Kristoffer Berg

Fair and Efficient Taxation

April 2021

Dissertation for the Ph.D. degree

Department of Economics

University of Oslo

Series of dissertations submitted to the
Faculty of Social Sciences, University of Oslo
No. 859

ISSN 1564-3991

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Acknowledgements

This dissertation is the product of four years of PhD studies at the University of Oslo. It is part of the research conducted at Oslo Fiscal Studies, which is funded by the Norwegian Research Council. I thank the Department of Economics and Oslo Fiscal Studies for the opportunity to conduct the PhD. I also thank Statistics Norway for data access.

My greatest thanks goes to my supervisor, Paolo G. Piacquadio. Our regular discussions on fairness and taxation have been both inspiring and essential for developing the ideas in this dissertation. He has also been a brilliant co-author for the paper ”The Equal-Sacrifice Social Welfare Function” (Chapter 2 of the dissertation). I am also grateful to my co-supervisor, Thor O. Thoresen, whose advice and comments are always detailed and helpful. He was also a great co-author on my first academic publication (Chapter 4 of the dissertation) ”Problematic Response Margins in the Estimation of the Elasticity of Taxable Income”. I would also like to thank my supervisor during the first year of the PhD, Gaute Torsvik, who assisted me in choosing the direction of my research.

My co-author Shafik Hebous deserves acknowledgement and thanks for providing me the opportunity to intern at the IMF and to write our paper ”Does a Wealth Tax Improve Equality of Opportunity?” (Chapter 3 of the dissertation). In my third year of the PhD, I visited the University of California, Berkeley, and I thank Emmanuel Saez for the opportunity and the interesting discussions during my time there.

Finally, my colleagues at the department and PhD program have been essential for both the academic and social life during these four years, and here I want to mention Torje Hegna, Morten Håvarstein, Andreas Myhre, Frikk Nesje and Kjetil Storesletten. A special thanks is due to Magnus E. Stubhaug, with whom I have had the pleasure of sharing an office with for four years. Our continuous discussions on economics and philosophy have been both enjoyable and motivating for completing the PhD.

Oslo, April 2021
Kristoffer Berg

Contents

Introduction and Summary **1**

I Revealing Inequality Aversion from Tax Policy **19**

II The Equal-Sacrifice Social Welfare Function **81**

III Does a Wealth Tax Improve Equality of Opportunity? **129**

IV Problematic Response Margins in the Estimation of the Elasticity of Taxable Income **161**

Introduction and Summary

1 Introduction

With increasing inequality and government debt, taxation is at the forefront of both academic and policy debates. Not only is taxation the main tool for raising government revenue, it is also the main redistributive tool for most governments. Now, as debt and inequality is likely to rise due to the global pandemic, taxation may be a policy response fit to address both issues. As the the International Monetary Fund (2020) writes about the longer term fiscal response to the pandemic *"...governments may need to consider raising progressive taxes on more affluent individuals and those relatively less affected by the crisis (including increasing tax rates on higher income brackets, high-end property, capital gains, and wealth)"*.

1.1 Utilitarianism's problematic tax policy implications

With this in mind, the criteria economists use to evaluate tax policies are crucial. Typically, the criterion has been utilitarianism. The standard utilitarian criterion consists in maximizing social welfare, W, where welfare is defined as the sum of individuals' utilities,

$$W = \sum_i u_i.$$

More general versions of the criterion allows for inequality aversion across utility levels by a concave transformation of each individual's utility, $G(u_i)$,

$$W = \sum_i G(u_i).$$

However, utility levels need to be defined. In the welfarist approach, utility levels describe the level of welfare that each individual has in a state of the world. The level of utility need however not be the level of welfare, but could describe society's judgement on how well-off the individual is considered to be (Fleurbaey, 2008 and Piacquadio, 2017). If these judgements are to respect the Pareto principle (it is an improvement if someone is made better off without anyone else being made worse off), they must respect individual's preferences over allocations. Since preferences are defined ordinally, the cardinality and comparability of utilities across individuals must come from somewhere else than individuals' preferences, such as society's judgement on how deserving each individual is.

Hence, a utilitarian criterion can be somewhat flexible. Still, it has problematic tax implications. First, a utilitarian criterion would recommend that all information available about individuals is exploited in setting taxes (commonly labelled as "tagging" after Akerlof (1978)). This implies that a utilitarian government would set different taxes for different characteristics, such as different taxes for women and men or the young and the old (Mankiw and Weinzierl, 2010). The reason is twofold: characteristics are informative about average income potential and about average differences in tax responsiveness. Take the example of gender. Women (in Norway) earn on average less than men, such that lower taxes on women and higher taxes on men would reduce (income) inequality. As the distortions involved in gender-based taxation are small (or non-existing), gender-based taxation would be more efficient in reducing inequality than increasing income taxes. In addition, women (in Norway) reduce the income they produce by more when their marginal tax rate goes up than men do. This means that the distortions of the income tax could be reduced by increasing marginal taxes for men while lowering them for women. Gender-based taxation is still contrary to many people's intuition that it is wrong for public policies to discriminate based on such characteristics. The point is then that discrimination in itself is not a concern for the standard utilitarian criterion.

Second, the utilitarian criterion implies "slavery of the talented" (Edgeworth, 1897). If redistribution can be achieved at no cost (the first-best) and there are income effects (utility is concave in consumption and convex in labour supply), the sum-of-utility utilitarian criterion selects equal consumption for all combined with higher labour supply for the more skilled. This appears too harsh on the highly skilled, as it makes them worse off than others for the benefit of society. The reason is that the utilitarian criterion does not consider that individuals' may deserve the income they produce. Hence, in optimal taxation under the standard utilitarian criterion, complete equality is only avoided because it would be too distortive to individual's incentives to work.

Third, standard utilitarian criteria do not distinguish between factors due to choice and circumstance. If individuals have different preferences, it is not obvious how to compare their utility levels (Fleurbaey, 2008). One approach is to say that individuals are responsible for their preferences (choice) but not for their ability (circumstance), which builds on work in political philosophy on luck egalitarianism and equality of opportunity. For the standard utilitarian criterion, the distinction between what is choice and what is circumstance is immaterial. Hence, it is irrelevant for end-state criteria whether someone is worse off due to their disadvantaged family background or their lack of effort. This does not account for the current focus in many societies on equality of opportunity and intergenerational mobility rather than just ex-post inequality (Chetty et al., 2014).

1.2 Political philosophy and economics

With utilitarianism's problematic implications, what are the alternatives? Crucial here is the regeneration of political philosophy after Rawls (1971). While utilitarianism had a dominant position in the field, new works on fairness, justice and equality were typically non-utilitarian. The focus was shifted to principles on different levels, just processes, opportunities and the meaning of equality. Key contributers were Nozick (1974) on libertarianism, Sen (1980) and Cohen (1989) on equality, Dworkin (1981) on luck egalitarianism, and Roemer (1986) and Arneson (1989) on equality of opportunity. The distinction between choice and circumstance became prevalent and opened up discussions about individual responsibility.

While political philosophy (and ethics more generally) became less utilitarian, welfare economics lost its central place in economics (Atkinson, 2011). When welfare criteria were used, they were typically still utilitarian. In optimal taxation, this was not the case before Mirrlees (1971). In fact, the classic Musgrave (1959) book presented two rivalling views on fair taxation. These were the "benefit principle" and the "ability to pay principle", where the former is the view that taxation should be in proportion to the benefits one receives from the state, while the latter is the principle that one should pay taxes according to one's ability to carry the tax burden. Later notable exceptions to the utilitarian tradition, partly inspired by Nozick's libertarianism, include Feldstein (1976) and Young (1988) with theories of horizontal equity and equal sacrifice in taxation.

Now, with the increasing focus on inequality after Piketty (2013), there is again an increasing interest in welfare criteria, such as the broad range of views considered in Saez and Stantcheva (2016) and the benefit-principle considered in Weinzierl (2018). Furthermore, there have been contributions developing and incorporating equality of opportunity criteria from political philosophy into formal models in economics (Fleurbaey and Maniquet, 2006 and 2018).

The three first chapters of the dissertation builds on these developments and considers fairness aspects in tax policies.

1.3 The relevance of pre-tax income

While utilitarianism requires information on preferences and feasible end-states, other criteria may use different types of information. For example, in deciding whether equals are treated equally, a benchmark situation where the equals are defined is required. This is also the case for more libertarian criteria, such as equality of sacrifice, where sacrifice must be measured with respect to a benchmark allocation. In both cases, a natural choice of a benchmark would be individuals' incomes in the absence of taxation.

Murphy and Nagel (2002) argue against criteria that are based on pre-tax incomes. They advance two main arguments. First, pre-tax income is not well-defined without a state to uphold property rights and basic security. Second, ownership is decided by law and not entitlements, such that there is no right to one's pre-tax income.

While it is certainly true that pre-tax incomes in the real world are different without any taxation, these counterfactual allocations are well-defined within our economic models. If different models require different criteria, as modern welfare economics may indicate, then the existence of pre-tax income is only a problem when the capacity to earn income depends on the state. Furthermore, the benchmark needs not be the case without any taxation, but for example a Nozickean minimal state, where security and property rights are protected. The question is then why one should have an entitlement to one's income in this minimal state. Murphy and Nagel (2002) argue against such a view, while others disagree. It is in any case an open normative question whether property rights derive from one's production in the laissez-faire.

As modern political philosophy and welfare economics have become more pluralist, the exploration of optimal policy under different normative assumptions should be a valuable endeavour. This is also an ambition of this dissertation.

1.4 Value judgements for fairness in taxation

Since the topic of the dissertation is fairness, one could question whether it involves problematic value judgements. I would like to argue otherwise.

The distinction between normative and positive statements has a long tradition in economics. Robbins (1932) argued that while economics dealt with facts, ethics is concerned with valuations, and there is a "logical gulf" between the two. In an influential paper, Friedman (1953) picks up the distinction, and argues that positive economics can contribute more to a consensus on economic policies than can normative analysis. In opposition to the normative/positive distinction in economics, Putnam (2002) argues that much of economics is "value laden", meaning that many of the concepts used depend on the values of the researchers.

My first point is that a strictly positive approach is insufficient, since normative principles can be useful for providing policy relevant research. Furthermore, as Atkinson (2011) makes clear, it is not the case that most economists can fully escape normative assumptions, but that we often limit ourselves to uncontroversial normative principles, such as Pareto efficiency. This leaves the weighing of different principles and their implications fully to the policy-makers. The problem is that this leads to a too large focus on efficiency rather than on ethical considerations both the public and policy makers may find relevant.

Conveniently, Mongin (2006) introduces a distinction between working with normative content and making judgements about the normative content. In other words, economists can work on normative issues without prescribing the normative conclusions. The economist can take for granted some ethical theory the policy maker endorses and work out the optimal tax system given this ethical theory, thereby avoiding the ethical value judgement, as the ethical value judgement is made when choosing which ethical theory to assume in the model.

A normative approach can also make positive economics more useful by clarifying the relationship between empirics and welfare. This can be seen from the work on sufficient statistics (Chetty, 2009). Sufficient statistics are parameters that can be estimated from observable data and are sufficient for making statements about welfare due to their relations to a welfare theory. Only by having a theory of what counts for the policy maker can such sufficient statistics be obtained, which in turn improves the usefulness of economists' empirical estimates for policy recommendations. When the relation between empirical estimates and welfare is unclear, it is often also unclear what the policy implications of the analysis are, and the empirical analyses are thereby also less useful for informing policy.

The normative content in optimal taxation is therefore useful for providing policy relevant research, and working with this normative content needs not imply controversial value judgements by the researcher. While choosing a welfare criterion requires a judgment, the judgement need not be based on the researcher's values. I would instead argue that it is problematic when normative assumption are not stated or discussed, as they may then involve hidden value judgements.

1.5 Contributions and relations between chapters

The dissertation contributes to fairness and efficiency aspects in tax policy. In the four chapters, I combine theory and empirics to study the taxation of individuals' income and wealth. The dissertation considers relevant fairness views, such as horizontal equity, equal sacrifice and intergenerational mobility. The main contribution is therefore to expand the fairness views considered in designing tax policy, building on recent trends in the field (Saez and Stantcheva, 2016, Weinzierl, 2018 and Fleurbaey and Maniquet, 2018), to consider a broader set of fairness views

In addition, the dissertation develops the Mirrlees (1971) optimal taxation framework and the inverse optimal taxation approach in Bourguignon and Spadaro (2012), presents new evidence on intergenerational effects of wealth (Boserup et al., 2018 and Fagereng et al., 2021) with a novel use of the design of the Norwegian wealth tax (partly building on Jakobsen et al., 2020), and estimates tax responses among the self-employed in Norway by accounting for the different behavioural margins and exploiting the elasticity of taxable income framework (Feldstein, 1995 and Saez et al., 2012).

Chapter 1 examines which principles can support actual tax policy. As the utilitarian criterion disregards discrimination, this chapter builds on Musgrave (1959) and Mankiw and Weinzierl (2010) to show how actual tax policy reflects both vertical equity, the priority on reducing income inequality, and horizontal equity, the priority on treating different characteristics equally. Chapter 2 develops a new welfare criterion for optimal taxation and derive its tax policy implications. The criterion combines efficiency with equal sacrifice (Mill, 1848 and Young, 1988). Contrary to utilitarianism, these social welfare functions prioritize different people according to their incomes in the laissez-faire. Chapter 3 measures the effect of wealth and wealth taxation on income in the next generation. These intergenerational effects are of particular importance for equality of opportunity criteria. Chapter 4 measures the behavioural effects of taxation for individuals that have a range of potential responses. It considers how the many potential response margins of the self-employed affects their tax responses, in order to provide a sufficient statistic for the efficiency effects of changing their marginal tax rates. This is then relevant independently of the welfare criterion.

The relationship between efficiency and fairness is therefore at the centre throughout the dissertation. However, while Chapter 1, 2 and 3 are focused on fairness, Chapter 4 is solely focused on efficiency. In addition, a key point in Chapter 2 is to preserve efficiency while prioritizing individuals that make larger sacrifice. Chapter 1 instead develops a criterion that violates efficiency in order to rationalize current policy choices. This shows the many potential relations between efficiency and fairness. Since what is fair taxation is a complex question, I believe the different ways of looking at fairness and efficiency in the dissertation highlights new considerations that may be significant both in developing economic models and in designing tax policy.

2 Summary of Chapters

Chapter 1: Revealing Inequality Aversion from Tax Policy

Single authored

This chapter shows how to reveal inequality aversion from observed tax policy when governments restrict the information they exploit. Governments have increasing access to information about individuals, but they exploit little of it in setting taxes. I build on the inverse optimal tax problem (Bourguignon and Spadaro, 2012), which considers the observed tax system, combined with information about the income distribution and how individuals react to taxes, to reveal the weight society assigns to each income level.

The main contribution is to map these 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. Here, I develop a theory and optimal tax algorithm to reveal the priority on each concern.

The theory has important implications. My first result is that by accounting for horizontal equity, the implied priority on vertical is lower. The reason is that the horizontal equity requirement 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. 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 when attributing the cost of not exploiting gender information to vertical equity.

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.

Chapter 2: The Equal-Sacrifice Social Welfare Function

Joint with Paolo G. Piacquadio, University of Oslo

How to share the tax burden? Standard economics investigates this question through the lenses of utilitarianism. However, contrary to widespread views about tax justice, the utilitarian criterion assumes workers are not entitled to the return of their work. We study how to share the tax burden when workers are entitled to their productivity. Our contribution is to axiomatically characterize an alternative to utilitarianism, namely the family of equal-sacrifice social welfare functions.

The idea of equal sacrifice is that taxes should be designed so that they impose an equal burden on each taxpayer (Mill, 1848). Three intuitive properties define equal sacrifice (Young, 1988). First, the more taxes an individual pays, the higher her sacrifice. Second, equality of sacrifice imposes larger taxes for higher-income individuals. Third, equality of sacrifice cannot make higher-income individuals poorer than lower-income ones. These properties rule out the utilitarian optimum, but are flexible enough to accommodate many views on sacrifice.

Each member of the family of equal-sacrifice social welfare functions is defined by two ethical choices. Views on tax progressivity are captured by the sacrifice function, while the trade off between equality of sacrifice and efficiency is captured by aversion to inequality in sacrifice.

We also apply our criteria to the US economy. We consider both a proportional definition of sacrifice and a progressive one. We conduct a standard Mirrleesian optimal taxation simulation. As is well-known, the utilitarian criterion recommends large redistribution with marginal tax rates above 60% (and up to 80%). Our criterion with proportional sacrifice justifies rates that are up to 20 percentage points lower, roughly in line with that of the current Californian tax system. For the case of progressive sacrifice, a larger degree of redistribution is called for: the optimal tax system involves marginal tax rates about 8 percentage points lower than the utilitarian recommendation and higher than that of the US tax system.

Chapter 3: Does a Wealth Tax Improve Equality of Opportunity?

Joint with Shafik Hebous, International Monetary Fund

This paper empirically studies the effects of parental wealth, and its taxation during childhood, on adult income in Norway. We ask the questions: Does parental wealth inequality impact next generation labour income inequality, and if so, does a tax on parental wealth affect the labour income distribution of the next generation? We tackle both questions empirically using detailed intergenerational data from Norway, focusing on effects on wages rather than capital income.

By exploiting the design of the Norwegian wealth tax, our analysis yields two main results. First, those who grow up in families with higher levels of net wealth tend to have higher labor incomes, controlling for the education and incomes of their parents as well as individual characteristics including education. The benchmark estimates suggest that an increase in net wealth of 1 million increases future annual wages of the children by 14,000.

Second, based on these point estimates, we estimate the counterfactual income distribution in 2017, in our sample, in the absence of the wealth tax to answer the question: What would have happened to the labor income distribution today had Norway not implemented a wealth tax in the late 1990s and early 2000s? Our results suggest that the wealth tax has made the labor income distribution less unequal—lowering the Gini coefficient by about 1 point. Third, results suggest that the intergenerational labor income mobility is influenced by the stock of parental wealth, with children from more wealthy families experiencing higher labor income mobility than those from less wealthy families.

In addition, our results indicate heterogeneous returns to labor, as higher levels of parental wealth are associated with a higher dispersion of labor income after controlling for individual and parents' characteristics. This evidence suggests that parental wealth is associated with higher labour risk taking.

Chapter 4: Problematic Response Margins in the Estimation of the Elasticity of Taxable Income

Joint with Thor O. Thoresen, Statistics Norway

We study the elasticity of taxable income among the the self-employed. The elasticity of taxable income (ETI) holds the promise of representing a summary measure of tax efficiency costs, which means that further information about the behavioral components of the ETI is not required for its use in tax policy design.

The ETI estimates for the self-employed obtained here are relatively small, in the range from 0.10 to 0.17. However, since there are response margins that can cause biases in the estimation of the elasticity, this paper warns against neglecting information about the composition of the behavior summarized by the ETI.

When using responses of the Norwegian self-employed to the tax reform of 2006, we discuss how three different response margins relate to the overall ETI: working hours, tax evasion and shifts in organizational form. We provide empirical illustrations of effects of each of these margins and argue that the standard procedure for estimating the ETI produces a biased estimate due to the organizational shift margin.

Our estimates suggest that effects on working hours are the dominant response margin summarized by the ETI, but we also attribute some of the overall tax response to tax evasion, for the latter effect obtaining evidence in support of tax evasion increasing in the marginal tax rate.

Importantly, we observe large changes in incentives for incorporation after the 2006-reform, and thereby in the composition of the self-employed. This creates an upward bias in the ETI estimation in our setting. We correct for this bias by deriving weights for the probability to change organizational form and exploiting these weights in the estimation of the ETI. The result is lower ETI estimates. According to one of the specifications, we find a reduction in the ETI estimate from 0.17 to 0.12 after the changed shifting patterns have been controlled for. Thus, this suggests a sizable bias in the standard ETI estimation due due to the selection problem induced by organizational shifts around the tax reform.

References

Akerlof, G. 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.

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

Atkinson, A. B. (2011). The restoration of welfare economics. *American Economic Review*, *101*(3), 157–61.

Boserup, S. H., Kopczuk, W., & Kreiner, C. T. (2018). Born with a silver spoon? danish evidence on wealth inequality in childhood. *The Economic Journal*, *128*(612), 514–544.

Bourguignon, F., & Spadaro, A. (2012). Tax–benefit revealed social preferences. *The Journal of Economic Inequality*, *10*(1), 75–108.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, *1*(1), 451–488.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, *99*(4), 906–944.

Dworkin, R. (1981). What is equality? part 2: Equality of resources. *Philosophy & Public Affairs*, 283–345.

Edgeworth, F. Y. (1897). The pure theory of taxation. *The Economic Journal*, *7*(25), 46–70.

Fagereng, A., Mogstad, M., & Ronning, M. (2021). Why do wealthy parents have wealthy children? *Journal of Political Economy*, (Forthcoming).

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

Feldstein, M. (1995). The effect of marginal tax rates on taxable income: A panel study of the 1986 tax reform act. *Journal of Political Economy*, *103*(3), 551–572.

Fleurbaey, M. (2008). *Fairness, responsibility, and welfare*. Oxford University Press.

Fleurbaey, M., & Maniquet, F. (2006). Fair income tax. *The Review of Economic Studies*, *73*(1), 55–83.

Fleurbaey, M., & Maniquet, F. (2018). Optimal income taxation theory and principles of fairness. *Journal of Economic Literature*, *56*(3), 1029–79.

Friedman, M. (1953). The methodology of positive economics. *Essays In Positive Economics*.

International Monetary Fund. (2020). World economic outlook, october: A long and difficult ascent. *World Economic Outlook Reports*.

Jakobsen, K., Jakobsen, K., Kleven, H., & Zucman, G. (2020). Wealth taxation and wealth accumulation: Theory and evidence from denmark. *The Quarterly Journal of Economics*, *135*(1), 329–388.

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

Mill, J. S. (1848). *Principles of political economy*. Parker.

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

Mongin, P. (2006). Value judgments and value neutrality in economics. *Economica*, *73*(290), 257–286.

Murphy, L., & Nagel, T. (2002). *The myth of ownership: Taxes and justice*. Oxford University Press.

Musgrave, R. A. (1959). *Theory of public finance: A study in public economy*. McGraw-Hill.

Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.

Piacquadio, P. G. (2017). A fairness justification of utilitarianism. *Econometrica*, *85*(4), 1261–1276.

Piketty, T. (2013). *Capital in the 21st century*. Harvard University Press.

Putnam, H. (2002). *The collapse of the fact/value dichotomy and other essays*. Harvard University Press.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Robbins, L. (1932). *The nature and significance of economic science*. MacMillan.

Roemer, J. E. (1986). Equality of resources implies equality of welfare. *The Quarterly Journal of Economics*, *101*(4), 751–784.

Saez, E., Slemrod, J., & Giertz, S. H. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, *50*(1), 3–50.

Saez, E., & Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, *106*(1), 24–45.

Sen, A. (1980). Equality of what? *The Tanner Lecture on Human Values*, *1*, 197–220.

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

Young, H. P. (1988). Distributive justice in taxation. *Journal of Economic Theory*, *44*(2), 321–335.

Chapter 1

Revealing Inequality Aversion from Tax Policy

Revealing Inequality Aversion from Tax Policy*

Kristoffer Berg†

April 6, 2021

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.

JEL: D63, H21, 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, Johannes Fleck, Emanuel Hansen, Bård Harstad, Shafik Hebous, Martin B. Holm, Harald Høyem, Kasper Kragh-Sørensen, Claus T. Kreiner, Etienne Lehmann, Ben Lockwood, Andreas Myhre, Frikk Nesje, Alex Rees-Jones, Marius Ring, Emmanuel Saez, Ragnhild Schreiner, Kjetil Storesletten, Magnus E. Stubhaug, Dmitry Taubinsky, Thor O. Thoresen, Gaute Torsvik, Ragnar Torvik, Danny Yagan, Matt Weinzierl, audiences at CESifo PEAC 2021, ESEWM 2020, IIPF 2020, NTA 2020, Frischsenteret, IRLE, NMBU, Norges Bank, NTNU, OsloMet, Oxford University CBT, Statistics Norway, UC Berkeley, Uppsala University, University of Amsterdam, University of Bergen and University of Oslo. This project has received funding from the 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

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

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 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

$$\sum_k \sum_i p_i^k (z_i^k - c_i^k) \geq R = 0 \text{ with multiplier } \gamma. \tag{3}$$

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.

1. The government has complete information. The first-best is obtained and tagging is irrelevant.

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

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

$$\Delta_1 \hat{g} = 0. \tag{4}$$

Then, taxes do not depend on 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), 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

$$\max_{c,z} W = \sum_i p_i G(V_i(c_i, z_i)), \tag{5}$$

s.t. enough revenue is raised

$$\sum_i z_i - c_i \geq 0 \text{ with multiplier } \gamma, \tag{6}$$

and the self-selection constraint

$$V_2(c_2, z_2) \geq V_2(c_1, z_1) \text{ with multiplier } \lambda. \tag{7}$$

The solution is

$$\Delta_2 c > 0 \text{ and } \Delta_2 g = -\lambda < 0, \tag{8}$$

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 the incentive compatability constraints are

$$V_2^k\left(c_2^k, z_2^k\right) \geq V_2^k\left(c_1^k, z_1^k\right) \ \forall k = f, m \text{ with multipliers } \lambda^{kk}. \tag{9}$$

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

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 self-selection constraints is the same as in Case 2.

$$\Delta_4 \hat{c} > 0 \text{ and } \Delta_4 \hat{g} = -\lambda < 0. \tag{11}$$

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

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

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.

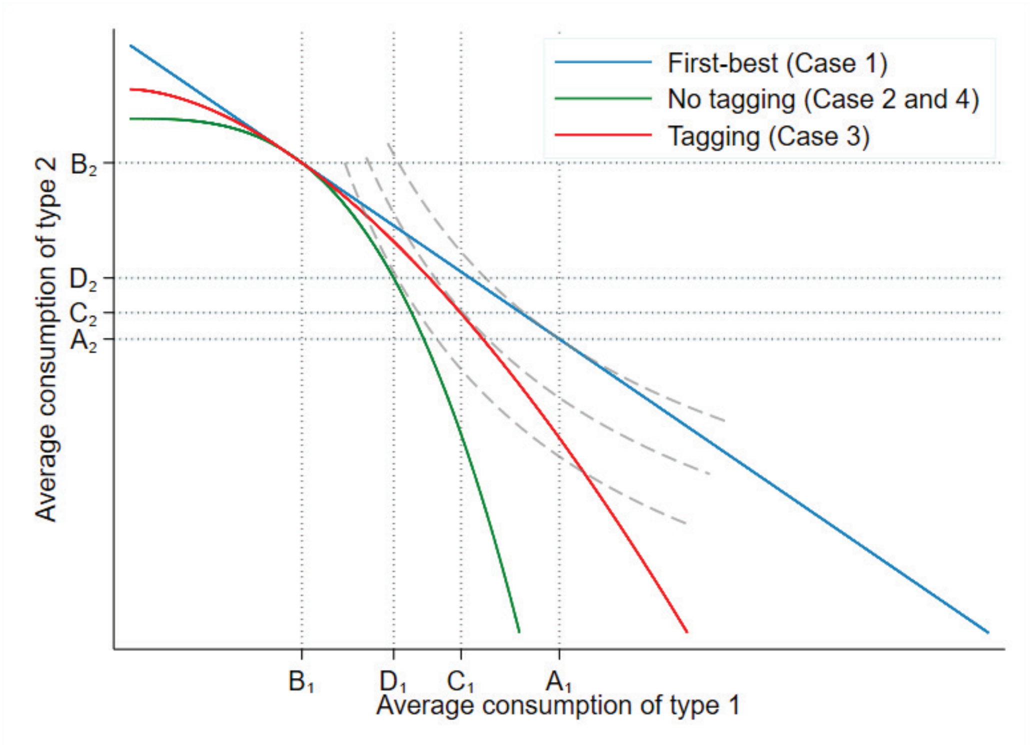
Average inequality aversion, the absolute value of the concavity of the social welfare function over types, averaged over genders, is measured by

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

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. 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

[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 $g(c(z)) = \pi\left(1 - \frac{v(z/w)}{c(z)}\right)$.

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 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]

[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.

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.

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, $\Delta W =$

[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.

$dmE_a\left(g_a(z)\right) - dmE_b\left(g_b(z)\right)$, 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 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. □

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

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

[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.

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.

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, 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 $g(z) = 1$.

[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.

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

[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.

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, $R = 0$. 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 $\alpha_k(z) = zh_k(z)/\left(1 - H_k(z)\right)$ is the characteristic-specific local Pareto parameter and $\bar{G}_k(z) = \int_{z\geq z_i}^{\bar{z}} g_k(z)h_k(z)dz/\left(1 - H_k(z)\right)$ is the characteristic-specific average marginal welfare weight above income level z.

Following the inverse optimum approach (Bourguignon and Spadaro 2012) one can infer marginal welfare weights at each income level, $g(z)$, from the actual tax schedule. *The inverse problem* is to find the marginal welfare weights $g_k(z)$ for which the *current* tax system is a solution to the optimal tax problem. It is simply to solve Equation 41 for $g_k(z)$. The marignal welfare weights from the inverse optimal problem are given by

$$g_k(z) = -\frac{1}{h(z)}\frac{d}{dz}\left[(1 - H(z))\left(1 - \frac{T_k'(z)}{1 - T_k'(z)}\rho_k(z)\varepsilon_k\right)\right]. \tag{42}$$

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

$$g_k(z) = 1 - \frac{T_k'(z)}{1 - T_k'(z)}\rho_k(z)\varepsilon_k, \tag{43}$$

where $\rho_k(z) = -\left(1 + zh_k'(z)/h_k(z)\right)$ is the characteristic-specific "elasticity of the income distribution" (Hendren 2020). 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

$$g(z) = 1 - \frac{T'(z)}{1 - T'(z)} \rho(z)\varepsilon(z), \tag{44}$$

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]

$$\tilde{g}(z) = g\left(z + c^{-1}(m)\right). \tag{45}$$

[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.

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. □

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.

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

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 $\not{k}$

$$dW = dmE_k\left(g(c_k(z))\right) - dmE_{\not{k}}\left(g(c_{\not{k}}(z))\right) = 0\ \forall k. \tag{46}$$

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 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:

[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.

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 $\Delta_k z \approx \epsilon_k\left(z/(1 - T_k'(z)\right)\Delta T_k'(z)$ and $g_k^t(z) = g(z + c_k^{-1}\left(\Delta m_k\right))$, where t denotes the number in the cycle of the algorithm. The behavioral response to the tax change creates the endogeneity, such that if there was no behavioral response to the new tax rates, the algorithm would be redundant, 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

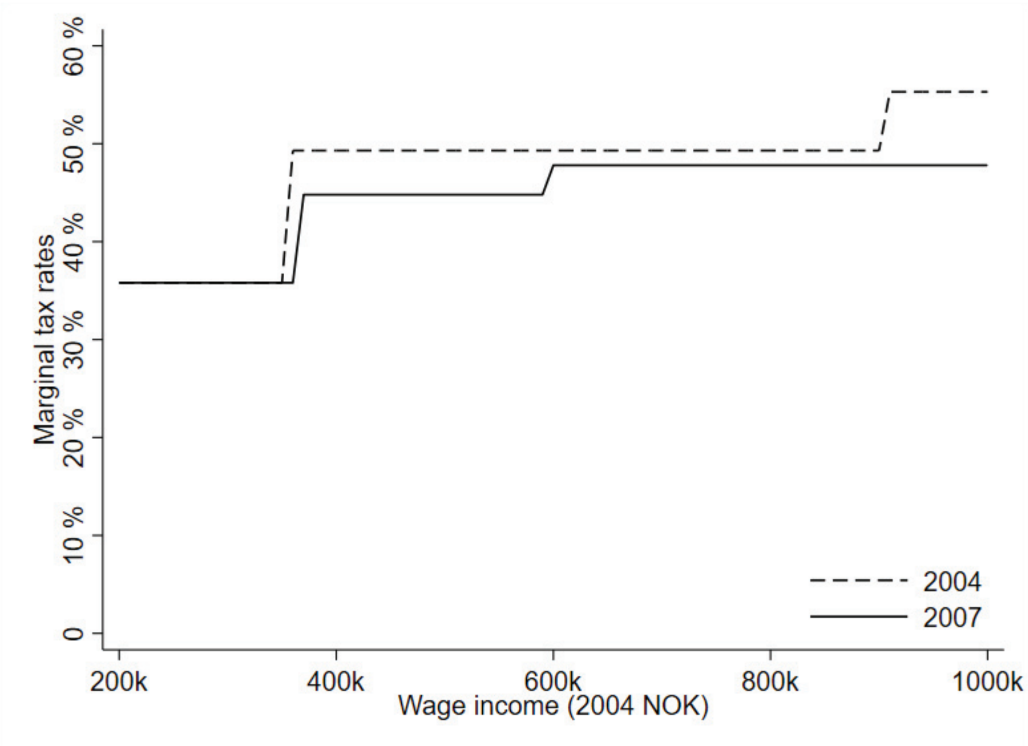
	Mean	Standard deviation
Wage income	541432.6	329576.9
Age	46.8	7.3
Share of males	57.4 %	
Share born in Norway	94.0 %	
Share with children	67.3 %	
Share married	61.0 %	
N	787722	

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.

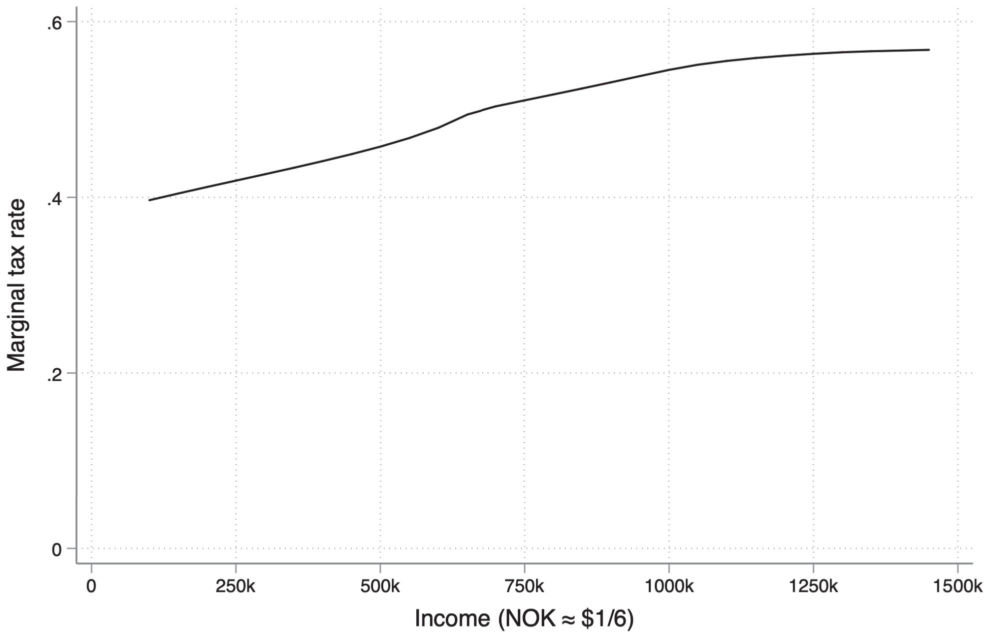
Figure 2: The 2006 tax reform



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

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

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

$$\varepsilon(z) = \frac{(1-\tau)}{z}\frac{\partial z}{\partial(1-\tau)}. \tag{47}$$

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

$$\begin{aligned}\Delta_3 \log(z_{i,t}) =& \alpha_t + \beta D_k \Delta_3 \log(1-\tau_{i,t}) + \theta \log(z_{i,t}) + \pi \Delta_1 \log(z_{i,t-1}) \\ & + \eta M_{i,t}' + \epsilon_{i,t},\end{aligned} \tag{48}$$

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

The resulting estimates are shown in Table 2. Although 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

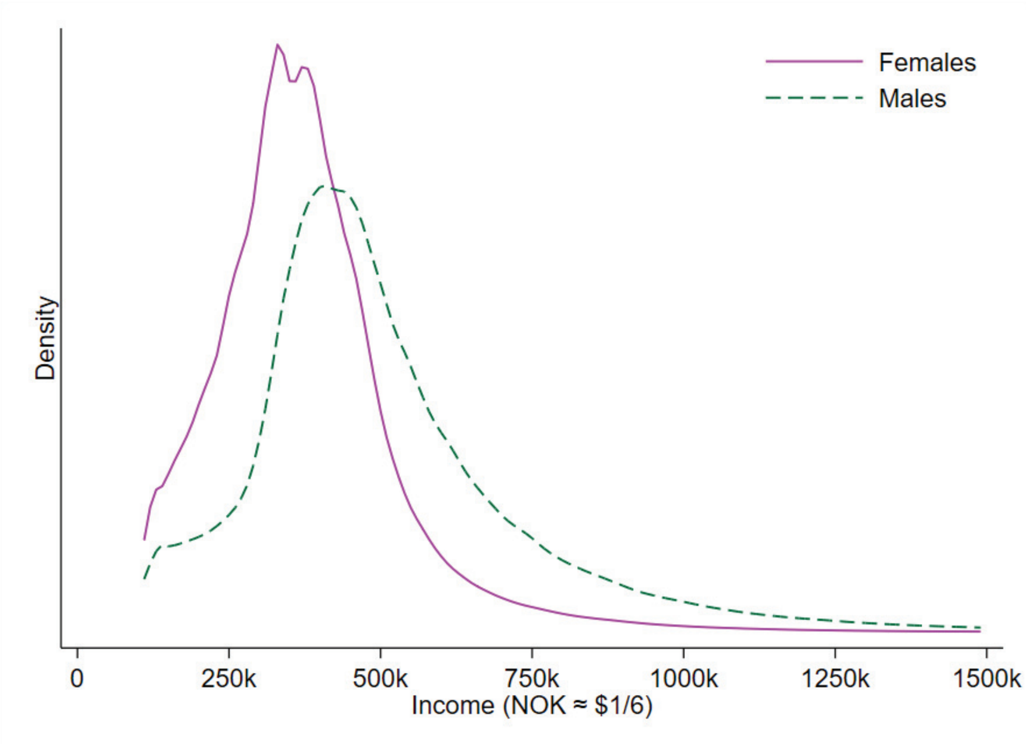
Notes: ETI estimates, average and separated by gender for wage earners. The estimation is a first-difference equation where the tax rate change is instrumented by the reform-induced tax rate change. Other controls are year-dummies, the log of base year income, the first-difference of the log of income 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.

