DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

PhD Thesis

Daniel Gerszon Mahler

Measuring What Matters?

Empirical Essays in Welfare Economics

Supervisor: David Dreyer Lassen

Date of submission: March 20, 2018

“[E]conomics is not ethics, though it borders on ethics; the line between them is a place where it behoves us to walk very delicately”

J.R. Hicks (1975)
The Scope and Status of Welfare Economics.

English Summary

A core purpose of economics is to provide knowledge about how to design economic policies. Any discussion about the appropriate design of economic policies hinges on some idea about what matters to society. Frequently, economic policies are judged by the extent to which they generate welfare, where welfare is understood as the satisfaction of preferences. Yet policymakers and ordinary citizens often have other ideas about what welfare is and about the appropriate objective of economic policies, invoking concepts such as freedom, opportunities, and justice. This dissertation contains six chapters, which cast light on whether applying some of these other ideas matters for the design of economic policies.

Chapter 1 introduces the dissertation. Chapter 2 analyzes if prospect theory, a theory known to apply to preference-based measures of welfare, also applies to happiness-based measures of welfare. Chapter 3 measures whether individuals have equal opportunities for obtaining welfare, and assesses whether this depends on how welfare is measured. Both chapters conclude that how welfare is measured matters little for policy design. Chapter 4 improves measures of equality of opportunity by using conditional inference regression trees.

The last two chapters deal with cases where there is a wedge between preference satisfaction and welfare. Chapter 5 analyzes attitudes towards paternalism in areas where some individuals lack information or willpower to act in their own best interest. Chapter 6 assesses voting outcomes when some individuals are altruistic in the sense that they do not vote for the outcome that maximizes their welfare. In both cases, the results indicate that using preferences may yield misleading policy evaluations.

The combined findings of this thesis suggest that using alternative measures of welfare hardly matters as long as internalities, externalities, and public goods are absent.

Danish Summary

En grundlæggende opgave for den økonomiske videnskab er at undersøge, hvordan man bør udforme økonomisk politik. En sådan undersøgelse kræver en idé om, hvad der er værdifuldt for samfundet. Økonomiske politikker bedømmes oftest på, hvorvidt de formår at skabe velfærd, hvor velfærd forstås som tilfredsstillelsen af præferencer. Politiske beslutningstagere og borgere har dog ofte andre opfattelser af, hvad velfærd er og, hvad der er værdifuldt for samfundet. Eksempelvis betones ofte værdier som frihed, muligheder og retfærdighed. Denne afhandlings seks kapitler belyser, hvorvidt anvendelsen af nogle af disse andre værdier har betydning for udformningen af økonomisk politik.

Kapitel 1 indleder afhandlingen. Kapitel 2 analyserer hvorvidt prospect theory – en teori, der anvendes til præferencebaserede velfærdsmål – tilmed holder med lykkebaserede velfærdsmål. I kapitel 3 måles, hvorvidt borgere har lige muligheder for at opnå velfærd, og om dette afhænger af, hvordan velfærd måles. Konklusionen i begge kapitler er, at måden, hvorpå man måler velfærd, har begrænset indflydelse på udformningen af økonomisk politik. Kapitel 4 forsøger at forbedre målingen af lige muligheder ved at drage nytte af conditional inference regression trees.

De sidste to kapitler behandler tilfælde, hvor der er et modsætningsforhold mellem præferencer og velfærd. Kapitel 5 behandler holdninger til paternalisme i situationer, hvor personer ikke besidder tilstrækkelig information eller viljestyrke til at handle i deres egen interesse. I kapitel 6 analyseres valgresultater, når nogle stemmer altruistisk, forstået som, at de ikke forsøger at maksimere deres velfærd. Konklusionen på begge kapitler er, at præferencebaserede velfærdsmål kan resultere i misvisende evalueringer af økonomisk politik.

Samlet set peger afhandlingen på, at det har lille betydning, hvorvidt der anvendes alternative velfærdsmål, så længe eksternaliteter, internaliteter og offentlige goder ikke er tilstede.

