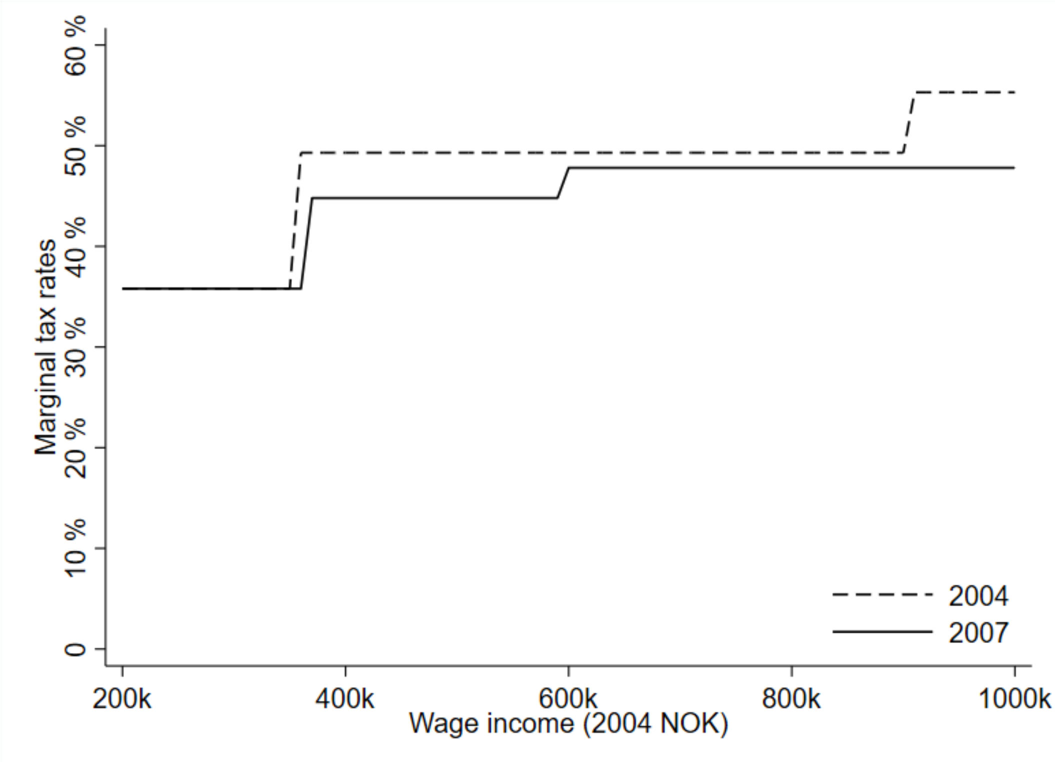
	Mean	Standard deviation
Wage income	541432.6	329576.9
Age	46.8	7.3
Share of males	57.4 %	
Share born in Norway	94.0 %	
Share with children	67.3 %	
Share married	61.0 %	
N	787722	

5.2 Tax system

The Norwegian tax system applies different tax rates to different types of incomes. Together with the other Nordic countries, it was characterised by a dual tax system with flat and relatively low rates on capital income combined with a progressive income tax schedule on labor earnings. More specifically, it was a combination of 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 individual tax rates, I employ the LOTTE tax-benefit calculator (Hansen et al. 2008).[20] 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. The resulting average marginal tax schedule over the income distribution is shown in Figure 3.

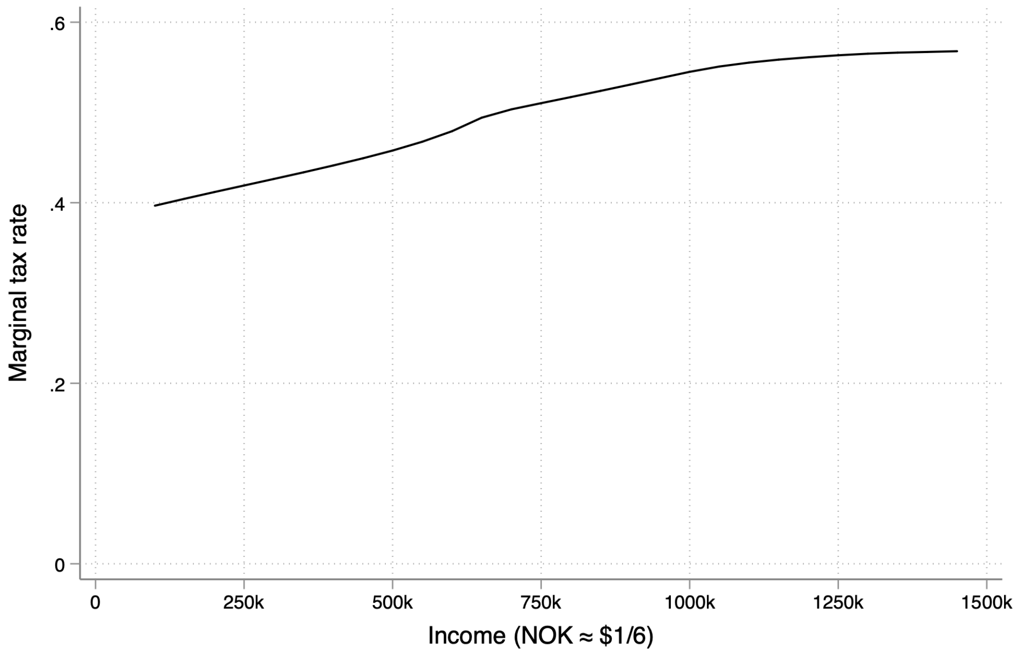
[20] I thank Bård Lian for assistance with the tax-benefit simulator.

Figure 2: The 2006 tax reform



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

Figure 3: Total marginal tax rates



Notes: Including VAT and income-dependant transfers for wage earners in 2010.

5.3 Elasticity of taxable income

The optimal tax rate depends on how individuals respond to tax changes. Since 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

$$\varepsilon(z) = \frac{(1-\tau)}{z}\frac{\partial z}{\partial(1-\tau)}. \tag{46}$$

In my setup, z is not total individual taxable income, but income for individuals who primarily obtain income from wage earnings. Since the Norwegian tax system is not comprehensive, different types of income face difference tax rates and my model does not address the optimal tax of different types of income, see Hermle and Peichl (2018) and Lefebvre, Lehmann, and Sicsic (2019) on how to account for different income types in optimal taxation.