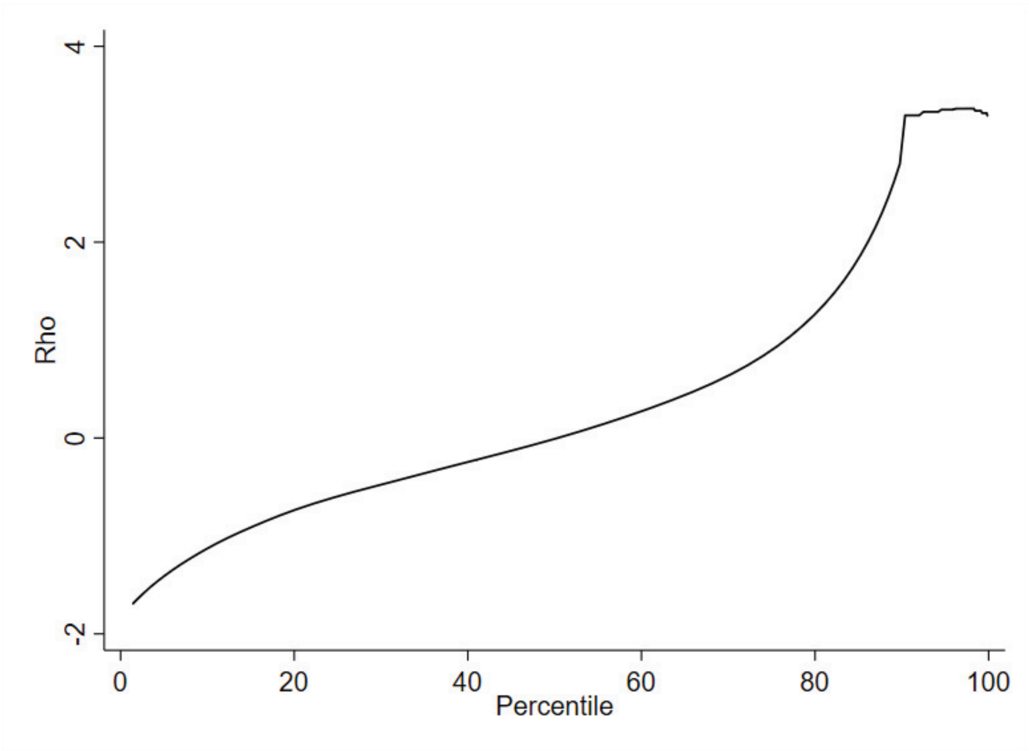
Appendix), I have estimated responses separately for the single and married, and the relative difference in response between females and males is equally large. The response among single females appears to be larger than among married females, although the difference is not statistically significant. 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.

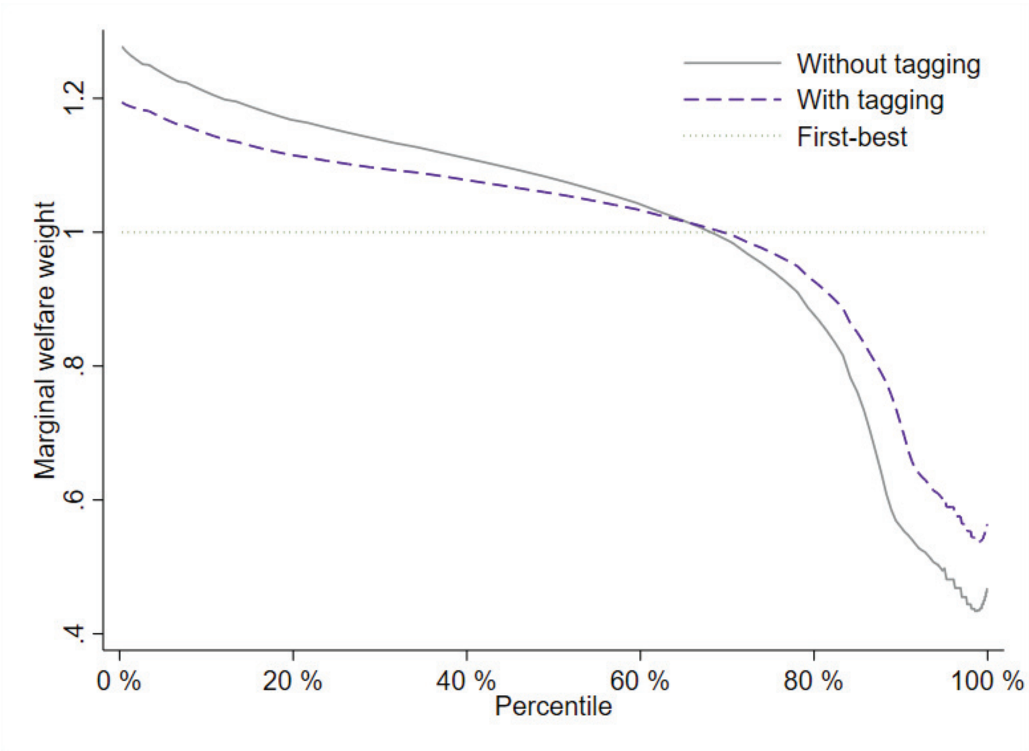
5.5 Marginal welfare weights and equity measures

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.

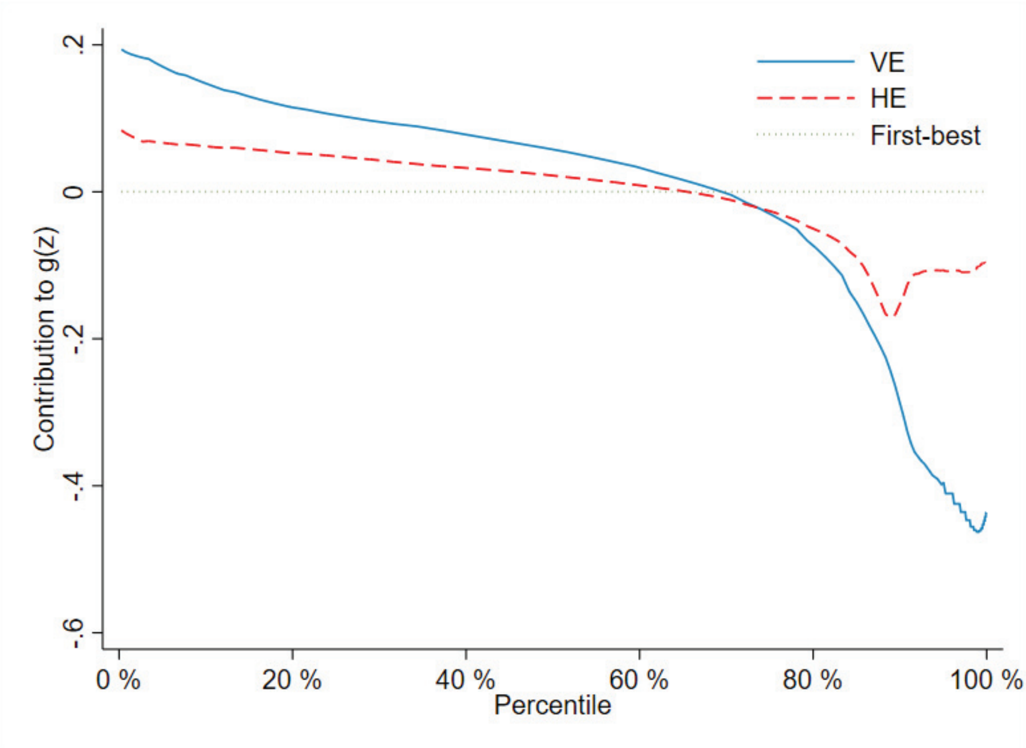
In Figure 8, the bias is estimated by the absolute value of the difference in marginal welfare weights relative to the actual marginal welfare weights, which measures the relative size of the bias to the measure of vertical equity by ignoring horizontal equity. The total difference in steepness between $g(z)$ 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



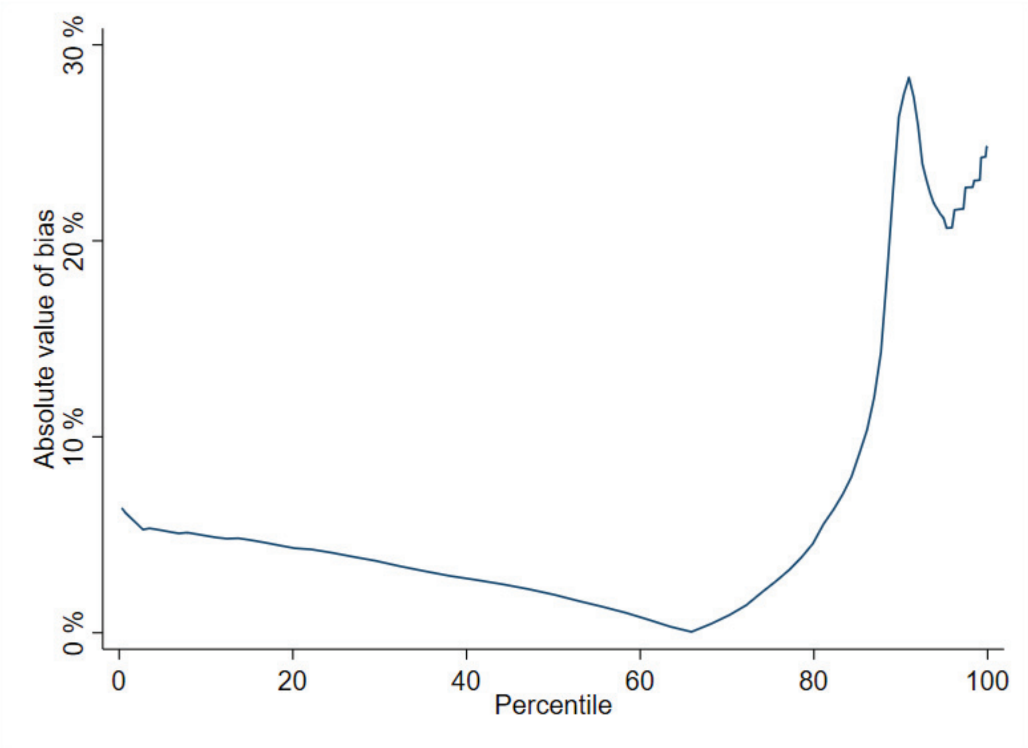
Notes: Inverse optimum marginal welfare weights without tagging, $g(z)$, and with tagging $\hat{g}(z)$ over the income distribution for wage earners in 2010.

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

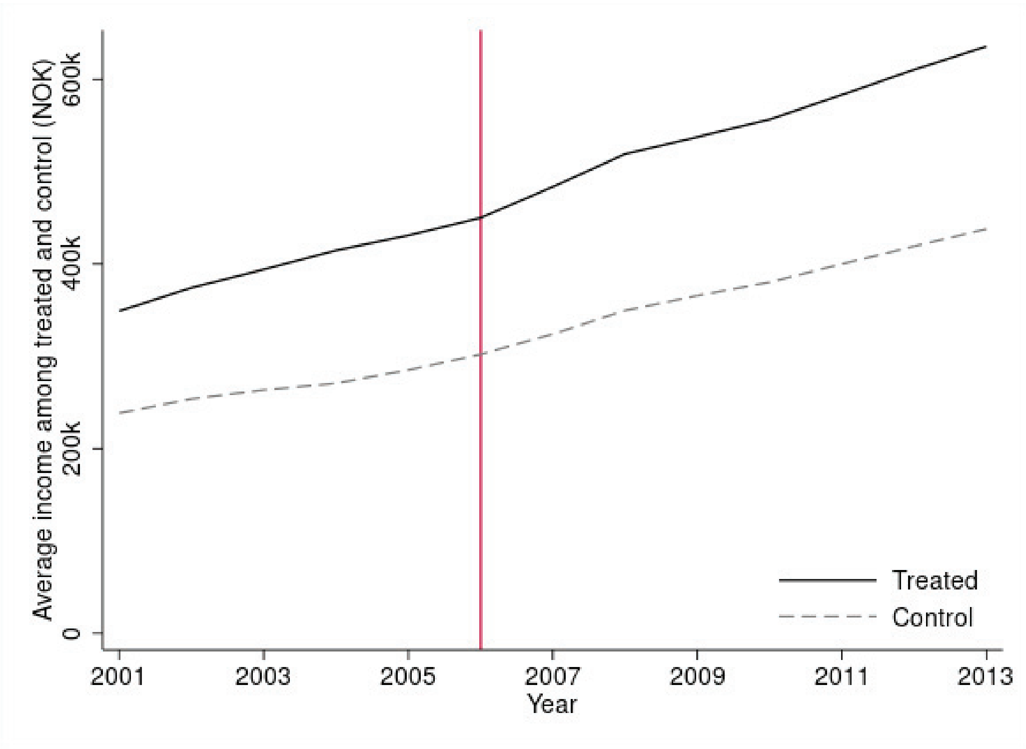


Table A2: Summary of treatment and control groups

	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

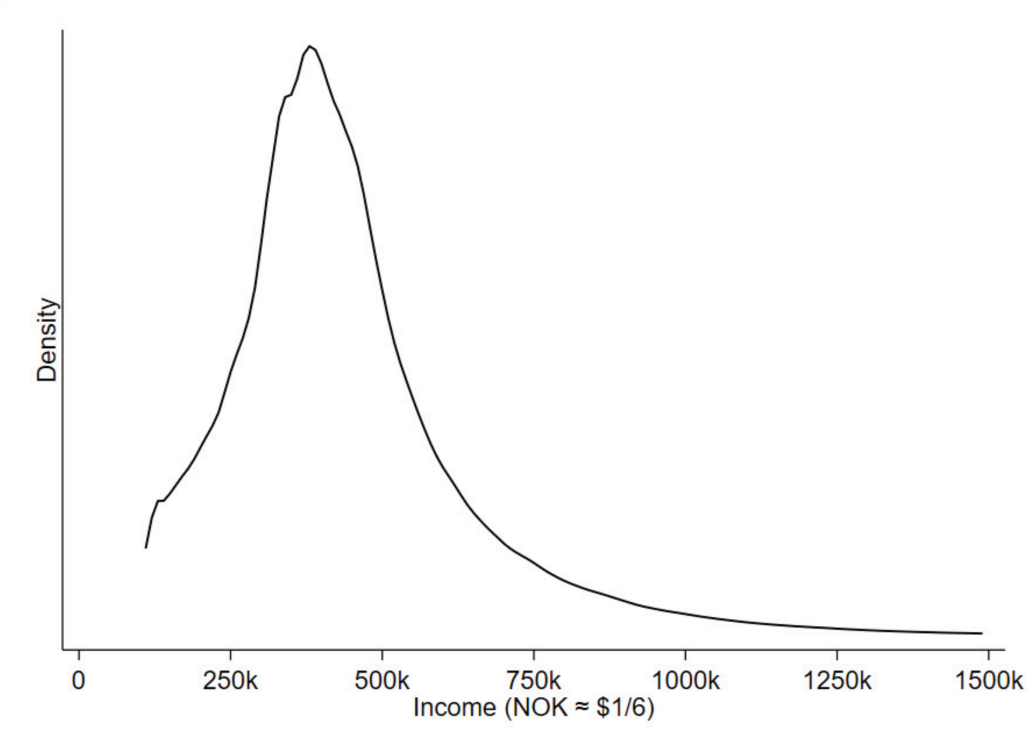
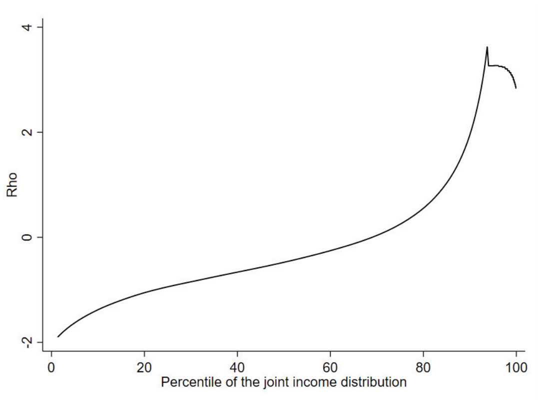
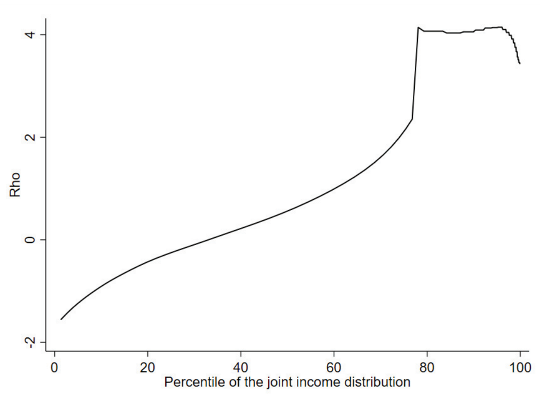


Figure A3: Elasticities of income distributions



(a) Males

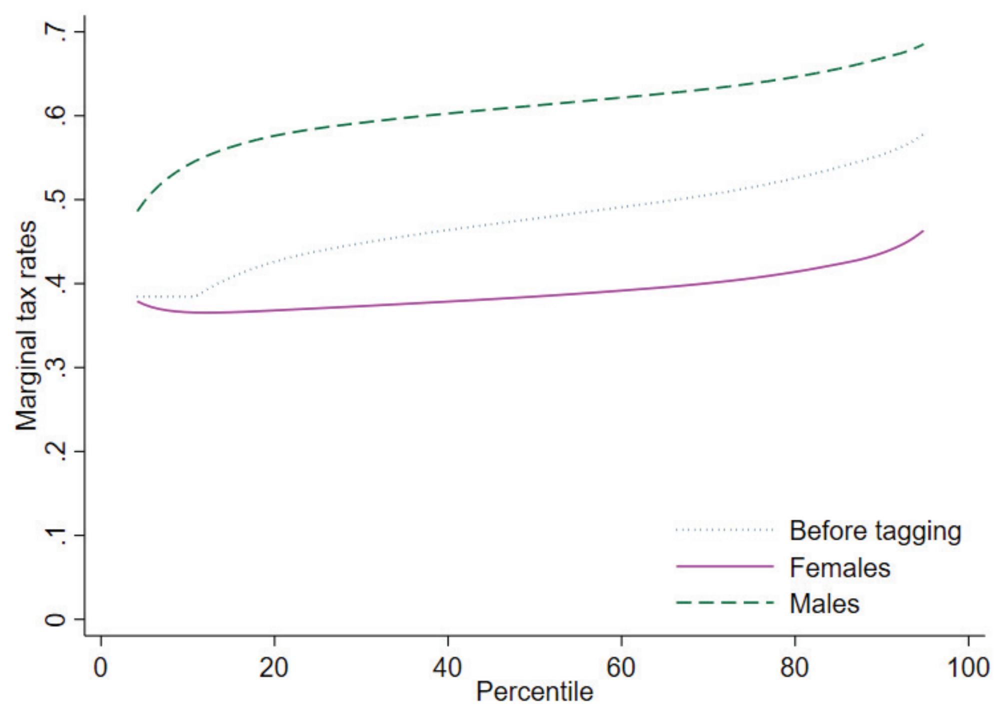


(b) Females

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



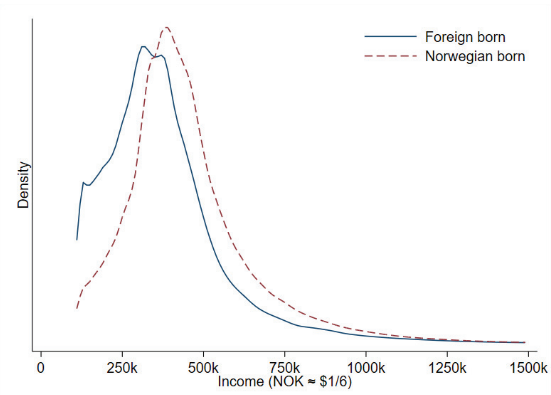
C Further applications: Immigration status and age

Table A7: ETI estimates by immigration status and age

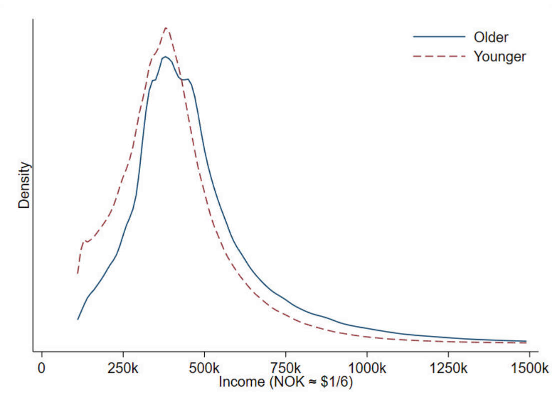
Sample	Norwegian born	Foreign born	Younger	Older
Tax treatment	0.076***	0.103***	0.150***	0.035***
	(0.002)	(0.010)	(0.004)	(0.003)
N	4,440,316	282,780	1,444,500	3,079,860

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

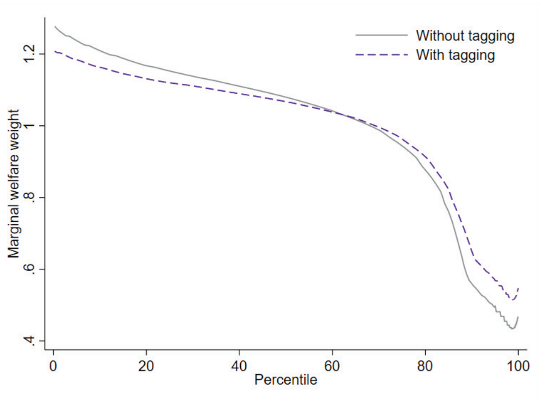


(a) Immigration

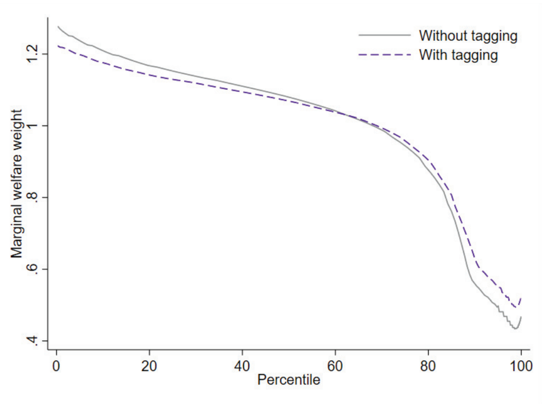


(b) Age

Figure A6: Marginal welfare weights



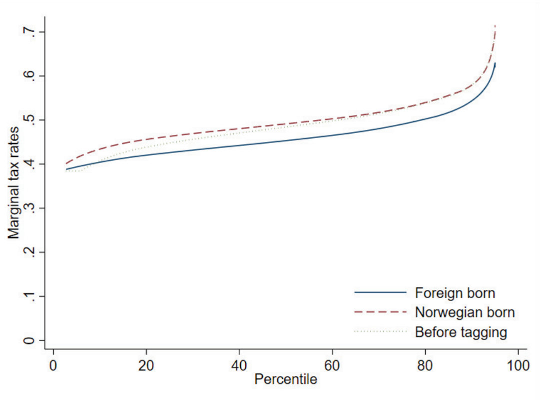
(a) Immigration



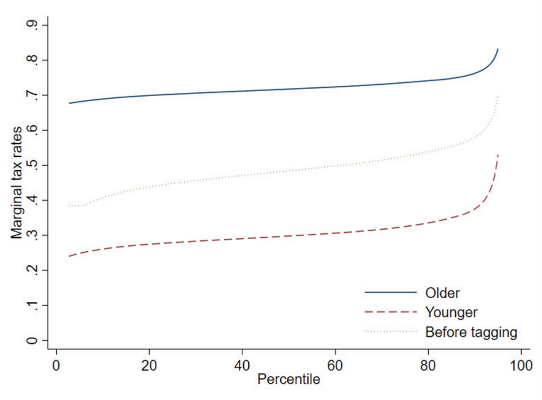
(b) Age

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

Chapter 2

The Equal-Sacrifice Social Welfare Function

Joint with Paolo G. Piacquadio

The Equal-Sacrifice Social Welfare Function*

Kristoffer Berg and Paolo G. Piacquadio

November 5, 2020

Abstract

How to share the tax burden? Standard economics investigates this question through the lenses of utilitarianism. However, contrary to widespread views about tax justice, the utilitarian criterion assumes workers are not entitled to the return of their work. In this paper, we study how to share the tax burden when workers are entitled to their productivity. We axiomatically characterize an alternative to utilitarianism, namely the family of equal-sacrifice social welfare functions. Each member of this family is defined by two ethical choices. Views on tax progressivity are captured by the sacrifice function, while the trade off between equality of sacrifice and efficiency is captured by aversion to inequality in sacrifice. We then illustrate our approach within Mirrleesian optimal income taxation. When sacrifice is proportional, our criterion delivers optimal taxes that are roughly in line with the US (Californian) tax system, with marginal tax rates about 20 percentage points lower than the utilitarian recommendation.

JEL codes: D63, H21, I31.

Keywords: equal-sacrifice principle, optimal income taxation, welfare criterion.

*The authors wish to thank Rolf Aaberge, Alberto Alesina, Geir Asheim, Marcus Berliant, Tomer Blumkin, Walter Bossert, Giacomo Brusco, Vidar Christiansen, Ashley Craig, Bård Harstad, Bas Jacobs, Claus Thustrup Kreiner, Etienne Lehmann, Edwin Leuven, Ben Lockwood, Frikk Nesje, Sergio Ocampo-Diaz, Tone Ognedal, Andreas Peichl, Uwe Thuemmel, Dominik Sachs, Emmanuel Saez, Dirk Schindler, Kjetil Storesletten, Valer Suteu, Thor O. Thoresen, and Matt Weinzierl for comments, in addition to the audiences at Osaka, ESWC2020, IIPF2020, NTA 2019, Heidelberg, and the Norwegian-German tax conference. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme ERC Starting Grant VALURED (Grant agreement No. 804104) and from the Research Council of Norway (RCN) through the research center Oslo Fiscal Studies.

1 Introduction

Rising inequality has brought the tradeoff between tax fairness and tax efficiency to the center of the economic debate. The key question is how to design the best—fair and efficient—tax system. Two ingredients are indispensable for answering this question. First, an economic model describes the workers, predicts their behavioral responses to taxes, and clarifies the information and policies available to the government. Second, a welfare criterion defines a ranking of alternatives in terms of their social desirability. Then, the optimal policy is the one that achieves the most desirable alternative among the feasible ones. In this paper, we propose and axiomatically characterize a new family of welfare criteria. We then illustrate these criteria in the context of optimal non-linear income taxation (Mirrlees, 1971).

Existing welfare criteria—such as utilitarianism—trade off efficiency with equality of outcomes.[1] In contrast, the distinctive feature of our criteria is to trade off efficiency with "equality of sacrifice." Equality of sacrifice is a well-known principle of fairness in taxation (Mill, 1848): taxes should be designed so that they impose an equal burden on each taxpayer.

To illustrate the difference, consider the utilitarian criterion in the context of income taxation. Andrea and Barbara work the same number of hours. Due to their different productivities, they earn different pre-tax incomes, $40,000 for Andrea and $60,000 for Barbara. Now consider the problem of optimally allocating a tax burden of $30,000 among them. For fixed labor supply, the utilitarian optimum tax scheme is to collect $5,000 from Andrea and $25,000 from Barbara, so that their after-tax incomes are equalized to $35,000. As Feldstein (1976) clarifies, utilitarianism implicitly assumes that all differences in productivity across individuals are undeserved: society is entitled to everyones' potential earnings.[2] Even more controversial, utili-

[1]Here, by equality of outcomes, we mean equality of some index of well-being of individuals that depends only on their consumption vector and/or on their utilities. In a one-dimensional setting—say income—the concern for equality of outcomes is generally captured by the (strict) Pigou-Dalton principle: an income transfer from a poorer individual to a richer one decreases the level of social welfare. For a generalization and interpretation of equality of outcomes in multidimensional settings, see Piacquadio (2017).

[2]All criteria trading off efficiency with equality of outcomes place no moral importance to individuals' types—here individuals productivities. This includes criteria such as generalized utilitarian, maximin, and rank-dependent utilitarianism. Similar counterintuitive implications hold also out of the optimum. Consider a tax scheme that taxes $10,000 from Andrea and $20,000 from Barbara, leaving them with an after-tax income of $30,000 and $40,000, respectively. Consider a different tax scheme that imposes no taxes on Andrea and imposes the entire $30,000 tax burden on Barbara,

tarianism recommends equal after-tax income combined with unequal labor supply. In fact, to maximize the common (utilitarian) good, the government wants high-skill individuals to supply more labor (while keeping equal after-tax income). This results in "slavery of the talented:" high-skill individuals are penalized for being more productive (Mill, 1848; Musgrave, 1959; Fleurbaey and Maniquet, 2018). Unsurprisingly, such a utilitarian dictatum conflicts with the ethical views on taxation held by the majority of the population (Schokkaert and Devooght, 2003; Weinzierl, 2014; Saez and Stantcheva, 2016).

The equal-sacrifice principle avoids the above shortcomings of utilitarianism and, thus, spawned a large interest and support in the early economic literature (Mill, 1848; Sidgwick, 1883; Edgeworth, 1897; Pigou, 1928; Vickrey, 1947). Three intuitive properties define equal sacrifice (Young, 1988). First, the more taxes an individual pays, the higher her sacrifice. Second, equality of sacrifice imposes larger taxes for higher-income individuals. Third, equality of sacrifice cannot make higher-income individuals poorer than lower-income ones. These properties rule out the utilitarian optimum, but are flexible enough to accommodate many views on sacrifice (Musgrave, 1959). To fix ideas, in figure 1 we represent pre- and after-tax incomes (in thousands of dollars) for a proportional definition of sacrifice and for a progressive definition of sacrifice (Young, 1990). With a proportional definition of sacrifice, both Andrea and Barbara pay 30% income tax and obtain an after-tax income of $c_A = 28$ and $c_B = 42$, respectively. With a progressive definition of sacrifice, Andrea pays less than 30% income tax and achieves an after tax income between $c_A \in (28, 35)$, while Barbara pays more than 30% income tax and achieves an after-tax income of $c_B \in (35, 42)$. For reference, the utilitarian optimum requires $c_A = c_B = 35$.

Independently of the definition of sacrifice, the equal-sacrifice principle has had a major drawback, which ultimately led to its demise and to the dominance of utilitarianism (see discussion in Weinzierl, 2018). Equal sacrifice—as a standard of perfection—cannot be adopted for second-best analysis. Behavioral responses, asymmetric information, and policy constraints are fundamental aspects of the real world, which make the first-best optimum unfeasible. For example, in the Mirrlees (1971) income tax model, equal-sacrifice tax schedules leave efficiency gains unreal-

leaving them with an after-tax income of $40,000 for Andrea and $30,000 for Barbara. All these criteria rank these tax schemes equally, despite that the second one forces the high-productivity individual to pay so much taxes as to end up with less than the low-productivity individual.

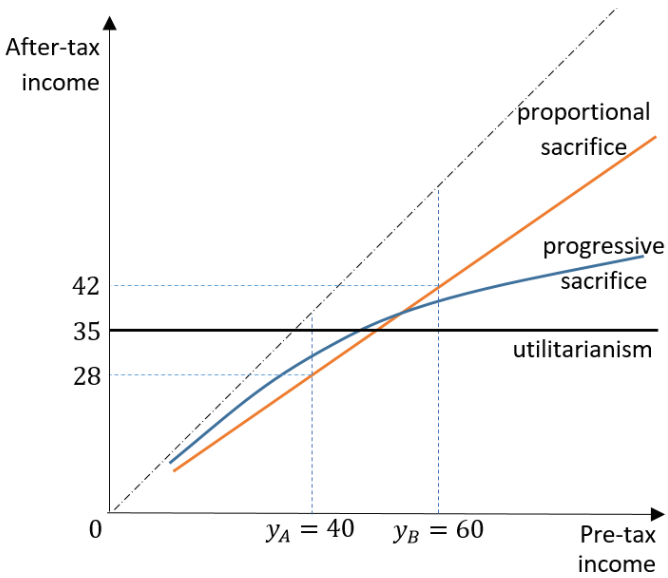


Figure 1: How to share the tax burden: utilitarian optimum and equality of sacrifice.

ized (Berliant and Gouveia, 1993; da Costa and Pereira, 2014). In this paper, we solve this issue. Our family of *equal-sacrifice social welfare functions* trade off efficiency with equality of sacrifice. Thus, they behave like the equal-sacrifice principle at the first-best optimum. In second-best settings—such as the Mirrleesian income taxation model—they compromise between efficiency losses and inequality in sacrifice.

The key breakthrough comes from axiomatically ruling out the controversial implications of utilitarianism. Utilitarianism compares individuals by their marginal utility of after-tax income. If i has a larger marginal utility than j, i is considered more deserving than j: thus, social welfare increases with a small transfer of income from j to i. In essence, utilitarianism is averse to inequality in marginal utilities. For utilitarianism, the pre-tax incomes of Andrea and Barbara—as well as their taxes payed—are irrelevant in determining how much after-tax income each deserves.

In contrast, we let society prioritize individuals depending on their pre-tax incomes and tax payments. As we show, this is equivalent to choosing an index of sacrifice and comparing individuals by their level of sacrifice. In fact, the equal-sacrifice social welfare function is averse to inequality in sacrifice. If i makes a larger sacrifice than j, i is more deserving than j and social welfare increases with a small transfer of income from j to i. Thus, whether Andrea or Barbara are more deserving of a transfer depends on how we compare their sacrifice due to taxes.

The similarity and differences between utilitarianism and our equal-sacrifice social welfare function are far reaching. Utilitarianism prioritizes individuals with larger marginal utility of consumption by maximizing the sum of utilities, which—tautologically—are the integral of individuals' marginal utilities. In contrast, the equal-sacrifice social welfare function prioritizes individuals with larger sacrifice. Our results show that, mirroring utilitarianism, the equal-sacrifice social welfare function is the sum of the integral of individuals' sacrifices. The following thought experiment clarifies this choice. Assume the government can assign a dollar to one individual only. The effect of giving a dollar to individual i is the derivative of the social welfare function with respect to the consumption of i. Since this derivative is, by construction, a measure of the sacrifice of i, the government would optimally assign the dollar to the individual making the largest sacrifice.[3]

The definitions of sacrifice singled out by our characterization include existing proposals as special cases, at fixed labor supply. Among others, our criteria can accomodate absolute and relative indices of sacrifice, progressive and regressive ones, indices based on utility or income. However, most existing indices of sacrifice cannot accommodate changes in labor supply: these indices do not respect how individuals trade off consumption and labor and, thus, violate the Pareto principle. In contrast, we impose the Pareto principle. This has two main implications. First, our approach can accommodate labor supply responses: we can compare individuals by their level of sacrifice at any allocation. Second, the integral of individuals' sacrifices are "utility functions" or—more precisely—numerical representations of individuals' preferences.

Beyond Pareto efficiency, our axiomatic analysis requires social preferences to be strictly concave, additively separable, and to satisfy three novel axioms disciplining inequalities of sacrifice. First, recall that equality of sacrifice requires that: (i) everyone should contribute by paying taxes; (ii) the contribution of a more productive worker should be larger than the contribution of a less productive one; and (iii) the

[3]Earlier attempts to extend equal sacrifice to complete rankings missed this point (see Musgrave (1959) and, more recently, Weinzierl (2012) and Jessen, Metzing, and Rostam-Afschar (2019)). Minimizing the sum of sacrifices leads to the government prioritizing individuals with the largest "marginal sacrifice," and not those with the largest level of sacrifice. In fact, when sacrifice is measured by the difference in utility due to taxes, minimizing the sum of sacrifices turns out to be equivalent to maximizing utilitarian welfare: both prioritize individuals with larger marginal utilities. The addition of a convex transformation of sacrifices introduces some prioritiy to those with highest sacrifice. However, unless one takes the limit case of infinite convexity, these criteria support policies that increase efficiency losses and, at the same time, move away from the equal-sacrifice allocation.

more productive should not contribute so much as to end up with a lower after-tax income. We impose that a violation of any of these conditions calls for redistribution. For example, if the more productive worker contributes less—violating (ii)—a transfer of income from the more productive to the less productive cannot decrease the level of social welfare.

We conclude our study by applying our criteria to the US economy. We consider both a proportional definition of sacrifice and a progressive one (as in figure 1). We conduct a standard Mirrleesian optimal taxation simulation, following the exercise by Mankiw, Weinzierl, and Yagan (2009). As is well-known, the utilitarian criterion recommends large redistribution with marginal tax rates above 60% (and up to 80%). Our criterion with proportional sacrifice justifies rates that are up to 20 percentage points lower, roughly in line with that of the current Californian tax system. For the case of progressive sacrifice, a larger degree of redistribution is called for: the optimal tax scheme involves marginal tax rates about 8 percentage points lower than the utilitarian recommendation and higher than that of the US tax system.

More generally, the second-best optimal equal-sacrifice tax scheme ultimately depends on ethical choices: in particular, the progressivity of the sacrifice function and society's aversion to inequality in sacrifice. Our results provide a mapping between these ethical choices and optimal tax systems. Thus, we cannot conclude whether the US tax system is optimal. However, we show that a proportional definition of sacrifice can (roughly) rationalize the current system. As a result, the large difference between the utilitarian optimum and the real-world tax schedules are not necessarily driven by the political influence of high-income individuals. Rather, the observed tax schedules might reflect the adoption of principles of distributive justice inspired by equal sacrifice.

Related literature

Historically, equal sacrifice was conceived as a standard of perfection (Mill, 1848), and the debate was centered on the definition of sacrifice (Musgrave, 1959). Closer to our contribution, Young (1988) has proposed an axiomatic characterization of equal sacrifice, building on intuitive requirements on the level of taxes and how these relate to individuals pre-tax incomes.[4] Young shows this is equivalent to choosing a utility function—unrelated to individuals' preferences—for the purpose of assessing individuals losses due to taxation.[5] However, doing so leads to a violation of the Pareto principle. Our approach extends Young's in two directions. First, it respects the preferences of individuals: labor supply choices are evaluated through the preferences of each individual. Second, it allows evaluating deviations from equal sacrifice by prioritizing individuals incurring a larger sacrifice. Thus, in a first-best setting, Young's equal-sacrifice solution—corrected to respect preferences—emerges. In an incomplete information setting, equality of sacrifice is often too costly, and our criteria compromise between inequality in sacrifice and efficiency losses.

A general approach to welfare criteria that can accommodate equal-sacrifice concerns is to let the social value of one more dollar for an individual—so called social marginal welfare weights—depend not only on the well-being of that individual, but also on some measure of her sacrifice (Saez and Stantcheva, 2016).[6] Our characterization clarifies which social marginal welfare weights emerge when abiding by principles of fairness inspired by equal sacrifice.

The theory of justice developed here also provides a modern interpretation of "justice as mutual advantage" (Gauthier, 1986), which relates to the axiomatic work on Nash's bargaining theory and its more recent extensions (Binmore, Rubinstein, and Wolinsky, 1986; Rubinstein, Safra, and Thomson, 1992). These theories unfold around two key ethical choices: the definition of equality and the comparative evaluation of inequalities. Crucially, bargaining theories of justice rely on cardinal and interpersonally comparable information about individuals' utilities to measure equality and evaluate inequalities. In contrast, here both the definition of equality and the evaluation of inequalities emerge from the axioms. Finally, the characterizations of bargaining theories of justice generally build on scale invariance, while here we remain closer to the utilitarian tradition and require an additively separable representation

[4]We follow the equal-sacrifice literature and use laissez-faire incomes as the reference for the measurement of sacrifice. Our results extend to different choices of the reference.