Acknowledgments

In the process of writing this dissertation, I have been fortunate to receive guidance and support from many people. I wish to extend some words of gratitude. Let me first thank my advisor, David, for guiding me through this whole endeavor. David has given me indispensable feedback on all parts of the process. Aside from his academic expertise, David has always kept me motivated, upbeat, and certain about the next steps to take.

I next want to thank the many others who have been kind to give me advice on several occasions. These include, but are not limited to, Ryan Enos, Francisco Ferreira, Christoph Lakner, Xavier Landes, Marco Piovesan, and Xavier Ramos. I also want to thank my fellow PhD students for many stimulating discussions during equally many walks around the lake and Friday afternoon P8 sessions. Thanks are also due to the institutions that hosted me during the PhD; the Institute for Quantitative Social Science at Harvard University, the Interdisciplinary Center for Economic Science at George Mason University, and the Poverty & Inequality Group at the World Bank. I have learned tremendously from my colleagues in all of these institutions.

Next, I want to give thanks to the co-authors I have had; Paolo Brunori, Felix Sebastian Døssing, Paul Hufe, Ohto Kanninen, David Dreyer Lassen, and Xavier Ramos. I am very grateful for the patience they have shown me, their cooperation, and mostly for many, many inspiring discussions. In this connection, I also want to offer my gratitude to Eugenio Peluso and Claudio Zoli, the organizers of the annual Winter School on Inequality and Social Welfare Theory in Canazei. I had the pleasure of participating in the winter school several times during my PhD, and it introduced me to many of my ensuing co-authors.

Penultimately, I want to thank two people who unknowingly guided me to pursue a PhD in this very topic. First, a family member of mine, Kirsten, who was one of the many people I approached to seek guidance on whether applying for a PhD was an appropriate choice for me. Her advice made it an easy choice: *"Daniel, I have known you should do a PhD since you were three years old"* [there goes my agency]. Second, my Japanese friend, Akifumi, whom I – upon having decided to do a PhD – inquired about whether researching the intersection between ethics and economics was a sound topic. The topic is unconventional, I argued, and may not be ideal for publications or future employability. *"Maybe so,"* he answered, *"but it is good for man."*

Finally, I want to thank my family, Fanny, Carsten, and Nathalie, and my beshert, Doreen, for their unconditional support. Without all of you, I would not be very far.

Copenhagen, January 2018

Contents

Chapter 1

Introduction

1.1 A Plurality of Moral Ends

One of the primary purposes of economics is to provide knowledge about how to design economic policies. Any discussion about the appropriate design of economic policies hinges on some idea about what society ought to pursue – some idea about what matters to society. If you open an introductory textbook in economics, chances are you will find a very specific idea about what matters when evaluating economic policies. Policies are judged by the extent to which they produce welfare, where welfare is understood as the satisfaction of revealed preferences. The underlying tenet behind such revealed preferences is that if an agent can choose between bundle A and B and chooses bundle A, then the agent must prefer A over B, and must be better off with A over B.

If you ask philosophers, politicians, and ordinary people alike, many make other judgments about what societies ought to pursue than the satisfaction of revealed preferences. This matters greatly for the economist's role as a policy advisor. If a policymaker has a different view on what welfare is than the economist, then the economist's assessment of the welfare effects of a reform will be of limited value. Likewise, if a policymaker cares for more than the distribution of welfare levels in society, economists ought to be able to speak to how reforms impact these other objectives. As Hausman and McPherson (1993) put it: *"Notions of fairness, opportunity, freedom, and rights are arguably of more importance in policy making than concerns about moving individuals up their given preference rankings. Insofar as economists want to assist in the formulation of policy towards such ends, they must link economic theory with such concerns."*

Once we venture beyond revealed preferences, a plurality of other possible ways to evaluate economic policies emerges. A plurality of different ways of conceptualizing what really matters. This dissertation is concerned with finding better ways to measure some of these other objectives, as well as exploring the degree to which going beyond revealed preferences matters for the design of economic policies.