There is a large literature estimating ETIs and estimates differ widely across countries (see the survey in Saez, Slemrod, and Giertz (2012)). Most comparable to the approach and setting here, Kleven and Schultz (2014) estimate ETIs in Denmark and obtain a response for wage earnings around 0.05. This is also similar to what Thoresen and Vattø (2015) find for wage earners in Norway, exploiting the same tax reform as here.

The difference is that I account for heterogeneity in tax responses across immutable observable characteristics. Here that is to estimate ETIs separatly for each gender. I estimate the ETI using using a standard difference panel data approach with a Weber (2014) style instrument and a Kopczuk (2005) type mean reversion control. See Table A2 in the Appendix for summary statistics for the "treatment" and "control" groups in the estimation of the elasticity of taxable income. 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, see Figure 2. The estimating equation is

$$\begin{aligned}\Delta_3 \log\left(z_{i,t}\right) =&\alpha_t + \beta D_k \Delta_3 \log\left(1-\tau_{i,t}\right) + \theta \log\left(z_{i,t}\right) + \pi \Delta_1 \log\left(z_{i,t-1}\right) \\ &+ \eta M'_{i,t} + \epsilon_{i,t},\end{aligned} \tag{47}$$

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 $\Delta_3 \log\left(1-\tau_{i,t}\right)$ is instrumented by the tax rate change that would have occurred had income stayed constant $\log\left(1-\tau_{i,t+3}\right) - \log\left(1-\tau^I_{i,t}\right)$, 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 $\log\left(z_{i,t}\right)$ and $\Delta_1 \log\left(z_{i,t-1}\right)$ are introduced as bias corrections (Kopczuk 2005).

The resulting estimates are shown in Table 2. Although the estimates are small compared to the US literature, the key point here is that females respond about twice as much to the reform than males.[21]

[21] This does not speak to why females and males respond differently. In a robustness (Table A5 in the

Table 2: ETI estimates

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

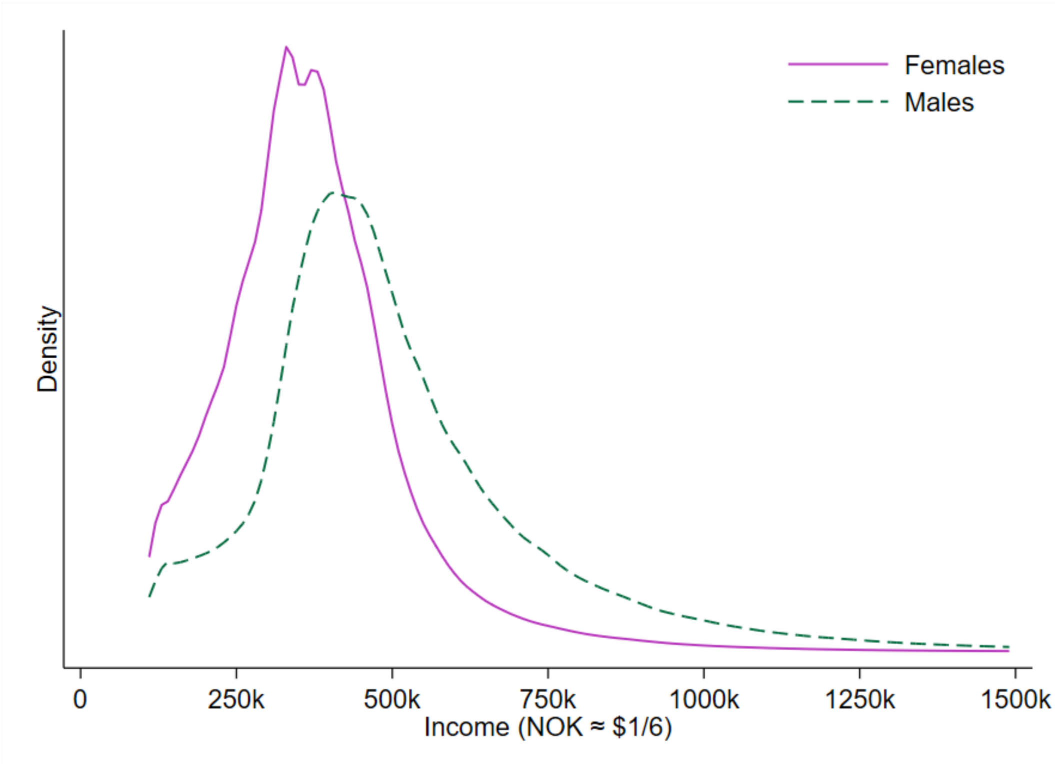
Notes: ETI estimates, average and separated by gender for wage earners. The estimation is a first-difference equation where the tax rate change is instrumented by the reform-induced tax rate change. Other controls are year-dummies, the log of base year income, the first-difference of the log of income in the year prior to the base year, educational field, educational level, family status, county of residence, age and gender. See Table A1 for more detailed results.

5.4 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 to apply 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 4 presents the Kernel estimates for the female and male income distributions, while Figure 5 shows the elasticity of the joint income distribution, ρ. Figure A2-A4 in the Appendix further describes the income distributions.