[5]Young (1990), Ok (1995), and Mitra and Ok (1996) build on this ordinal version of equal sacrifice and discuss the relationship between equal sacrifice and progressivity. Chambers and Moreno-Ternero (2017) introduce a concern for poverty. Stovall (2020) provides an improved characterization. He also suggests a generalization of equal sacrifice, allowing the utility functions to be person-specific.

[6]Alternative criteria for income taxation are reviewed in Fleurbaey and Maniquet (2018). The recent literature mostly addresses the issue of preference heterogeneity (such as Fleurbaey and Maniquet, 2006, and, in an abstract setting, Piacquadio, 2017), which, in the context of equal sacrifice, we leave to future work.

of the welfare criterion.[7]

The rest of the paper is organized as follows. Section 2 contains an illustration of the criterion and a comparison with utilitarianism. Section 3 presents the formal model and the axioms. Section 4 discusses the characterization result. Section 5 explores the implications of the criterion with a simulation of the optimal tax system for the US economy. Section 6 briefly concludes. All the proofs are in the appendix.

2 A simple illustration

We illustrate our approach in a Mirrleesian model with quasi-linear utilities. Individuals' preferences over consumption c and labor supply ℓ are represented by a utility function $u(c, \ell) = c - v(l)$ with $v', v'' > 0$. Individuals differ in their labor market productivity: each individual i is characterized by the wage rate $w_i > 0$.

2.1 The proportional-sacrifice social welfare function

We next introduce a simple version of our welfare criterion. The first step is to define how to measure and compare the sacrifice of any two individuals i and j. A natural starting point is the **laissez-faire allocation**. At the laissez-faire allocation, no taxes are levied and each individual i maximizes her utility over the budget set $B_i \equiv \{(c_i, \ell_i) | c_i \leq w_i \ell_i\}$. Let $\left(\bar{c}_i, \bar{\ell}_i\right)$ denote the **laissez-faire bundle** of i. The idea is that, at the reference laissez-faire allocation, no individual makes any sacrifice.[8]

Assume instead individual i consumes c_i and works the laissez-faire labor supply $\bar{\ell}_i$. Then, her **proportional sacrifice** is $(\bar{c}_i - c_i) / \bar{c}_i$, that is, the ratio of the (implicit) tax contribution $\bar{c}_i - c_i$ to the pre-tax income $\bar{c}_i = w_i \bar{\ell}_i$.[9] The proportional sacrifice is represented in figure 1. Assume individual i has higher laissez-faire consumption than individual j, i.e., $\bar{c}_i > \bar{c}_j$. When $c_i = \bar{c}_i$ and $c_j = \bar{c}_j$ both individuals make no

[7]Scale invariance requires the ranking of alternatives remain unchanged when rescaling alternatives. Scale invariance is logically independent of additive separability. Interestingly, the Nash and the Kalai–Smorodinsky bargaining solutions are characterized based on scale invariance, but also satisfy separability (on a subdomain of alternatives).

[8]Piketty and Saez (2013) have emphasized how utilitarianism fails to ensure laissez-faire prevails even when all agents have the same productivity level (see also Jacquet and Van de Gaer, 2011, and Fleurbaey and Maniquet, 2018).

[9]If this definition of sacrifice was extended to any level of labor supply, it would necessarily be independent of the utility cost of working and thus would lead to violations of the Pareto principle.

sacrifice. The smaller their consumption, the larger their sacrifice. When sacrifice is proportional, i and j make the same sacrifice when $c_i/\bar{c}_i = c_j/\bar{c}_j$.

The key fairness principle for the aggregation across individuals is the following. When two individuals incur the same level of sacrifice, society ought to be indifferent between assigning a marginal increase in consumption to either of them. This ethical stand leads to the **proportional-sacrifice social welfare function**, formally characterized as a special case in Section 4. For each individual i, define the **equivalent consumption** at (c_i, ℓ_i) as the level of consumption k that makes the individual indifferent between the bundle (c_i, ℓ_i) and consuming k while working the laissez-faire labor supply $\bar{\ell}_i$. Formally, $e_i(c_i, \ell_i) = k$ if and only if $u\left(k, \bar{\ell}_i\right) = u(c_i, \ell_i)$. The proportional-sacrifice social welfare function is defined as

$$W^p \equiv \sum_i \bar{c}_i^\gamma \frac{\left[e_i(c_i, \ell_i)\right]^{1-\gamma}}{1-\gamma}, \text{ with } \gamma > 0.$$

The parameter γ is a free ethical parameter and measures the willingness of society to avoid inequalities in the level of sacrifice incurred by individuals. At the limit for $\gamma = 0$, society is indifferent to such inequalities and social welfare simplifies to the simple sum of individuals' utilities. As γ increases, society is less and less willing to trade off inequalities in sacrifice against a larger sum of consumption. At the limit for $\gamma \to \infty$, society attributes full priority to the individual with the largest sacrifice.

The equivalent consumption function e_i is a representation of the preferences of individual i: $e_i(c_i, \ell_i) = u(c_i, \ell_i) + v\left(\bar{\ell}_i\right)$. Thus, society maximizes the sum of weighted and transformed equivalent consumptions of individuals.

The weight attached to the equivalent consumption of each individual depends, through the laissez-faire bundle, on her skill level. This dependence is crucial to ensure equal consideration for all individuals when they incur the same sacrifice. To see this, note that the social marginal welfare weight (Saez and Stantcheva, 2016) of an individual at bundle $\left(c_i, \bar{\ell}_i\right)$ is

$$\frac{\partial W^p}{\partial c_i} = \frac{\partial}{\partial c_i}\left(\bar{c}_i^\gamma \frac{\left[e_i\left(c_i, \bar{\ell}_i\right)\right]^{1-\gamma}}{1-\gamma}\right) = \left(\frac{c_i}{\bar{c}_i}\right)^{-\gamma}.$$

The factor $\bar{c}_i^\gamma$—placing a larger weight on the utilities of high-skill individuals—is key in achieving equal concern for sacrifice. When two individuals i and j incur the same sacrifice $(\bar{c}_i - c_i/)\,\bar{c}_i = (\bar{c}_j - c_j)\,/\bar{c}_j$, also $c_i/\bar{c}_i = c_j/\bar{c}_j$, and society is indifferent between allocating a marginal increase in consumption to either i or j.

2.2 A comparison with utilitarianism: first best

We first assume away the asymmetric information problem: the government covers the budget requirement R by levying an individual-specific lump-sum tax T_i. By the quasi-linear utility function and lump-sum taxation, the labor supply is at the laissez-faire level. Thus, the maximization problem of a (generalized) utilitarian society (with isoelastic inequality aversion $\rho \geq 0$) simplifies to

$$\begin{array}{rl} \max_{\{T_i\}} & \sum \frac{\left[u\left(w_i\bar{\ell}_i - T_i, \bar{\ell}_i\right)\right]^{1-\rho}}{1-\rho} \\ \text{s.t.} & \sum_i T_i \geq R. \end{array}$$

Similarly, the first-best maximization problem for the proportional-sacrifice social welfare function simplifies to

$$\begin{array}{rl} \max_{\{T_i\}} & \frac{(w_i\bar{\ell}_i)^{\gamma}(w_i\bar{\ell}_i - T_i)^{1-\gamma}}{1-\gamma} \\ \text{s.t.} & \sum_i T_i \geq R. \end{array}$$

The first-best optimum for the utilitarian criterion is instructive. If $\rho = 0$, the distribution of consumption is irrelevant since social welfare is linear in consumptions. Thus, the lump-sum taxes are not uniquely defined. Instead, when $\rho > 0$ and small (formally at the limit for $\rho \to 0$), the optimal lump-sum taxes are set to equalize the levels of consumption. This redistribution is extreme: at the optimum, high-skill individuals achieve a lower level of utility than low-skill ones. Utilitarianism "forces" high-skill individuals to produce for the sake of providing more consumption to low-skill individuals. Only at the limit for $\rho \to \infty$, when the criterion is "Rawlsian," are utilities equalized.

In contrast, the first-best optimum for the proportional-sacrifice social welfare function requires the lump-sum tax to be a fixed proportion of the laissez-faire income, independently of $\gamma > 0$.[10] Combining the first-order conditions on the lump-sum taxes

[10] Note that $\gamma = 0$ is excluded and emerges only as a limit case. The reason is technical. When $\gamma = 0$, the criterion is insensitive to the distribution of individuals' sacrifice and thus the notion of sacrifice itself cannot be singled out from the axioms.

of i and j leads to

$$\frac{T_i^*}{T_j^*} = \frac{w_i \bar{\ell}_i}{w_j \bar{\ell}_j} = \frac{\bar{c}_i}{\bar{c}_j} = \frac{c_i^*}{c_j^*}.$$

Therefore, the high-skill individuals combine a larger labor supply with a larger consumption. In contrast, low-skill individuals work less and consume less. This correlation between consumption and labor supply emerges from the proportional-sacrifice social welfare function attributing relatively more weight to the high-skill individuals. All individuals must contribute to the tax burden so as to incur the same level of sacrifice, here measured as a proportion of laissez-faire consumption.

Nevertheless, the proportional-sacrifice optimum does not ensure that the utility of high-skill individuals be higher than that of low-skill individuals. This is an immediate implication of efficiency and the size of the budget of the government. To illustrate, consider the extreme case of the budget of the government being equal to the laissez-faire income ($R = \sum_i w_i \bar{\ell}_i$). At the optimum, each individual's tax burden is her laissez-faire income. Then, consumption is zero and equal across individuals, while the labor supply is unchanged and penalizes (in terms of utility) the high-skill individuals (who supply more labor). Consider instead the other extreme, when the government need not raise money ($R = 0$). The utilitarian optimum equalizes consumptions, while the proportional-sacrifice optimum avoids any taxation. Then, with equal sacrifice high-skill individuals are better off than low-skill ones.

2.3 A comparison with utilitarianism: second best

Assume now that types are private information of the individuals and consider the case of two individuals i and j, with $w_i > w_j$. The government sets a tax schedule T associating a level of taxes $T(y)$ to each level of income y. Let $y_i \equiv w_i \ell_i$ and $y_j \equiv w_j \ell_j$. For both social welfare functions, the government's problem has the same structure as in Stiglitz (1982). Thus, the government maximizes its objective subject to the budget requirement

$$T(y_i) + T(y_j) = y_i + y_j - c_i - c_j \leq R,$$

and subject to the incentive compatibility constraints

$$ICC_i: \ u\left(c_i, \frac{y_i}{w_i}\right) \geq u\left(c_j, \frac{y_j}{w_i}\right),$$

$$ICC_j: \ u\left(c_j, \frac{y_j}{w_j}\right) \geq u\left(c_i, \frac{y_i}{w_j}\right).$$

With a utilitarian government, ICC_i is always binding. The government would like the high-skill individuals to achieve a lower utility than the low-skill ones. However, the high-skill individuals can always mimic the low-skill ones and achieve a higher utility (the utility cost of earning y_j is smaller). Thus, the incentive compatibility constraint of the high-skill individuals is always binding. As a result, the optimal tax schedule requires $T'(y_i) = 0$ and $T'(y_j) > 0$: the labor supply choice of the high-skill individuals is undistorted, while the labor supply choice of the low-skill individuals is distorted downward. The government trades off labor supply distortions against information rents.

With a proportional-sacrifice government, two cases can emerge. When the government budget requirement is low (R is below a threshold $\bar{R} > 0$), neither ICC_i nor ICC_j bind. The government need not distort labor supply choices and proportional income taxation can be implemented. When, instead, the government budget requirement is large ($R > \bar{R}$), ICC_i is binding. The government agrees that after-tax income ought to be larger for the high-skill individual. Nevertheless, the large tax rate needed to cover the budget would push the high-skill individual to mimic the low-skill ones. To avoid it, the government distorts the labor supply of the low-skill individuals downwards.

Thus, second-best optimal policies might not be very different across criteria. The difference will depend on the budget requirement, on the distribution of wages in the population, and on individuals' behavioral responses.

3 Model and axioms

3.1 Model

The set of individuals is $I \subset \mathbb{N}$; it is finite and satisfies $|I| \geq 3$. Individuals differ by their labor skills, reflected in their wage rates: for each $i \in I$, let $w_i > 0$ denote the wage rate of individual i.

Each individual $i \in I$ supplies labor $\ell_i \geq 0$, earns income $y_i \equiv w_i \ell_i$, and consumes $c_i \geq 0$. Her preferences are represented by a utility function $u(c_i, \ell_i)$, which is continuous, increasing in c_i, decreasing in ℓ_i, and strictly concave. We assume consumption is an essential good, that is, $\lim_{c \to 0} u_c = \infty$.

An allocation $a \equiv \left(\{c_i, \ell_i\}_{i \in I}\right)$ specifies a bundle (c_i, ℓ_i) for each individual $i \in I$. Let A be the set of all allocations.

Social preferences are a complete, transitive, and continuous preference relation $\succsim$ on the set of allocations A. For each pair $a, a' \in A$, $a \succsim a'$ means that a is socially at least as desirable as a'. The asymmetric and symmetric counterparts of $\succsim$ are denoted $\succ$ and $\sim$. Social welfare can be represented by a continuous social welfare function $W : A \to \mathbb{R}$. Thus, for each pair $a, a' \in A$, $a \succsim a'$ if and only if $W(a) \geq W(a')$.

3.2 "Sum-of-utilities" social welfare functions

As standard, we require social preferences to satisfy the Pareto principle. In other words, if individuals are made better off, social welfare cannot decrease.

Efficiency: *For each pair $a, a' \in A$, if $u(c_i, \ell_i) \geq u(c_i', \ell_i')$ for each $i \in I$ and $u(c_i, \ell_i) > u(c_i', \ell_i')$ for some $i \in I$, then $a \succsim a'$.*

Next, we impose inequality aversion on social preferences by requiring social preferences to be strictly convex.[11]

Inequality aversion: *For each pair $a, a' \in A$ and each $\beta \in (0,1)$, $a \sim a'$ implies $\beta a + (1-\beta) a' \succ a$.*

Finally, we impose that social welfare comparisons do not depend on the bundle assigned to an unconcerned individual. Denote by (a_i, a_{-i}) the allocation $a \in A$ that assigns $a_i \equiv (c_i, \ell_i)$ to individual i and $a_{-i} \equiv (c_j, \ell_j)_{j \in I \setminus \{i\}}$ to the other individuals.

Separability: *For each $a, a' \in A$, each $i \in I$, and each $\bar{a}_i = \left(\bar{c}_i, \bar{\ell}_i\right)$, $(a_i, a_{-i}) \succsim \left(a_i, a'_{-i}\right)$ if and only if $(\bar{a}_i, a_{-i}) \succsim \left(\bar{a}_i, a'_{-i}\right)$.*

[11] Convexity is significantly weaker than what is generally assumed in the literature, where this condition is supplemented with some form of symmetry or anonymity. In fact, most social welfare functions satisfy strict convexity. Strict convexity—rather than convexity—avoids a technical issue: when social preferences are linear, inequalities are irrelevant and the axioms cannot identify how to measure inequalities in sacrifice. Convex social preferences then emerge as a limit case.

Efficiency, *inequality aversion*, and *separability* imply the social welfare function belongs to a very general class of criteria. By *efficiency*, society evaluates individuals through their own preferences: W can be written as a function of the utilities achieved by each individual. By *inequality aversion*, social preferences are strictly convex with respect to the allocation and, thus, W is strictly concave in its arguments. By *separability*, the assignment of individual i does not matter for how society trades off the utility of individuals j and k; thus, W is additively separable.

Let a social welfare function $W : A \to \mathbb{R}$ be a **sum-of-utilities** social welfare function if there exist real-valued functions $(P_i)_{i \in I}$ such that for each $a \in A$,

$$W(a) = \sum_{i \in I} P_i \left(u \left(c_i, \ell_i \right) \right), \tag{1}$$

with $P_i \left(u \left(c_i, \ell_i \right) \right)$ continuous, strictly increasing, and strictly concave for each $i \in I$.

The choice of the functions $(P_i)_{i \in I}$ will be determined by later axioms. We refer to the functions $(P_i)_{i \in I}$ as **Pareto functions** to highlight these generalize the more standard Pareto weights (for which each P_i is defined by a multiplicative constant). As a result, the family of sum-of-utilities social welfare functions is significantly more general than usual. Unlike in Mirrlees (1971), the functions $(P_i)_{i \in I}$ need not be equal across individuals. Unlike in weighted utilitarianism (d'Aspremont and Gevers, 1977; Maskin, 1978), the functions $(P_i)_{i \in I}$ need not be increasing affine transformations. As we discuss in the following, this degree of freedom is necessary to incorporate principles of justice inspired by equality of sacrifice. Before doing so, we formalize the implication of the above axioms. The proof is in Appendix A.

Lemma 1. *Social preferences $\succsim$ satisfy efficiency, inequality aversion, and separability if and only if $\succsim$ can be represented by a sum-of-utilities social welfare function.*

3.3 Averting unequal sacrifice

In this subsection, we introduce axioms that ultimately *(i)* restrict the admissible definitions of sacrifice, *(ii)* discipline how social preferences ought to compare distributions of sacrifice, and thus, jointly with the previous axioms, *(iii)* characterize the Pareto functions $(P_i)_{i \in I}$ consistent with social aversion to inequality in sacrifice.[12]

[12] A different approach is to start with a cardinally measurable and interpersonally comparable index of sacrifice of individuals. With such rich information, one could just require that social welfare

To start with, define the **laissez-faire allocation** $\bar{a}$. At the laissez-faire allocation, each individual freely chooses how much labor to supply and consumes her entire income. Formally, the laissez-faire bundle of each individual $i \in I$ is $\left(\bar{c}_i, \bar{\ell}_i\right)$ such that $u\left(\bar{c}_i, \bar{\ell}_i\right) \geq u\left(c_i, \ell_i\right)$ for each $\left(c_i, \ell_i\right)$ with $c_i \leq \ell_i w_i$. Crucially, since individuals entirely appropriate the returns from their own work, individuals make no sacrifice.[13]

Let individual $i's$ **(implicit) tax burden** at the bundle $\left(c_i, \bar{\ell}_i\right)$ be measured by the difference in consumption with the laissez-faire allocation, that is, $b_i \equiv \bar{c}_i - c_i$. By definition of laissez-faire, an individual working $\bar{\ell}_i$ has a gross income of $\bar{y}_i \equiv \bar{\ell}_i w_i$. At the laissez-faire allocation, individual i would consume the entire income $\bar{c}_i = \bar{y}_i$. At the bundle $\left(c_i, \bar{\ell}_i\right)$ instead, individual i works the same time $\bar{\ell}_i$ but consumes c_i. Then, the tax burden is the difference between these consumption levels.[14] Intuitively, each individual's sacrifice increases the larger her tax burden is.

The first principle of equal sacrifice tells us that society should avert situations whereby one individual makes a sacrifice while another individual does not. In other words, individuals should solidarily bear the cost of taxation. We state this ideal in the form of a transfer principle. More precisely, assume that at an allocation $a \in A$, individual i has a positive tax burden $b_i > 0$ (i consumes less than at the laissez-faire bundle), while individual j has a negative tax burden $b_j < 0$ (j consumes more than at the laissez-faire bundle). This distribution of the tax burden is unfair according to the equal-sacrifice principle. Then, *ceteris paribus*, a transfer of consumption from i to j increases further the tax burden of i while decreasing that of j. This distribution of the tax burden is even more unfair and thus social welfare cannot be higher.

Tax solidarity: *For each pair $a, a' \in A$, each pair $i, j \in I$, and each $\varepsilon > 0$, such*

decreases when sacrifice is transfered from a low-sacrifice individual to a high-sacrifice individual. However, without a theory of how to measure sacrifice at each allocation, the corresponding welfare criterion would not be applicable. Our approach is more ambitious. Here, the index of sacrifice emerges endogenously from the axioms as a way to represent the social ranking of allocations.

[13] This is a natural choice in the Mirrleesian framework. In general, however, the no-sacrifice allocation is a more controversial choice. For example, with general equilibrium effects on wages, the level of taxation can affect the relative productivity of individuals. More drastically, the government might not be able to ensure property rights without taxation, making the laissez-faire allocation undefined. Importantly, our results extend to alternative choices of no-sacrifice allocations when the axioms are modified accordingly.

[14] The fixed labor supply can be interpreted as a *ceteris paribus* assumption. Its importance is easily explained: when paying a positive income tax, individuals may adjust labor supply upward to compensate for the lost income. If labor supply is allowed to vary, the extent of this income effect will matter for measuring the tax burden. In some situations, the individual might end up consuming more than at the laissez-faire (when leisure is a Giffen good), but is still worse off.

that:

- $b'_i - \varepsilon = b_i \geq 0 \geq b_j = b'_j + \varepsilon$;
- $\ell_i = \ell'_i = \bar{\ell}_i$ *and* $\ell_j = \ell'_j = \bar{\ell}_j$; *and*
- $(c_k, \ell_k) = (c'_k, \ell'_k) = (\bar{c}_k, \bar{\ell}_k)$ *for each* $k \in I / \{i, j\}$;

then, $a \succsim a'$.

Tax solidarity implies that (together with *efficiency*) no taxation is optimal when the government's budget is $R = 0$. The intuition is immediate. The $R = 0$ budget condition means taxation is not needed, but can be introduced for the sake of redistribution. However, *tax solidarity* tells us that redistribution away from the laissez-faire allocation cannot improve social welfare. Thus, when $R = 0$, the laissez-faire allocation is optimal and no taxation should be adopted. Moreover, no individual makes any sacrifice and equal sacrifice is achieved.[15]

The next ethical principle deals with a different type of unfairness. Without loss of generality, assume individual i's consumption at the laissez-faire allocation is larger than j's, that is, $\bar{c}_i \geq \bar{c}_j$. At an allocation $a \in A$, individual i has a smaller tax burden than j does, that is, $0 \leq b_i < b_j$; labor supply is that of the laissez-faire allocation. Individuals earn incomes $w_i \bar{\ell}_i = \bar{c}_i \geq \bar{c}_j = w_j \bar{\ell}_j$ and consume $c_i > c_j$. Crucially, $b_i < b_j$ implies the difference in earnings is smaller than the difference in consumption: the tax burden imposed on individuals exacerbates inequality. Consider now increasing further the tax burden of j, while further reducing that of i. This transfer of consumption makes the allocation more unfair and thus cannot improve social welfare.

Fair burden: *For each pair* $a, a' \in A$, *each pair* $i, j \in I$ *with* $\bar{c}_i \geq \bar{c}_j$, *and each* $\varepsilon > 0$, *such that:*

- $0 \leq b'_i + \varepsilon = b_i < b_j = b'_j - \varepsilon$;
- $\ell_i = \ell'_i = \bar{\ell}_i$ *and* $\ell_j = \ell'_j = \bar{\ell}_j$; *and*

[15] This axiom rules out redistribution motives. However, it captures the ideal of equal sacrifice. The interpretation is that individuals entirely deserve the wage rate their labor supply gives, not more nor less. The fathers of the equal-sacrifice principles suggest to deal with poverty by ensuring individuals have a subsistence consumption. Within our framework, a possible solution is to introduce a threshold of consumption s and measure the tax burden by $b_i \equiv \max\{\bar{c}_i, s\} - c_i$. Then, at laissez-faire allocation, lower-skill individuals have a positive "burden" (when $s > \bar{c}_i$) and redistribution is optimal, even when $R = 0$.

- $(c_k, \ell_k) = (c'_k, \ell'_k) = \left(\bar{c}_k, \bar{\ell}_k\right)$ *for each* $k \in I/\{i,j\}$;

then, $a \succsim a'$.

Fair burden deals with situations whereby the sacrifice of individual i (who is more productive) is too small relative to that of some other less productive individual.

Next, we discipline how social welfare deals with situations whereby the sacrifice of individual i is too large. As before, assume individual i's consumption at the laissez-faire allocation is larger than j's, that is, $\bar{c}_i \geq \bar{c}_j$. At allocation $a \in A$, individual i's consumption is smaller than j's, that is, $c_i < c_j$; labor supply is that of the laissez-faire allocation. Individuals earn incomes $w_i\bar{\ell}_i = \bar{c}_i \geq \bar{c}_j = w_j\bar{\ell}_j$ and consume $c_i \leq c_j$. The sacrifice of i is so large that, net of the sacrifice, the consumption of i is smaller than that of j. Consider now making the sacrifice of i even harsher by reducing her consumption for the benefit of j. This change makes the allocation more unfair and cannot improve social welfare.

Fair reward: *For each pair* $a, a' \in A$, *each pair* $i, j \in I$ *with* $\bar{c}_i \geq \bar{c}_j$, *and each* $\varepsilon > 0$, *such that:*

- $c'_i + \varepsilon = c_i < c_j = c'_j - \varepsilon$;
- $\ell_i = \ell'_i = \bar{\ell}_i$ *and* $\ell_j = \ell'_j = \bar{\ell}_j$; *and*
- $(c_k, \ell_k) = (c'_k, \ell'_k) = \left(\bar{c}_k, \bar{\ell}_k\right)$ *for each* $k \in I/\{i,j\}$;

then, $a \succsim a'$.

We represent the three axioms introduced above in figure 2. On the Cartesian plane, the consumptions of individuals i and j are represented on the axes. Consumption levels $\bar{c}_i$ and $\bar{c}_j$ are those of the laissez-faire allocation.[16] Without loss of generality, here i is again the individual consuming more at the laissez-faire allocation, that is, $\bar{c}_i \geq \bar{c}_j$. Not represented, the labor supply choices are those of the laissez-faire allocation.

[16]These consumption levels can be interpreted as legitimate claims of individuals. In a "claims problem," the objective is to fairly allocate an endowment that is not sufficient to cover the claims (Thomson, 2019). Interestingly, the "path of awards of a rule"—that is, the locus of assignments associated to each level of endowment—is related to our sacrifice function—that is, the locus of allocations where individuals incur the same sacrifice (a similar point emerges also in Chambers and Moreno-Ternero, 2017 and Stovall, 2020). Our focus on complete rankings and the framework (multidimensional and with individuals' choices) make the axiomatic analysis very different. For a related approach in the context of intergenerational justice, see Piacquadio (2020).

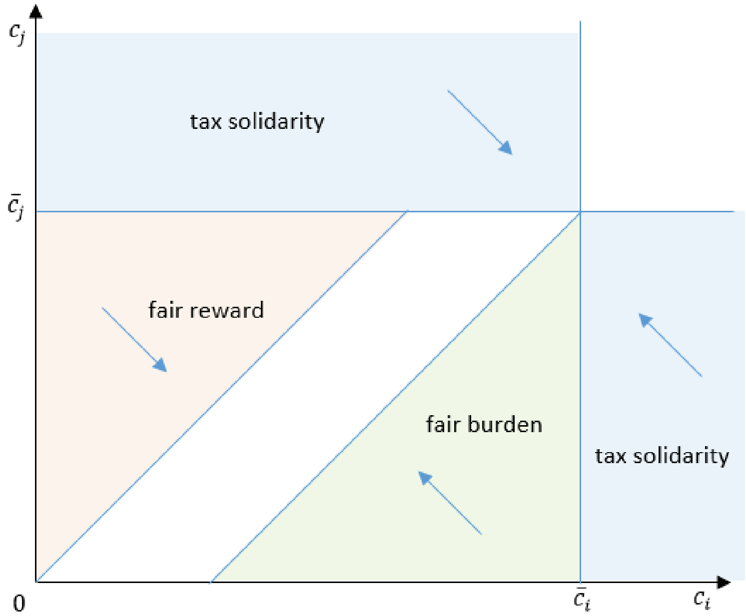


Figure 2: Equal-sacrifice principles.

The northwest and southeast areas from the laissez-faire consumptions are those where *tax solidarity* applies. These areas are characterized by one individual making a sacrifice, while the other does not. The arrow pointing toward the laissez-faire consumptions $(\bar{c}_i, \bar{c}_j)$ represents the direction of increasing social welfare.

Southwest of the laissez-faire consumptions is the area where both individuals make a sacrifice. *Fair burden* applies below the 45 degree line from the laissez-faire consumptions. In this area, the tax burden of j is larger than that of i. Again, the arrow suggests that transferring consumption from i to j increases social welfare.

Finally, the portion of the area above the 45 degree line from the origin is such that the consumption of j is larger than that of i. In this area, *fair reward* applies. In this area, the tax burden of i is so large that her consumption is now smaller than j's. The arrow points to reducing j's consumption for the benefit of i's. *Fair reward* suggests this transfer increases social welfare.

To further illustrate our axioms, it is helpful to contrast them to the Pigou-Dalton transfer principle, satisfied by the utilitarian criterion. The Pigou-Dalton transfer principle requires that a transfer from a poorer to a richer individual reduces social welfare. Let utility be additively separable in consumption and leisure. Then in the above graph, the arrow of increasing social welfare for the Pigou-Dalton principle always points to the 45 degree line, independently of the laissez-faire income of the individuals. Thus, the standard Pigou-Dalton transfer principle agrees with the implication of both *fair burden* and *fair reward*, while these remain weaker.[17] Such a weakening is necessary to accommodate the view expressed by survey respondents (see Saez and Stantcheva, 2016). They overwhelmingly support the view that a family earning \$50,000 and paying \$15,000 of taxes is more deserving than an (otherwise identical) family earning \$40,000 and paying \$5,000. This means that—in contrast to the Pigou-Dalton transfer principle and utilitarianism—the arrow of social improvement for the survey respondents points away from an equal after-tax income of \$35,000 to the benefit of the higher-income earner. Similar survey evidence is shown in Schokkaert and Devooght (2003) and Weinzierl (2014).

4 The sacrifice-based welfare criteria

4.1 Comparisons of sacrifice

First, we identify a counterfactual consumption level that, when combined with an individuals' laissez-faire labor supply, ensures the same level of well-being. For each allocation $a \in A$ and each individual $i \in I$, let the **equivalent consumption of $\boldsymbol{i}$ at $\boldsymbol{a}$** be the level of consumption $e_i(c_i, \ell_i)$ such that

$$e_i(c_i, \ell_i) = k \iff u(c_i, \ell_i) = u\left(k, \bar{\ell}_i\right),$$

where $\bar{\ell}_i$ is the labor supply at the laissez-faire allocation.

Next, we define the **sacrifice function** $S : \mathbb{R}_+ \times \mathbb{R}_{++} \to \mathbb{R}$. This function measures the sacrifice made by each individual in an interpersonally comparable way. Let individual $i \in I$ be assigned the bundle $\left(c_i, \bar{\ell}_i\right)$; then, i's sacrifice is given by $S(c_i, \bar{c}_i)$. More generally, i's sacrifice at bundle (c_i, ℓ_i) is given by $S(e(c_i, \ell_i), \bar{c}_i)$. Let the sacrifice function S be decreasing in the first argument, increasing in the second argument, and continuous. Furthermore, it satisfies the following restrictions:

- [*zero sacrifice normalization*] $x = y$ implies $S(x, y) = 0$;

[17] As the graphical representation suggests, the axiom that conflicts with the Pigou-Dalton transfer principle and, thus, rules out utilitarianism is *tax solidarity*.

- [*slope bound for positive sacrifice*] whenever $S(x, y) = S(x', y') > 0$, then $|x - x'| \leq |y - y'|$.

Let $\mathcal{S}$ be the domain of these functions. Importantly, the sacrifice function is ordinal, because it represents only an ordering of levels of sacrifice incurred by any two individuals. However, this does not exclude defining the sacrifice function based on a cardinal utility function of individuals, as originally suggested by Mill (1848).

In figure 3, we illustrate how the sacrifice function works. On the left part of the Cartesian plane, individual i faces the wage level w_i and chooses the utility-maximizing bundle $\left(\bar{c}_i, \bar{\ell}_i\right)$. Similarly, individual j with wage level w_j chooses the utility-maximizing bundle $\left(\bar{c}_j, \bar{\ell}_j\right)$. The bundles $\left(\bar{c}_i, \bar{\ell}_i\right)$ and $\left(\bar{c}_j, \bar{\ell}_j\right)$ constitute the laissez-faire allocation. The corresponding levels of consumption $\bar{c}_i$ and $\bar{c}_j$ are reported by the horizontal axis of the right part of the Cartesian plane, where we represent the sacrifice function through its **isosacrifice curves**. At the 45 degree line, the level of sacrifice is 0. The level of sacrifice decreases with the assigned consumption and increases with the laissez-faire consumption, making the isosacrifice curves increasing. The slope bound implies that, for positive levels of sacrifice, the slope of the sacrifice function cannot exceed 1. Equal sacrifice is progressive if the sacrifice function—for positive levels of sacrifice—is concave (Young, 1990).

Let individual i be assigned the bundle (c_i, ℓ_i). Her level of utility is the same as if she was assigned her equivalent consumption $e_i(c_i, \ell_i)$ and the laissez-faire labor supply $\bar{\ell}_i$. The implicit tax burden of i is given by the difference between $\bar{c}_i$ and $e_i(c_i, \ell_i)$. Interpersonal comparisons of sacrifice are made through the isosacrifice curve of level $S(e_i(c_i, \ell_i), \bar{c}_i)$. Individual i makes a larger sacrifice than j whenever $S(e_i(c_i, \ell_i), \bar{c}_i) \geq S(e_j(c_j, \ell_j), \bar{c}_j)$.

4.2 Pareto functions

Given a sacrifice function, we need to define the Pareto functions that are consistent with such a sacrifice function.

For each i and each bundle (c_i, ℓ_i), denote by $\beta_i(c_i, \ell_i)$ the **social marginal welfare weight** of i at bundle (c_i, ℓ_i). For a "sum-of-utility" social welfare function,

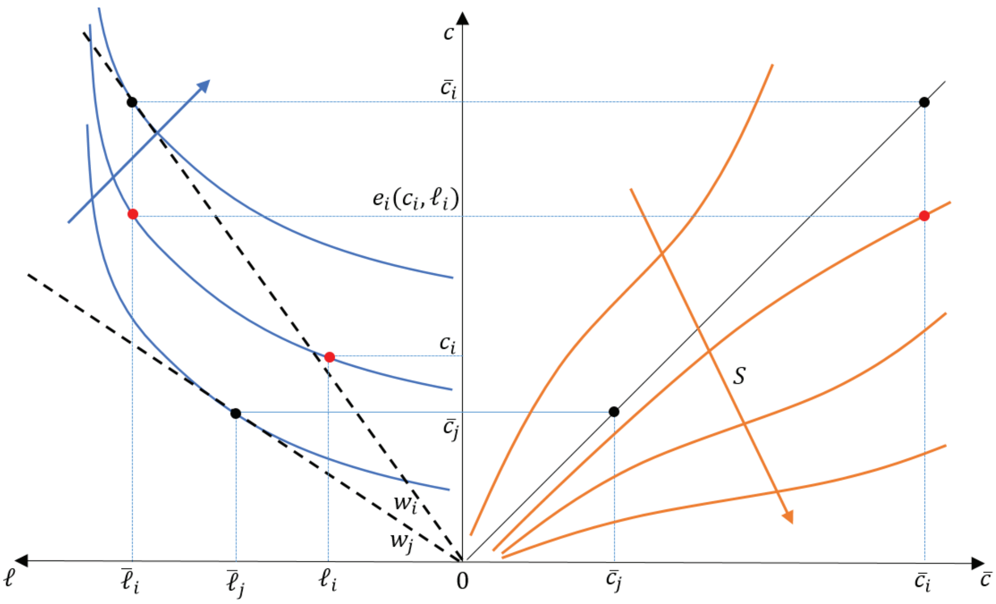


Figure 3: Isosacrifice curves and comparability in terms of sacrifice.

18

$$\beta_i(c_i, \ell_i) \equiv \frac{\partial}{\partial c_i} P_i(u(c_i, \ell_i)).$$

Then, we impose the Pareto functions $(P_i)_{i \in I}$ satisfy the following requirement. There exists a sacrifice function $S \in \mathcal{S}$ and a real-valued increasing function g such that for each i and each $c_i \in \mathbb{R}_+$:

1. $\beta_i\left(c_i, \bar{\ell}_i\right) = g\left(S\left(c_i, \bar{c}_i\right)\right) > 0$; and

2. $P_i(u(c_i, \ell_i))$ is strictly concave in its arguments.

To explain, since the social marginal welfare weights are positive, the Pareto functions are increasing. Thus, the contributions to social welfare of individuals are representations of their preferences. Condition 1 also imposes equality of social marginal welfare weights when the level of sacrifice incurred by individuals is the same, as identified by the sacrifice function S. Furthermore, since g is increasing, the social marginal welfare weights are higher for individuals incurring a larger level of sacrifice. Condition 2 ensures concavity of individuals' contributions to social welfare and thus aversion to inequality in sacrifice.

[18]We adopt the convention that $\frac{\partial}{\partial c_i} P_i(u(c_i, \ell_i))$ denotes the one-sided derivative at the boundary of the consumption space (when consumption is 0).

When the above conditions are satisfied, we say the Pareto functions $(P_i)_{i\in I}$ are **consistent** with the sacrifice function S.

4.3 The welfare criterion

The **equal-sacrifice social welfare function** $W : A \to \mathbb{R}$ is defined by setting for each $a \in A$,

$$W(a) \equiv \sum_{i\in I} P_i(u(c_i, \ell_i)),$$

where the Pareto functions $(P_i)_{i\in I}$ are consistent with a sacrifice function $S \in \mathcal{S}$.

Our main result shows the above ethical principles characterize the family of equal-sacrifice social welfare functions.

Theorem 1. *Social welfare $\succsim$ satisfies efficiency, inequality aversion, separability, tax solidarity, fair burden, and fair reward if and only if it can be represented by an equal-sacrifice social welfare function.*

4.4 The proportional-sacrifice social welfare function

In Section 2, we introduced the proportional-sacrifice social welfare functions, a special case of the family of equal-sacrifice social welfare functions. Two ethical choices characterize these criteria. First, the sacrifice function S is proportional, that is, $S(c, \bar{c}) = (\bar{c} - c)/\bar{c}$. Second, the social attitude toward inequality in sacrifice is captured by a unique parameter $\gamma > 0$. In the following, we provide a characterization of these criteria.

The following principle strengthens *fair burden* to deal with situations in which the sacrifice of individual i (who is better off at the laissez-faire allocation) is relatively too small as opposed to that of some other individual j. In these cases, a regressive transfer from j to i cannot improve social welfare.