One might argue that, although revealed preferences are central to introductory textbooks in economics, economic policy suggestions are not frequently based upon revealed preferences. I would argue otherwise. Revealed preferences lie behind policy suggestions coming from game theory, experimental economics, mechanism design, structural models, and micro-founded macroeconomics – to give a few examples. Revealed preferences also lie behind notions of efficiency, the literature on optimal taxation, and policies based on cost-benefit analysis. Moreover, it can be argued that the use of the GDP as a proxy for societal welfare and the dominant approach to poverty measurement have justifications in revealed preferences (see Fleurbaey (2009) and Ravallion (2015) for discussions on this). That is, policies that try to spur growth or reduce poverty can be seen as having roots in revealed preferences. Explicitly or implicitly, revealed preferences lie behind many economic policy recommendations.

Why are revealed preferences so popular? First and foremostly because they are considered to be value free, that is, free of moral judgments (Boulding, 1969). Individuals themselves decide what to pursue in life and all there is left for the economist is to clarify how these desires can be fulfilled. However, one can easily argue against this conviction. In particular, revealed preferences can be seen to rely on at least three moral judgments: 1) that the maximization of welfare is the only objective of policies, 2) that the satisfaction of preferences constitutes welfare, and 3) that preferences can be revealed through choices.

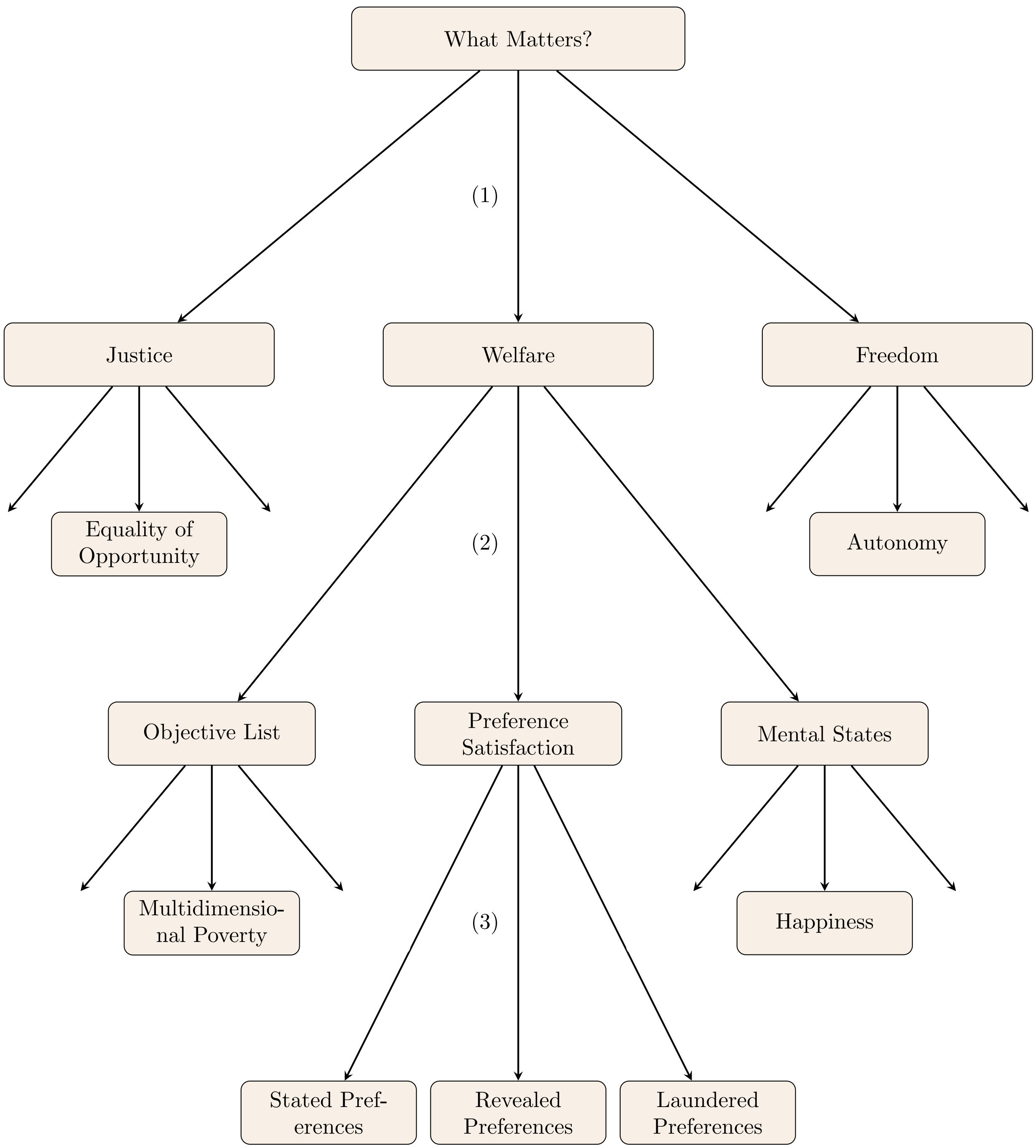
Each of these three assumptions can and have been contended. In some sense, whereas the use of preferences is motivated on the grounds that individuals have different views over what bundles of goods give them well-being, we often fail to acknowledge that people also have different judgments about what ought to be the objective of government policies (see Benjamin *et al.* (2014), Cappelen *et al.* (2007) & Weinzierl (2014) for some notable exceptions to this claim). Figure 1.1 tries to provide a simple overview of some of these other objectives. The three moves from the top of Figure 1.1 – from *what matters* to *revealed preferences* – reflect the three moral judgments I claim are needed for the satisfaction of revealed preferences to be the rightful objective of economic policies.[1]