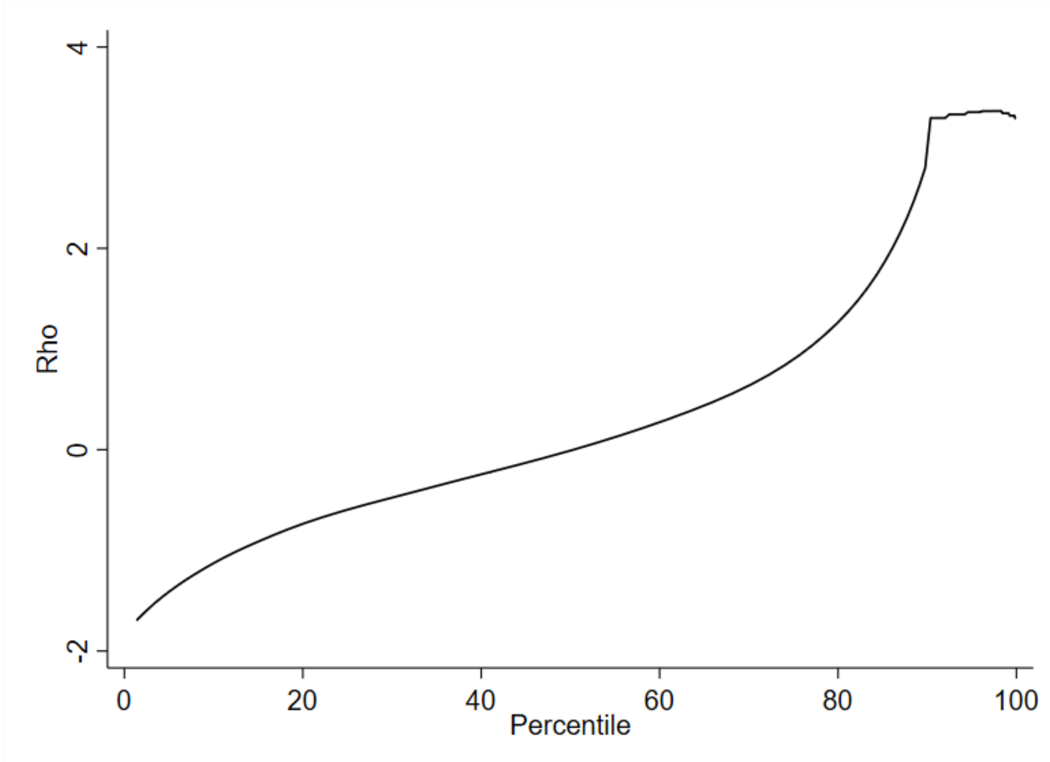
Appendix), I have estimated responses separately for the single and married, and the relative difference in response between females and males is equally large. The response among single females appears to be larger than among married females, although the difference is not statistically significant. My speculation is that the difference in tax response is driven by labor market characteristics and career choices. More females than males work part time, especially in health care, and this makes it possible for females to respond to tax changes. For full-time workers in Norway, the margins on which to respond to tax changes are more limited due to restrictions in working hours.

Figure 4: Income distributions by gender



Notes: Adaptive kernel estimates of the female and male income distributions for wage earners in Norway in 2010.

Figure 5: Elasticity of the joint income distribution



Notes: Local elasticity of the income distribution estimates derived from the adaptive kernel estimate of the joint income distribution for wage earners in Norway in 2010.

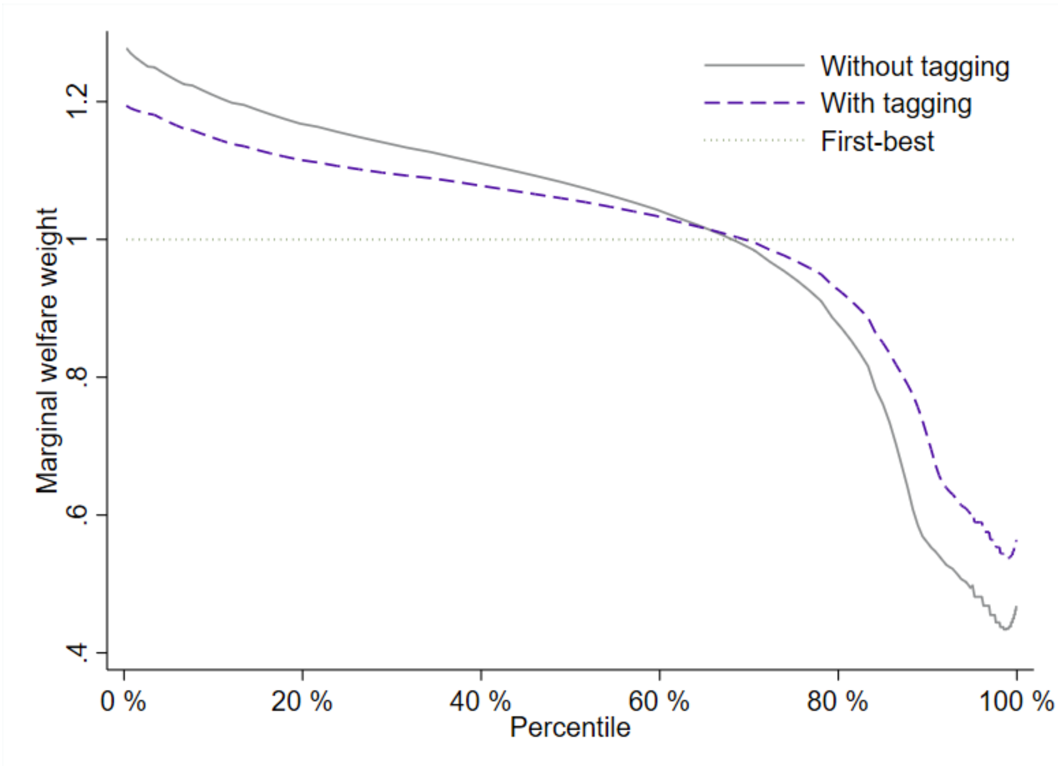
5.5 Marginal welfare weights and equity measures

Using the results above, Figure 6 presents marginal welfare weights with tagging $g_k(z) = 1 - T_k'(z)\,(1 - T_k'(z))\,\rho_k(z)\varepsilon_k$, averaged over characteristics at each income level to obtain $\hat{g}(z)$, and without tagging $g(z) = 1 - T'(z)/\,(1 - T'(z))\,\rho(z)\varepsilon(z)$.