Fair relative burden: *For each pair $a, a' \in A$, each pair $i, j \in I$ with $\bar{c}_i \geq \bar{c}_j$, and each pair $\alpha, \varepsilon > 0$, such that:*

- $c'_i + \varepsilon = c_i > \alpha\bar{c}_i$ *and* $\alpha\bar{c}_j > c_j = c_j - \varepsilon$*;*
- $\ell_i = \ell'_i = \bar{\ell}_i$ and $\ell_j = \ell'_j = \bar{\ell}_j$; and
- $(c_k, \ell_k) = (c'_k, \ell'_k) = (\bar{c}_k, \bar{\ell}_k)$ *for each* $k \in I/\{i, j\}$*;*

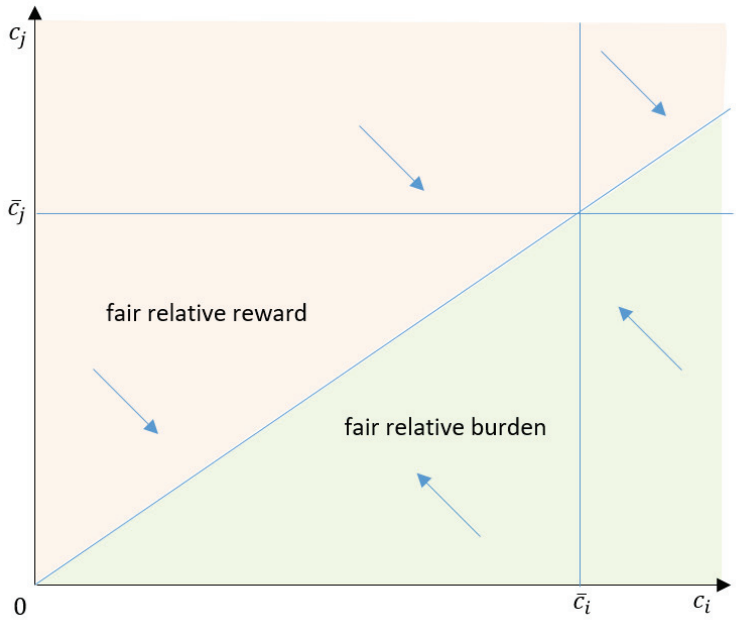


Figure 4: Relative equal-sacrifice principles.

then, $a \succsim a'$.

The next principle strengthens *fair reward* and disciplines situations in which the sacrifice of individual i (who is better off at the laissez-faire allocation) is relatively too large as opposed to that of some other individual j. In these cases, a regressive transfer from i to j cannot improve social welfare.

Fair relative reward: *For each pair* $a, a' \in A$*, each pair* $i, j \in I$ *with* $\bar{c}_i \geq \bar{c}_j$*, and each pair* $\alpha, \varepsilon > 0$*, such that:*

- $c'_i + \varepsilon = c_i < \alpha\bar{c}_i$ *and* $\alpha\bar{c}_j < c_j = c_j - \varepsilon$*;*
- $\ell_i = \ell'_i = \bar{\ell}_i$ and $\ell_j = \ell'_j = \bar{\ell}_j$; and
- $(c_k, \ell_k) = (c'_k, \ell'_k) = \left(\bar{c}_k, \bar{\ell}_k\right)$ *for each* $k \in I / \{i, j\}$*;*

then, $a \succsim a'$.

We represent these principles in figure 4.

Finally, we introduce a weak form of scale invariance with respect to the consumption of individuals. Assume all individuals work at the laissez-faire labor supply. Then, proportional changes in consumptions do not affect how society ranks two allocations.

Scale invariance: *For each $a, a', a'', a''' \in A$ and each $\kappa > 0$ such that:*

- $c_i = \kappa c_i''$ *and* $c_i' = \kappa c_i'''$ for each $i \in I$;
- $\ell_i = \ell_i' = \ell_i'' = \ell_i''' = \bar{\ell}_i$ for each $i \in I$;

then, $a \succsim a'$ *if and only if* $a'' \succsim a'''$.

The following lemma summarizes the relationships between the above axioms. The proof is omitted.

Lemma 2. *For a social welfare* $\succsim$*:*

1. fair relative burden implies fair burden;

2. fair relative reward implies fair reward;

3. fair relative burden and fair relative reward imply tax solidarity;

4. fair burden and scale invariance imply fair relative burden;

5. fair reward and scale invariance imply fair relative reward;

6. tax solidarity and scale invariance imply fair relative burden and fair relative reward.

We can now characterize the proportional-sacrifice social welfare function.

Theorem 2. *Social welfare* $\succsim$ *satisfies efficiency, inequality aversion, separability, tax solidarity and scale invariance if and only if it can be represented by a social welfare function* W^p *such that for each* $a \in A$,

$$W^p(a) \equiv \sum_{i \in I} P_i\left(u\left(c_i, \ell_i\right)\right),$$

where, for each $i \in I$ *and each* $c_i > 0$, P_i *satisfies*

$$\frac{\partial}{\partial c_i} P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) \equiv \beta_i\left(c_i, \bar{\ell}_i\right) = \left(\frac{c_i}{\bar{c}_i}\right)^{-\gamma},$$

for some $\gamma > 0$.

This result states that, for the proportional-sacrifice social welfare function, the Pareto functions $(P_i)_{i \in I}$ need to be consistent with the proportional-sacrifice function $S(c, \bar{c}) = (\bar{c} - c)/\bar{c}$. Moreover, their derivatives at the laissez-faire labor supply need to be a power transformation of $1 - S(c, \bar{c})$.

4.5 The (log-) progressive-sacrifice social welfare function

Next, we present another welfare criterion which builds on a progressive—rather than proportional—definition of equal sacrifice. Let the sacrifice function be **log-progressive** if:

$$S^{pr}(c,\bar{c}) = \frac{\ln(1+\bar{c}) - \ln(1+c)}{\ln(1+\bar{c})}.$$

The progressivity of this sacrifice function is best assessed by a numerical example. According to the log-progressive sacrifice function, the following situations correspond to (approximately) the same level of sacrifice, where income is measured in thousands of dollars:

	A	*B*	*C*	*D*
Pre-tax income	25	50	100	500
Taxes	6	15	33	215
After-tax income	19	35	67	285
Average tax rate	26%	30%	33%	43%

As before, let inequality aversion in sacrifice be captured by the unique parameter $\gamma > 0$. Then, the corresponding equal-sacrifice social welfare function is:

$$W^{pr} = \sum_{i \in I} P_i^{pr}(u(c_i, \ell_i)),$$

where, for each $i \in I$, P_i^{pr} satisfies

$$\frac{\partial}{\partial c_i} P_i^{pr}\left(u\left(c_i, \bar{\ell}_i\right)\right) = \left[\frac{\ln(1+c_i)}{\ln(1+\bar{c}_i)}\right]^{-\gamma}.$$

A comparison with the proportional-sacrifice social welfare function highlights that $c_i/\bar{c}_i = 1 - S(c_i, \bar{c}_i)$ is here simply replaced by $\ln(1+c_i)/\ln(1+\bar{c}_i) = 1 - S^{pr}(c_i, \bar{c}_i)$.

5 A simulation exercise

Next, we turn to the continuous version of the Mirrlees model for a simulation exercise. Individuals' wage rates are now continuously distributed according to $f(w)$ on $w \in$

$[w^b, w^t]$.[19] We assume a standard additively separable utility function $U(c_i, l_i) = \left(c_i^{1-\rho} - 1\right) / (1-\rho) - \alpha l_i^{\sigma} / \sigma$. Let the tax function be denoted by $T : \mathbb{R} \to \mathbb{R}$. For each $i \in I$, after-tax income is $y_i - T(y_i)$. Following Saez (2001), the first-order condition for optimal marginal tax rates $T'(y(w))$ is given by

$$\frac{T'(y(w))}{1 - T'(y(w))} = \frac{1 + \varepsilon^u(w)}{\varepsilon^c(w)} \frac{U_c(w)}{w f(w)} \int_w^{w^t} \left(\left(1 - \frac{P'_\theta (U(\theta)) U_c(\theta)}{\lambda} \right) \frac{1}{U_c(\theta)} f(\theta) d\theta \right),$$

where $\varepsilon^u(w)$ and $\varepsilon^c(w)$ are, respectively, the uncompensated and compensated labor supply elasticities. The marginal utility of consumption is $U_c(w)$ and the marginal social welfare weight of increasing consumption for the agent with skill w is $P'_w(U(w))U_c(w)$. Since the structure of the optimal tax problem is unaffected by the Pareto functions P_w chosen for each individual, all the standard results apply, including no negative tax rates and the zero marginal tax rate at the upper limit. Moreover, the necessary condition is also sufficient when the single-crossing condition is satisfied, that is pre-tax income is non-decreasing in wage rates.

Unfortunately, little more can be said analytically about optimal tax rates from the Mirrlees problem. Hence, we turn to numerical simulations. Our objective is to highlight the difference between the optimal tax schedules under the utilitarian and the equal-sacrifice social welfare functions.

We first compute the marginal social welfare weights for our criteria needed for the optimal tax formula. Equivalent consumption is given by

$$e_i(c_i, l_i) = \left((1-\rho) \left(U(c_i, l_i) + v(\bar{l}_i) \right) + 1 \right)^{1/(1-\rho)}.$$

The marginal social welfare weights for the proportional-sacrifice criterion are

$$\frac{\partial}{\partial c_i} W^p = \left(\frac{e(c_i, l_i)}{\bar{c}_i} \right)^{-\gamma} \frac{\partial e(c_i, l_i)}{\partial c_i} = (1-\rho)^{-1} \left(\frac{e(c_i, l_i)}{\bar{c}_i} \right)^{-\gamma} \left(\frac{e(c_i, l_i)}{c_i} \right)^{\rho}.$$

These marginal social welfare weights differ from those obtained in Section 2. Here, the additional term $\partial e(c_i, l_i) / \partial c_i$ differs from 1 (when $\ell_i \neq \bar{\ell}_i$) due to income effects, which were excluded earlier. When individuals distort their labor supply downwards ($l_i < \bar{l}_i$), the income effects amplify the well-being change of an increase

[19] The continuity of the welfare criterion with respect to the types of individuals—here identified by their wage rate—ensures that the continuous version of our criterion is well-defined.

in consumption: the indifference curves are not parallel, but rather become "closer" to each other with a decrease in labor supply. Society accepts the implications of these income effects by *efficiency*.

The marginal social welfare weights for the log-progressive sacrifice criterion are

$$\frac{\partial}{\partial c_i} W^{pr} = \left(\frac{e(c_i, l_i)}{\bar{c}_i} \right)^{-\gamma} \frac{\partial e(c_i, l_i)}{\partial c_i} = (1-\rho)^{-1} \left(\frac{\ln(1+e(c_i,l_i))}{\ln(1+\bar{c}_i)} \right)^{-\gamma} \left(\frac{e(c_i, l_i)}{c_i} \right)^{\rho}.$$

We set the same utility parameters as in Mankiw et al. (2009): $\rho = 1.5$, $\alpha = 2.55$, and $\sigma = 3$.[20] We also use the same income distribution parameters (for the US in 2007). However, we deviate from their study by including an exogenous revenue requirement, R, equal to 30% of total laissez-faire income.[21]

Our exercise consists in comparing the *utilitarian social welfare function* (with logarithmic concavity)

$$W^U = \int_w^{w^t} \ln\left[u\left(c_\theta, \ell_\theta\right)\right] f(\theta) d\theta$$

with the *proportional-sacrifice social welfare function* (when $\gamma = 1$)

$$W^p = \int_w^{w^t} \bar{c}_\theta \ln\left[e\left(c_\theta, \ell_\theta\right)\right] f(\theta) d\theta$$

and the *progressive-sacrifice social welfare function* (when $\gamma = 1$)

$$W^{pr} = \int_w^{w^t} \left(\int_{-\infty}^{e(c_\theta, \ell_\theta)} \frac{\ln(1+\bar{c}_\theta)}{\ln(1+x)} dx \right) f(\theta) d\theta,$$

where the θ subscript refers to the individual i with wage $w_i = \theta$.[22]

A few preliminary remarks are in order. First, the goal of this exercise is simply to illustrate the applicability of our results and argue that the policy implications of our proposal are reasonable. This ensures that one cannot reject our axioms and equal-sacrifice criterion by Rawls' reflective equilibrium argument. Second, the concern

[20] We thank Gregory Mankiw, Matthew Weinzierl, and Danny Yagan for making their data and code available.

[21] As a reminder, in the absence of a revenue requirement, the equal-sacrifice social welfare function would optimally select the laissez-faire outcome.

[22] The equal sacrifice social welfare functions are obtained by integrating the sacrifice function over consumption, i.e., $\int_{-\infty}^{c} \left(1 - S\left(x, \bar{c}\right)\right)^{-\gamma} dx$ with $\gamma = 1$, and evaluating the integral at the equivalent consumption, $c = e\left(c_\theta, \ell_\theta\right)$.

for the different types of inequality is set to logarithmic ($\gamma = 1$). The logarithm is recognized as a middle ground level of concavity, between linear (no concern for inequality) and infinite (full priority to the worst-off). We leave to future research the analysis of the inverse optimal taxation problem, whereby the ethical parameters of the welfare criteria are set to match the observed tax system (see Bourguignon and Spadaro, 2012).

The results are summarized by the graphs in figure 5, representing the marginal tax rate, the average tax rate, the after-tax incomes, and the utility schedule for the above criteria. The optimal tax system derived with the proportional-sacrifice social welfare function is less redistributive than the one derived with the progressive-sacrifice social welfare function, which is less redistributive than the one derived with the utilitarian criterion. Crucially, this does not imply that the equal-sacrifice criteria are insensitive to the well-being of low-income individuals. Rather, these criteria are characterized by a lower willingness of society to transfer to the worst-off individuals.

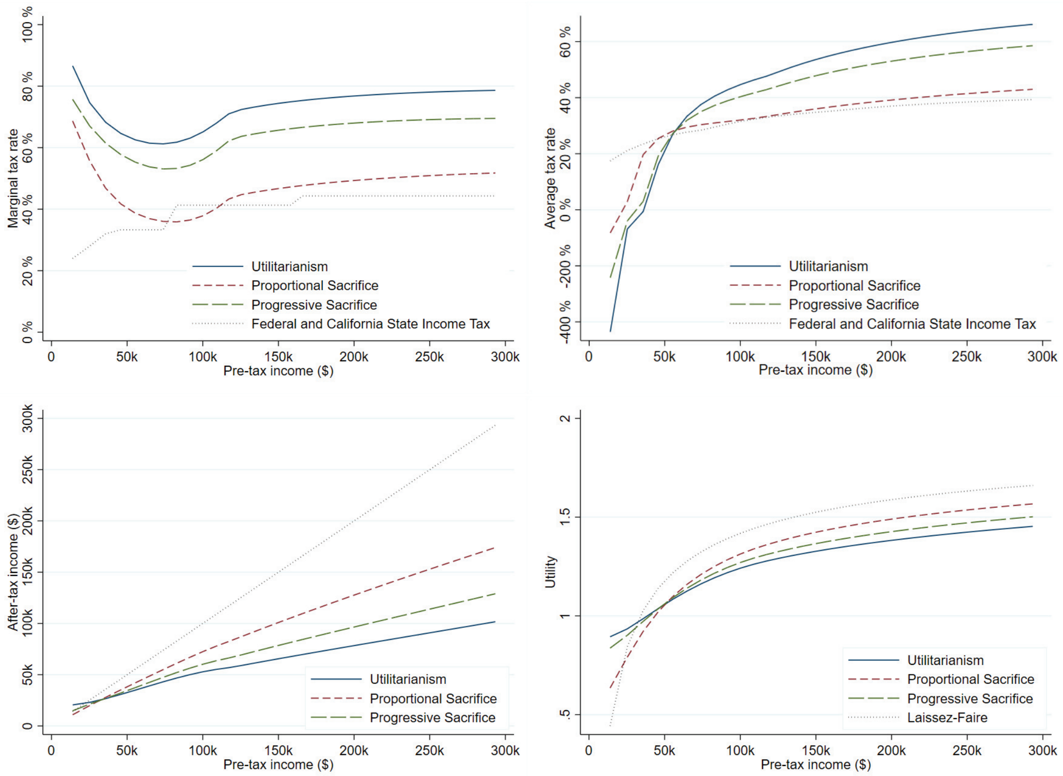


Figure 5: Utilitarian and equal-sacrifice criteria second-best optima: marginal tax rates, average tax rates, after-tax incomes, and utility levels.

This lower concern for redistribution appears to be consistent with real-world income taxes. In particular, the utilitarian second-best policy supports marginal tax rates above 60% for all individuals (and up to 80% for high-income individuals). In contrast, the proportional-sacrifice criterion supports marginal tax rates that are about 20 percentage points lower.[23] These marginal tax rates are similar to those observed in the US tax system. Here, we plot the 2007 marginal tax rates for the combined federal and California state taxes on the income of singles (since California sets the highest state income taxes, it provides the strongest case for the tax schedule implied by utilitarianism). Note that, for low incomes, the discrepancy of marginal tax rates with the US tax system can be explained by the absence of an extensive labor supply margin (see the discussion in Diamond and Saez, 2011).[24] A middle ground is given by the tax schedule supported by the progressive-sacrifice criterion.

The average tax rates are also informative. Utilitarianism suggests subsidies (negative average taxes) ought to be distributed to the bottom 35% of the population; the progressive-sacrifice criterion recommends subsidies to the bottom 30% of the population; the proportional-sacrifice criterion does so for only the bottom 15% of the population. In fact, utilitarianism implies a lump-sum transfer more than 5 times larger than that of the proportional-sacrifice criterion and twice as large as that of the progressive-sacrifice criterion. The flat tax provides an intuition about the size of the revenue requirement.

Contrary to what one might expect, the tax systems implied by our equal-sacrifice criteria redistribute on net to the lowest income levels. This is due to the presence of income effects in individuals' utility. The incentive compatibility constraint forces downward labor supply distorsions for the low-skill individuals. At such low labor supply, two contrasting forces define their second-best after-tax income. On the one hand, the negative sacrifice (due to the large lump-sum transfer) suggests society ought to increase their taxes for the benefit of higher income individuals. On the other hand, the income effects amplify the welfare effect of changes in their after-tax

[23]Note that the optimal tax schedule for the utilitarian criterion is not very sensitive to inequality aversion in utilities. At the extreme of no inequality aversion, the utilitarian social welfare function is the sum of (untransformed) utilities. The corresponding optimal tax schedule has marginal tax rates that are about 5 percentage points lower than with the log-utilitarian criterion.

[24]When the extensive margin is accounted for, optimal marginal tax rates at the lowest income levels are lower (Saez, 2002; Jacquet, Lehman, and van der Linden, 2013). We expect the introduction of an extensive margin to bring the equal-sacrifice optimal tax rates close to the US tax system also for low incomes. We leave this exercise to future research.

income. Thus, while the equity motive suggests an additional dollar be given to high-skill individuals, the efficiency motive dominates for the lowest income earners. This result disappears when society is infinitely averse to inequality in sacrifice ($\gamma \to \infty$), at which point the criterion requires minimizing the largest sacrifice.

The pre-tax/after-tax income schedule provides a different representation of the effect of the identified tax schedules. The apparently large reduction of after-tax income moving from, say, proportional sacrifice to utilitarianism is explained by two factors. First and most obviously, the distribution of pre-tax incomes is very skewed: a large number of low-wage earners benefit from increases in the level of taxation. Second, an increase in taxes is accompanied by a reduction in labor supply. As an implication, consider an individual earning $\$100,000$ in pre-tax income with the utilitarian tax schedule. The same individual supplies more labor and, thus, earns more with the proportional sacrifice criterion. Thus, labor supply responses mitigate the utility cost of taxation, as reflected by the relatively flat utility schedule.

6 Conclusion

The optimal choice of income taxation is a key question in public economics. The answer requires combining a positive model of the economy—capturing behavioral choices of individuals—with normative aspects—reflecting ethical principles about how to compare benefits and losses of individuals. However, since the seminal contribution of Mirrlees (1971), the literature has mostly advanced by considering richer models of the economy, while the normative criterion was generally utilitarian.

The utilitarian criterion is subject to a number of criticisms. Among those, Edgeworth (1897) highlights the utilitarian criterion imposes too strong a motivation to redistribute: with inelastic earnings, the optimal taxation policy is to tax income at 100% and redistribute the tax revenues equally across individuals.

The solution we propose in this paper builds on an old and well-known theory of fairness in taxation, namely equal sacrifice. Defined as a standard of perfection, equal sacrifice was unable to provide reasonable policy guidance: in second best settings, the equalization of sacrifice leads to inefficient tax schedules. In contrast, our results establish it is possible to construct a social welfare function combining fairness considerations based on the equal-sacrifice principle with a concern for efficiency. The main result of the paper is the axiomatic characterization of a family of equal-sacrifice social welfare functions, which prioritizes individuals making a larger sacrifice. Our criterion redeems utilitarianism's counterintuitive instances of a redistributive motive and thus can have large impacts on optimal tax policy.

To speak to those impacts, we show that second-best optimal tax policy differs most from the utilitarian one when the government's budget requirement is not too large and when the labor-supply elasticity is small. Then, we demonstrate how to apply the criterion in a Mirrlees model. In a numerical simulation for the US economy, we show that our equal-sacrifice social welfare function implies lower tax rates and less redistribution than utilitarianism does. Moreover, when sacrifice is defined as a proportion of taxes payed, our criterion roughly rationalizes the observed Californian tax schedule.

We conclude with a remark on how to bring our results to different settings. We believe our axiomatic approach innovatively shows how to bridge the gap between approaches to first-best distribution of resources—as addressed in the theory of fair allocations—and fine-grained welfare criteria—trading off equity and efficiency considerations and, thus, able to assess second-best policies. This is particularly important in settings where differences in marginal utilities are not enough information to accommodate widespread views on distributive justice. Examples include other taxation problems, as well as the provision of public goods, the allocation of health services, matching problems, and so on.

References

Berliant, M. and M. Gouveia (1993): "Equal sacrifice and incentive compatible income taxation," *The Journal of Public Economics*, 51, 219–240.

Binmore, K., A. Rubinstein, and A. Wolinsky (1986): "The Nash bargaining solution in economic modelling," *The RAND Journal of Economics*, 17, 176–188.

Bourguignon, F. and A. Spadaro (2012): "Tax–benefit revealed social preferences," *The Journal of Economic Inequality*, 10, 75–108.

Chambers, C. P. and J. D. Moreno-Ternero (2017): "Taxation and poverty," *Social Choice and Welfare*, 48, 153–175.

da Costa, C. E. and T. Pereira (2014): "On the efficiency of equal sacrifice income tax schedules," *The European Economic Review*, 70, 399–418.

d'Aspremont, C. and L. Gevers (1977): "Equity and the informational basis of collective choice," *The Review of Economic Studies*, 44, 199–209.

Diamond, P. and E. Saez (2011): "The case for a progressive tax: from basic research to policy recommendations," *Journal of Economic Perspectives*, 25, 165–90.

Edgeworth, F. Y. (1897): "The pure theory of taxation," *The Economic Journal*, 7, 46–70.

Feldstein, M. (1976): "On the theory of tax reform," *The Journal of Public Economics*, 6, 77–104.

Fleurbaey, M. and F. Maniquet (2006): "Fair income tax," *The Review of Economic Studies*, 73, 55–83.

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

Gauthier, D. (1986): *Morals by agreement*, Oxford: Oxford University Press.

Gorman, W. M. (1968): "The structure of utility functions," *The Review of Economic Studies*, 35, 367–390.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013): "Optimal redistributive taxation with both extensive and intensive responses," *The Journal of Economic Theory*, 148, 1770–1805.

Jacquet, L. and D. Van de Gaer (2011): "A comparison of optimal tax policies when compensation or responsibility matter," *The Journal of Public Economics*, 95, 1248–1262.

Jessen, R., M. Metzing, and D. Rostam-Afschar (2019): "Optimal taxation under different concepts of justness," *Working Paper*.

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): "Optimal taxation in theory and practice," *The Journal of Economic Perspectives*, 23, 147–74.

Maskin, E. (1978): "A theorem on utilitarianism," *The Review of Economic Studies*, 45, 93–96.

Mill, J. S. (1848): *Principles of political economy*, London: John W. Parker.

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

Mitra, T. and E. A. Ok (1996): "Personal income taxation and the principle of equal sacrifice revisited," *The International Economic Review*, 37, 925–948.

Musgrave, R. (1959): *Theory of public finance: A study in public economy*, New York: McGraw-Hill.

Ok, E. A. (1995): "On the principle of equal sacrifice in income taxation," *The Journal of Public Economics*, 58, 453–467.

Piacquadio, P. G. (2017): "A fairness justification of utilitarianism," *Econometrica*, 85, 1261–1276.

——— (2020): "The ethics of intergenerational risk," *Journal of Economic Theory*, 186, 104999.

Pigou, A. C. (1928): *A study in public finance*, Read Books Ltd.

Piketty, T. and E. Saez (2013): "Optimal labor income taxation," in *Handbook of public economics*, Elsevier, vol. 5, 391–474.

Rubinstein, A., Z. Safra, and W. Thomson (1992): "On the interpretation of the Nash bargaining solution and its extension to non-expected utility preferences," *Econometrica*, 60, 1171–1186.

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

——— (2002): "Optimal income transfer programs: intensive versus extensive labor supply responses," *The Quarterly Journal of Economics*, 117, 1039–1073.

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

Schokkaert, E. and K. Devooght (2003): "Responsibility-sensitive fair compensation in different cultures," *Social Choice and Welfare*, 21, 207–242.

Sidgwick, H. (1883): *The principles of political economy*, London: Macmillan.

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

Stovall, J. E. (2020): "Equal sacrifice taxation," *Games and Economic Behavior*, 121, 55–75.

Thomson, W. (2019): *How to divide when there isn't enough: From Aristotle, the Talmud, and Maimonides to the axiomatics of resource allocation*, vol. 62, Cambridge University Press.

Vickrey, W. (1947): *Agenda for progressive taxation*, New York: Ronald Press.

Weinzierl, M. (2012): "Why do we redistribute so much but tag so little? The principle of equal sacrifice and optimal taxation," *Working Paper.*

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

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

Young, H. (1988): "Distributive justice in taxation," *The Journal of Economic Theory*, 44, 321–335.

Young, H. P. (1990): "Progressive taxation and equal sacrifice," *The American Economic Review*, 80, 253–266.

A Proof of Lemma 1

Inequality aversion implies that any subset of individuals is "strictly essential:" for each $I' \subseteq I$ and each $\{a_i^*\}_{i \in I \backslash I'}$, allocations $a, a' \in A$ with $a_i = a_i' = a_i^*$ for each $i \in I \backslash I'$ are not all indifferent. By continuity of the social preferences, *separability*, and *strict essentiality*, Theorem 1 and Theorem 2 in Gorman (1968) apply and prove the existence of a representation $W(a) = \sum_{i \in I} H_i(c_i, \ell_i)$, where H_i is continuous for each $i \in I$. By *efficiency*, $H_i(c_i, \ell_i)$ is an order preserving transformation of $u(c_i, \ell_i)$. Thus, there exist a continuous function P_i such that $H_i(c_i, \ell_i) = P_i(u(c_i, \ell_i))$. Substituting gives the result.

B Proof of Theorem 1

Part 1. We first show the equal-sacrifice social welfare function satisfies the axioms. Let the sacrifice function be $S \in \mathcal{S}$ and let the individual Pareto functions $(P_i)_{i \in I}$ be consistent with S. Then, the equal-sacrifice social welfare function is

$$W(a) \equiv \sum_{i \in I} P_i(u(c_i, \ell_i)).$$

Efficiency. Since social marginal welfare weights are positive, the Pareto functions are increasing. Then, for each $i \in I$ and each pair $(c_i, \ell_i), (c_i', \ell_i')$, $u(c_i, \ell_i) \geq u(c_i', \ell_i')$ if and only if $P_i(u(c_i, \ell_i)) \geq P_i(u(c_i', \ell_i'))$. Consider a pair of allocations $a, a' \in A$ such that $u(c_i, \ell_i) \geq u(c_i', \ell_i')$ for each $i \in I$ and $u(c_i, \ell_i) > u(c_i', \ell_i')$ for some $i \in I$. Thus, also $P_i(u(c_i, \ell_i)) \geq P_i(u(c_i', \ell_i'))$ for each $i \in I$ and $P_i(u(c_i, \ell_i)) > P_i(u(c_i', \ell_i'))$ for some $i \in I$. Then, $\sum_{i \in I} P_i(u(c_i, \ell_i)) = W(a) > W(a') = \sum_{i \in I} P_i(u(c_i', \ell_i'))$ and $a \succ a'$. This proves the equal-sacrifice social welfare function satisfies *efficiency*.

Inequality aversion. By construction, for each $i \in I$, $P_i(u(c_i, \ell_i))$ is strictly concave in its arguments. It follows that $W(a)$ is strictly concave in its arguments and *inequality aversion* holds.

Separability. *Separability* follows from the additivity of the function W: the bundle of an unconcerned individual is irrelevant for the ranking of two allocations.

Tax solidarity. Consider a pair of allocations $a, a' \in A$ satisfying the requirements in the definition of *tax solidarity*. These allocations are such that for some pair of individuals $i, j \in I$ and some $\varepsilon > 0$: $b_i' - \varepsilon = b_i \geq 0 \geq b_j = b_j' + \varepsilon$; $\ell_i = \ell_i' = \bar{\ell}_i$ and $\ell_j = \ell_j' = \bar{\ell}_j$; and $(c_k, \ell_k) = (c_k', \ell_k') = (\bar{c}_k, \bar{\ell}_k)$ for each $k \in I/\{i,j\}$. By definition, $b_i \equiv \bar{c}_i - c_i$ and $b_i' \equiv \bar{c}_i - c_i'$. Thus, $c_i' = \bar{c}_i - b_i' = \bar{c}_i - b_i - \varepsilon$ and, substituting for b_i, $c_i' = c_i - \varepsilon$. Similarly, $c_j' = c_j + \varepsilon$. Substituting and since individuals $k \in I/\{i,j\}$ are unaffected, we may write

$$W(a) - W(a') = P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) - P_i\left(u\left(c_i - \varepsilon, \bar{\ell}_i\right)\right) + P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) - P_j\left(u\left(c_j + \varepsilon, \bar{\ell}_j\right)\right).$$

Now, by first-degree Taylor expansion and concavity of individuals' contributions to social welfare,

$$P_i\left(u\left(c_i - \varepsilon, \bar{\ell}_i\right)\right) \leq P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) - \varepsilon \beta_i\left(c_i, \bar{\ell}_i\right),$$

where, for memory, $\beta_i\left(c_i, \bar{\ell}_i\right) = \frac{\partial}{\partial c_i} P_i\left(u\left(c_i, \bar{\ell}_i\right)\right)$; and

$$P_j\left(u\left(c_j + \varepsilon, \bar{\ell}_j\right)\right) \leq P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) + \varepsilon \beta_j\left(c_j, \bar{\ell}_j\right),$$

where $\beta_j\left(c_j, \bar{\ell}_j\right) = \frac{\partial}{\partial c_j} P_j\left(u\left(c_j, \bar{\ell}_j\right)\right)$. Thus,

$$W(a) - W(a') \geq \varepsilon\left[\beta_i\left(c_i, \bar{\ell}_i\right) - \beta_j\left(c_j, \bar{\ell}_j\right)\right].$$

Finally, since $S(c_i, \bar{c}_i) > 0 > S(c_j, \bar{c}_j)$ and since g is increasing, $\beta_i\left(c_i, \bar{\ell}_i\right) = g(S(c_i, \bar{c}_i)) > g(S(c_j, \bar{c}_j)) = \beta_j\left(c_j, \bar{\ell}_j\right)$. Thus, $W(a) \geq W(a')$ and $a \succsim a'$. This proves *tax solidarity* holds.

Fair burden. Consider a pair of allocations $a, a' \in A$ satisfying the requirements in the definition of *fair burden*. These allocations are such that for some pair of individuals $i, j \in I$ with $\bar{c}_i \geq \bar{c}_j$ and some $\varepsilon > 0$: $0 \leq b_i' + \varepsilon = b_i < b_j = b_j' - \varepsilon$; $\ell_i = \ell_i' = \bar{\ell}_i$ and $\ell_j = \ell_j' = \bar{\ell}_j$; and $(c_k, \ell_k) = (c_k', \ell_k') = (\bar{c}_k, \bar{\ell}_k)$ for each $k \in I/\{i,j\}$. Substituting for b_i, b_i', b_j, and b_j', $c_i' = c_i + \varepsilon$ and $c_j' = c_j - \varepsilon$. Substituting and since individuals $k \in I/\{i,j\}$ are unaffected, we may write

$$W(a) - W(a') = P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) - P_i\left(u\left(c_i + \varepsilon, \bar{\ell}_i\right)\right) + P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) - P_j\left(u\left(c_j - \varepsilon, \bar{\ell}_j\right)\right).$$

Now, by first-degree Taylor expansion and concavity of individuals' contributions to social welfare,

$$P_i\left(u\left(c_i + \varepsilon, \bar{\ell}_i\right)\right) \leq P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) + \varepsilon\beta_i\left(c_i, \bar{\ell}_i\right),$$

and

$$P_j\left(u\left(c_j - \varepsilon, \bar{\ell}_j\right)\right) \leq P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) - \varepsilon\beta_j\left(c_j, \bar{\ell}_j\right).$$

Thus,

$$W(a) - W(a') \geq \varepsilon\left[\beta_j\left(c_j, \bar{\ell}_j\right) - \beta_i\left(c_i, \bar{\ell}_i\right)\right].$$

Finally, since $c_i \leq \bar{c}_i$, $S(c_i, \bar{c}_i) \geq 0$. Furthermore, $c_i - c_j \geq \bar{c}_i - \bar{c}_j$. Thus, by the slope restriction on S, $S(c_i, \bar{c}_i) \leq S(c_j, \bar{c}_j)$. Then, since g is increasing, $\beta_i\left(c_i, \bar{\ell}_i\right) = g\left(S(c_i, \bar{c}_i)\right) \leq g\left(S(c_j, \bar{c}_j)\right) = \beta_j\left(c_j, \bar{\ell}_j\right)$. Thus, $W(a) \geq W(a')$ and $a \succsim a'$. This proves *fair burden* holds.

Fair reward. Consider a pair of allocations $a, a' \in A$ satisfying the requirements in the definition of *fair reward*. These allocations are such that for some pair of individuals $i, j \in I$ with $\bar{c}_i \geq \bar{c}_j$ and some $\varepsilon > 0$: $c'_i + \varepsilon = c_i < c_j = c'_j - \varepsilon$; $\ell_i = \ell'_i = \bar{\ell}_i$ and $\ell_j = \ell'_j = \bar{\ell}_j$; and $(c_k, \ell_k) = (c'_k, \ell'_k) = \left(\bar{c}_k, \bar{\ell}_k\right)$ for each $k \in I/\{i,j\}$. Since individuals $k \in I/\{i,j\}$ are unaffected, we may write

$$W(a) - W(a') = P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) - P_i\left(u\left(c_i - \varepsilon, \bar{\ell}_i\right)\right) + P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) - P_j\left(u\left(c_j + \varepsilon, \bar{\ell}_j\right)\right).$$

Now, by first-degree Taylor expansion and concavity of individuals' contributions to social welfare,

$$P_i\left(u\left(c_i - \varepsilon, \bar{\ell}_i\right)\right) \leq P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) - \varepsilon\beta_i\left(c_i, \bar{\ell}_i\right),$$

and

$$P_j\left(u\left(c_j + \varepsilon, \bar{\ell}_j\right)\right) \leq P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) + \varepsilon\beta_j\left(c_j, \bar{\ell}_j\right).$$

Thus,

$$W(a) - W(a') \geq \varepsilon\left[\beta_i\left(c_i, \bar{\ell}_i\right) - \beta_j\left(c_j, \bar{\ell}_j\right)\right].$$

Finally, S is decreasing in the first argument and increasing in the second: $c_i < c_j$

and $\bar{c}_i \geq \bar{c}_j$ imply that $S(c_i, \bar{c}_i) > S(c_j, \bar{c}_j)$. Thus, since g is increasing, $\beta_i\left(c_i, \bar{\ell}_i\right) = g(S(c_i, \bar{c}_i)) > g(S(c_j, \bar{c}_j)) = \beta_j\left(c_j, \bar{\ell}_j\right)$. Thus, $W(a) \geq W(a')$ and $a \succsim a'$. This proves *fair reward* holds.

Part 2. We now show social preferences satisfying the axioms admit a representation by means of an equal-sacrifice social welfare function.

The proof is divided in several steps.

Step 1. *Assume social preferences $\succsim$ satisfy the axioms. Then, there exists real-valued increasing and strictly concave functions $(h_i)_{i\in I}$ such that social welfare W representing $\succsim$ is defined by setting for each $a \in A$,*

$$W(a) = \sum_{i\in I} h_i(e_i(c_i, \ell_i)).$$

Proof. By Lemma 1, there exist real-valued increasing functions $(f_i)_{i\in I}$ such that $f_i(u(c_i, \ell_i))$ is strictly concave for each $i \in I$ and such that, for each pair $a, a' \in A$, $a \succsim a'$ if and only if

$$W(a) = \sum_{i\in I} f_i(u(c_i, \ell_i)) \geq \sum_{i\in I} f_i(u(c_i', \ell_i')) = W(a').$$

Next, for each $i \in I$, $e_i(c_i, \ell_i)$ is the consumption-equivalent representation of preferences of i. Thus, there exists a real-valued increasing function h_i such that $h_i(e_i(c_i, \ell_i)) = f_i(u(c_i, \ell_i))$ for each (c_i, ℓ_i). This shows social preferences may be represented by the social welfare function W, as defined above.

It remains to show the functions $(h_i)_{i\in I}$ are strictly concave. By definition of the consumption-equivalent representation of preferences, for each $i \in I$ and each $c_i \in \mathbb{R}_+$, $h_i(c_i) = f_i\left(u\left(c_i, \bar{\ell}_i\right)\right)$. Since $f_i(u(c_i, \ell_i))$ is strictly concave, also h_i is strictly concave. □

Next, for each $c_i \in \mathbb{R}_+$, denote $h_i'\left(c_i^-\right)$ and $h_i'\left(c_i^+\right)$ the left and right first-order derivatives, respectively, of h_i at c_i. Let $\bar{A}$ be the set of allocations $a \in A$ such that $\ell_i = \bar{\ell}_i$ for each $i \in I$. Then, for each $a \in \bar{A}$, $W(a) = \sum_{i\in I} h_i(c_i)$. Let the choice correspondence C be defined as follows: for each $k \geq 0$, $C(k)$ is the set of consumption vectors $(c_i)_{i\in I}$ with $\sum_{i\in I} c_i \leq k$ that maximize W. Let $\bar{k} \equiv \sum_{i\in I} \bar{c}_i$. The following steps characterize the properties of C (with a slight abuse of notation, we shall use C also to denote the choice function, after showing the correspondence C is single-valued).

Step 2. *The choice correspondence C satisfies the following properties:*

1. *it is non-empty, single-valued, and continuous with respect to k;*

2. *it is strictly monotonic, $k > k'$ implies $C(k) \gg C(k')$;*

3. $C(\bar{k}) = (\bar{c}_i)_{i \in I}$*;*

4. $(c_i)_{i \in I} = C(k)$ *implies* $c_i > c_j \Longleftrightarrow \bar{c}_i > \bar{c}_j$ *for each* $i, j \in I$*;*

5. *for* $k \leq \bar{k}$, $(c_i)_{i \in I} = C(k)$ *implies* $c_i - c_j < \bar{c}_i - \bar{c}_j$ *for each* $i, j \in I$.

Proof. 1. W is increasing, continuous, and strictly concave, and so is $\sum_{i \in I} h_i(c_i)$. Thus, the choice correspondence C is non-empty, single-valued, and continuous with respect to k.

2. Let $(c_i)_{i \in I} = C(k)$ and $(c'_i)_{i \in I} = C(k')$. By contradiction of strict monotonicity, assume $k > k'$ and $C(k) \not\gg C(k')$. Then, there exists a pair of individuals $i, j \in I$ such that $c'_i \leq c_i$ and $c'_j > c_j$. At the optima, $h'_i(c_i^-) \geq h'_j(c_j^+)$ and $h'_i(c_i^+) \leq h'_j(c_j^-)$ and, similarly, $h'_i(c_i'^-) \geq h'_j(c_j'^+)$ and $h'_i(c_i'^+) \leq h'_j(c_j'^-)$. By strict concavity, $h'_i(c_i'^-) \geq h'_i(c_i'^+) \geq h'_i(c_i^-) \geq h'_i(c_i^+)$ and $h'_j(c_j^-) \geq h'_j(c_j^+) > h'_j(c_j'^-) \geq h'_j(c_j'^+)$. Combining these conditions leads to the following contradiction:

$$h'_i(c_i^-) \geq h'_j(c_j^+) > h'_j(c_j'^-) \geq h'_i(c_i'^+) \geq h'_i(c_i^-).$$

3. By contradiction of $C(\bar{k}) = (\bar{c}_i)_{i \in I}$, assume $(\bar{c}_i)_{i \in I} \neq C(\bar{k}) \equiv (c_i)_{i \in I}$. Then, $\sum_{i \in I} h_i(c_i) > \sum_{i \in I} h_i(\bar{c}_i)$. At $(c_i)_{i \in I}$, the tax burden of each individual $i \in I$ is $b_i \equiv \bar{c}_i - c_i$. Since $\bar{k} = \sum_i \bar{c}_i$, $\sum_{i \in I} b_i = 0$. Let $\overrightarrow{b}$ be the reordered vector of tax burdens of individuals: $\overrightarrow{b} \equiv (b_{(1)}, ..., b_{(|I|)})$ is such that $b_{(1)} \leq b_{(2)} \leq ... \leq b_{(|I|)}$, where (i) is the individual that, after permutation, occupies the i'th place in the order of tax burdens. Since $(\bar{c}_i)_{i \in I} \neq (c_i)_{i \in I}$, $\overrightarrow{b} \neq 0$ and thus $b_{(1)} < b_{(|I|)}$.

We next apply *tax solidarity* a finite number of times to show that $(\bar{c}_i)_{i \in I}$ is socially at least as desirable as $(c_i)_{i \in I}$, leading to a contradiction. The process is iterative and indexed by t. Let $c^t \equiv (c_i^t)_{i \in I}$ and let $\overrightarrow{b}^t$ be the corresponding reordered vector of tax burdens. Let $c^1 \equiv (c_i)_{i \in I}$. At each step t, three cases emerge.

Case (i). $\left|b^t_{(1)}\right| < \left|b^t_{(|I|)}\right|$. Let c^{t+1} be such that $c^{t+1}_{(1)} = \bar{c}_{(1)}$, $c^{t+1}_{(|I|)} = c^t_{(|I|)} + b^t_{(1)}$, and $c^{t+1}_{(i)} = c^t_{(i)}$ for each i such that $(i) \neq (1), (|I|)$. Since $b^t_{(1)} < 0$, this is a transfer from $(|I|)$ to (1).

Case (ii). $\left|b^t_{(1)}\right| > \left|b^t_{(|I|)}\right|$. Let c^{t+1} be such that $c^{t+1}_{(1)} = c^t_{(1)} - b^t_{(|I|)}$, $c^{t+1}_{(|I|)} = \bar{c}_{(|I|)}$, and $c^{t+1}_{(i)} = c^t_{(i)}$ for each i such that $(i) \neq (1), (|I|)$. Since $b^t_{(|I|)} > 0$, this is again a transfer from $(|I|)$ to (1).

Case (iii). $\left|b^t_{(1)}\right| = \left|b^t_{(|I|)}\right| = 0$. Let $c^{t+1} = c^t$.

The process converges in a finite number of iterations: $c^{(|I|)} = (\bar{c}_i)_{i \in I}$. For each $t = 1, ..., (|I|)$, let the allocation $a^t \in A$ assign to each individual the bundle $\left(c^t_i, \bar{\ell}_i\right)$. Then, for each iteration t, if $a^t \neq a^{t-1}$, cases *(i)* or *(ii)* apply and, by *tax solidarity*, $a^t \succsim a^{t-1}$. Otherwise, $a^t = a^{t-1}$ and thus $a^t \sim a^{t-1}$. By the representation result in Step 1, this implies $\sum_{i \in I} h_i\left(c^t_i\right) \geq \sum_{i \in I} h_i\left(c^{t-1}_i\right)$. Thus, also $\sum_{i \in I} h_i\left(\bar{c}_i\right) \geq \sum_{i \in I} h_i\left(c_i\right)$. This is a contradiction.

4. The proof is similar to that of 3, where *fair reward* is applied.

5. The proof is similar to that of 3, where *fair burden* is applied. □

We next construct a specific function $S : \mathbb{R}_+ \times \mathbb{R}_{++} \to \mathbb{R}$. Step 3 proves S is a sacrifice function.

First, for each $i \in I$ and each $k \geq 0$, let $S\left(c_i, \bar{c}_i\right) = \bar{k} - k$ if and only if $\left(c_i\right)_{i \in I} = C\left(k\right)$. Second, we complete the sacrifice function *linearly* for non-observed levels of laissez-faire consumption. Reorder individuals in increasing order of laissez-faire consumption, that is, $(i) \leq (j)$ if $\bar{c}_i \leq \bar{c}_j$. Let $\bar{c}_0 = 0$. Set $y \in \mathbb{R}_{++}$. Then, two cases emerge: either (1) there exists $i \in I$ such that $\bar{c}_{(i-1)} \leq y \leq \bar{c}_{(i)}$ or (2) $y > \bar{c}_{(|I|)}$.

Case 1. Let $i \in I$ be such that $\bar{c}_{(i-1)} \leq y \leq \bar{c}_{(i)}$ and let $\alpha \in [0, 1]$ be such that $y = \alpha \bar{c}_{(i-1)} + (1 - \alpha)\, \bar{c}_{(i)}$. Then, for each $x \in \mathbb{R}_+$, $S\left(x, y\right) = \bar{k} - k$ if and only if $x = \alpha c_{(i-1)} + (1 - \alpha)\, c_{(i)}$ where $S\left(c_{(i-1)}, \bar{c}_{(i-1)}\right) = S\left(c_{(i)}, \bar{c}_{(i)}\right) = \bar{k} - k$.

Case 2. Let (i) be such that $\bar{c}_{(|I|)} - \bar{c}_{(i)}$ is positive and smallest for all $i \in I$.[25] Let $\alpha > 1$ be such that $\left(y - \bar{c}_{(i)}\right) = \alpha \left(\bar{c}_{(|I|)} - \bar{c}_{(i)}\right)$. Then, for each $x \in \mathbb{R}_+$, $S\left(x, y\right) = \bar{k} - k$ if and only if $\left(x - c_{(i)}\right) = \alpha \left(c_{(|I|)} - c_{(i)}\right)$ where $S\left(c_{(i)}, \bar{c}_{(i)}\right) = S\left(c_{(|I|)}, \bar{c}_{(|I|)}\right) = \bar{k} - k$.

Step 3. *The function S is a sacrifice function. That is, S satisfies the following conditions:*

1. *a) decreasing in the first argument, b) increasing in the second argument, and c) continuous;*

[25] When $\bar{c}_{(|I|)} = \bar{c}_{(|I|-1)}$, such individual (i) differs from $(|I| - 1)$.

2. $x = y$ *implies* $S(x, y) = 0$*; and*

3. $S(x, y) = S(x', y') > 0$ *implies* $|x - x'| \leq |y - y'|$.

Proof. 1a) For each i, the function $S(c_i, \bar{c}_i)$ is decreasing in c_i by strict monotonicity of $C(k)$: more precisely, let $k < k'$; then, $(c_i)_{i \in I} = C(k) \ll C(k') = (c'_i)_{i \in I}$, $c_i < c'_i$, and $S(c_i, \bar{c}_i) = \bar{k} - k > S(c'_i, \bar{c}_i) = \bar{k} - k'$. For each $y \in \mathbb{R}_+$, $S(x, y)$ is decreasing in x as it is constructed as a linear combination of functions $(S(c_i, \bar{c}_i))_{i \in I}$ which are decreasing in the first variable.

1b) Property 4 of Step 3 states that: $(c_i)_{i \in I} = C(k)$ implies $c_i > c_j \iff \bar{c}_i > \bar{c}_j$ for each $i, j \in I$. By construction of S, this implies that $S(x, y) = S(x', y')$ with $y < y'$ if and only if $x < x'$. Since S is decreasing in the first argument, $S(x, y) < S(x, y')$.

1c) Since $C(k)$ is continuous in k, for each i, the function $S(c_i, \bar{c}_i)$ is continuous in c_i. Continuity of S then follows by construction.

2) By construction, $S(\bar{c}_i, \bar{c}_i) = \bar{k} - \bar{k} = 0$ for each $i \in I$. Now, for each $y \in \mathbb{R}_{++}$, either there exists $i \in I$ such that $\bar{c}_{(i-1)} \leq y \leq \bar{c}_{(i)}$ or $y > \bar{c}_{(|I|)}$. In the first case, $S(y, y) = 0$ since, by definition of S, $S\left(\bar{c}_{(i-1)}, \bar{c}_{(i-1)}\right) = S\left(\bar{c}_{(i)}, \bar{c}_{(i)}\right) = 0$ and $y = \alpha \bar{c}_{(i-1)} + (1 - \alpha)\bar{c}_{(i)}$ for some $\alpha \in [0, 1]$. In the second case, let (i) be such that $\bar{c}_{(|I|)} - \bar{c}_{(i)}$ is positive and smallest for all $i \in I$. Then, $S(y, y) = 0$ since, by definition of S, $S\left(\bar{c}_{(i)}, \bar{c}_{(i)}\right) = S\left(\bar{c}_{(|I|)}, \bar{c}_{(|I|)}\right) = 0$ and $\left(y - \bar{c}_{(i)}\right) = \alpha\left(\bar{c}_{(|I|)} - \bar{c}_{(i)}\right)$ for some $\alpha > 1$.

3) By contradiction, let $k \equiv S(x, y) = S(x', y') > 0$ and $|x - x'| > |y - y'|$. Without loss of generality, let $x > x'$ and $y > y'$. By construction, the implicit function $S(x, y) = k$ is piecewise linear: it may change slope only in correspondence to $y = \bar{c}_i$ with $(i) = 2, ..., (|I|) - 1$. By the mean value theorem, $x - x' > y - y'$ implies there exists a pair $i, j \in I$ such that $c_i - c_j > \bar{c}_i - \bar{c}_j$ with $S(c_i, \bar{c}_i) = S(c_j, \bar{c}_j) = k$. Clearly, c_i and c_j belong to $(c_m)_{m \in I} = C(k)$. Thus, $c_i - c_j > \bar{c}_i - \bar{c}_j$ is a violation of *fair reward* (as shown above). □

The proof is completed by the following step, which shows that the Pareto functions $P_i = f_i$ are consistent with the sacrifice function S.

Step 4. *For each $i \in I$, let the Pareto function of i be P_i such that $P_i(u(c_i, \ell_i)) \equiv h_i(e_i(c_i, \ell_i))$. The Pareto functions $(P_i)_{i \in I}$ are consistent with S. That is, for each $i \in I$, the social marginal welfare weights β_i satisfy $\beta_i\left(c_i, \bar{\ell}_i\right) = g(S(c_i, \bar{c}_i)) > 0$, where g is a real-valued increasing function, equal across individuals, and such that individuals' contributions to social welfare $(P_i(u(\cdot, \cdot)))_{i \in I}$ are strictly concave in their arguments.*

Proof. Strict concavity of $P_i(u(\cdot,\cdot))$ immediately follows from Step 1. Strict concavity and efficiency also imply g is increasing and social marginal welfare weights are positive.

Finally, we show g is equal across individuals. By contradiction, assume not. Then, there exists a pair $i, j \in I$ and $c_i, c_j \in \mathbb{R}_+$ with $S(c_i, \bar{c}_i) = S(c_j, \bar{c}_j)$ such that $\beta_i\left(c_i, \bar{\ell}_i\right) \neq \beta_j\left(c_j, \bar{\ell}_j\right)$. By construction, c_i and c_j belong to $(c_m)_{m \in I} = C(k)$ for some $k \geq 0$. However, $\beta_i\left(c_i, \bar{\ell}_i\right) \neq \beta_j\left(c_j, \bar{\ell}_j\right)$ imply $\frac{\partial h_i\left(c_i, \bar{\ell}_i\right)}{\partial c_i} \neq \frac{\partial h_j\left(c_j, \bar{\ell}_j\right)}{\partial c_j}$. This contradicts that $C(k)$ maximizes social welfare W (from Step 1) among the vectors $(c'_m)_{m \in I}$ such that $\sum_{m \in I} c'_m \leq k$. □

C Proof of Theorem 2

Part 1. We show that the criterion satisfies the axioms.

Proof. Since the criterion is a special case of the equal-sacrifice social welfare function, it satisfies *efficiency*, *inequality aversion*, *separability*, and *tax solidarity*. To show that it satisfies *scale invariance*, compute the marginal rate of substitution between the consumption of any two individuals $i, j \in I$ at allocation $a \in A$, such that $\ell_k = \bar{\ell}_k$ for each $k \in I$. By definition of W^p, this marginal rate of substitution is given by $-\beta_i\left(c_i, \bar{\ell}_i\right) / \beta_j\left(c_j, \bar{\ell}_j\right)$. Let $\kappa > 0$. Consider now allocation $a' \in A$ such that $c_k = \kappa c'_k$ and $\ell'_k = \bar{\ell}_k$ for each $k \in I$. The marginal rate of substitution between the consumption of i and j at allocation a' is $-\beta_i\left(c'_i, \bar{\ell}_i\right) / \beta_j\left(c'_j, \bar{\ell}_j\right)$. Since

$$\left(\frac{c_i}{\bar{c}_i}\frac{\bar{c}_j}{c_j}\right)^{-\gamma} = \left(\frac{\kappa c_i}{\bar{c}_i}\frac{\bar{c}_j}{\kappa c_j}\right)^{-\gamma} = \left(\frac{c'_i}{\bar{c}_i}\frac{\bar{c}_j}{c'_j}\right)^{-\gamma},$$

the marginal rates of substitution are the same at any proportional change κ of consumptions of individuals. This implies that social preferences are homothetic with respect to consumption at laissez-faire labor supply and, thus, *scale invariance* holds. □

Part 2. We show that the axioms imply the criterion.

Step 1. *The social welfare $\succsim$ is sum-of-utilities.*

Proof. By Lemma 1, *efficiency*, *inequality aversion*, and *separability*, imply that social welfare $\succsim$ is sum-of-utilities. Thus, there exist Pareto functions $(P_i)_{i \in I}$ such that social welfare is represented by W such that for each $a \in A$,

$$W(a) = \sum_{i \in I} P_i \left(u \left(c_i, \ell_i \right) \right). \qquad \square$$

Step 2. *There exists a decreasing function $g : \mathbb{R} \to \mathbb{R}$ such that for each $i \in I$ and each $c_i \geq 0$, if $P_i \left(u \left(c_i, \bar{\ell}_i \right) \right)$ is differentiable at c_i, then*

$$\frac{\partial}{\partial c_i} P_i \left(u \left(c_i, \bar{\ell}_i \right) \right) = g \left(\frac{c_i}{\bar{c}_i} \right).$$

Proof. For each $i \in I$ and each $c_i \geq 0$, define the function $f_i \left(\frac{c_i}{\bar{c}_i} \right) \equiv \frac{P_i \left(u \left(c_i, \bar{\ell}_i \right) \right)}{\bar{c}_i}$. Clearly, f_i is strictly increasing, continuous, and strictly concave since $P_i \left(u \left(c_i, \ell_i \right) \right)$ satisfies these.

Let a pair $a, a' \in A$ be such that for a pair $i, j \in I$ and a $\varepsilon > 0$ the following conditions hold: *(i)* $b_i' - \varepsilon = b_i \geq 0 \geq b_j = b_j' + \varepsilon$; *(ii)* $\ell_i = \ell_i' = \bar{\ell}_i$ *and* $\ell_j = \ell_j' = \bar{\ell}_j$; and *(iii)* $(c_k, \ell_k) = (c_k', \ell_k') = \left(\bar{c}_k, \bar{\ell}_k \right)$ for each $k \in I / \{i, j\}$. By *tax solidarity*, $a \succsim a'$. Rewrite condition (i) in terms of consumptions: $c_i = c_i' + \varepsilon$, $c_j = c_j' - \varepsilon$, and $\bar{c}_i - c_i \geq 0 \geq \bar{c}_j - c_j$. By the representation W, this implies that $W(a) - W(a') \geq 0$, and, substituting,

$$\begin{aligned} P_i \left(u \left(c_i, \bar{\ell}_i \right) \right) - P_i \left(u \left(c_i - \varepsilon, \bar{\ell}_i \right) \right) + \\ P_j \left(u \left(c_j, \bar{\ell}_j \right) \right) - P_j \left(u \left(c_j + \varepsilon, \bar{\ell}_j \right) \right) \geq 0 \end{aligned} \tag{2}$$

Substituting, we can write

$$\begin{aligned} \bar{c}_i f_i \left(\frac{c_i}{\bar{c}_i} \right) - \bar{c}_i f_i \left(\frac{c_i - \varepsilon}{\bar{c}_i} \right) + \\ \bar{c}_j f_j \left(\frac{c_j}{\bar{c}_j} \right) - \bar{c}_j f_j \left(\frac{c_j + \varepsilon}{\bar{c}_j} \right) \geq 0. \end{aligned}$$

Assume f_i and f_j are differentiable at $c_i / \bar{c}_i$ and $c_j / \bar{c}_j$, respectively. Then, dividing by ε and taking the limit for $\varepsilon \to 0$, yields

$$\left. \frac{\partial f_i(x)}{\partial x} \right|_{x = c_i / \bar{c}_i} \geq \left. \frac{\partial f_j(x)}{\partial x} \right|_{x = c_j / \bar{c}_j}. \tag{3}$$

Since f_i and f_j are strictly increasing, these are differentiable almost everywhere.

Thus, (3) holds for almost all $c_i/\bar{c}_i \leq 1 \leq c_j/\bar{c}_j$ and, symmetrically, the reverse inequality holds for almost all $c_i/\bar{c}_i \geq 1 \geq c_j/\bar{c}_j$. Thus, if the functions are differentiable at 1,

$$\left.\frac{\partial f_i(x)}{\partial x}\right|_{x=1} = \left.\frac{\partial f_j(x)}{\partial x}\right|_{x=1}.$$

Next, given $a \in A$, denote by $a(\kappa) \in A$ the allocation with consumption rescaled by a factor $\kappa > 0$. Then, by *scale invariance* and using the sum-of-utilities representation, $W(a) \geq W(a')$ if and only if $W(a(\kappa)) \geq W(a'(\kappa))$ for each $\kappa > 0$. Thus, equation (3) holds almost everywhere for each $c_i/\bar{c}_i \leq \kappa \leq c_j/\bar{c}_j$ and each $\kappa > 0$. It follows

$$\left.\frac{\partial f_i(x)}{\partial x}\right|_{x=\kappa} = \left.\frac{\partial f_j(x)}{\partial x}\right|_{x=\kappa}$$

holds almost everywhere for each $\kappa > 0$. Thus, the functions f_i and f_j have the same derivatives (where these are defined). This implies that there exists a decreasing function $g : \mathbb{R} \to \mathbb{R}$ such that for each $i \in I$ and each $c_i \geq 0$, if $P_i\left(u\left(c_i, \bar{\ell}_i\right)\right)$ is differentiable at c_i, then

$$\frac{\partial}{\partial c_i} P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) = g\left(\frac{c_i}{\bar{c}_i}\right). \qquad \square$$

<u>Step 3.</u> *For each $i \in I$ and each $c_i > 0$, $P_i\left(u\left(c_i, \bar{\ell}_i\right)\right)$ is differentiable.*

Proof. By contradiction, assume $P_i\left(u\left(c_i, \bar{\ell}_i\right)\right)$ is not differentiable at $\tilde{c}_i > 0$ for individual $i \in I$. Then, left and right derivative at $\tilde{c}_i$ are such that $\lim_{x \to x^-} g(x) \neq \lim_{x \to x^+} g(x)$, where $x = \tilde{c}_i/\bar{c}_i > 0$. By continuity and almost everywhere differentiability of W, there exists a pair $i, j \in I$ and a pair $a, a' \in A$ such that: *(i)* $c_i/\bar{c}_i > c'_i/\bar{c}_i = x = c'_j/\bar{c}_j > c_j/\bar{c}_j$; *(ii)* $c_k = c'_k$ for each $k \neq i, j$; *(iii)* $\ell_k = \ell'_k = \bar{\ell}_k$ *for each* $k \in I$; *(iv)* $W(a) = W(a')$; and *(v)* W is differentiable at a. Define $\Delta W \equiv W(a) - W(a')$. By the previous definitions and *(iv)*

$$\Delta W = P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) - P_i\left(u\left(c'_i, \bar{\ell}_i\right)\right) + P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) - P_j\left(u\left(c'_j, \bar{\ell}_j\right)\right) = 0.$$

For each $\kappa > 0$, define $\Delta W(\kappa) \equiv W(a(\kappa)) - W(a'(\kappa))$ and, substituting,

$$\Delta W(\kappa) = P_i\left(u\left(\kappa c_i, \bar{\ell}_i\right)\right) - P_i\left(u\left(\kappa c'_i, \bar{\ell}_i\right)\right) + P_j\left(u\left(\kappa c_j, \bar{\ell}_j\right)\right) - P_j\left(u\left(\kappa c'_j, \bar{\ell}_j\right)\right).$$

By *scale invariance*, $\Delta W(\kappa) = 0$ for each $\kappa > 0$. Differentiating $\Delta W(\kappa)$ with respect to κ and evaluating at $\kappa = 1$, gives

$$\frac{\partial \Delta W(\kappa)}{\partial \kappa} = 0.$$

Thus, this derivative is the same for $\kappa \to 1^+$ and $\kappa \to 1^-$, contradicting the statement that $P_i\left(u\left(c_i, \bar{\ell}_i\right)\right)$ is not differentiable at $c_i = \tilde{c}_i$.

We have thus established that for each $i \in I$ and each $c_i > 0$,

$$\frac{\partial}{\partial c_i} P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) = g\left(\frac{c_i}{\bar{c}_i}\right). \qquad \square$$

Step 4. *The function g is a strictly decreasing power function. That is, for each $i \in I$ and each $c_i > 0$,*

$$\frac{\partial}{\partial c_i} P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) = \left(\frac{c_i}{\bar{c}_i}\right)^{-\gamma},$$

for some $\gamma > 0$.

Proof. Let a pair $a, a' \in A$ and a pair $i, j \in I$ be such that: *(i)* $c_i, c_i', c_j, c_j' > 0$; *(ii)* $c_k = c_k'$ for each $k \neq i, j$; *(iii)* $\ell_k = \ell_k' = \bar{\ell}_k$ for each $k \in I$; and *(iv)* $W(a) = W(a')$. Then:

$$P_i\left(u\left(c_i, \bar{\ell}_i\right)\right) + P_j\left(u\left(c_j, \bar{\ell}_j\right)\right) = P_i\left(u\left(c_i', \bar{\ell}_i\right)\right) + P_j\left(u\left(c_j', \bar{\ell}_j\right)\right). \tag{4}$$

By *scale invariance*, for each $\kappa > 0$,

$$P_i\left(u\left(\kappa c_i, \bar{\ell}_i\right)\right) + P_j\left(u\left(\kappa c_j, \bar{\ell}_j\right)\right) = P_i\left(u\left(\kappa c_i', \bar{\ell}_i\right)\right) + P_j\left(u\left(\kappa c_j', \bar{\ell}_j\right)\right). \tag{5}$$

For $\xi \in \mathbb{R}$, let $a(\xi)$ be a smooth path through A, which satisfies *(i)*-*(iii)* for all $\xi \neq 0$ and such that $a(\xi) = a$. Thus equations (4) and (5) are satisfied when a is replaced by $a(\xi)$ for each $\xi \in \mathbb{R}$.

Differentiate with respect to ξ, evaluate at $\xi = 0$, and simplify to get:

$$g\left(\frac{c_i}{\bar{c}_i}\right) \left.\frac{\partial c_i(\xi)}{\partial \xi}\right|_{\xi=0} + g\left(\frac{c_j}{\bar{c}_j}\right) \left.\frac{\partial c_j(\xi)}{\partial \xi}\right|_{\xi=0} = 0,$$

$$g\left(\frac{\kappa c_i}{\bar{c}_i}\right) \left.\frac{\partial c_i(\xi)}{\partial \xi}\right|_{\xi=0} + g\left(\frac{\kappa c_j}{\bar{c}_j}\right) \left.\frac{\partial c_j(\xi)}{\partial \xi}\right|_{\xi=0} = 0.$$

Since g is strictly positive, we can combine these equations as

$$\frac{g\left(\frac{\kappa c_i}{\bar{c}_i}\right)}{g\left(\frac{c_i}{\bar{c}_i}\right)} = \frac{g\left(\frac{\kappa c_j}{\bar{c}_j}\right)}{g\left(\frac{c_j}{\bar{c}_j}\right)}. \tag{6}$$

Define the function $\lambda(\kappa)$ as the right-hand side of (6). By strict concavity of W, λ is continuous. Since g is strictly positive, also λ is. Substituting in (6) and taking the log transformation gives:

$$\ln g\left(\frac{\kappa c_i}{\bar{c}_i}\right) - \ln g\left(\frac{c_i}{\bar{c}_i}\right) = \ln \lambda(\kappa). \tag{7}$$

Equation (7) holds for each $c_i > 0$. Define $x \equiv c_i/\bar{c}_i$ and the function f such that $f(\ln x) = g(x)$ for each $x > 0$. Substituting and rearranging gives:

$$\ln f(\ln x + \ln \kappa) - \ln f(\ln x) = \ln \lambda(\kappa).$$

Divide by $\ln \kappa$ and take the limit for $\kappa \to 1$:

$$\frac{d \ln f(\ln x)}{d \ln x} = \lim_{\alpha \to 1} \frac{\ln \lambda(\kappa)}{\ln \kappa}.$$

By differentiability of W (see Step *3*), the limit on the right-hand side exists and is finite. Let $\gamma \equiv -\lim_{\alpha \to 1} \frac{\ln \lambda(\kappa)}{\ln \kappa}$. Then, integrating with respect to x gives:

$$f(\ln x) = g(x) = \eta x^{-\gamma},$$

for some integrating constant $\eta \in \mathbb{R}$. Since $g(x) > 0$ for each $x \geq 0$, $\eta > 0$. Set $\eta = 1$, since it is irrelevant for social welfare. Since W is strictly concave, $\gamma > 0$. □

Chapter 3

Does a Wealth Tax Improve Equality of Opportunity?

Joint with Shafik Hebous

Does a Wealth Tax Improve Equality of Opportunity?*

Kristoffer Berg† and Shafik Hebous‡

April 7, 2021

Abstract

Does parental wealth inequality impact next generation *labor* income inequality? And does a tax on parental wealth affect the labor income distribution of the next generation? We tackle both questions empirically using detailed intergenerational data from Norway, focusing on effects on wages rather than capital income. Results suggest that a net wealth of NOK 1 million increases wages of the children by NOK 14,000. Children of wealthy parents also have a higher labor income mobility. The estimated hypothetical wage distribution without the wealth tax is more unequal. Moreover, suggestive evidence indicates parental wealth is associated with higher labor risk taking.

Keywords: *Wealth Tax, Equality of Opportunity, Parental Wealth, Income Mobility, Inequality, Redistribution*

JEL Classification: D31, D63, H24

*We received helpful comments from Giacomo Brusco, Kasper Kragh-Sørensen, Edwin Leuven, Michael Keen, Paolo G. Piacquadio, Kjetil Storesletten, Magnus E. Stubhaug, Marius Ring, and Thor O. Thoresen, and seminar and conferences participants at the IMF, University of Oslo. The views expressed here are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management. This project has received funding from the Research Council of Norway (RCN) through the research center Oslo Fiscal Studies.

†University of Oslo, kristoffer.berg@econ.uio.no

‡International Monetary Fund, shebous@imf.org

1 Introduction

At the heart of the current debate about sharp and increasing wealth inequality is its potential impact on income inequality in the next generation—an aspect reflecting equality of opportunity.[1] If children of wealthy parents are not only more likely to earn higher capital income—as suggested in the literature[2]—but also higher *labor* income than peers from less wealthy families with otherwise similar characteristics, then parental wealth entails a privilege that reduces equal prospects of earning income. In this sense, parental wealth can affect intergenerational income mobility through affecting wages.

The debate on wealth inequality has triggered a strong interest in—and a growing recent literature on—wealth taxation. Thus far, however, the literature has not provided evidence regarding the question: does a tax on parental wealth affect the *labor* income distribution of the next generation? Arguably, it is a challenging question to answer, not least in face of demanding data requirement to establish links between parental wealth, children income when grown up, and a real-world wealth tax.

This paper empirically studies the effects of parental wealth, and its taxation during childhood, on adult income in Norway. The wealth tax in Norway—currently one of the few wealth taxes in the world—has a relatively broad coverage, providing the advantage of studying a wide spectrum of taxpayers beyond the superrich.[3] Our research design focuses on cohorts born during 1978-1980 and estimates the effect of taxing the wealth of their parents in the late 1990s (i.e., when they were at advanced stages in school) on their income in 2013-2017 (i.e., during adulthood). We focus on three outcomes for these cohorts: i) the level of wage; ii) the position on the labor income distribution; and iii) position on the labor income distribution relative to that of their parents—a measure of intergenerational income mobility. We relate parental wealth to one of these outcome variables using OLS. Next, we use an instrumental variable (IV) approach to address concerns about potential effects on wages that are correlated with parental wealth and left uncontrolled for in the OLS (i.e., potential omitted variable bias due to unobserved confounders).

The IV identification of the causal effect of parental wealth on the income of the children relies on two sources of variation: i) changes to the wealth tax rate in the late 1990s; and ii) different levels of taxation of the same level of wealth, depending on the marital status. Specifically, we exploit that the wealth tax threshold and deduction of a married couple filing jointly were higher than that for single filers with the same level of wealth. Thus, we estimate an IV model using tax changes

[1] See, e.g., Piketty and Zucman (2014), Smith et al. (2020), and Boserup et al. (2018).
[2] Fagereng et al. (2021)
[3] See Scheuer and Slemrod (2020).

(due to marital status differences and tax law changes) as an instrument for changes in net parental wealth.[4] To address potential concerns about the exclusion restriction (including the possibility that divorce can directly affect future income of the children), we separately estimate the direct divorce effect on the income of the children from a sample of taxpayers that are out of the scope of the wealth tax throughout the entire sample period, and adjust our IV estimation accordingly. As we will explain later, this strategy plausibly provides an upper bound of the divorce effect.

Our analysis yields two main results, consistent across the three considered outcome variables and different rich sets of covariates to control for potential omitted variables bias. First, those who grow up in families with higher levels of net wealth tend to have higher labor incomes, controlling for the education and incomes of their parents as well as individual characteristics including education. The benchmark estimates suggest that a net wealth of 1 million NOK in Norway increases future annual wages of the children by 14,000 NOK, ceteris paribus. Second, based on these point estimates, we estimate the counterfactual income distribution in 2017, in our sample, in the absence of the wealth tax to answer the question: What would have happened to the labor income distribution today had Norway not implemented a wealth tax in the late 1990s and early 2000s? Our results suggest that the wealth tax has made the labor income distribution less unequal—lowering the Gini coefficient by about 1 point. Moreover, results suggest that the intergenerational *labor* income mobility is influenced by the stock of parental wealth, with children from more wealthy families experiencing higher labor income mobility than those from less wealthy families.

Extending the analysis to account for heterogeneous effects across wealth levels suggests that the impact of the wealth tax on the labor income of the children is higher at middle levels of wealth. Intuitively for the superrich, capital income plays a key role diminishing the importance of employment income whereas low levels of parental wealth do not appear to strongly increase the chances of improving the children position on the labor income distribution.

The questions as to how much and how (if at all) the income distribution should be made more equal require normative analysis as, ultimately, optimal redistribution polices are dependent on society's preferences and the social welfare function. The positive analysis in our study, however, does inform policymakers by providing empirical evidence that a wealth tax is one policy instrument that can lower the next generation income inequality.

Our paper leaves it for future research to closely study different mechanisms through which the parental stock of wealth impacts the labor income of their children. However, we do provide

[4] A similar strategy was used in Jakobsen et al. (2020) who focus on the elasticity of capital with respect to the abolished Danish wealth tax.

empirical evidence pointing to directions for further research. Our results mute a potential education channel when we control for higher education and the field of the study of the children.[5] This prompts us to think of further mechanisms beyond human capital formation. We provide novel empirical evidence suggesting that one of those mechanisms is the risk profiles of decisions related to labor income. For example, wealth may act as a private safety net. Results indicate heterogeneous returns to labor, as higher levels of parental wealth are associated with a higher dispersion of labor income after controlling for individual and parents' characteristics. This finding complements recent evidence on intergenerational earning dynamics (Halvorsen et al. (2021)) and heterogeneous returns to capital as one explanation of intergenerational correlation in wealth levels (Fagereng et al. (2021) and Benhabib and Bisin (2018)). In this context, our results explicitly point to the role of the heterogeneity of labor income (in addition to capital income), associated with different levels of parental wealth, in driving heterogeneous total wealth returns.

Our study links three strands of literature. The first is the empirical literature on wealth taxation, which—as surveyed in Scheuer and Slemrod (2021)—mainly looks at two broad aspects: the behavioral (both real and evasion) responses as well as the revenue potential of various wealth tax designs (e.g., Bjørneby et al. (2020); Brülhart et al. (2019); Duran-Cabré et al. (2019); Jakobsen et al. (2020); Ring (2020); Seim (2017); Saez and Zucman (2019); and Zoutman (2018)). This literature does not look at the intergenerational aspects of parental wealth. Secondly, a strand of the literature looks at intergenerational or regional income mobility but with a focus on describing patterns in the data without linking them to a wealth tax (e.g., Chetty, Hendren, Kline, Saez, and Turner (2014); Corak (2013); Lee and Solon (2009); and Thoresen (2009)). Finally, a related growing literature studies specific mechanisms of inequality of opportunity. For example, a series of papers—including Chetty et al. (2020), Chetty, Hendren, Kline, and Saez (2014), and Chetty et al. (2018)—relate the distribution of students' earnings in their thirties to their parents' *incomes*. They document, inter alia, that low- and middle-income students attend selective schools at much lower rates than their peers from higher-income families with the same test scores, but those that attend these schools have similar long-term outcomes. This suggests that college attendance patterns have an upward effect on income mobility.

This paper proceeds as follows. Section II summarizes the Norwegian wealth tax during the sample period. Section III presents the identification approach. Section IV discusses the results. Section V concludes.

[5] Note also that affordability and access to high quality education is facilitated by the public education system in Norway.

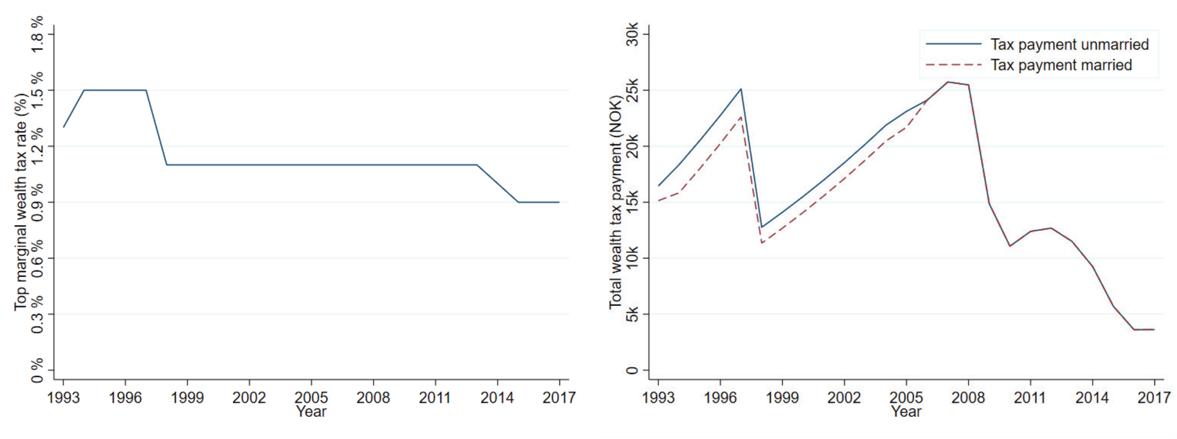
2 Norwegian Wealth Tax and Data

2.1 Norwegian Wealth Tax

Today, Norway is one of a few OECD countries that levies a tax on the net wealth of individuals.[6] The marginal wealth tax rate has varied considerably over time from a three-step progressive rate in the mid-90s—reaching a rate of 1.5 percent—to a flat rate from the early 2000s—currently set at 0.85 percent (left panel of Figure 1). One specific feature in the Norwegian wealth tax is its relatively low threshold, implying a significant number of taxpayers. The tax threshold currently is a net wealth above NOK 1.5 million (about USD 174,000)—which is doubled for married couples—compared to NOK 125,000 and NOK 150,000 for singles and married couples, respectively, in the early 1990s. In 1993, about 18 percent of Norwegian taxpayers were subject to the wealth tax, while in 2017 the number had dropped to 10 percent.[7]

Figure 1: *Wealth Tax Rates and Payments*

(a) *Top Marginal Wealth Tax Rate* **(b)** *Tax Payments: Married vs Unmarried*



To illustrate differences in taxing the wealth based on marital status in the 1990s and early 2000s, the right panel of Figure 1 shows tax payments over time for married and unmarried couples that have the same the level of equally distributed wealth. For illustration, couples start with NOK 500,000 in 1993 (roughly USD 40,000 using 1993 exchange rate) and we increase the wealth at a predetermined rate of 5 percent annually. Figure 1 displays larger differences in tax payments between married and unmarried parents before 2006, which we exploit in our identification strategy

[6]The Norwegian wealth tax was introduced in 1892. Currently, in OECD countries, in addition to Norway, e.g., Switzerland and Spain have a wealth tax. Ongoing discussions about a wealth tax are taking place in several countries including the United States, Argentina, and South Africa.