In line with Proposition 4, tagging decreases the average steepness of the welfare weight schedule. It also shows that a gender tag would have a visible effect on the welfare weight schedule. Tags that reveal more of an individual's productivity would imply even larger differences. Figure 7 presents the decomposition of marginal welfare weights into the contribution from vertical and horizontal equity. VE is scaled as a deviation from first-best welfare weights, $VE(z) = \hat{g}(z) - 1$, to compare the relative contribution of each form of equity. The steepness of the inverse optimum welfare weights from the actual tax system reflects both the contribution from the vertical equity concern and horizontal equity. Horizontal equity is particularly important at upper and lower points of the income distribution. The reason is that the steepness of the marginal welfare weight schedule from the actual tax system is higher in these parts of the income distribution. Horizontal equity contributes in the same direction as vertical equity over the whole income distribution. Hence, if horizontal equity is ignored here, the contribution from vertical equity is overestimated in all parts of the income distribution.

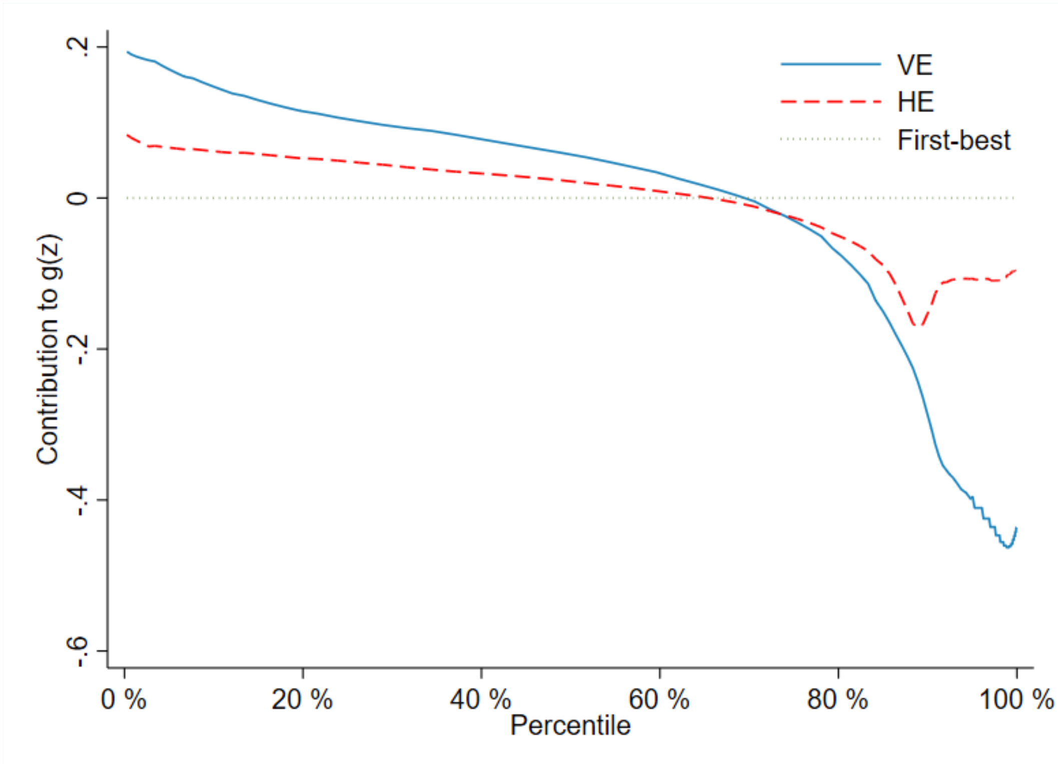
In Figure 8, the bias is estimated by the absolute value of the difference in marginal welfare weights relative to the actual marginal welfare weights, which measures the relative size of the bias to the measure of vertical equity by ignoring horizontal equity. The total difference in steepness between $g(z)$ and $\hat{g}(z)$, which measures aggregate total bias to vertical equity, $-E\left(g'(z) - \hat{g}'(z)\right)$, is 32 percent of the average steepness in the actual tax system, $-E\left(g'(z)\right)$. Then, if inequality aversion is measured by the average steepness, it is overestimated by 32 percent by ignoring the concern for horizontal equity. Appendix B presents optimal taxes by gender, showing that males on average face about 20 percentage points higher marginal tax rates than females, mainly due to the large difference in taxable income elasticities. Another illuminating comparison in the relative marginal welfare weight at the different income levels. In the actual tax system, society is indifferent between \$100 to an individual with income at the 90th percentile and \$63 to an individual with income at the 10th percentile. In the tax system with tagging, society is indifferent between \$100 to an individual with income at the 90th percentile and \$75 to an individual with income at the 10th percentile. Hence, the priority on vertical equity implies a relative weight of 1.34, while including the priority on horizontal equity increases the relative weight to 1.59.

Figure 6: Marginal welfare weights with and without tagging



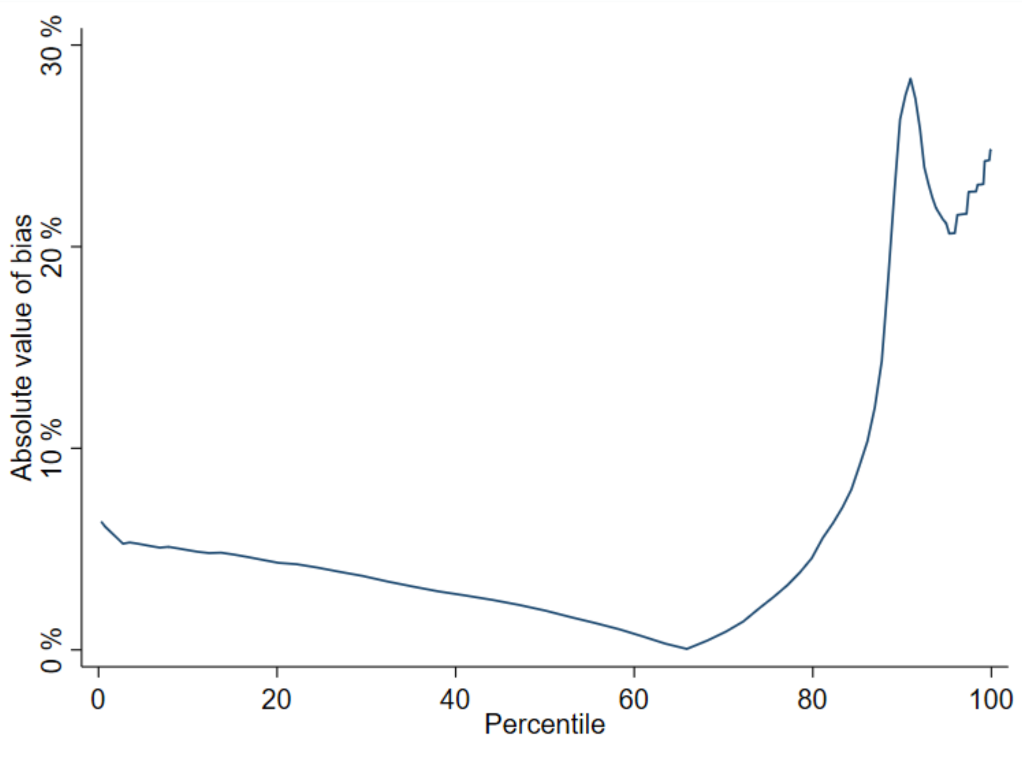
Notes: Inverse optimum marginal welfare weights without tagging, $g(z)$, and with tagging $\hat{g}(z)$ over the income distribution for wage earners in 2010.