[7]See Bjørneby et al. (2020) and Ring (2020) for detailed descriptions of the Norwegian wealth tax.

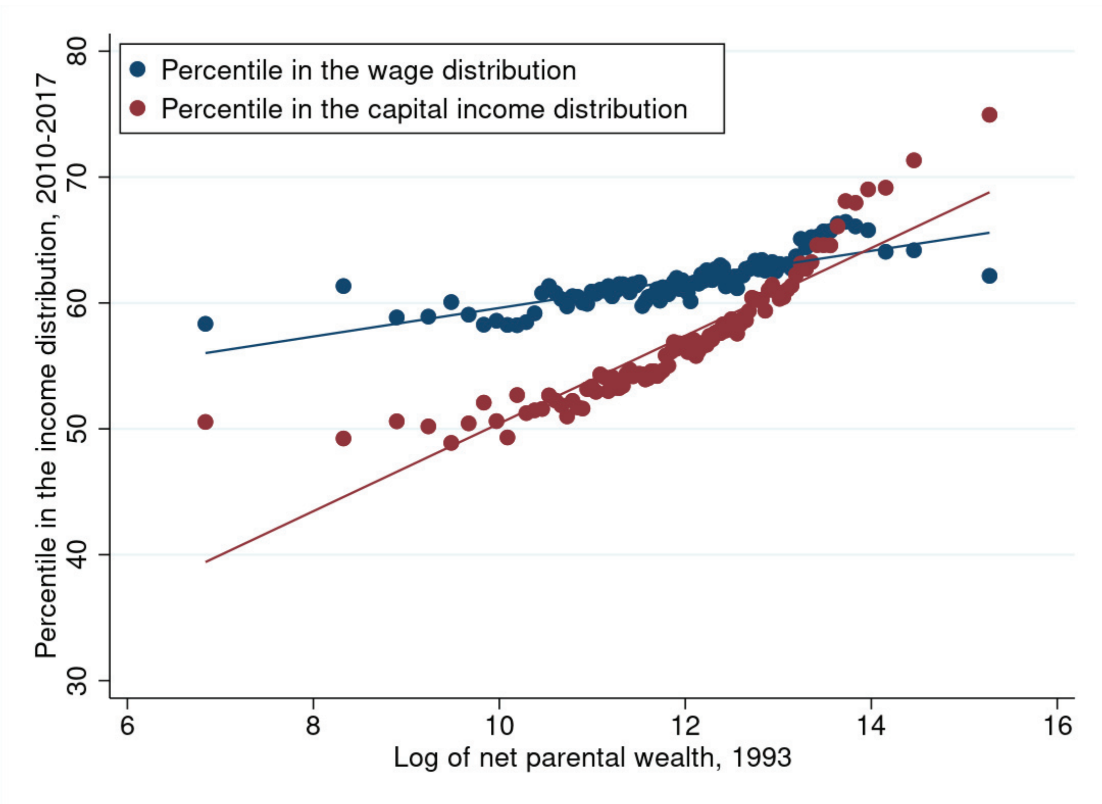
(see also Table A.1).

2.2 Data

The source of the data is Statistics Norway's databases including the Income Statistics for Families and Persons which contains the Register of Tax Returns and other detailed information on individuals, enabling us to link parents with their children and trace their different sources of income and their net wealth since 1993. Moreover, the database contains information about education levels, including the field of study, the place of birth, and other characteristics. The Appendix describes the definitions of all variables and presents detailed descriptive statistics in the sample distinguishing between married and unmarried taxpayers (Tables A.2, A.3, and A.4). In 1993, the average net financial wealth was 85 percent of the average wage. By 2017, the average net financial wealth had risen to 135 percent of the average wage. Unsurprisingly, wealth is concentrated at the top 10 percent wealthiest owned about half of all (positive) net wealth in Norway in 2017. Halvorsen et al. (2021) present a rich set of stylized facts about intergenerational earnings in Norway.

Figure 2 visualizes the main finding of the paper. It presents graphical evidence showing the correlation between parental wealth in 1993-1999 and the percentiles of the income distribution of the children in 2013-2017. The correlation patterns are estimated separately for wages and capital income, controlling for characteristics including: parental wages in 1993-1999, birth in an urban area, age of the wage earner, age and education of the parents. Figure 2 shows that high parental wealth—during childhood—is associated with a better position in the labor income distribution when grown up. Furthermore, confirming existing studies in the literature, the upward slopping relationship is also observed between parental wealth and capital income.

Figure 2: *Parental Wealth, Wage, and Capital Income*



Note: The binned scatterplot shows the estimated relationship between net parental wealth in 1993 and the position on the labor income or capital income distribution in 2010-2017, controlling for parents and individual's characteristics including education.

3 Empirical Approach

3.1 IV Estimation and Clarifying the Potential Bias

The main IV identification strategy is to exploit differences in the taxation of wealth of married, unmarried, and divorced parents throughout the 1990s and early 2000. The idea is to use these exogenous changes in parental wealth to account for confounding variables that impact wage outcomes of children. Furthermore, we explicitly account for a potential direct effect of divorce on the income of the children.

Let Y_i be the outcome (wages), X_i is the stock of parental wealth during childhood, Z_i is the instrument (parental divorce) and C_i is a confounder (unobserved variable that affects wages and is related to parental wealth). For illustrating the main point, we can safely drop the time dimension here. Assume random assignment of Z_i, which in our context means that divorce is unrelated to C_i

(i.e., Z_i and C_i are independent), the second stage equation is

$$Y_i = \alpha + \beta X_i + \epsilon_i, \tag{1}$$

where the error term is

$$\epsilon_i = \delta Z_i + \phi C_i. \tag{2}$$

The *exclusion restriction* is that $\delta = 0$, which may not hold if parental divorce directly affects earnings of the children when grown up (and we account for this possibility as described below). The first-stage is

$$X_i = \theta + \gamma Z_i + \sigma_i, \tag{3}$$

where σ_i contains all factors that affect parental wealth other than divorce. Using 2SLS, we obtain $\hat{X}_i = \hat{\theta} + \hat{\gamma} Z_i$ from the first stage estimation, and next the IV-estimator replaces X_i by $\hat{X}_i$.

The *instrument relevance* holds if $\gamma \neq 0$, and the IV-estimator $\hat{\beta}^{IV} = \hat{cov}(Y_i, Z_i)/\hat{cov}(X_i, Z_i)$ is then

$$\hat{\beta}^{IV_1} = \frac{\hat{cov}\left(\beta(\hat{\theta} + \hat{\gamma} Z_i) + \delta Z_i + \phi C_i, Z_i\right)}{\hat{\gamma}\hat{var}\left(Z_i\right)} = \beta + \frac{\delta}{\hat{\gamma}} + \frac{\hat{cov}\left(\phi C_i, Z_i\right)}{\hat{\gamma}\hat{var}\left(Z_i\right)} = \beta + \frac{\delta}{\hat{\gamma}}, \tag{4}$$

where the last step follows from random assignment.

Thus, if $\delta < 0$ and $\gamma < 0$ then the bias, b is positive and β^{IV} will overestimate the effect, $\beta^{IV} = \beta + b$. Hence, to relax a priori assumption $\delta = 0$, we estimate δ to correct for the potential bias in the IV estimator of the causal effect of parental wealth on wages of the children.

3.2 Accounting for the Potential Bias

We estimate δ from a sample of individuals with wealth below the tax threshold, which means divorce does not affect their tax payments at all. The estimation equation of the direct effect of divorce is:

$$Y_j = \alpha + \beta X_j + \delta Z_j + \epsilon_j. \tag{5}$$

Under random assignment, the OLS estimator $\hat{\delta}$ identifies δ, allowing us to difference out the direct effect of Z_i on Y_i.

The adjusted second stage in the IV estimator is:

$$Y_i^{pred} = Y_i - \hat{\delta} Z_i = \alpha + \beta X_i + \theta C_i, \tag{6}$$

where Y_i^{pred} is the variation in Y_i that remains after accounting for the direct effect of Z_i. If δ is equal across samples, such that $Y_i^{pred} = Y_j^{pred}$, then $\hat{\delta}$ is an unbiased estimate of the true δ in our sample of interest. Hence, using the corrected values, Y_i^{pred}, the IV-estimator now identifies β:

$$\hat{\beta}^{IV_2} = \frac{cov\left(\beta(\hat{\theta} + \hat{\gamma}Z_i) + \phi C_i, Z_i\right)}{\hat{\gamma} var\left(Z_i\right)} = \beta. \tag{7}$$

If $\hat{\delta}$ is larger for the low parental wealth sample, which is a plausible assumption, then our strategy identifies an upper bound estimate of the divorce effect on income of the children—although this assumption is not needed for the validity of our adjustment.

3.3 Specifications

OLS and IV

Our sample includes three cohorts born during 1978-1980, who are 14-16 years old in 1993, 19-21 in 1998 and 38-40 in 2017. Individual i at time t has wage $wage_{i,t}$ and parents p with total net wealth $netwealth_{p,t}$. The OLS specification is:

$$wage_{i,t} = \alpha_t + \beta\, netwealth_{p_i,1998} + \gamma\, netwealth_{p_i,1993} + \theta\, controls_{p_i,1993} + \delta\, controls_i + \epsilon_{i,t}, \tag{8}$$

where $controls_{p_i,1993}$ is a vector of characteristics of i's parents including wage, education, age, and marital status in $t-20$; $controls_i$ is another vector of characteristics of individual i, including the age and whether the individual is born outside of Norway; α_t are year-dummies; and $\epsilon_{i,t}$ are error terms.

As described above, since net parental wealth may be associated with unobservable features of each family, we also instrument $netwealth_{p_i,1998}$ by the change in wealth tax payments, which occurs because of changes in tax rules and marital status, while holding wealth and income constant. $\Delta taxpayment_{p_i,t} = taxrule_t(netw_{p_i,1993}, marriage_{p_i,1998}) - taxrule_{1993}(netw_{p_i,1993}, marriage_{p_i,1993})$, where $taxrule_t$ are the tax rules for net wealth in each year. Increases in tax payments due to changes in the rules reduce net wealth in 1998, conditional on net wealth in 1993. The difference in taxation of the same level of wealth derives from different tax treatments based on marital status and changes to the marital status. Hence, the IV specification is

$$\begin{aligned} wage_{i,t} = \alpha_t + \beta\left(netwealth_{p_i,1998} = \Delta taxpayment_{p_i,1998}\right) + \gamma\, netwealth_{p_i,1993} + \\ \eta\, marriage_{p_i,1993} + \theta\, controls_{p_i,1993} + \delta\, controls_i + \epsilon_{i,t}. \end{aligned} \tag{9}$$

We control for the marital status of the parents in 1993. Importantly, in this specification the sources of variation in the tax treatment of the same level of wealth are changes in tax rules for the

the married and unmarried and divorce (since we condition on both parents being alive in 1998). Furthermore, we control for the initial wealth levels of the parents, such that we estimate the effect of changes in parental wealth due to exogenous tax changes and their impact on wages 19 years later.

Instrument Relevance and Validity

The effect of exogenous changes in parental net wealth, β, is identified if the change in tax payments affect net parental wealth in 1998 (relevance) and is unrelated to wages other than through net wealth in 1998 (exclusion). As reported in Table A.5 in the Appendix, the $F-$statistics and R^2 form the first stage regressions support the relevance of the instrument passing the Stock-Yogo cutoffs.

As discussed above, to address concerns that the exclusion restriction may not hold, we employ a differences-in-differences design. The approach is to estimate the effect of parental divorce on wages of the children for those that not subject to the wealth tax, and use these estimates to adjust our the IV estimator as follows:

$$\begin{aligned} wage_{\underline{i},t} = \alpha_t + \xi\, divorce_{p_i,1998} + \beta\, netwealth_{p_i,1998} + \gamma\, netwealth_{p_i,1993} + \\ \eta\, marriage_{p_i,1993} + \theta\, controls_{p_i,1993} + \delta\, controls_{\underline{i}} + \epsilon_{\underline{i},t}, \end{aligned} \tag{10}$$

where $divorce_{p_i,1998}$ is a dummy that is equal to one when the parents divorce in the period 1994-1998 and zero otherwise. $\underline{i}$ is an individual with parental wealth between NOK 0 and NOK 50,000 during 1993-1998, whereas $\bar{i}$ are individuals above NOK 50,000. The estimation results are reported in Table A.6 in the appendix.

Next, for all levels of parental wealth, we linearly predict wages using the estimation results from Equation 10. This predicted wage is then subtracted from the observed wage for $i = \bar{i}$, obtaining $\hat{wage}_{\bar{i},t}$:

$$\begin{aligned} \hat{wage}_{\bar{i},t} = \alpha_t + \beta\left(netwealth_{p_i,1998} = \Delta\, taxpayment_{p_i,1998}\right) + \gamma\, netwealth_{p_i,1993} + \\ \eta\, marriage_{p_i,1993} + \theta\, controls_{p_i,1993} + \delta\, controls_{\bar{i}} + \epsilon_{\bar{i},t}. \end{aligned} \tag{11}$$

To summarize, if the direct effect of divorce on wages of the children is independent of parental wealth, our approach identifies the effect of exogenous changes in parental wealth on wage outcomes. If instead the direct effect is higher at lower levels of wealth, then our approach to account for it is using an upper-bound estimate of the direct effect (thereby lowering the wages of children of divorced parents that pay the wealth by the same amount as for those that do not pay the wealth tax). Hence, in this case, the true effect for the wealthy is between our non-adjusted and adjusted approaches.

Other Outcome Variables

In addition to the levels of wage of the children, we consider two other dependent variables: i) The position of the child in the wage distribution (percentiles). This variable is particularly suitable for our IV strategy because it is unlikely that divorce directly affects the *percentile* in the wage distribution of the children of parents with wealth *more than* children with low parental wealth; ii) A measure of intergenerational income mobility defined as the child position on the wage distribution relative to the parents.

4 Results

4.1 Main Results

Table 1 shows our main results. In columns 1-3, the variable of interest is total net wealth of the parents. The first column displays OLS estimation results, whereas the second column shows the IV estimation results without adjusting for the direct divorce effect on children income. Column 3 adjusts the IV model for this effect as described in Section 3. The dependent variable in the first row is the level of wages. The OLS yields an estimate suggesting that a net parental wealth of NOK 1 million in Norway increases future annual labor income of the children by NOK 2,710. The IV and adjusted IV estimates are larger, at NOK 16,700 and NOK 14,000 respectively. The adjusted-IV point estimate is only slightly smaller than the IV indicating to a relatively low potential bias from a violation of the exclusion restriction. Columns 4-6 repeat columns 1-3 but using only the net *financial* wealth of the parents. Estimates are rather similar ranging from NOK 2,550 (OLS), NOK 17,900 (IV), to NOK 10,100 (adjusted IV).

In the second row of Table 1, the dependent variable is the percentile of the child on the wage distribution. All estimation methods suggest that a net parental wealth has a positive impact on the position of the child on the labor income distribution. The third row shows the results for the intergenerational income mobility measure as the dependent variable. Again, the three estimation models give the same finding that net parental wealth has a positive effect on the income of the child relative to the income of the parents. Redoing the analysis using total income instead of wages yields higher estimates, in particular at the top (see Appendix, Table A.7), which is intuitive as wealth also generates capital income.

Table 1: *Main Results*

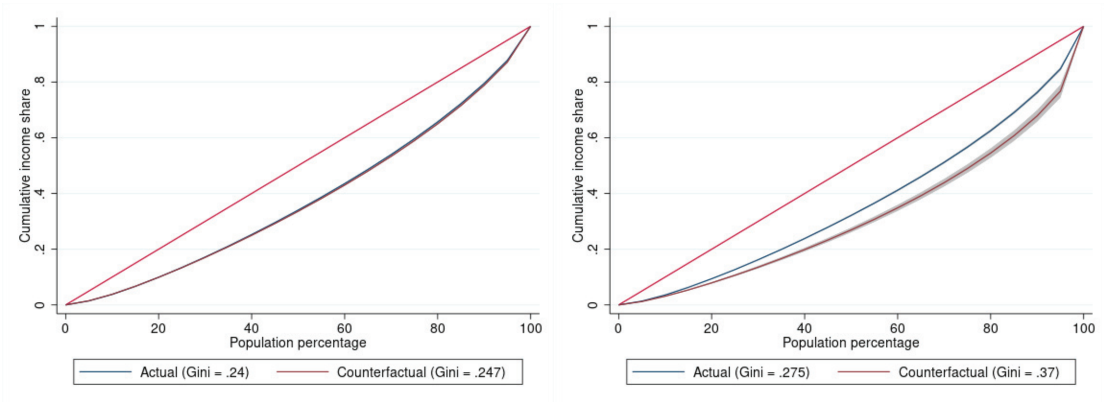
Estimator	OLS	IV	Adjusted IV	OLS	IV	Adjusted IV
Effect of	Net parental wealth			Parental financial wealth		
On wage level	0.00271***	0.0167***	0.0140***	0.00255***	0.0179***	0.0101***
	(0.000356)	(0.00121)	(0.00119)	(0.000395)	(0.00124)	(0.00103)
On wage percentile	0.000208***	0.00135***	0.00113***	0.000218***	0.00145***	0.000745***
	(0.0000276)	(0.000104)	(0.000107)	(0.0000333)	(0.000105)	(0.000890)
On wage mobility	0.000649***	0.00317***	0.00295***	0.000633***	0.00335***	0.00206***
	(0.0000638)	(0.000180)	(0.000191)	(0.0000869)	(0.000190)	(0.000158)
Sample restrictions	0<PW	0<PW	200<PW	0<PW	0<PW	200<PW
N	480,971	480,971	270,995	480,971	480,971	270,995

*Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All monetary amounts are measured in NOK 1000. PW is net parental wealth divided by the number of siblings in 1993. The first 3 columns show the effects of the change in net parental wealth from 1993 to 1998 instrumented by the wealth tax change. Columns 4 to 6 show the effects for the change in parental net financial wealth from 1993 to 1998 instrumented by the wealth tax change. All estimation include controls and results are shown in the Appendix. Controls include wages, education and age of father and mother, parental wealth and marital status in 1993, whether the individual earns mainly capital income, age, whether the individual is born in an urban area, and year dummies. Mobility outcomes are measured in percentiles from father's wage income in 1993 to children's wage income in 2010-2017.*

All reported specifications control for characteristics of the parents, including wages, education levels, marital status, and ages. Regarding individuals, controls include age, a dummy for being born in an urban area, and whether earning is mainly capital income. Unsurprisingly, education of the parents is positively associated with wages of the children. The Appendix reports the full results for the controls (Table A.8).

Figure 3 presents the estimated counterfactual distribution of wages in the absence of the Norwegian wealth tax, based on Equation 11 and corresponding to the estimates in column 3 of Table 1. We compute the Gini coefficients of the counterfactual and observed distributions of wages, and find that the latter is less unequal with a Gini coefficient of close 0.24 compared to the counterfactual Gini coefficient in the our sample close to 0.25.

Figure 3: *Counterfactual Income Distribution in the Absence of a Wealth Tax*



(a) *Wage Income Inequality* **(b)** *Total Income Inequality*

4.2 Extensions

Heterogeneity: To explore heterogeneous effects, Table 2 presents estimation results for three ranges of net parental wealth. For the lower range (NOK 100,000 to 500,000), there is a combination of treated and untreated taxpayers by tax changes over time, and the estimates in this range are insignificant. The effect becomes significant at the middle range of wealth (between NOK 500,000 and 1.2 million). In the upper range, the effect becomes smaller but remains significant at the 1-percent level. This pattern is intuitive as at the very top of the wealth distribution, capital income becomes more important than labor income. Similarly, the effects of net parental wealth on the percentiles of the labor income distribution of the children and on their income mobility are the highest for the middle range of wealth (second and third rows of Table 2), and the effect remains significant, but smaller, at the very top. Additionally, the estimated counterfactual wage distribution in the absence of the wealth tax looks very similar after taking the heterogeneous effects into account (A.1).

Table 2: *Heterogeneous Effects across Wealth Levels*

Strategy	Adjusted IV		
Effect of	Net parental wealth		
On wage level	5.137	0.0380***	0.00790***
	(8.097)	(0.0116)	(0.00284)
On wage percentile	0.294	0.00226***	0.000519***
	(0.468)	(0.000712)	(0.000243)
On wage mobility	0.867	0.00528***	0.000321
	(1.368)	(0.00147)	(0.000374)
On total income	15.54	0.122***	0.131***
	(24.37)	(0.0359)	(0.0391)
Sample restrictions	100<PW<500	500<PW<1200	1200<PW
N	205,030	50,586	15,470

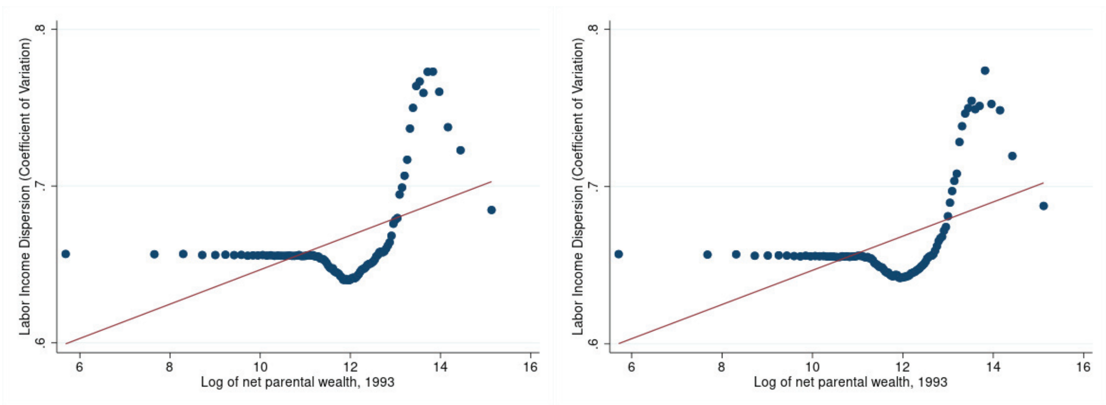
*Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All monetary amounts are measured in NOK 1000. PW is net parental wealth divided by the number of siblings in 1993. The effect is the change in net parental wealth from 1993 to 1998 instrumented by the predicted wealth tax change given parental wealth in 1993. Controls are wages, education and age of father and mother, parental wealth and marital status in 1993, whether the individual earns mainly capital income, age, whether the individual is born in an urban area and year dummies. Mobility outcomes are measured in percentiles from father's wage income in 1993 to children's wage income in 2010-2017.*

Underlying mechanism: There can be various mechanisms behind our findings. For example, wealth can potentially affect human capital formation, and thus wages, possibly through: i) affordability of private education (particularly relevant for countries with higher private provision of education and less relevant for Norway); and ii) decision to invest in human capital (e.g., to attend a graduate school or not). We do replicate the main results of Table 1 after controlling for higher education and the field of study (science, business, etc) of the children. Unsurprisingly, having a higher degree positively impacts wages. Importantly, the effects of net parental wealth on the wages of the children, the position on the wage distribution, and intergenerational income mobility, are very similar—slightly smaller after controlling for the education of the children (Table A.9 in the Appendix shows). This indicates that there are other mechanisms beyond the level of education of the children through which wealth affects intergenerational labor income.

As a first assessment to trigger further research on the linkages between the stock of parental wealth and wages of the children, we compute a measure of dispersion (the coefficient of variation) of labor income corresponding to bins of the stock of parental wealth, controlling for the education of the children. This measure is indicative of "risk-taking" in the sense that wage earners' decisions can also be associated with a risk profile (e.g., via occupational choices)—for instance a graduate with a business administration degree from a wealthy family may take different career decisions, internalizing the wealth of the parents, from someone with the same degree but zero parental wealth. Next, we estimate the relationship between the wage dispersion measure and parental net wealth controlling for individual and parents' characteristics such as the level of education. We visualize the results here in Figure 4 and report the IV estimates in the Appendix (A.10).

The results in Figure 4 and the IV estimates in Table A.10 suggest a strong correlation between net parental wealth and dispersion in the returns to labor.[8] This finding indicates a novel mechanism related to the recent literature on the concentration of wealth within families across generations. That literature points out to determinants such as financial risk-taking by investors and direct wealth transfers through bequest, inter alia (Fagereng et al., 2021). Thus the findings suggest that in addition to the set of reasons that generally operate through increasing capital income of the children, parental wealth appears to affect their risk-taking behavior—potentially through occupational choices, among other things—generating larger labor income dispersion for high levels of net parental wealth. This finding is also consistent with the hypothesis that parental wealth acts as an insurance in the form of a private safety net (Pfeffer and Rodems (2021)).

Figure 4: *Labor Income Dispersion and Parental Wealth Levels*



(a) *Labor Income Dispersion* **(b)** *Labor Income Dispersion, Controlling for Education*

Note: The The binned scatterplot shows the estimated relationship between net parental wealth in 1993 and wage dispersion in 2010-2017, controlling for parents and individual's characteristics including education. The measure of wage dispersion is the coefficient of variation defined as the ratio of standard deviation to the mean (averaged within each bin of wealth).

[8]Additionally, we do the same estimation for capital income, and also find that higher wealth is associated with higher dispersion of capital income (i.e., risk-taking), broadly in line with Fagereng et al., 2021 (Appendix, Figure A.2).

5 Conclusion

The discussion on wealth inequality stresses that parental wealth is a significant predictor of future wealth of the children through mechanisms such as wealth transfers and returns to wealth through links operating via *capital* income. Our findings add one more aspect to this discussion. Namely, using exogenous variations in parental net wealth, we find that children from wealthy families tend to have higher *labor* income. The analysis suggests that a wealth tax brings the income of the children closer to their peers from less wealthy families. This finding contributes to the debate on wealth taxation. It does not state that the wealth tax is the only, or the optimal, policy tool to influence intergenerational income inequality, but the results suggest that in the absence of the Norwegian wealth tax, intergenerational income mobility would have been lower.

The results of this paper based on Norwegian data are indicative for other countries. If wealth entails a "privilege effect" on the income of the children in a country with a relatively strong provision of public goods—especially health and education—, this raises the question whether this effect is even more pronounced in countries with lower provision of public goods. Our analysis does lend support to one—and thus far neglected—mechanism through which parental wealth impacts the income of the children. Results indicate heterogeneous returns to labor in the form of positive correlation between wage dispersion and parental net wealth. This finding suggests that the risk profile of occupational choice is influenced by the stock of parental wealth, contributing to the literature that attempts to explain why wealthy parents tend to have well-off children. Future research can shed light on further mechanisms.

References

Benhabib, J., & Bisin, A. (2018). Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature*, *56*(4), 1261–1291.

Bjørneby, M., Markussen, S., & Røed, K. (2020). Does the wealth tax kill jobs?

Boserup, S. H., Kopczuk, W., & Kreiner, C. T. (2018). Born with a silver spoon? danish evidence on wealth inequality in childhood. *The Economic Journal*, *128*(612), F514–F544.

Brülhart, M., Gruber, J., Krapf, M., & Schmidheiny, K. (2019). Behavioral responses to wealth taxes: Evidence from switzerland.

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). The opportunity atlas: Mapping the childhood roots of social mobility.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the united states. *The Quarterly Journal of Economics*, *135*(3), 1567–1633.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, *135*(3), 1553–1623.

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review*, *104*(5), 141–47.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, *27*(3), 79–102.

Duran-Cabré, J. M., Esteller-Moré, A., & Mas-Montserrat, M. (2019). Behavioural responses to the (re) introduction of wealth taxes. evidence from spain.

Fagereng, A., Mogstad, M., & Ronning, M. (2021). Why do wealthy parents have wealthy children? *Journal of Political Economy*, (Forthcoming).

Halvorsen, E., Ozkan, S., & Salgado, S. (2021). Earnings dynamics and its intergenerational transmission: Evidence from norwa. *Working paper*.

Jakobsen, K., Jakobsen, K., Kleven, H., & Zucman, G. (2020). Wealth taxation and wealth accumulation: Theory and evidence from denmark. *The Quarterly Journal of Economics*, *135*(1), 329–388.

Lee, C.-I., & Solon, G. (2009). Trends in intergenerational income mobility. *Review of Economics and Statistics*, *91*(4), 766–772.

Pfeffer, F., & Rodems, R. (2021). Avoiding material hardship: The buffer function of wealth.

Piketty, T., & Zucman, G. (2014). Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics*, *129*(3), 1255–1310.

Ring, M. A. (2020). Wealth taxation and household saving: Evidence from assessment discontinuities in norway. *Working paper*.

Saez, E., & Zucman, G. (2019). Progressive wealth taxation. *Brookings Papers on Economic Activity*.

Scheuer, F., & Slemrod, J. (2020). Taxation and the superrich. *Annual Review of Economics*, *12*(1), 189–211.

Scheuer, F., & Slemrod, J. (2021). Taxing our wealth. *Journal of Economic Perspectives*, *35*(1), 207–230.

Seim, D. (2017). Behavioral responses to wealth taxes: Evidence from sweden. *American Economic Journal: Economic Policy*, *9*(4), 395–421.

Smith, M., Zidar, O., & Zwick, E. (2020). Top wealth in america: New estimates and implications for taxing the rich.

Thoresen, T. O. (2009). Income mobility of owners of small businesses when boundaries between occupations are vague. *CESifo Working Paper Series No. 2633*.

Zoutman, F. T. (2018). The elasticity of taxable wealth: Evidence from the netherlands.

Appendix: Further Results

Table A.1: *Thresholds and Deductions in the Norwegian Wealth Tax 1993-2017*

	Singles Threshold 1 NOK	Threshold 2 NOK	Threshold 3 NOK	Married Threshold 1 NOK	Threshold 2 NOK	Threshold 3 NOK
1993	120,000	235,000	.	150,000	260,000	.
1994	120,000	235,000	530,000	150,000	260,000	570,000
1995	120,000	235,000	530,000	150,000	260,000	570,000
1996	120,000	235,000	530,000	150,000	260,000	570,000
1997	120,000	235,000	530,000	150,000	260,000	570,000
1998	120,000	540,000	.	150,000	580,000	.
1999	120,000	540,000	.	150,000	580,000	.
2000	120,000	540,000	.	150,000	580,000	.
2001	120,000	540,000	.	150,000	580,000	.
2002	120,000	540,000	.	150,000	580,000	.
2003	120,000	540,000	.	150,000	580,000	.
2004	120,000	540,000	.	150,000	580,000	.
2005	151,000	540,000	.	181,000	580,000	.
2006	200,000	540,000	.	400,000	1,080,000	.
2007	220,000	540,000	.	440,000	1,080,000	.
2008	350,000	540,000	.	700,000	1,080,000	.
2009	470,000	.	.	940,000	.	.
2010	700,000	.	.	1,400,000	.	.
2011	700,000	.	.	1,400,000	.	.
2012	750,000	.	.	1,500,000	.	.
2013	870,000	.	.	1,740,00	.	.
2014	1,000,000	.	.	2,000,000	.	.
2015	1,200,000	.	.	2,400,00	.	.
2016	1,400,000	.	.	2,800,000	.	.
2017	1,480,000	.	.	2,960,000	.	.

Until 2006, married couples share one basic allowance and a joint threshold. From 2006, married couples share twice the threshold of singles on their total wealth. The threshold for singles and married is therefore the same independently of the distribution of couple wealth after 2006.

Table A.2: *Summary Statistics, Individuals, 2017*

Strategy	Mean in 1993		
	All	Married 1993-98	Unmarried and divorced 1993-98
Wage	513,155	526,999	454,621
	(372,967)	(372,580)	(368,756)
Capital income	31,770	31,591	32,677
	(445,701)	(363,359)	(694,284)
Total income	576,325	591,325	512,873
	(603,270)	(540,422)	(815,183)
Number of siblings	1.94	1.90	2.10
	(1.16)	(1.11)	(1.34)
Born in urban area	0.145	0.151	0.119
	(0.352)	(0.358)	(0.323)
Sample restrictions	PW>0	PW>0	PW>0
N	63,533	51,318	12,127

Standard deviation in parentheses. All monetary amounts are measured in NOK 1000. PW is net parental wealth in 1993.

Table A.3: *Summary Statistics, Parents (Main Variables)*

Strategy	Mean		
	All	Married 1993-98	Unmarried and divorced 1993-98
Married 1993	0.856		
	(0.351)		
Divorce 1993-1998	0.0552		
	(0.228)		
Net wealth, 1993	466,032	462,658	479,031
Median	233,756	252,307	153,497
	(6,420,489)	(1,096,826)	(14,450,127)
Net wealth, 1998	835,655	908,852	524,426
Median	387,268	437,132	135,752
	(4,461,313)	(5,530,931)	(8,018,887)
Financial wealth, 1993	320,826	302,948	393,668
Median	122,150	127,644	91,918
	(5,956,855)	(1,099,872)	(13,452,354)
Financial wealth, 1998	575,301	596,188	484,025
Median	138,048	150,026	84,916
	(6,703,878)	(6,639,670)	(6,974,936)
Wealth tax payment, 1993	3533	3256	4688
Median	0	0	0
	(83,412)	(13,966)	(188,735)
Wealth tax payment 1998-rules, 1993	2955	2624	4369
Median	0	0	0
	(98,126)	(11,210)	(224,388)
Sample restrictions	PW>0	PW>0	PW>0
N	63,533	51,318	12,127

Standard deviation in parentheses. All monetary amounts are measured in NOK 1000. PW is net parental wealth in 1993.

Table A.4: *Summary Statistics, Parents (Further Variables)*

Strategy	Mean in 1993		
	All	Married 1993-98	Unmarried and divorced 1993-98
Mother's wage	105,846	107,698	98,093
	(83,144)	(81,344)	(89,831)
Father's wage	199,051	213,627	136,577
	(158,989)	(153,314)	(164,258)
Mother's capital income	7307	6762	9605
	(62,510)	(55,654)	(85,728)
Father's capital income	30,340	32,100	22,565
	(291,760)	(286,937)	(311,184)
Mother's total income	127,329	130,281	114,951
	(117,451)	(111,127)	(140,487)
Father's total income	277,991	298,465	190,182
	(372,076)	(370,284)	(363,899)
Mother higher education	0.228	0.230	0.219
	(0.420)	(0.421)	(0.414)
Father higher education	0.173	0.182	0.131
	(0.378)	(0.386)	(0.338)
Sample restrictions	PW>0	PW>0	PW>0
N	63,533	51,318	12,127

Standard deviation in parentheses. All monetary amounts are measured in NOK 1000. PW is net parental wealth in 1993.