Figure 7: Contribution of vertical and horizontal equity



Notes: The contribution of VE and HE to inverse optimum marginal welfare weights for the actual tax system, $g(z)$, over the income distribution for wage earners in 2010. $VE = \hat{g}(z) - 1$ and $HE(z) = g(z) - \hat{g}(z)$.

Figure 8: The bias to VE from ignoring HE



Notes: The absolute value of the relative difference in VE at each level of the income distribution if HE is ignored: $|(g(z) - \hat{g}(z))/\hat{g}(z)|$.

6 Conclusion

Governments do not exploit all the relevant available information when setting taxes. This cannot be explained by standard (utilitarian) 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 standard measurement of vertical equity. Since horizontal equity increases the cost of redistribution, standard inverse optimum marginal welfare weights overestimate the role of vertical equity in supporting the current tax system. In the empirical application to gender neutral taxation in Norway, I estimate that, by one measure, implicit inequality aversion is overestimated by about 30 % when horizontal equity is ignored. More generally, it shows that the instruments governments employ to reduce inequality (such as tagging or not tagging), matter for how redistributive one should consider their tax policy to be.

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.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. “Intergenerational mobility and preferences for redistribution”. *American Economic Review* 108 (2): 521–54.

Arneson, Richard J. 1989. “Equality and equal opportunity for welfare”. *Philosophical Studies* 56 (1): 77–93.

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, Washington.

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. 2014. "Tax-benefit revealed social preferences in Europe and the US". *Annals of Economics and Statistics*: 257–289.

Bastani, Spencer. 2013. "Gender-based and couple-based taxation". *International Tax and Public Finance* 20 (4): 653–686.

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.

Basu, Kaushik. 1980. *Revealed Preference of Government*. CUP Archive.

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.

Dworkin, Ronald. 2002. *Sovereign Virtue: The Theory and Practice of Equality*. Harvard University Press.

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. 2006. "Fair income tax". *The Review of Economic Studies* 73 (1): 55–83.

— . 2018. "Optimal income taxation theory and principles of fairness". *Journal of Economic Literature* 56 (3): 1029–79.

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

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

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

Hermle, Johannes, and Andreas Peichl. 2018. "Jointly optimal taxes for different types of income".

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* X (X).

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

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.

Lambert, Peter, and Xavier Ramos. 1995. "Vertical redistribution and horizontal inequity". *Institute for Fiscal Studies Working Paper*.

Lefebvre, Marie-Noëlle, Etienne Lehmann, and Michaël Sicsic. 2019. "Comprehensive or separate income tax? A sufficient statistics approach". *Working paper*.

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.

McFadden, Daniel. 1975. "The revealed preferences of a government bureaucracy: Theory". *The Bell Journal of Economics*: 401–416.

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.

Negishi, Takashi. 1960. "Welfare economics and existence of an equilibrium for a competitive economy". *Metroeconomica* 12 (2-3): 92–97.

Parfit, Derek. 1991. *Equality or priority*. University of Kansas, Department of Philosophy.

Piacquadio, Paolo Giovanni. 2017. "A fairness justification of utilitarianism". *Econometrica* 85 (4): 1261–1276.

Roemer, John E. 2009. *Equality of opportunity*. Harvard University Press.

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.

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*.

Stiglitz, Joseph E. 1982. "Self-selection and Pareto efficient taxation". *Journal of Public Economics* 17 (2): 213–240.

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.

Ziesemer, Vinzenz. 2019. "Why we do not tag to tax: Pareto optimality versus horizontal equity in the age of data". *Working paper*.

A Equivalent consumption formulation

An alternative approach to the one in Section 3 is to assume that the government assigns the same marginal welfare weight to the same *equivalent consumption levels*, accounting also for individuals' different labor supply levels, rather than just their consumption levels (Fleurbaey and Maniquet 2006 and Piacquadio 2017). This requires choosing a specific utility function and assuming the government has information on labor supply to use in assigning marginal welfare weights, but which it cannot exploit in setting tax rates. This is not entirely implausible, as some countries have register data on working hours (which is the case for Norway, even though the data are imperfect), but exploiting these in setting taxes is not incentive compatible if individuals can easily manipulate their reported labor supply.

A.1 Equity principles

Assume the government knows the characteristic-specific utility functions $u_k(c_i, l_i)$. Hence, equivalent consumption, e_i, is the consumption level combined with a fixed labor supply $\tilde{l}$ that makes the individual as well off as with their actual consumption and labor supply, $u_k(c_i, l_i) = u_k\left(e_i, \tilde{l}\right)$. The relevant sorting property is $\partial u_k(c_i, l_i)/\partial z_i \geq 0$, which also implies $\partial e_i(c_i, l_i)/\partial z \geq 0$. A redistributive government has $g'(e) \leq 0$ for all e and $g'(e) < 0$ for some e.

All main results (Proposition 1-8) hold for any equivalent consumption representation with no income effects, with e in place of c. The proofs are equivalent to the ones in Section 3 and 4. The key difference is the information requirement, as the equivalent consumption formulation requires that the government knows the characteristic-specific utility functions and each individual's labor supply, since the marginal welfare weight is $g(e_k(z))$. I do not expect the difference in results between the consumption and equivalent consumption formulations be large, mainly because variations in working hours are limited in Norway.

B Summary statistics and detailed results

For the purpose of the summary statistics and visualizing the difference-in-difference strategy, 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. In the elasticity estimation by regression, all variations in tax rates and income levels are exploited.

Table A1: Income over time for treated and control

	Wage income treated	Wage income control
2001	349126.7	238619.1
2002	374352.9	253870.1
2003	394173.2	263546.3
2004	414865.8	270962.9
2005	431045.4	285201.5
2006	450112.8	302176.8
2007	483712.6	324189.6
2008	519275.7	349745.5
2009	537719.8	365800.2
2010	556762	380328.8
N	22,081	110,880

Figure A1: Tax treatment

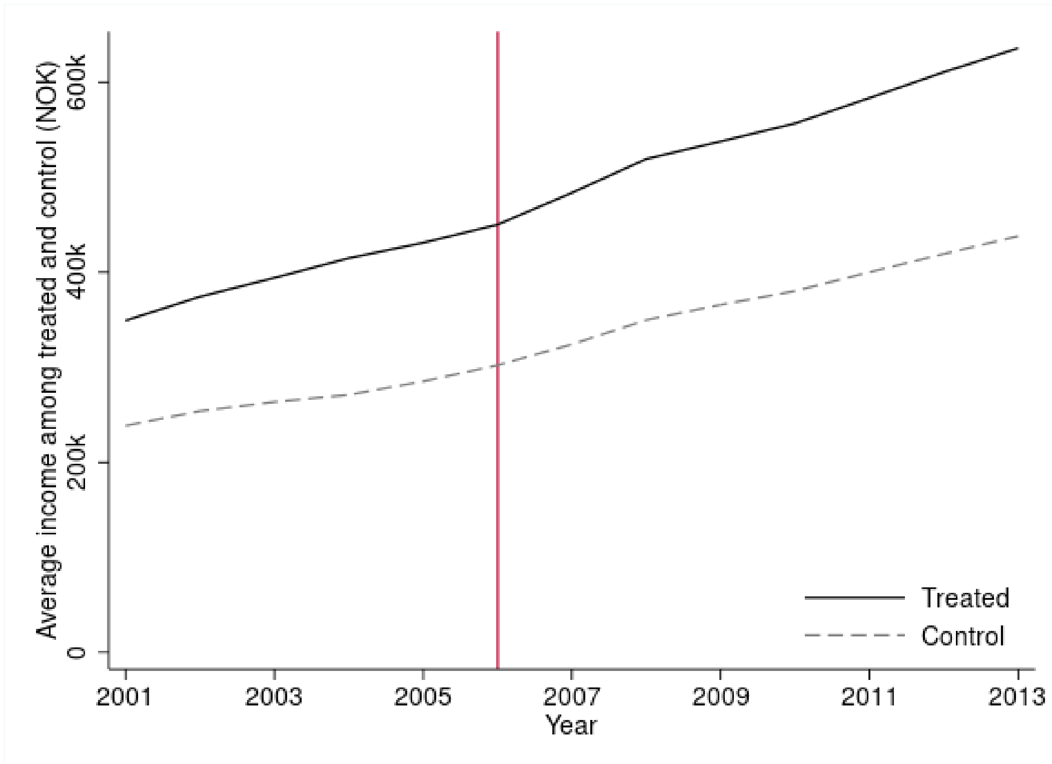


Table A2: Summary of treatment and control groups

	Mean	
	Treated	Control
Age	40.9	40.4
Male	67.7 %2	36.1%
Born in Norway	94.5%	93.7 %
Children	69.7 %	71.9 %
Married	56.5 %	56.7 %
N	22,081	110,880

Table A3: Summary of treatment and control groups by gender

	Mean			
	Males		Females	
	Treated	Control	Treated	Control
Age	40.4	39.5	41.8	40.9
Born in Norway	94.7%	92.5 %	93.9%	94.4 %
Children	70.0 %	63.8 %	69.2 %	76.6 %
Married	57.7 %	49.7 %	54.0 %	60.6 %
N	14,887	38,707	7,101	68,387

Table A4: ETI estimates by gender

Sample	Full	Male	Female
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Taxable income for wage earners is the dependent variable. 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, educational level and county of residence.

Table A5: ETI estimates by gender and marital status

Sample	Single male	Married male	Single females	Married female
Tax treatment	0.062***	0.043***	0.121***	0.097***
	(0.004)	(0.005)	(0.005)	(0.006)
N	1,362,246	1,347,980	989,143	1,023,727

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Taxable income for wage earners is the dependent variable. 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, educational level, age, birth country, children and county of residence.

Table A6: ETI estimates interacted by age and gender

Sample	Younger males	Older males	Younger females	Older females
Tax treatment	0.094***	0.027***	0.223***	0.036***
	(0.005)	(0.004)	(0.007)	(0.004)
N	870,069	1,721,265	574,431	1,358,595

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Taxable income for wage earners is the dependent variable. The young are from 25 to 40 years old and the older from 41 to 64 years old. 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, educational level, birth country, children, marital status and county of residence.