Table A.5: *First-Stage IV Estimation Results*

Effect of	Instrument
On net parental wealth	-21031.6***
	(1997.9)
t-value	-10.53
F-value	447.14
R^2	0.625
On parental financial wealth	-8983.5****
	(902.4)
t-value	-9.96
F-value	309.13
R^2	0.510
Sample restrictions	0<PW
N	480,971

*Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All monetary amounts are measured in NOK 1000. PW is net parental wealth divided by the number of siblings in 1993. The effect for Divorce IV is the predicted change in wealth tax payments given parental wealth in 1993 as a percentage of parental wealth in 1993 on the change in parental wealth between 1993 and 1998. Controls are wages, education and age of father and mother, parental wealth and marital status in 1993, whether the individual earns mainly capital income, age, whether the individual is born in an urban area and year dummies. Mobility outcomes are measured in percentiles from father's wage income in 1993 to children's wage income in 2010-2017.*

Table A.6: *Direct Effect of Divorce on Wages of Children*

Strategy Effect of	OLS Parental divorce
On wage	-20.58***
	(2.221)
On wage percentile	-2.112***
	(0.243)
On wage mobility	-0.951***
	(0.263)
On total income	-42.41***
	(3.007)
Sample restrictions	0<PW<100
N	161,897

*Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All monetary amounts are measured in NOK 1000. PW is net parental wealth divided by the number of siblings in 1993. The effect is parental divorce in the period 1993 to 1998. Controls are wages, education and age of father and mother, parental wealth and marital status in 1993, whether the individual earns mainly capital income, age, whether the individual is born in an urban area and year dummies. Mobility outcomes are measured in percentiles from father's wage income in 1993 to children's wage income in 2010-2017.*

Table A.7: *Total Income*

Strategy	OLS	IV	Adjusted IV	OLS	IV	Adjusted IV
Effect of	Net parental wealth			Parental financial wealth		
On total income	0.0329***	0.0879***	0.0871***	0.0352***	0.0887***	0.0796***
	(0.000877)	(0.0116)	(0.0126)	(0.0101)	(0.0110)	(0.0122)
Sample restrictions	0<PW	0<PW	200<PW	0<PW	0<PW	200<PW
N	480,971	480,971	270,995	480,971	480,971	270,995

*Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All monetary amounts are measured in NOK 1000. PW is net parental wealth divided by the number of siblings in 1993. The effect is the change in net parental wealth from 1993 to 1998 instrumented by the wealth tax change. Controls are wages, education and age of father and mother, parental wealth and marital status in 1993, whether the individual earns mainly capital income, age, whether the individual is born in an urban area and year dummies. Mobility outcomes are measured in percentiles from father's wage income in 1993 to children's wage income in 2010-2017.*

Table A.8: *Main Results with Effect of Controls*

Strategy Effect on	OLS	IV Wage	Adjusted IV
Change in net parental wealth 1993-1998	0.00271***	0.0167***	0.0140***
	(0.000356)	(0.00121)	(0.00119)
Net parental wealth 1993	0.00133***	0.00773***	0.00650***
	(0.000182)	(0.000560)	(0.000551)
Parents married in 1993	35.35***	33.85***	34.74***
	(1.194)	(1.209)	(1.250)
Earning mainly capital income 2013-2017	-442.9***	-449.4***	-459.9***
	(0.930)	(1.153)	(1.698)
Father's wage 1993	0.176***	0.142***	0.134***
	(0.00342)	(0.00471)	(0.00593)
Mother's wage 1993	0.179***	0.184***	0.181***
	(0.00548)	(0.00567)	(0.00763)
Born in an urban area	15.81***	16.76***	15.75***
	(1.252)	(1.258)	(1.841)
Father has higher education	23.37***	25.06***	25.78***
	(1.316)	(1.341)	(1.863)
Mother has higher education	23.72***	21.16***	17.64***
	(1.085)	(1.127)	(1.465)
Sample restrictions	PW>0	PW>0	PW>0
N	481319	481319	292673

*Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All monetary amounts are measured in NOK 1000. PW is net parental wealth divided by the number of siblings in 1993. The first effect is the change in net parental wealth from 1993 to 1998 instrumented by the wealth tax change. Mobility outcomes are measured in percentiles from father's wage income in 1993 to children's wage income in 2010-2017.*

Table A.9: *Results: Controlling for Educational Level and Field*

Strategy	OLS	IV	Adjusted IV	OLS	IV	Adjusted IV
Effect of	Net parental wealth			Parental financial wealth		
On wage level	0.00196***	0.0133***	0.0117***	0.00226***	0.0112***	0.00777***
	(0.000310)	(0.00103)	(0.00104)	(0.000238)	(0.000904)	(0.000897)
On wage percentile	0.000209***	0.00130***	0.00108***	0.000248***	0.00106***	0.000670***
	(0.0000272)	(0.000101)	(0.000104)	(0.0000219)	(0.0000873)	(0.0000862)
On wage mobility	0.000652***	0.00313***	0.00291***	0.000602***	0.00281***	0.00209***
	(0.0000637)	(0.000177)	(0.000188)	(0.0000507)	(0.000161)	(0.000155)
Sample restrictions	0<PW	0<PW	200<PW	0<PW	0<PW	200<PW
N	480,971	480,971	270,995	480,971	480,971	270,995

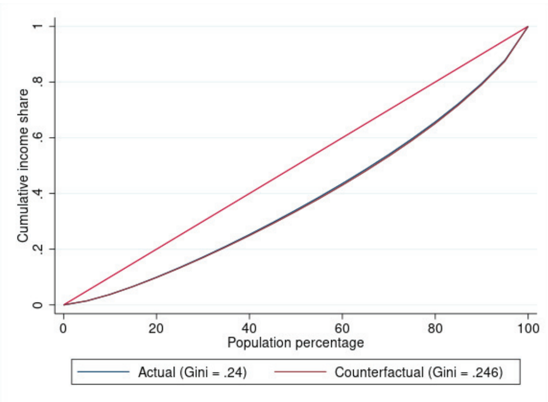
Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. *All monetary amounts are measured in NOK 1000. PW is net parental wealth divided by the number of siblings in 1993. We control for a dummy on whether the individual has completed higher education and a dummy on whether the higher education field was in business or science. The first effect is the change in net parental wealth from 1993 to 1998 instrumented by the wealth tax change. The second effect is the change in parental financial wealth from 1993 to 1998 instrumented by the wealth tax change. Controls are wages, education and age of father and mother, parental wealth and marital status in 1993, whether the individual earns mainly capital income, age, whether the individual is born in an urban area and year dummies. Mobility outcomes are measured in percentiles from father's wage income in 1993 to children's wage income in 2010-2017.*

Table A.10: *Labor Earnings Dispersion*

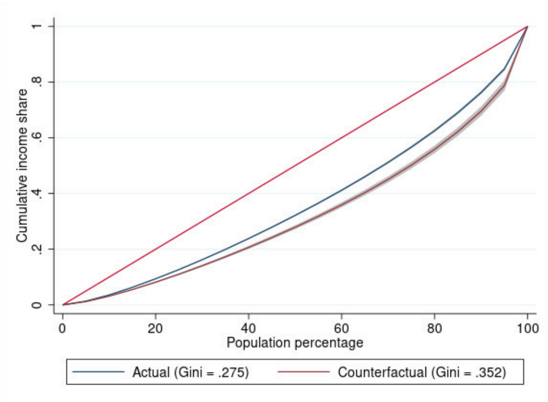
Strategy	OLS (no controls)	OLS	IV
Effect of	Net parental wealth		
Wage income dispersion	0.0000213***	$1.85 * 10^{-6}$***	0.0000135***
	($7.77*10^{-7}$)	($1.19*10^{-7}$)	($9.35*10^{-7}$)
Total income dispersion	$2.19 * 10^{-6}$***	0.000136***	0.0000770***
	($5.01*10^{-7}$)	($7.13*10^{-7}$)	($4.46*10^{-6}$)
N:	481,319	481,319	481,319

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. *All monetary amounts are measured in NOK 1000. The effect of parental wealth on dispersion is calculated by constructing 100 groups of parental wealth and calculating dispersion within each of these groups, before running an OLS regressions of parental wealth in 1993 on the income dispersion measure. The parental wealth groups are constructed beginning with a bin of 100,000 and increasing it by a factor of* $bin^{1.1}$, *to measure dispersion also at the top.*

Figure A.1: *Income Inequality (Considering Heterogeneous Effects of Parental Wealth on Income)*

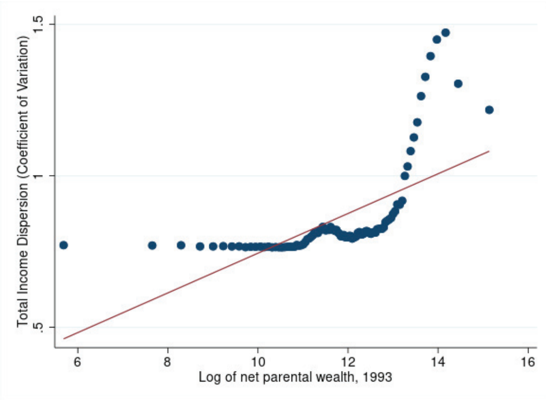


(a) *Wage Income*

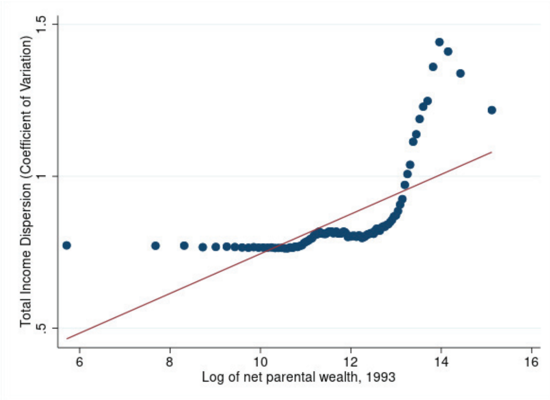


(b) *Total Income*

Figure A.2: *Total Income Dispersion and Parental Wealth Levels*



(a) *Total Income Dispersion*



(b) *Total Income Dispersion, Controlling for Education*

Note: For each bin of the logarithm of net parental wealth in 1993, the figure shows the estimated relationship between net parental wealth and capital income dispersion, controlling for parents and individual's characteristics including education. The measure of total income dispersion is the coefficient of variation defined as the ratio of standard deviation to the mean (averaged within each bin of wealth).

Chapter 4

Problematic Response Margins in the Estimation of the Elasticity of Taxable Income

Joint with Thor O. Thoresen

Published in *International Tax and Public Finance* 27: 721-752 (2020)

International Tax and Public Finance (2020) 27:721–752
https://doi.org/10.1007/s10797-019-09576-3

Problematic response margins in the estimation of the elasticity of taxable income

Kristoffer Berg[1] · **Thor O. Thoresen**[2,3]

Published online: 30 October 2019

Abstract
The elasticity of taxable income (ETI) holds the promise of representing a summary measure of tax efficiency costs, which means that further information about the behavioral components of the ETI is not required for its use in tax policy design. However, since there are response margins that can cause biases in the estimation of the elasticity, this paper warns against neglecting information about the composition of the behavior summarized by the ETI. When using responses of the Norwegian self-employed to the tax reform of 2006 for illustration, we discuss how three different response margins relate to the overall ETI: working hours, tax evasion and shifts in organizational form. We provide empirical illustrations of effects of each of these margins and argue that the standard procedure for estimating the ETI produces a biased estimate due to the organizational shift margin.

Keywords Elasticity of taxable income · Self-employed · Tax evasion · Organizational shift

JEL Classification H24 · H26 · H31 · J22

✉ Thor O. Thoresen
tot@ssb.no

Kristoffer Berg
kristoffer.berg@econ.uio.no

1 Oslo Fiscal Studies, Department of Economics, University of Oslo, Oslo, Norway

2 Oslo Fiscal Studies, Department of Economics, Statistics Norway, University of Oslo, Oslo, Norway

3 CESifo, Munich, Germany

1 Introduction

After Feldstein (1995) it has become widespread to obtain estimates of income responses to tax changes by analyzing panel data over a tax reform period, exploiting the variation in changes in marginal net-of-tax rates across individuals to obtain estimates of the elasticity of taxable income (ETI). In the most straightforward version of the empirical strategy, one identifies a "control group" that represents the change in income which would have occurred to the "treatment group", if the tax reform did not take place. As the ETI in principle captures all tax-induced responses, and as estimates can be derived by standard econometric tools, obtaining estimates of the ETI from micro-data has become a popular empirical strategy for measuring the efficiency costs of taxation (Saez et al. 2012).

In the case when private and social costs of changes in the marginal tax rate are equal, the ETI is considered to be a "sufficient statistic" for welfare analysis, as the optimal tax rate is a simple function of the ETI (Feldstein 1999; Saez 2001; Chetty 2009). Then, the behavioral anatomy of the response does not matter for the measurement, and the ETI represents a summary measure of tax efficiency costs. One should, however, be cautious in the practical implementation of the approach. One reason is that the social implications of the behavioral responses to tax changes differ to the extent there are external effects involved. Externalities may arise because the ETI captures valuable activities, such as charitable giving, or because it reflects detrimental activities, such as tax evasion. Nevertheless, the ETI literature includes contributions on how ETI estimates still can be used to measure tax efficiency effects in the presence of behavioral diversities with different social costs; see Chetty (2009) and Slemrod and Gillitzer (2014).[1]

Further, it is well established that the ETI is a function of the environment from which it is derived, and it is therefore subject to policy control (Slemrod 1996; Slemrod and Kopczuk 2002; Giertz 2009; Fack and Landais 2016). This means that policy-makers often have a range of policy instruments to control different margins of the response, and it implies that the broader tax system design influences the overall ETI through the components of behavioral response.[2]

In the present study, we direct attention to another implication of multiple response margins in the ETI literature, namely that the econometric identification of the ETI is sensitive to what type of response margins are involved. There are well-known econometric challenges concerning the identification of the ETI, given that net-of-tax rate depends on income and therefore is clearly endogenous. Several studies have evidently contributed to how the practitioner still may proceed despite these difficulties, including Auten and Carroll (1999), Moffitt and Wilhelm (2000), Gruber and Saez (2002), Kopczuk (2005) and Weber (2014). However, here we draw

[1] Chetty (2009) differentiates between tax sheltering as transfers to other agents in the economy and real resource costs.

[2] See Doerrenberg et al. (2017) on the use of the ETI as a sufficient statistic in the presence of deduction possibilities.

attention to the possibility that response margins cause bias in the estimation of the ETI.

In the following we discuss empirical evidence of three separate response dimensions—working hours, tax evasion and organizational shifts—arguing that the latter margin likely is a source of bias, given the conventional method to derive ETI estimates. Contributions in the literature have considered measurement problems in estimating effects of reforms originating from changes of organizational form; see Slemrod (1996, 1998), Gordon and Slemrod (2000) and Saez (2004).[3] Here, we set the potential bias into the perspective of the measurement of the overall ETI. Thus, the main message of the present study is that the behavioral anatomy of the ETI may matter as there are response margins that can cause estimation bias.

We discuss the various underlying behavioral responses empirically by employing micro-data on the Norwegian self-employed, exploiting the tax changes due to the tax reform of 2006 in the identification. The behavior of the self-employed is interesting as it is typically assumed that they have wider scope for behavioral response than the wage earners (Heim 2010).[4] Although the share of self-employed in proportion to the total workforce is low in Norway, around 7% (Parker 2009; OECD 2019),[5] their role in the economy receives considerable attention, as illustrated by the considerations in the design of the tax system (which we soon will return to).[6]

Before explaining further why some response dimensions may represent sources of estimation bias and others may not, let us briefly restate the standard method of obtaining ETI estimates. The ETI provides an intensive margin response, which is conventionally identified by addressing information on taxable income over a period where there is variation in the net-of-tax rate (1 minus the marginal tax rate) generated by a tax reform. Thus, inspired by Feldstein (1995), a great majority of empirical studies of the ETI have used panel data in the identification,[7] where first differenced income for each individual in the panel is regressed against an expression for the change in the net-of-tax rate. To allow for the new tax prices to be absorbed by the agents, it is standard to use a 3-year span from pre-reform to post-reform.

[3] In the US context, after TRA 1986, one saw that taxpayers moved from Subchapter C, which includes corporate income tax on profits, toward Subchapter S, implying that profits are taxed directly at the individual level (Saez 2004).

[4] Whereas estimates of the ETI for wage earners have been obtained for a wide selection of countries, see Auten and Carroll (1999) and Gruber and Saez (2002) for the USA, and Aarbu and Thoresen (2001), Blomquist and Selin (2010), Kleven and Schultz (2014) and Matikka (2018) for Norway, Sweden, Denmark and Finland, respectively, there are relatively few studies of the ETI for the self-employed. Exceptions include Wu (2005), Blow and Preston (2002), Heim (2010), Kleven and Schultz (2014). Note also that Saez (2010), le Maire and Scherning (2013) and Bastani and Selin (2014) estimate taxable income elasticities for the self-employed, but use bunching techniques in the identification.

[5] Some simplified calculations based on income statistics (Statistics Norway 2014) suggest that approximately 4–5 percent (measured both at the household level and at the individual level) of total (gross) income comes from business income.

[6] Although self-employment rates are higher in many other countries, they have been falling in most countries over time (OECD 2019).

[7] However, Lindsey (1987) used repeated cross sections. See also Goolsbee (1999).

Following Auten and Carroll (1999), Moffitt and Wilhelm (2000) and Gruber and Saez (2002), most studies use an instrument for the tax change based on statutory tax changes, obtained by letting the tax law at time t and time $t+3$ (mechanically) be applied to the same pre-reform income, using a two-stage-least-squares (2SLS) procedure.

It follows from the standard data selection criteria of the ETI framework that data on the self-employed are established by conditioning on being self-employed in both periods, t and $t+3$. This is an innocuous sample selection condition if the tax changes do not induce taxpayers to move out of the personal income tax base. However, several studies, as Slemrod (1995, 1996, 1998), Gordon and Slemrod (2000), Goolsbee (2000), Saez (2004), Thoresen and Alstadsæter (2010), Edmark and Gordon (2013) and Harju and Matikka (2014), advise against ignoring organizational shifts when discussing tax responses.

Moreover, the organizational shift aspect is clearly critical in the present context, given that we use the Norwegian tax reform of 2006 in the identification of effects, and the tax schedule prior to the 2006-reform is known to have included incentives to shift organizational form, as shown by Thoresen and Alstadsæter (2010). They show that in particular high-income business owners moved out of self-employment and took advantage of the lower taxation of dividend income. As the Norwegian tax reform of 2006 involved tax changes meant to abolish these incentives (Sørensen 2005), both through a reduction in the marginal tax rate on labor income and increased taxation of dividends and capital gains, the composition of the self-employed in the data used to estimate the ETI is likely influenced by the reform. In other words, as high-income taxpayers were overrepresented among those who shifted out of self-employment prior to the reform (Thoresen and Alstadsæter 2010), and as the reform reversed these incentives, we get a non-random change in the treatment group because of self-selection—a case of incidental truncation. If adequate measures are not taken, we are in danger of erroneously attributing increases in income to standard income responses to lower marginal tax rates, whereas it is a sample selection effect and therefore should be characterized as a source of bias in the estimation of the ETI.

We are able to investigate effects of organizational shifts on the ETI because of the richness in the data we have available. The main data source is the yearly Income Statistics for Families and Persons, which is based on information from administrative registers (such as the Register of Tax Returns), covers the whole population, includes a large set of control variables, and can be turned into a panel data set through personal ID numbers. Observations from an unbalanced panel of self-employed over the period from 2001 to 2010 are used in the analysis. Further, we combine the income data with three other data sources in order to explore the extent of organizational shifts: information from the Business and Enterprise Register, the Shareholder Register and the End of the Year Certificate Register. By combining information from these data sources, we can establish whom among the taxpayers has moved out of self-employment to be an employee and shareholder in the same firm. We consider a difference in the patterns of these movements from the pre-reform to the post-reform period as corroborative evidence of a measurement problem in the estimation of the ETI, expected to cause biased estimation results.

In contrast, a change in working hours represents a conventional component of the ETI and causes no bias. We estimate a working hours tax elasticity by employing repeated cross-sectional data, derived from the Labor Force Surveys. Correspondingly, we categorize tax evasion as a standard component of the ETI, and illustrate the effect of tax evasion empirically by using the so-called expenditure approach (Pissarides and Weber 1989) for identification of the tax evasion component, using consumption and income data in combination.

The paper is organized as follows: In Sect. 2, we present the Norwegian tax schedule and the reform of 2006, which is used in the identification of the ETI. Further, in Sect. 3, the empirical approaches to obtain estimates of the effects of different response margins and the overall ETI are presented, before estimation results for the different response margins are presented in Sect. 4. Section 5 concludes the paper.

2 The Norwegian dual income tax and the reform of 2006

The reform of the Norwegian dual income tax schedule in 2006 is used to obtain tax response estimates. A dual income tax schedule combines a low proportional tax rate on capital income and progressive tax rates on labor income, and was introduced in Norway by the tax reform of 1992. Thus, as the system involves separate rate schedules for different income components, there are certainly incentives for a variety of behavioral effects when reforming the system, as in the 2006-reform.

The dual income tax proliferated throughout the Nordic countries in the early 1990s, and the Norwegian version had a flat 28% tax rate levied on corporate income, capital and labor income, coupled with a social security contribution and a progressive surtax applicable to labor income. The post-1992 schedule involved a system for mitigating corporate double taxation of dividends which effectively eliminated the personal dividend tax. The capital gain tax system exempted gains attributable to retained earnings taxed at the corporate level. Given the low flat tax rate of 28% on capital and corporate income and an additional progressive schedule on high labor income, there were obvious incentives for taxpayers to recharacterize labor income as capital income. To limit such tax avoidance, the 1992-reform introduced the so-called split model for the self-employed, partnerships and closely held firms[8]: the split model involved rules for dividing business income into capital and labor income by imputing a return to business assets and attributing the residual income to labor. Between 1992 and 2004, both the threshold for the second tier of the surtax and marginal rates increased, resulting in the statutory marginal tax rates as shown for 2004 (the last year before the reform) in Fig. 1, with 55.3% at the maximum.[9] Thus, as the self-employment labor income, for large income intervals, was taxed by more than 50% at the margin, a substantial number of taxpayers moved out of the "split model" and established their firm under an incorporated form, to take

[8] The latter is defined as businesses in which the active owner holds more than two-thirds of the shares.
[9] Use 1$ = 6.42 Norwegian kroner (NOK) and 1€ = 8.05 NOK (2006-rates) to convert to US dollars and Euros.

advantage of the lower tax on dividends and capital gains (Thoresen and Alstadsæter 2010). Recall that it is the reduced incentives to incorporate after the 2006-reform, which in the present study is expected to create bias in the estimation of the ETI.

The 1990s saw increasing pressure on the dual income tax system, resulting in numerous "patches".[10] For example, a distinction in the tax treatment between liberal professions (lawyers, dentists, doctors and other independent contractors delivering services to the public) and other professions was introduced, and kept as part of the tax schedule until the split model was eliminated in 2004.[11] In Fig. 1, which describes schedules before and after the reform, the remarkable system for non-liberal professions prior to the reform in 2006 is also described.[12] Note that 2005 is not treated as a pre-reform year, as the tax reform was phased in that year.

The reform of 2006 emerged as an attempt to create a system that would prevent taxpayers from transforming labor income into capital income to benefit from the lower rate applied to the latter; see Sørensen (2005) for the wider background to the reform and steps taken to adjust the dual income tax. Harmonization of the marginal tax rates on capital income and labor income is achieved by cutting top marginal tax rates on the wage part; see Fig. 1. This tax cut represents an increase in the net-of-tax rate for most taxpayers and is the tax change we use to derive ETI estimates for the self-employed here. Business income from a sole proprietorship activity in excess of the risk-free return allowance, calculated on the invested capital, is taxed as imputed wage income. The other initiative to curb the incentives to shift income comes from increases in the taxation of dividends and capital gains. The combination of the corporate tax and the personal capital income tax means that dividends and capital gains are taxed at 48.2% at the maximum after the reform in 2006, above a rate-of-return allowance, that is, on profits above a risk-free rate of return.[13]

3 Problematic responses?

3.1 Estimation of the overall ETI

In this section, we discuss to what extent various response margins reflected in the overall ETI represent sources of estimation bias, or if they can be seen as conventional components of the ETI. Estimates of the overall ETI for the self-employed are few, compared to both results for wage earners (see Footnote 4) and to the literature on how tax changes affect decisions to enter or exit self-employment; see

[10] Christansen (2004) sees this as resulting from political games motivated in part by the concerns of politicians of various colors with special interest groups.

[11] This particular schedule represents a separate opportunity for identification of response to tax changes, but, as seen in the figure, it only applies to very large incomes.

[12] In 2000 the share of the self-employed belonging to the liberal professions was 42% (Ministry of Finance 2003).

[13] Thus, this is a clear example of policy-makers having access to several tools for tax optimization—recall that the ETI in this perspective can be seen as controlled by the policy-makers (Slemrod and Kopczuk 2002).

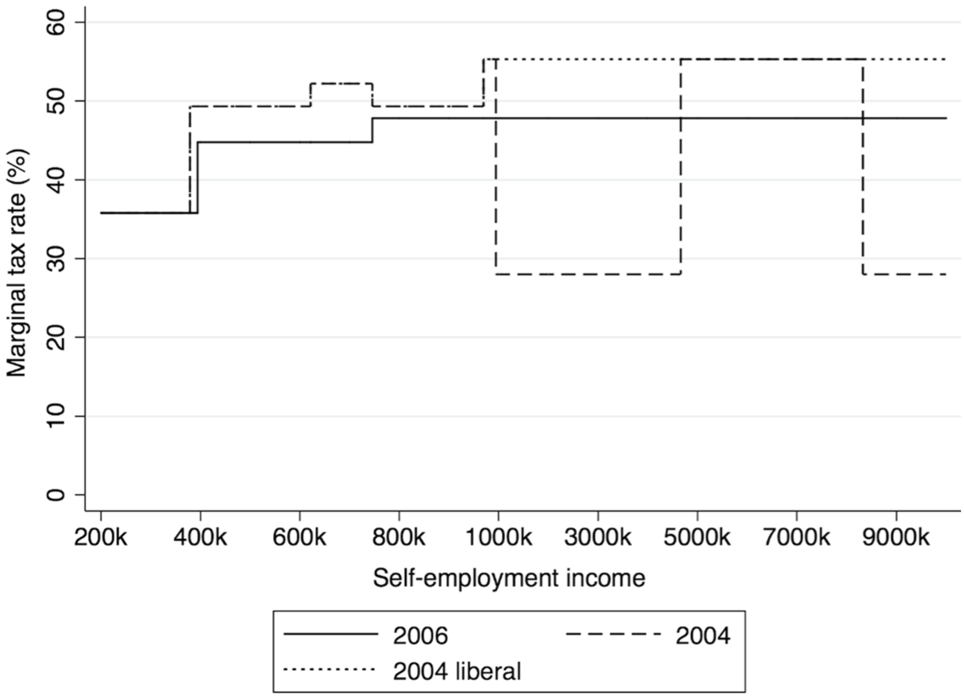


Fig. 1 Marginal tax rates for the self-employed in 2004 (liberal and non-liberal businesses) and 2006. *Note*: There is a break in the horizontal axis at 1,000,000 NOK (1 mill. NOK ≈ \$ 156,000, ≈ € 124,000, in 2006). The "liberal" professions include lawyers, dentists, doctors and other similar professions

reviews in Parker (2009) and Heim (2010). Heim (2010) and Kleven and Schultz (2014) provide ETI estimates for the self-employed by using the same methodology as employed here, whereas Saez (2010), le Maire and Scherning (2013) and Bastani and Selin (2014) obtain ETI estimates by using bunching techniques.

Subsequent to Feldstein (1995), a standard estimation procedure for the identification of the ETI has been developed, benefitting from contributions by, among others, Auten and Carroll (1999), Moffitt and Wilhelm (2000), Gruber and Saez (2002), Kopczuk (2005) and Weber (2014).[14] Recall that in the estimation of the elasticity, $e = \frac{1-\tau}{x}\frac{\delta x}{\delta(1-\tau)}$ (τ is the marginal tax rate, x is income), the main data source is income panel data, covering a period with assorted variation in the net-of-tax rate across individuals. Using 3-year differences, the estimated equation can be specified as

$$\log\left(\frac{x_{i,t+3}}{x_{it}}\right) = \alpha_t + \beta \log\left(\frac{1-\tau_{i,t+3}}{1-\tau_{it}}\right) + B_i'\theta + M_{it}'\eta + \rho \log x_{it} + \varepsilon_{it}, \quad (3.1)$$

where x_{it} and $x_{i,t+3}$ are taxable income for individual i before and after the reform (t and $t+3$), $1-\tau_{it}$ and $1-\tau_{i,t+3}$ are the corresponding net-of-tax rates, α_t is a time-specific effect, B_i is a vector of individual observed characteristics that are time-invariant (but may change relationship with income over time), and M_{it} is a vector of observed time-variant variables. β and ρ are parameters, whereas θ and η are vectors

[14] Note that there is another acronym too: Goolsbee (1999) refers to studies in this field as belonging to the "New Tax Responsiveness" (NTR) literature.

of parameters and the error term, ε_{it}, is assumed to be independently and identically distributed.

As already noted, the marginal tax rate in this setup is clearly endogenous, and studies typically employ the change in net-of-tax rates based on fixed first period income as an instrument in an IV regression; see Auten and Carroll (1999) and Gruber and Saez (2002). The instrument is obtained by letting the tax rate in year $t+3$ be applied to income in year t (base year), inflated by the average income growth. This means that $\log\left(\frac{1-\tau_{i,t+3}}{1-\tau_{it}}\right)$ is instrumented by $\log\left(\frac{1-\tau^{I}_{i,t+3}}{1-\tau_{it}}\right)$, where $\tau^{I}_{i,t+3}$ symbolizes the marginal tax rate in year $t+3$ when applied to income of year t.

The difficulty with this representation of the tax change is that $\log\left(\frac{1-\tau^{I}_{i,t+3}}{1-\tau_{it}}\right)$ is likely correlated with the differenced error in Eq. (3.1); see discussion in Moffitt and Wilhelm (2000). Mean reversion stems from individuals with temporarily high levels of income in period t, and therefore mistakenly placed in the treatment group with large reductions in marginal tax rates, returning to their normal income levels in period $t+3$. To account for the mean reversion bias, Auten and Carroll (1999) suggest including $\log x_{it}$, log of base year income, as an additional control variable; see Eq. (3.1).

Further, Gruber and Saez (2002) propose adding a ten-piece spline in the log of base year income (each decile of the income distribution) to account for mean reversion and (exogenous) developments in the income distribution, while Kopczuk (2005) suggests including splines in the lagged base year income and in the deviation of lagged base year income from base year income to separately estimate the mean reversion and exogenous trend components. These approaches can be seen as

$$\log\left(\frac{x_{i,t+3}}{x_{it}}\right) = \alpha_t + \beta\log\left(\frac{1-\tau_{i,t+3}}{1-\tau_{it}}\right) + B_i'\theta + M_{it}'\eta + \mu \text{ Splines } \log x_{it} + \varepsilon_{it}, \tag{3.2}$$

in the Gruber and Saez specification, and

$$\begin{aligned}\log\left(\frac{x_{i,t+3}}{x_{it}}\right) = \alpha_t + \beta\log\left(\frac{1-\tau_{i,t+3}}{1-\tau_{it}}\right) + B_i'\theta + M_{it}'\eta + \phi \text{ Splines } \log x_{it} \\ + \pi \text{ Splines } \log\left(\frac{x_{i,t-1}}{x_{it}}\right) + \varepsilon_{it},\end{aligned} \tag{3.3}$$

in the Kopczuk version. It follows that μ, ϕ and π are vectors of parameters.

Finally, Weber (2014) criticizes the use of first-year income as a basis for the instrument,[15] and suggests using higher lags of base year income instead. The main argument is that an instrument constructed from the appropriate lag is orthogonal to the error term, which in turn renders the mean reversion control superfluous. Consequently, we show estimation results for a version where $\log\left(\frac{1-\tau_{i,t+3}}{1-\tau_{it}}\right)$ is instrumented by income of the year preceding the base year. In addition to estimation results of

[15] Blomquist and Selin (2010) also address this problem.

the latter specification, in Sect. 4, we shall present ETI estimates based on Eqs. (3.1), (3.2) and (3.3), using 2SLS and controlling for a number of individual characteristics (included in B_i' and M_{it}').

It follows from this exposition that the exogeneity of the tax change instrument is the key condition for consistent estimation of the ETI. Estimation bias appears when there are systematic differences across income groups correlated with, but not caused by, the tax reform under investigation. As we soon will return to, we shall use data for the period 2001–2010 in the estimation, thus employing observations from outside the reform period too. But in terms of the conventional ETI framework, when the last pre-reform year (2004) is used to establish the tax change variable, consistent estimation of the ETI relies on the people experiencing no or small changes in net-of-tax rates, in the present case mostly taxpayers with low and median income, representing a valid control group for the main targets of the reform.

As revealed by this brief review, there are obvious methodological weaknesses and challenges in the standard procedure of obtaining ETI estimates. Here, however, we would like to draw attention to additional problems in the estimation of the elasticity, namely that some of the underlying response margins may cause inconsistent estimates.

3.2 Working hours response

Let us start with what we believe is a less problematic response margin—the response in working hours. To obtain empirical evidence about this response margin is usually challenging. Scarcity of data sets with a panel dimension on working hours partly explains why we see fewer studies (along the same lines as described here) with changes in working hours as the dependent variable.[16] However, cross-sectional data can also be used to obtain ETI estimates, as emphasized by Saez et al. (2012), and here we use ten cross sections from the Labor Force Surveys (Statistics Norway 2003), covering the period 2001–2010, to identify the response in working hours to the tax change.

Thus, the ambition is to obtain an elasticity estimate, $e_h = \frac{1-\tau}{h}\frac{\delta h}{\delta(1-\tau)}$, derived from repeated cross sections. Given that we have access to information about working hours through cross-sectional data (2001–2010), the identification strategy relies on assigning individuals to the treatment and control groups, exploiting that all observations in the repeated cross sections are linked to the panel income data. An instrument (grouping variable) for the net-of-tax rate change is obtained to predict the net-of-tax rate in a first stage of a 2SLS regression. This is done by letting pre-reform income, income over the period 2002–2004, be taxed by the tax laws of 2004 and 2007, obtaining a dummy variable, D_i, that differentiates between taxpayers exposed to an increase in the net-of-tax rate or not. When also introducing a dummy variable for the post-reform period, Q_t, the variable D_iQ_t is the excluded variable in

[16] An important reason for income being the preferred measure is that it reflects the overall efficiency costs of taxation, as made clear by Feldstein (1995, 1999).

the first stage. This means that we essentially estimate a cross-sectional difference-in-differences model. The main regression can be seen as

$$\log h_{it} = \alpha + \lambda_t + \gamma D_i + \delta\left(1 - \tau_{it}\right) + B_i'\theta + M_{it}'\eta + \omega_{it}, \tag{3.4}$$

where h_{it} is working hours for individual i in the cross section at time t, and $\left(1 - \tau_{it}\right)$ is the variable predicted by the first stage. α is a constant, and λ_t symbolizes calendar year. As for the estimation of the overall ETI (see Sect. 3.1), B_i and M_{it} refer to individual characteristics (but here the distinction between time-invariant and time-variant characteristics is not important), and ω_{it} is the error term. It follows that the identification of the effect of net-of-tax rate on working hours benefits from the tax treatment variable being detached from the dependent variable.[17]

3.3 Contribution from tax evasion

Next, we would like to see how the tax evasion component relates to the ETI for the self-employed. The self-employed are known to be disproportionately more involved in tax evasion than wage earners, which has led Heim (2010) to distinguish between income reporting and real effects in the discussion of the ETI for the self-employed. In fact, in many studies the identification of the tax evasion component relies on wage earners not evading tax, while the self-employed do. But are there reasons to caution against the tax evasion dimension in terms of estimation inconsistency? We argue that the tax evasion response does not represent a source of bias. In the following we show how we identify this component of the ETI.

It is not obvious how changes in marginal tax rates affect tax evasion, and thereby, it is uncertain whether the tax evasion component of the overall ETI estimate holds a negative or positive sign. The theoretical literature, as Allingham and Sandmo (1972) and Yitzhaki (1974), offers no clear answers,[18] and empirical findings are mixed (Freire-Serén and Panadés 2013). Some of the early studies, such as Clotfelter (1983), find increased tax evasion for higher marginal tax rates. More recently, Kleven et al. (2011) obtain a very small positive relationship, based on a randomized tax enforcement experiment in Denmark, whereas Gorodnichenko et al. (2009) find a strong positive relationship.

Nevertheless, it seems that the reasoning in the self-employment ETI literature (Heim 2010; Doerrenberg and Duncan 2014) is based on a perspective where reported income is increasing in the net-of-tax rate, i.e., that tax evasion is increasing in the marginal tax rate. This means that ETI estimates for the self-employed are larger than for wage earners if there is a discernible effect on tax evasion from a reduction in the marginal tax rate.[19] We obtain an estimate of the tax evasion

[17] There are likely mean reversion effects in work, as people have temporarily high income in the year used to measure the change in the net-of-tax rate due to the reform (in 2004), because they have high working hours. But such effects are expected to be equally present pre- and post-reform and should therefore not bias estimates.

[18] In the seminal model of Allingham and Sandmo (1972), a tax increase has two contradicting effects on tax evasion: the return to cheating goes up, but at the same time it lowers (full compliance) post-tax income, which most likely makes people more risk averse.

[19] See also Kuka (2014) on obtaining a tax evasion component, but with the use of bunching techniques.

component by addressing estimates of tax evasion before and after the 2006-reform, using the so-called expenditure approach (Pissarides and Weber 1989). This method builds on one group reporting income correctly and another not, but both groups reporting food expenditures truthfully. Thus, this part of the analysis involves the use of consumption data from the Survey of Consumer Expenditure (Holmøy and Lillegård 2014). Under the assumption that the two groups share the same preferences for food, given a set of observable characteristics, estimates on the degree of underreporting among evading households are obtained by exploiting observations on income and food expenditures. More precisely, a common point of departure is the log-linear Engel function, $\log C_h = Z_h'\psi + \xi \log Y_h^*$, where $\log C_h$ is the log of food expenditure for household h, Z_h is a set of observable household characteristics, and $\log Y_h^*$ is the log of "true" disposable household income.[20] A standard assumption is that underreporting takes place at a constant fraction, such that $Y_h^* = kY_h$, where Y_h is the reported income, and there is underreporting if $k > 1$. Here, as in Engström and Holmlund (2009), the following reduced form specification is employed[21]

$$\log C_h = Z_h'\psi + \mu \log Y_h + \kappa SE_h + u_h, \tag{3.5}$$

where SE_h is a dummy for being self-employed and u_h is the error term. A positive κ suggests that the self-employed underreport income, and the number which can be used to multiply reported self-employment income to obtain "true income", is given by $k = \frac{\kappa}{\mu}$, the relationship between the shift parameter, κ, and the slope of the Engel curve, μ. It follows that estimates of k before and after the 2006-reform are used to give an estimate of the tax evasion component of the ETI.

3.4 Organizational shifts generate measurement problem

Now, we direct attention to how we obtain information about a dimension that potentially imposes bias in the estimation of the ETI, namely organizational shifts. As already discussed in Introduction, we expect that the panel data of the self-employed reflect that individuals likely have responded to the tax changes by changing their incorporation decision. Given that high-income taxpayers were overrepresented among those who shifted out of self-employment prior to the reform (Thoresen and Alstadsæter 2010), and because the 2006-reform substantially reduced incentives to incorporate, a different set of business owners remain in the self-employment data sample after the reform (compared to counterfactual, with no changes in incorporation incentives).[22] We interpret the evidence presented in Papini (2018) in support of this, as he finds clear effects of tax-induced organizational shifts after the 2006 reform, when using the differentiated payroll tax schedule of Norway in the

[20] Thus, reflecting that the household is the economic unit in the consumption data.

[21] As both income and k are assumed to be stochastic according to Pissarides and Weber (1989), there are more complications involved when obtaining estimates of k, discussed with respect to Norwegian data in Nygård et al. (2019).

[22] One may employ a balanced panel of self-employed individuals for the whole time period 2001–2010 in the empirical investigation, but this would not eliminate the bias from sample selection.

identification. Then, the sample attrition is a result of non-random self-selection, or incidental truncation, and the organizational shift response margin is a source of bias in the identification of the ETI. This effect has been addressed in several studies from the USA. For instance, at the same time of the Tax Reform Act of 1986, which has been used in several studies of the ETI in the USA as it gave substantial reductions in the top marginal tax rate, numerous shifts of business income from so-called C corporations to so-called S corporations are reported (Slemrod 1996; Saez 2004; Gordon and Slemrod 2000).[23] Here, we go further and relate our findings explicitly to the measurement of the overall ETI.

We explore the extent of organizational shifts before and after the tax reform by utilizing information from three different registers: the Business and Enterprise Register, the Shareholder Register and the End of the Year Certificate Register. By combining information from these three data sources with the income data, individuals are linked to companies, in terms of ownership, employment and transfers of dividends. In turn, these data are used to distinguish between individuals who move out of self-employment because of a "real" change in occupation (i.e., decide to take on paid employment), and those who turn up as wage earners because they have decided to run their business as an incorporated firm.

In practice, we measure the effects of changes in the patterns of organizational shifts by introducing weights in the calculation of the ETI, where the weights are derived from an estimation of the probability to leave self-employment. We employ inverse probability weighting, which is an alternative to the Heckman approach for handling non-random selection; see Wooldridge (2002, 2010). Thus, we estimate a probit model for incorporation, letting it be explained by marginal tax rates faced by the individual (as self-employed), wage income, capital income and other observable characteristics (such as education, gender, birth country, etc.). The probability of incorporation can be seen as

$$s_{it} = a_t + \xi\tau_{it} + \rho Q_t\tau_{it} + \zeta\tau_{it-1} + \sigma w_{it-1} + \chi c_{it-1} + Z'_{it}\varphi + \upsilon_{it}, \tag{3.6}$$

where s_{it} is a dummy for shifting to an incorporated business in year t, α_t is the time-specific effect, τ_{it} and τ_{it-1} are the income tax rates faced by the self-employed in year t and $t-1$, respectively, and Q_t is a dummy variable for the after-reform period, which is interacted with τ_{it} to allow for the effects of the income tax to differ before and after the tax reform. Further, w_{it-1} and c_{it-1} are wage income and capital income in year $t-1$, respectively, Z'_{it} is a group of other potential predictors of shifting, including gender, age, education, county and birth country. υ_{it} is the error term and ξ, ρ, ς, σ, χ and φ are parameters.

The consistency of the inverse probability weighting estimation hinges on the variables included in the shifting equation predicting shifting sufficiently well, such that conditional on these variables, shifting is independent of the unobservables affecting the tax response in Eq. 3.1. We cannot be fully confident that this assumption holds in the present context, but, nevertheless, use our estimates for the

[23] See Christansen and Tuomala (2008) for a discussion of consequences of income shifting for optimal taxation.

empirical illustration of the impact of the organizational choice dimension without further qualification. Wooldridge (2002, 2010) provide more details on the assumptions required for consistency of this technique.