Figure A2: The joint income distribution

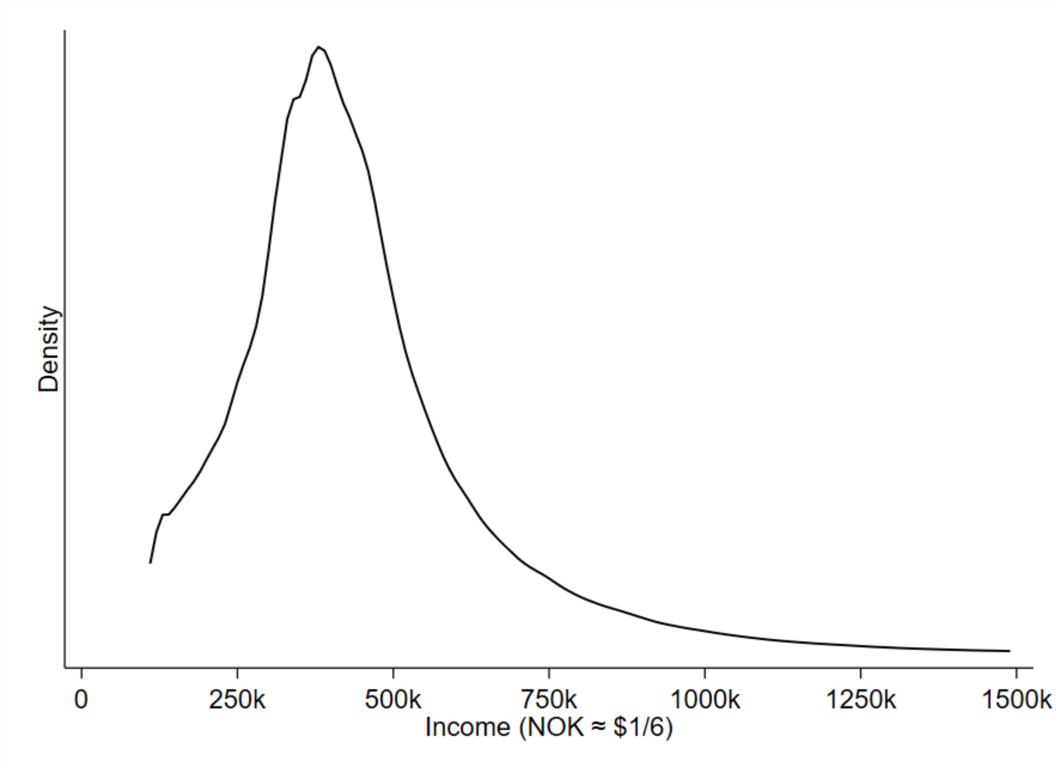
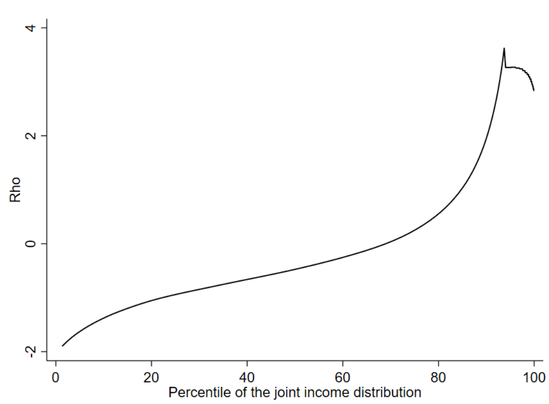
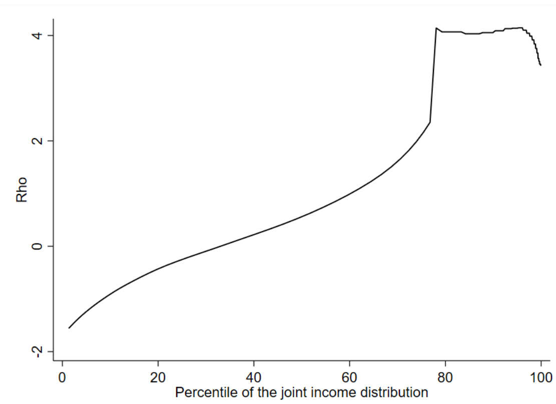


Figure A3: Elasticities of income distributions



(a) Males

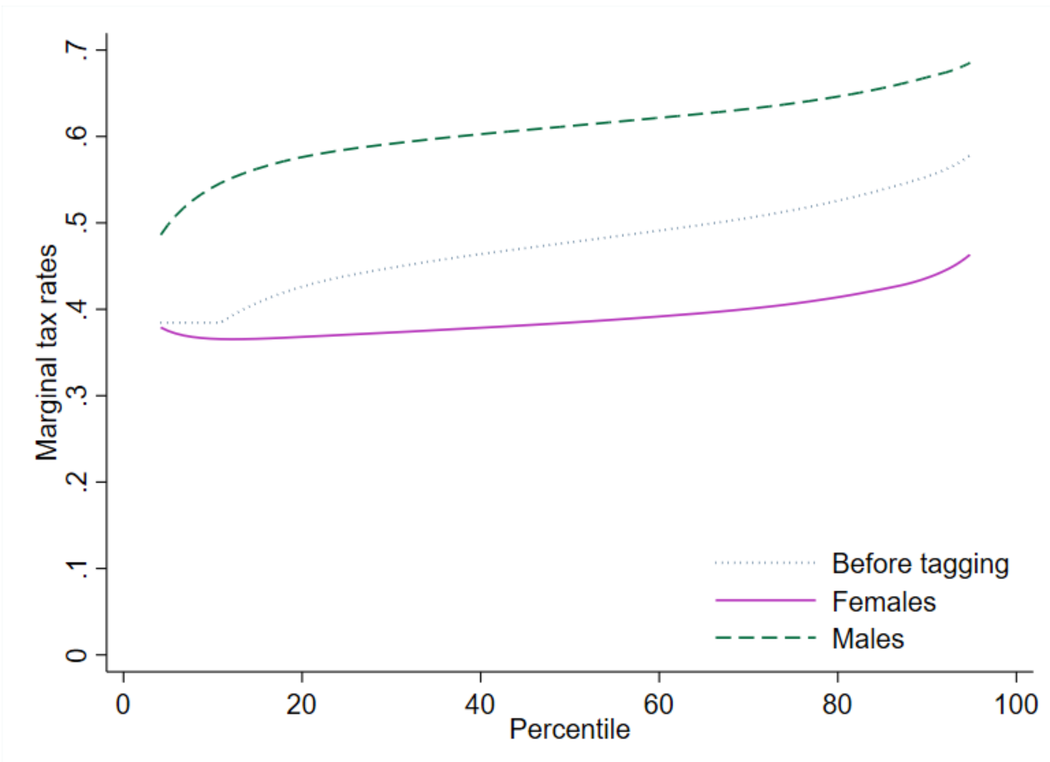


(b) Females

B.1 Gender-specific taxes

When tagging is introduced, females and males face different lump sum transfers and marginal tax rates. The optimal gender-specific transfer from males to females is roughly NOK 50,000. Marginal tax rates are depicted in Figure 13, where females face significantly lower tax rates than males, due to differences in income distributions and differences in elasticities. Differences in elasticities are the main driver, and tax rates are particularly high for males as they respond very little to tax changes.

Figure A4: Marginal tax rates with and without tagging



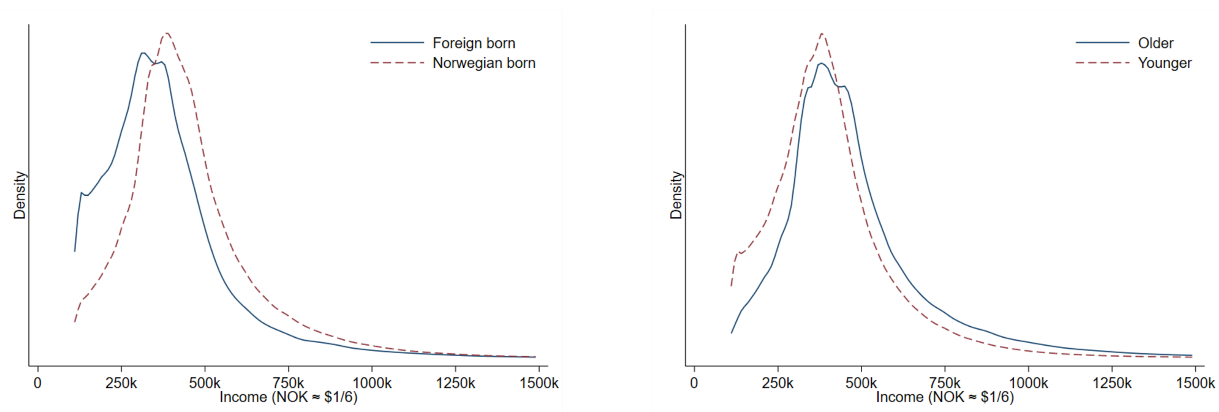
C Further applications: Immigration status and age

Table A7: ETI estimates by immigration status and age

Sample	Norwegian born	Foreign born	Younger	Older
Tax treatment	0.076***	0.103***	0.150***	0.035***
	(0.002)	(0.010)	(0.004)	(0.003)
N	4,440,316	282,780	1,444,500	3,079,860

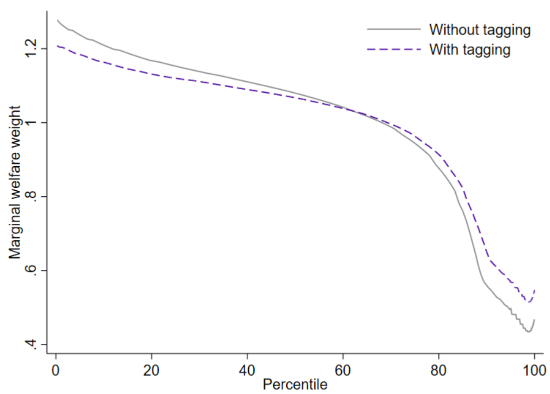
Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Taxable income for wage earners is the dependent variable. The young are from 25 to 40 years old and the older from 41 to 64 years old. Controls are year-dummies, the log of base year income, the first-difference of the log of income in the year prior to the base year, educational field, educational level, age, birth country, children, marital status, gender and county of residence.

Figure A5: Income distributions

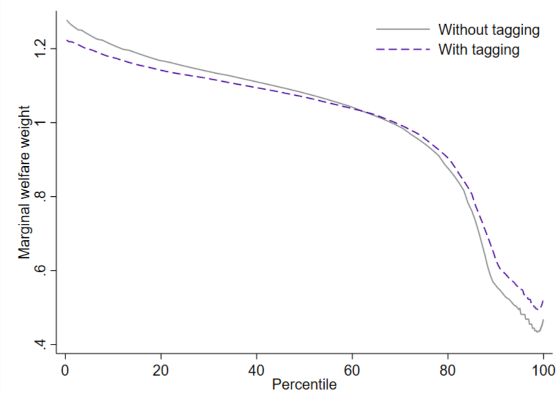


(a) Immigration (b) Age

Figure A6: Marginal welfare weights



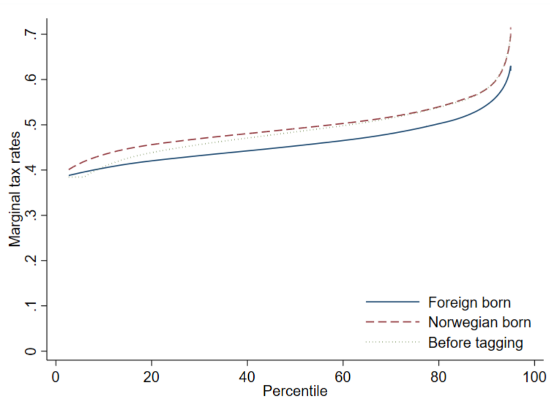
(a) Immigration



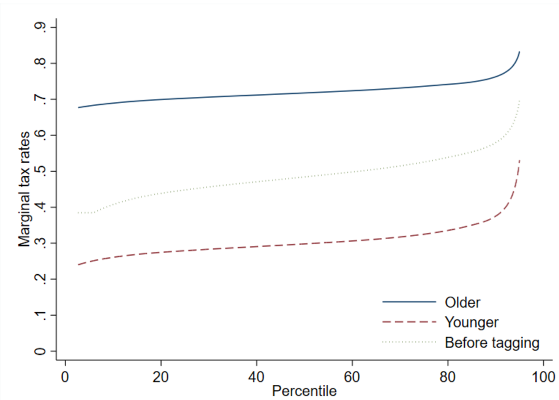
(b) Age

For the immigration status tag, the bias to inequality aversion is 23 % while for the age based tag, the bias is 17 %.

Figure A7: Marginal tax rates



(a) Immigration



(b) Age