4 The overall ETI and its components

4.1 The ETI of the Norwegian self-employed

As already noted, there are numerous studies of the responsiveness of wage earners using the standard method to derive estimates of the ETI, whereas there are relatively few estimates of the ETI for the self-employed. Two recent studies of the ETI for the self-employed are Heim (2010) and Kleven and Schultz (2014). Heim suggests that the overall elasticity is around 0.9 for the USA, and identifies a "real" elasticity part of approximately 0.4 when controlling for tax evasion.[24] Kleven and Schultz, using data for Denmark, find that the total elasticity of taxable income is about twice as large for the self-employed compared to the wage earners. However, both elasticity estimates are relatively small, and approximately 0.1 for the self-employed.[25]

In the present study, we benefit from having access to large administrative data sets, close to 60,000 self-employed each year, based on information from income tax returns (Statistics Norway 2005) and other administrative registers. Self-employment is defined by conditioning on both self-employment income being higher than wage income and yearly income being larger than two "basic amounts", where the basic amount is a concept of the Norwegian Social Insurance Scheme, corresponding to 62,161 NOK in 2006 ($9700; €7700).[26] See Tables 7 and 8 in the Appendix for summary statistics. As we use data for the period 2001–2010, we have access to information about 400,000–500,000 3-year differences in the estimation of the ETI. This also means that observations from periods without any major changes in the net-of-tax rates are included. Thus, there is no clearly identified control group of this analysis, also because the identification relies on differential tax changes for the treated taxpayers.

The main issue in this type of study is that the identification of the effect of the net-of-tax rate often becomes blurred, as both the mean reversion control and the tax change instrument depend on income. This problem is alleviated here by including periods both with and without tax changes in the estimation, and it is also reduced by

[24] Heim (2010) distinguishes between a real response part and an evasion part by adopting estimates of Clotfelter (1983) and Joulfaian and Rider (1998) for the latter.

[25] Of course, one should not necessarily find similar response estimates across countries and across studies. One obvious source of variation in estimates is the size of the tax reform used in the identification of effects, as discussed by Chetty (2012). However, as the literature seems to suggest stronger responses in the USA than in the Scandinavian countries, one should take a closer look at explanations in the future. See also Kleven (2014).

[26] Also, those with higher negative self-employment income than other types of income are included among the self-employed.

Table 1 ETI estimation results

	(1)	(2)	(3)	(4)
Net-of-tax rate	0.126***	0.103***	0.169***	0.173***
	(0.017)	(0.018)	(0.018)	(0.015)
Age	0.010***	0.009***	0.006***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.085***	0.080***	0.062***	0.060***
	(0.001)	(0.002)	(0.003)	(0.003)
Children	0.019***	0.018***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
Married	0.015***	0.014***	0.016***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)
Norwegian born	0.023***	0.021***	0.016***	0.016***
	(0.003)	(0.004)	(0.004)	(0.004)
Log of period t (Auten/Carroll)	X			
Splines of log of period t income (Gruber/Saez)		X		
Splines of log $t-1$ income and log deviation between $t-1$ and t incomes (Kopczuk)			X	X
Number of observations	406,375	406,375	347,196	337,329

Estimation based on instrumental variable estimation (using 2SLS), with 3-year differences, corresponding to Eqs. 3.1, 3.2 and 3.3. In column (4) the instrument is based on lagged income (period $t-1$), as suggested by Weber (2014). Additional control variables in all regressions: age squared, dummy variables for educational field, length of education, county and years. Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

the tax burden depending on other characteristics than income alone. With respect to the latter, information about type of profession, given the different tax treatment of liberal and non-liberal professions (see Fig. 1), is used, and it is also helpful that marginal tax rates are lower for people located in the northern part of Norway.

Table 1 presents estimation results for four different specifications—the first three, columns (1)–(3), corresponding to Eqs. (3.1)–(3.3); see Sect. 3. Recall that we also estimate the model when defining the tax change instrument based on period $t-1$ income, as suggested by Weber (2014); see column (4).[27] The table demonstrates that results to some extent are sensitive with respect to the mean reversion control technique used. However, all estimates point to relatively small effects, in the range from 0.10 to 0.17. These estimates are not far from those Kleven and Schultz (2014) found for Denmark, and as them, we find results which indicate that the self-employed are somewhat more tax responsive than the wage earners. We use findings of Thoresen and Vattø (2015) as evidence for the tax responsiveness of Norwegian

[27] Note that the specification behind the results of column (4) also includes a mean reversion control, to control for possible heterogeneity in income developments across groups. Note also that in the preferred specification of Weber (2014), the instrument was based on 2-, 3- and 4-year lags, whereas a 1-year lag is used here.

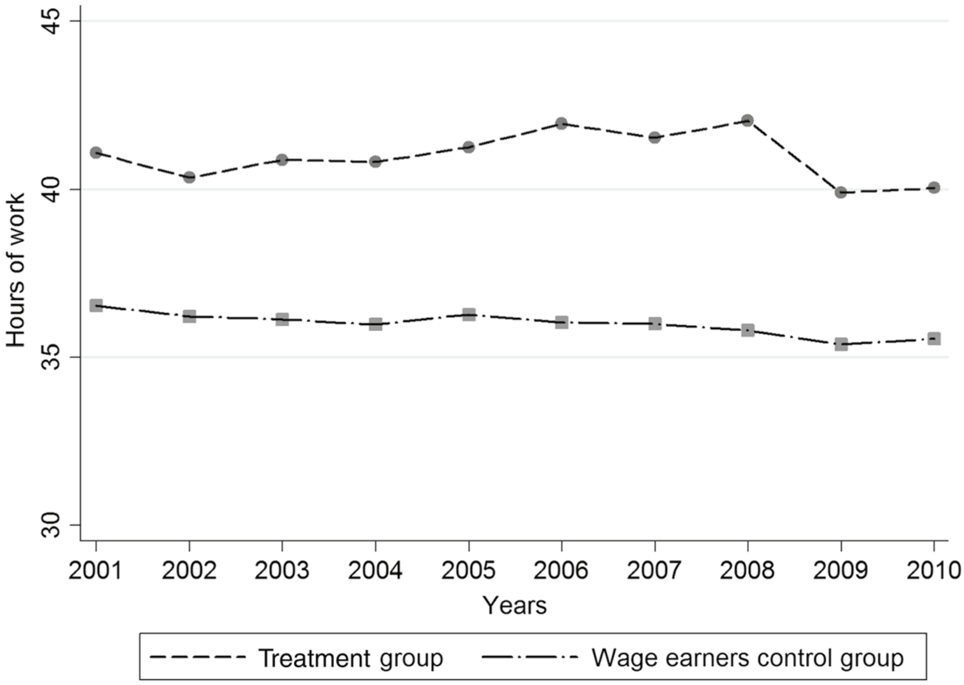


Fig. 2 Average working hours for the treatment group and the control group. Both wage earners and self-employed in the control group. *Note*: Whereas the data used in the estimation of the working hours dimension are from 2002–2004 (pre-reform) and 2007–2009 (post-reform), some additional pre-reform and post-reform years are added here

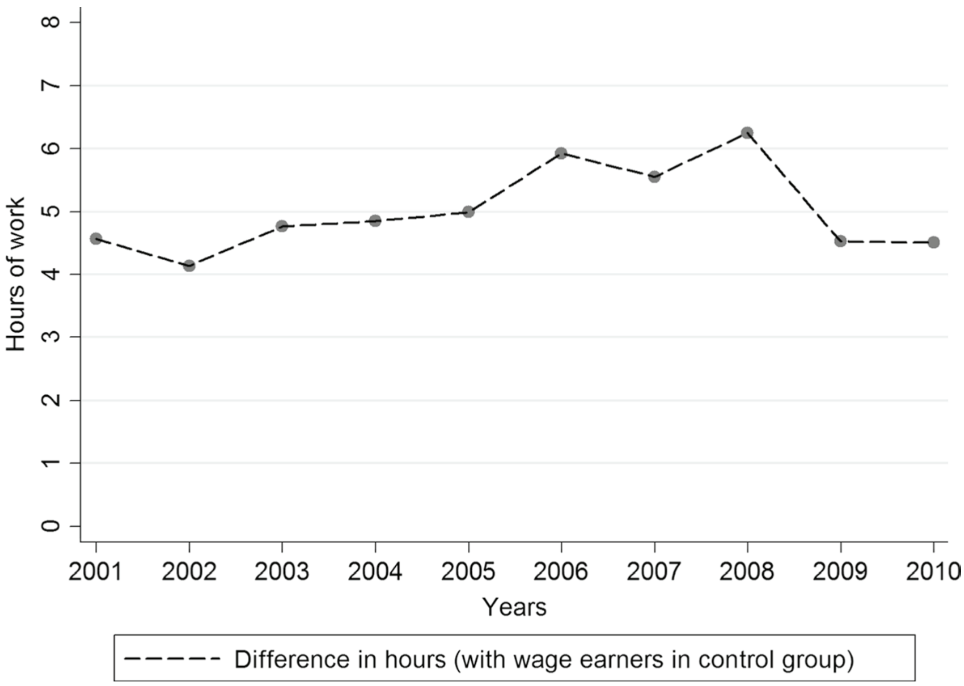


Fig. 3 Difference in average working hours between the treatment group and the control group. Both wage earners and self-employed in the control group. *Note*: Whereas the data used in the estimation of the working hours dimension are from 2002–2004 (pre-reform) and 2007–2009 (post-reform), some additional pre-reform and post-reform years are added here

wage earners (for the same tax reform)—their main ETI estimate for wage earners is 0.06.

4.2 Estimation results for working hours

As explained in Sect. 3.2, due to constraints in the access to information about hours of work for the self-employed, estimates of this response component are obtained by using information from repeated cross sections, derived from the Labor Force Surveys (Statistics Norway 2003). As the Labor Force Survey consists of approximately 22,000 observations per year in total, it follows that the evidence with respect to responses in working hours is based on a smaller data set than the one used to obtain estimates of the overall ETI.

Recall that estimates of responses in working hours are obtained by using an instrument for the net-of-tax rate based on dividing the sample into treatment and control groups.[28] The relation is estimated using data for three pre- and post-reform years, 2002–2004 and 2007–2009. In one of the specifications we also include wage earners (who experience no tax changes) in the control group. More information about the data can be found in the Appendix, Tables 9 and 10. Figures 2 and 3 provide some support for a common trend prior to the reform and some increases in the hours of work among the treated after the reform, but the graphical evidence is unclear.[29]

Table 2 presents response estimates for the two alternative specifications, dependent on the definition of the control group. As explained in Sect. 3.2, δ in Eq. (3.4) is the working hours elasticity estimate that we compare to the overall ETI. The estimated response ranges from 0.20 to 0.23, but only the tax treatment estimate in column (2) is significantly different from zero. In other words, only when adding wage earners to the control group do we obtain a statistically significant result for the tax treatment variable. However, we see that the point estimate of the regression for the self-employed only, reported in column (1), is almost identical to this estimate.

The lack of clear identified effects on working hours prevents us from making any strong statements about the relationship between the hours of work elasticity and the overall ETI. However, we note that the point estimates for working hours are close to the overall ETI estimates, suggesting that a large part of the ETI reflects hours of work adjustments to the tax changes.

4.3 Less tax evasion after the reform?

Next, we add the tax evasion component to ETI response account, by examining to what extent the overall ETI estimate is influenced by changes in the income

[28] Note that Saez et al. (2012) argue that repeated cross-sectional analysis may be preferable to panel data studies in some contexts.

[29] We may conjecture that the reform is too small to see very strong effects by visual inspection. See also Chetty (2012) on how adjustment frictions and the size of the reform may influence elasticity estimates.

Table 2 Estimation results for working hours and net-of-tax rate regression

	(1) Small control group (self-employed)	(2) Large control group
Net-of-tax rate	0.198	0.234*
	(0.193)	(0.131)
Treated	0.038	0.126***
	(0.024)	(0.017)
Constant	3.445***	3.687***
	(0.114)	(0.057)
Age	0.009**	-1.10×10^{-3}
	(0.004)	(0.001)
Male	0.160***	0.103***
	(0.012)	(0.004)
Number of children	−0.023**	−0.017***
	(0.010)	(0.002)
Married	−0.001	−0.015***
	(0.009)	(0.002)
Norwegian born	0.048**	0.011***
	(0.020)	(0.003)
Number of obs.	3664	64,900

Estimation based on difference-in-differences technique; see Eq. 3.4. Additional control variables: age squared, dummy variables for educational field, length of education, county and years. Robust standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

reporting caused by the tax reform.[30] Table 3 presents separate estimation results for the coefficient k, before and after the reform, which gives the number by which the average self-employed person's income has to be multiplied in order to obtain the "true" income. As discussed in Sect. 3.3, we are inclined to expect a reduction in tax evasion from lower marginal tax rates, and in accordance with this, we see a 2.5 percentage point reduction in k when moving from the pre-reform to the post-reform tax schedule.[31] Note that the difference in the estimate of k is not significant, even though we observe a clear reduction in the self-employment parameter estimate.[32] However, to illustrate the implication of the point estimate for k in terms of the overall ETI, a "back-of-the-envelope" calculation suggests that the tax evasion component of the ETI is approximately 0.04, which is approximately one-fourth of the

[30] Tables 12 and 13 in the Appendix provide more information about the data used in this part of the analysis, which primarily are from the Survey of Consumer Expenditure.

[31] Weaknesses in the empirical approach are admitted, although we do not believe them to affect results. For example, ideally, we would like to use a measure of permanent income when estimating the relation between consumption and income, as done in Nygård et al. (2019).

[32] We use the delta method to calculate standard errors for k, based on parameter estimates of κ and μ.

Table 3 Estimation results for parameters used to calculate tax evasion before and after the reform

	Before reform 2003–2004	After reform 2006–2007
Income	0.597***	0.554***
	(0.043)	(0.036)
Self-employed	0.109**	0.087**
	(0.046)	(0.044)
Age	0.033***	0.045***
	(0.007)	(0.007)
Male	−0.026	−0.097***
	(0.027)	(0.030)
Children under 7	0.115***	0.117***
	(0.017)	(0.022)
High school	0.029	0.099***
	(0.045)	(0.036)
Higher education	0.028	0.149***
	(0.048)	(0.039)
Constant	1.940***	2.237***
	(0.492)	(0.454)
Tax evasion	1.182**	1.157**
Number of observations	2221	2041

Additional control variables: age squared and dummy variables for regions. Standard errors in parentheses

*$p<0.10$, **$p<0.05$, ***$p<0.01$

overall ETI estimate. This estimate is obtained by calculating the percentage change in income evaded due to the reform. Then, the "evasion elasticity" is derived by dividing this figure by the percentage change in the net-of-tax rate, when restricting to self-employed with higher net-of-tax rates (those assumed to react), and multiplying and dividing with the (calculated) tax evasion and income reported before the reform, respectively.

This suggests that the tax evasion response is smaller than the working hours response, but again, as the reduction in tax evasion is not statistically significant, one should be cautious in putting too much emphasis on the estimate of the tax evasion component. However, we find it believable that the tax evasion represents a positive component of the ETI, when the estimation of the ETI is obtained from a reduction in tax.

4.4 Implications of organizational shifts

Recall that, in contrast to responses in working hours and in terms of changes in tax evasion, we assert that organizational shifts represent a source of bias in the

Table 4 Self-employed in year t who have incorporated in year $t+3$, 2001–2010

T	Number	Percent of self-employed
2001	1784	2.06
2002	1845	2.06
2003	2416	2.83
2004	2913	3.17
2005	3257	3.43
2006	2415	2.61
2007	2162	2.34
2008	1993	2.24
2009	2469	2.82
2010	2198	2.55

estimation of the ETI.[33] To obtain information about the extent of organizational shifts over the reform period, information from the Business and Enterprise Register (*Virksomhet og foretaksregisteret*) (Hansson 2007), the Shareholder Register (*Aksjonærregisteret*) (Statistics Norway 2015) and the End of the Year Certificate Register (*Lønns- og trekkoppgaveregisteret*) (Aukrust et al. 2010) is used. By establishing a longitudinal data set, we can verify if the self-employed have moved their business activities from self-employment to an incorporated firm, and assess to what extent these movements have been altered by the reform, thereby representing a source of estimation bias.

This part of the analysis is constrained by information from the Shareholder Register only being available from 2004 and onward, which implies that 2004 is the first year with information about the owner/employment combination in incorporated firms. Table 4 presents the number of self-employed in year t who in year $t+3$ run an incorporated business.[34] The definition of self-employed in year t is the same as in the ETI estimation. The taxpayer is included among those who have shifted organizational form if, conditional on being assigned to self-employment in year t, she is observed in year $t+3$ with ownership of more than 50% of the shares in an incorporated business, in combination with higher wage income or shareholder income than self-employment income. See Tables 14 and 15 in the Appendix for more information about the "shifters".

As expected, the figures of Table 4 show that there are large yearly changes in the incorporation rates. As explained in Sect. 3.4, to illustrate to what extent the estimates of the ETI are biased, we obtain estimates of the ETI when accounting for weights, where the weights are derived from an estimation of the probability to leave self-employment; see Eq. (3.6). Results for both OLS and probit estimations

[33] See Papini (2018) for a more thorough analysis of effects of the tax reform of 2006 on organizational shifts.

[34] Three-year shifts allow us to estimate inverse probability weights for all years 2001–2010.

Table 5 Estimation results for organizational shift probability model. The probability to incorporate is the dependent variable

	OLS	Probit
Self-employment tax rates	0.108***	2.074***
	(0.005)	(0.084)
Self-employment tax rates in $t-1$	0.117***	2.101***
	(0.004)	(0.748)
Post- vs. pre-reform self-employment tax rates	0.059***	1.131***
	(0.007)	(0.123)
Capital income in $t-1$	7.76×10^{-10}	5.49×10^{-9}
	(5.61×10^{-10})	(4.35×10^{-9})
Wage income in $t-1$	4.05×10^{-8}***	3.83×10^{-7}***
	(4.00×10^{-9})	(5.04×10^{-8})
Number of observations	432,357	432,357

Other variables included in the regression, but not shown in the table are: age, year dummies, dummies for educational level and educational field, male, marriage, children and Norwegian born

*$p<0.10$, **$p<0.05$, ***$p<0.01$

are shown in Table 5. We see that higher income tax rate as self-employed is associated with a higher probability of incorporating. Further, we also observe there is a effect of taxation of the self-employed after the reform.

Next, we use the estimated probabilities from this model to reweight the sample by inverse probability weighting to estimate a "shift-robust" ETI. By correcting for the selection bias by inverse probability weighting, our estimates are informative about the tax response without shifting effects. The results of these estimations are presented in Table 6. We see that the estimated ETI is lower for all specifications when accounting for the weights reflecting probabilities for incorporation. Although the difference between the unweighted and the weighted estimates is small, the difference in the Weber specification is significant at the 10% level (according to a simple t test). As the selection model is simple, this represents corroborative evidence of the ETI being smaller when accounting for the change in shift incentives. We therefore conclude that the ETI estimates of Table 1 likely are biased upward due to organizational shift patterns. This is the most important finding of the present study.

5 Conclusion

The sufficient statistic interpretation of the ETI has received a lot of attention in applied public finance. A major attraction of the approach is that one does not need to address the behavioral anatomy of the ETI. However, in this paper we warn against neglecting the effects of various response dimensions, as some responses can create biases in the estimation of the ETI. Access to several data sets, mainly from Norwegian administrative registers, has been essential for this analysis probing

Table 6 Estimates of the ETI, unweighted and weighted by inverse probabilities of incorporation

	Unweighted (as in Table 1)			Weighted		
	Gruber/Saez mean reversion	Kopczuk mean reversion	Weber instr./Kopczuk mean reversion	Gruber/Saez mean reversion	Kopczuk mean reversion	Weber instr./ Kopczuk mean reversion
Net-of-tax rate	0.103***	0.169*	0.173***	0.096*	0.139***	0.124***
	(0.018)	(0.018)	(0.015)	(0.025)	(0.025)	(0.021)
Number of observations	405,152	345,441	335,906	406,375	347,196	337,329

See Sect. 3 for further details behind the different specifications. The weighted estimations account for the fact that the weights are estimated

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

deeper into the various effects underlying the overall ETI, providing empirical estimates of margins assumed to be standard components of the ETI and a dimension, organizational shift, which we expect to cause problems.

The ETI estimates for the self-employed obtained here are relatively small, in the range from 0.10 to 0.17, which is close to findings for Denmark, reported in Kleven and Schultz (2014), and considerably smaller than found for the USA by Heim (2010). Further, our estimates suggest that effects on working hours are the dominant response margin summarized by the ETI, but we also attribute some of the overall tax response to tax evasion, for the latter effect obtaining evidence in support of tax evasion increasing in the marginal tax rate.

However, the main message of the present study is that such estimates are in danger of being misleading if not controlling for confounding factors in the identification of the ETI. We see large changes in incentives for incorporation after the 2006-reform, and thereby in the composition of the self-employed, which we argue represents a source of upward bias in the ETI. When we derive weights for the probability to change organizational form and exploit these weights in the estimation of the ETI, we find lower ETI estimates. For example, according to one of the specifications, we see a reduction in the ETI estimate from 0.17 to 0.12 after the changed shifting patterns have been controlled for. Thus, this suggests a sizable bias in the naïve ETI estimation.

Finally, we assert that more investigations of the multiple behavioral components of the ETI benefit the understanding of it, both in a national and an international context. We have illustrated that some behavioral margins are more problematic than others, which suggests that one should carefully investigate which responses are involved. Such examinations are demanding with respect to data, but with increased access to larger and richer data sources in the future, we expect to see more studies addressing problematic response margins.

Acknowledgements This work is part of the research carried out at Oslo Fiscal Studies, which is supported by the Research Council of Norway. The authors thank Odd Erik Nygård for providing the consumption data. We have benefited from comments by Wiji Arulampalam, Spencer Bastani, Evelina Gavrilova, Laurence Jacquet, Katharina Jenderny, Andrea Papini, Terje Skjerpen, Stefan Tübbicke, Trine Vattø, and seminar participants of the ETI workshop in Mannheim (August 2015), Skatteforum 2015 (Bergen, September 2015), the Norwegian-German Seminar (Munich, November 2015), the OFS workshop "Self-Employment/Entrepreneurship and Public Policy" (Oslo, September 2016), and the 73rd Annual Congress of the International Institute of Public Finance (Tokyo, August 2017).

Funding The funding was provided by Norges Forskningsråd (NO).

Appendix: Summary statistics

Income data

See Tables 7 and 8.

Table 7 Average income and net-of-tax rate, 2001–2010

Year	Reported income	Net-of-tax rate	Self-employed individuals
	Mean	Mean	Number
2001	322,751	0.573	59,491
	(321,121)	(0.074)	
2002	340,197	0.580	60,415
	(301,468)	(0.074)	
2003	342,212	0.586	59,996
	(304,839)	(0.073)	
2004	372,889	0.582	61,611
	(326,168)	(0.074)	
2005	396,224	0.592	61,965
	(409,372)	(0.064)	
2006	430,018	0.600	63,053
	(436,778)	(0.050)	
2007	467,953	0.594	64,038
	(438,840)	(0.052)	
2008	477,578	0.595	63,582
	(405,790)	(0.052)	
2009	477,194	0.598	63,547
	(412,038)	(0.052)	
2010	493,658	0.598	63,528
	(445,654)	(0.051)	

Standard deviations in parentheses

Table 8 Descriptive statistics for control variables in the estimation of the ETI

Characteristic	Mean	Length of education	Percentage in sample	Educational field	Percentage in sample
Male	0.76	No education	0.1	General	32.9
Age	46.3	Primary school	0.1	Human. and arts	4.3
Children	0.59	Secondary school	19.3	Teaching	2.1
Married	0.58	High school, started	25.1	Social science and law	3.6
Norwegian born	0.93	High school, completed	28.5	Business and administration	9.0
		High school, supplement	2.3	Science, crafts and technology	24.4
		University, undergrad	13.0	Health, social and sports	11.3
		University postgrad	10.0	Agriculture and fishery	5.8
		Research degree	0.3	Transp., security and services	5.6
		Unknown	1.3	Unknown	1.8
Number of obs.			578,884		

Working hours data

See Tables 9, 10 and 11.

Table 9 Hours of work and net-of-tax rate in groups, before and after the reform

	Treated		Small control group		Large control group	
	Hours of work	Net-of-tax rate	Hours of work	Net-of-tax rate	Hours of work	Net-of-tax rate
Before reform, 2002–2004	41.3	0.517	42.2	0.618	36.2	0.625
	(9.1)	(0.055)	(9.6)	(0.053)	(6.3)	(0.047)
After reform, 2007–2009	40.7	0.558	41.0	0.615	35.7	0.624
	(9.1)	(0.043)	(9.7)	(0.044)	(6.6)	(0.038)
Number of obs.	2462	1360	45,466			

Standard deviations in parentheses

Table 10 Descriptive statistics for control variables involved in the working hours estimation, 2002–2009

	Treated Mean	Small control group Mean	Large control group Mean
Male	0.80	0.78	0.49
Age	48.0	47.3	40.9
Child	0.63	0.61	0.57
Married	0.61	0.62	0.50
Norwegian born	0.93	0.95	0.94
Length of education	**Treated Percentage**	**Small control group Percentage**	**Large control group Percentage**
No education	0.0	0.0	0.1
Primary school	0.1	0.0	0.1
Secondary school	15.0	20.3	14.3
High school, started	16.2	30.9	18.5
High school, completed	24.3	33.3	34.0
High school, supplement	2.0	2.6	3.4
University, undergrad	18.0	9.1	28.8
University, postgrad	22.0	3.5	3.7
Research degree	0.7	0.0	0.2
Unknown	1.7	0.5	0.9
Educational field	**Treated Percentage**	**Small control group Percentage**	**Large control group Percentage**
General	22.6	34.8	24.5
Humanities and arts	4.2	3.8	4.4
Teaching	2.2	2.3	7.5
Social science and law	7.4	1.2	2.2
Business and administration	8.7	11.2	
Science, crafts and technology	21.6	30.0	26.1
Health, social and sports	24.9	5.6	15.6
Agriculture and fishery	3.0	4.7	1.1
Transport, security and services	3.5	5.6	3.1
Unknown	2.0	0.9	1.5
Number of observations	2462	1360	45,466

Table 11 Placebo test, using comparison of 2001 and 2002 versus 2003 and 2004 in the identification of the response in working hours

	(1) Log	(2) Log, large control group
Tax treatment	0.065	−0.074
	(0.318)	(0.224)
Treated	0.016	0.074*
	(0.052)	(0.009)
Constant	3.467***	3.626***
	(0.204)	(0.105)
Number of obs.	1386	23,607

Additional control variables: age, age squared, dummy variables for male, children, married, Norwegian born, dummies, educational field, length of education, county and years. Robust standard errors in parentheses

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Expenditure data

See Tables 12 and 13.

Table 12 Average income and food consumption, self-employed and wage earners, 2003–2007

	Self-employed		Wage earners		Self-employed individuals
	Income	Food consumption	Income	Food consumption	
	Mean	Mean	Mean	Mean	Number
2003	472,001	49,013	454,463	44,432	99
	(239,176)	(25,259)	(250,193)	(23,896)	
2004	494,889	51,956	484,997	43,204	95
	(220,883)	(24,270)	(779,194)	(22,739)	
2005	680,560	52,252	508,431	46,586	77
	(967,065)	(28,047)	(478,743)	(25,987)	
2006	542,039	57,406	503,499	47,970	83
	(270,838)	(32,358)	(266,367)	(28,095)	
2007	653,805	60,977	550,958	51,493	90
	(440,945)	(41,758)	(285,587)	(30,057)	

Standard deviations in parentheses

Organizational shifts

See Tables 14 and 15.

Table 13 Descriptive statistics for control variables used in the estimation of the tax evasion equation

	Self-employed Mean	Wage earners Mean
Male	0.75	0.71
Age	46.8	46.1
Number of children under 7	0.35	0.38
High school	0.52	0.49
Higher education	0.30	0.36
Geographical area: South	0.13	0.14
West	0.19	0.17
East	0.30	0.29
North	0.10	0.13
Center	0.13	0.11
Number of observations	444	4896

Table 14 Average income and average net-of-tax rate for business owners who shift organizational form, 2001–2010

	Reported income Mean	Net-of-tax rates Mean	Number of observations
2001	502,385	0.535	1784
	(485,805)	(0.071)	
2002	481,508	0.545	1845
	(324,073)	(0.075)	
2003	571,487	0.5742	2416
	(434,606)	(0.076)	
2004	596,808	0.537	2913
	(442,140)	(0.075)	
2005	621,724	0.555	3257
	(504,011)	(0.065)	
2006	681,428	0.571	2415
	(572,034)	(0.049)	
2007	738,690	0.563	2162
	(719,009)	(0.050)	
2008	765,311	0.562	1993
	(588,156)	(0.048)	
2009	737.201	0.569	2469
	(609,355)	(0.051)	
2010	755,972	0.565	2198
	(522,541)	(0.049)	

Standard deviations in parentheses

Table 15 Summary statistics for business owners who shift organizational form

	Mean
Male	0.85
Age	44.4
Children	0.60
Married	0.53
Norwegian born	0.92
Length of education	**Percentage**
No education	0.1
Primary school	0.1
Secondary school	13.1
High school, started	14.9
High school, completed	35.5
High school, supplement	3.7
University, undergrad	16.6
University, postgrad	14.4
Research degree	0.6
Unknown	1.2
Educational field	**Percentage**
General	20.5
Humanities and arts	3.7
Teaching	2.1
Social science and law	3.5
Business and administration	13.7
Science, crafts and technology	35.3
Health, social and sports	11.2
Agriculture and fishery	3.3
Transport, security and services	5.2
Unknown	1.7
Number of observations	23,452

References

Aarbu, K. O., & Thoresen, T. O. (2001). Income responses to tax changes—Evidence from the Norwegian tax reform. *National Tax Journal, 54*(2), 319–338.

Allingham, M., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics, 1,* 323–338.

Aukrust, I., Aurdal, P. S., Bråthen, M., & Køber, T. (2010). Registerbasert sysselsettingsstatistikk. *Dokumentasjon*, *Notater* 8/2010, Statistics Norway.

Auten, G., & Carroll, R. (1999). The effect of income taxes on household income. *Review of Economics and Statistics, 81*(4), 681–693.

Bastani, S., & Selin, H. (2014). Bunching and non-bunching at kink points of the swedish tax schedule. *Journal of Public Economics, 109,* 36–49.

Blomquist, S., & Selin, H. (2010). Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates. *Journal of Public Economics, 94,* 878–889.

Blow, L., & Preston, I. (2002). *Deadweight loss and taxation of earned income: Evidence from tax records of the UK self-employed.* Institute for fiscal studies working paper WP 02/15.

Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *American Economic Journal: Economic Policy, 1*(2), 31–52.

Chetty, R. (2012). Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply. *Econometrica, 80*(3), 969–1018.

Christansen, V. (2004). Norwegian income tax reforms. *CESifo DICE Report, 3*(2004), 9–14.

Christansen, V., & Tuomala, M. (2008). On taxing capital income with income shifting. *International Tax and Public Finance, 15*(4), 527–545.

Clotfelter, C. T. (1983). Tax evasion and tax rates: An analysis of individual returns. *Review of Economics and Statistics, 65*(3), 363–373.

Doerrenberg, P., & Duncan, D. (2014). Experimental evidence on the relationship between tax evasion opportunities and labor supply. *European Economic Review, 68*(May), 48–70.

Doerrenberg, P., Peichl, A., & Siegloch, S. (2017). The elasticity of taxable income in the presence of deduction possibilities. *Journal of Public Economics, 151,* 41–55.

Edmark, K., & Gordon, R. H. (2013). The choice of organizational form by closely-held firms in Sweden: Tax versus non-tax determinants. *Industrial and Corporate Change, 22*(1), 219–243.

Engström, P., & Holmlund, B. (2009). Tax evasion and self-employment in a high-tax country: Evidence from Sweden. *Applied Economics, 41*(19), 2419–2430.

Fack, G., & Landais, C. (2016). The effect of tax enforcement on tax elasticities: Evidence from charitable contributions in France. *Journal of Public Economics, 133,* 23–40.

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

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics, 81*(4), 674–680.

Freire-Serén, M. J., & Panadés, J. (2013). Do higher tax rates encourage/discourage tax compliance? *Modern Economy, 4,* 809–817.

Giertz, S. H. (2009). The elasticity of taxable income: Influences on economic efficiency and tax revenues, and implications for tax policy". In A. D. Viard (Ed.), *Tax policy lessons from the 2000s* (pp. 101–136). Washington: AEI Press.

Goolsbee, A. (1999). Evidence on the high-income Laffer curve from six decades of tax reforms. *Brookings Papers on Economic Activity, 1999,* 1–47.

Goolsbee, A. (2000). What happens when you tax the rich? Evidence from executive compensation. *Journal of Political Economy, 108*(2), 352–378.

Gordon, R. H., & Slemrod, J. B. (2000). Are 'Real' responses to taxes simply income shifting between corporate and personal tax bases? In J. Slemrod (Ed.), *Does Atlas shrug? The economic consequences of taxing the rich* (pp. 240–281). Cambridge: Harvard University Press.

Gorodnichenko, Y., Martinez-Vazquez, J., & Peter, K. S. (2009). Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia. *Journal of Political Economy, 117*(3), 504–554.

Gruber, J., & Saez, E. (2002). The elasticity of taxable income: Evidence and implications. *Journal of Public Economics, 84*(1), 1–32.

Hansson, A.-K. (2007). Bedrifts- og foretaksregisteret. Regler og rutiner for ajourhold av Bof. Notater 2/2007, Statistics Norway (**in Norwegian**).

Harju, J., & Matikka, T. (2014). The elasticity of taxable income and income-shifting: What is "Real" and what is not? *International Tax and Public Finance, 23*(4), 640–669.

Heim, B. T. (2010). The responsiveness of self-employment income to tax rate changes. *Labour Economics, 17*(6), 940–950.

Holmøy, A., & Lillegård, M. (2014). Forbruksundersøkelsen 2012. Dokumentasjonsrapport. Notater 2014/17. Statistics Norway (**in Norwegian**).

Joulfaian, D., & Rider, M. (1998). Differential taxation and tax evasion by small business. *National Tax Journal, 51*(4), 675–687.

Kleven, H. J. (2014). How can scandinavians tax so much? *Journal of Economic Perspectives, 28*(4), 77–98.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica, 79*(3), 651–692.

Kleven, H. J., & Schultz, E. A. (2014). Estimating taxable income responses using Danish Tax Reforms. *American Economic Journal: Economic Policy, 6*(4), 271–301.

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics, 89,* 2093–2119.
Kuka, E. (2014). EITC and the self-employed: Real effects or reporting effects? *Public Finance Review, 42*(6), 691–719.
le Maire, D., & Scherning, B. (2013). Tax bunching, income shifting and self-employment. *Journal of Public Economics, 107,* 1–18.
Lindsey, L. B. (1987). Individual taxpayer responses to tax cuts: 1982–1984—With implications for the revenue maximizing tax rate. *Journal of Public Economics, 33*(2), 173–206.
Matikka, T. (2018). Elasticity of taxable income. Evidence from changes in municipal income tax rates in Finland. *The Scandinavian Journal of Economics, 120*(3), 271–301.
Ministry of Finance. (2003). Skatteutvalget. Norges offentlige utredninger (**in Norwegian**).
Moffitt, R. A., & Wilhelm, M. O. (2000). Taxation and labor supply decisions of the affluent. In J. Slemrod (Ed.), *Does atlas shrug? The Economic Consequences of taxing the rich* (pp. 93–234). New York: Russell Sage Foundation.
Nygård, O., Slemrod, J., & Thoresen, T. O. (2019). Distributional implication of joint tax evasion. *The Economic Journal, 129*(620), 1894–1923.
OECD. (2019). *Self-employment rate (indicator)*. https://www.oecd-ilibrary.org/employment/self-employment-rate/indicator/english_fb58715e-en.
Papini, A. (2018). *Tax incentives and choice of organisational form of small businesses. Identification through a differentiated payroll tax schedule*. Oslo: University of Oslo.
Parker, S. C. (2009). *The economics of entrepreneurship*. Cambridge: Cambridge University Press.
Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics, 39,* 17–32.
Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies, 68*(1), 205–229.
Saez, E. (2004). Reported incomes and marginal tax rates, 1960–2000: Evidence and policy implications. In J. M. Poterba (Ed.), *Tax policy and the economy* (Vol. 18, pp. 117–174). Cambridge: MIT Press.
Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy, 2*(3), 180–212.
Saez, E., Slemrod, J., & Giertz, S. H. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature, 50*(1), 3–50.
Slemrod, J. (1995). Income creation or income shifting? Behavioral responses to the Tax Reform Act of 1986. *American Economic Review: Papers & Proceedings, 85*(2), 175–180.
Slemrod, J. (1996). High-income families and the tax changes of the 1980s: The anatomy of behavioral response. In M. Feldstein & J. Poterba (Eds.), *Empirical foundations of household taxation* (pp. 169–192). Chicago: University of Chicago Press.
Slemrod, J. (1998). Methodological issues in the measuring and interpreting taxable income elasticities. *National Tax Journal, 51*(4), 773–788.
Slemrod, J., & Gillitzer, C. (2014). *Tax systems*. Cambridge: MIT Press.
Slemrod, J., & Kopczuk, W. (2002). The optimal elasticity of taxable income. *Journal of Public Economics, 84*(1), 91–112.
Sørensen, P. B. (2005). Neutral taxation and shareholder income. *International Tax and Public Finance, 12*(6), 777–801.
Statistics Norway. (2003). Labour force survey 2001. *Official Statistics of Norway.*
Statistics Norway. (2005). Income statistics for persons and families 2002–2003. *Official Statistics of Norway.*
Statistics Norway. (2014). http://www.ssb.no/en/inntekt-ogforbruk/statistikker/i_us/aar/2014-12-17?fane=om#content.
Statistics Norway. (2015). http://www.ssb.no/en/virksomheter-foretak-og-regnskap/statistikker/aksjer/aar-forelopige/2015-06-26.
Thoresen, T. O., & Alstadsæter, A. (2010). Shifts in organizational form under a dual income tax system. *FinanzArchiv: Public Finance Analysis, 66*(4), 384–418.
Thoresen, T. O., & Vattø, T. E. (2015). Validation of the discrete choice labor supply model by methods of the new tax responsiveness literature. *Labour Economics, 37,* 38–53.
Weber, C. (2014). Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. *Journal of Public Economics, 117,* 90–103.
Wooldridge, J. M. (2002). Inverse probability weighted m-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal, 1*(2), 117–139.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge: MIT Press.
Wu, S.-Y. (2005). The effect on taxable income from privately held businesses. *Southern Economic Journal, 71*(4), 891–912.
Yitzhaki, S. (1974). A note on income tax evasion: A theoretical analysis. *Journal of Public Economics, 3,* 201–202.