Let us start from the bottom of Figure 1.1, and assume that welfare/well-being is all that matters (I use the two interchangeably), and that preference satisfaction is a good manifestation of welfare. Even within this framework, it need not be the case that preferences can or ought to be revealed through choice. There are at least two reasons why revealed preferences may be ill-advised. Firstly, individuals may suffer from behavioral biases, such as a lack of willpower, which prohibit them from choosing what is in their best interest. If individuals are aware of their behavioral biases, one could instead use stated preferences as a measure of welfare. *Stated preferences* assume that individuals' stated choices reflect what is in their best interest. Suppose for example, that I due to a lack of willpower choose to eat pizza rather than going to the gym, knowing that I would be better off if I went to the gym. If asked, I might understand my lack of willpower and state that going to the gym would be better for me. In this case, stated preferences solve the first issue with revealed preferences.

Chances are that individuals' stated preferences are neither fully rational nor fully informed. If I mistakenly think that eating pizza is better for me than going to the gym, then

[1] The figure is not intended as a comprehensive scheme over what matters in life. Many important things that matter are not included. The figure is intended as a simple overview of the moral ends this dissertation deals with. As such, the figure is a useful tool for tying the different chapters together. The only claim I make is that the ends reported in Figure 1.1 may be important for purposes of economic policymaking.

Figure 1.1: A Plurality of Moral Ends



stated preferences will not solve the problem. Instead, we could try to "clean" the preferences for all behavioral shortcomings. This is the purpose of *laundered preferences*, which are – loosely speaking – the preferences individuals would have if fully rational and fully informed (Goodin, 1986). Laundered preferences do not deal with the second reason why revealed preferences may be ill-advised, which is that people may have preferences for things that do not involve their welfare. A silly example can cast light on this issue. If asked whether I prefer that people in year 2150 are starving or are not starving, I clearly prefer the latter. However, under no account will I be better off if/when this preference of mine is realized. By year 2150, I will be long gone.

There are many other notions of welfare we could adopt instead that can handle this issue. In the philosophical literature, defining welfare as the satisfaction of preferences is one among three main alternatives (Parfit, 1984; Griffin, 1986). Our normal identification of welfare with the satisfaction of preferences *"is so automatic and ubiquitous that [we] seldom realize how controversial it is"* (Hausman and McPherson, 1997). Going back to Figure 1.1, this takes us to the second level of assumptions – the arrow from *welfare* to *preference satisfaction*. To deal with the problem that people can have preferences over aspects that do not involve their own welfare, we could instead resort to mental state theory. *Mental state theory* claims that what matters is not what individuals choose or what they prefer, but whether individuals are happy or feel satisfied with their lives – the quality of their mental states. I will not be happier when my preference regarding hunger in 2150 is realized, so this notion of welfare addresses the second concern of using preferences. This approach has spurred a large literature in happiness economics, which seeks to determine how to design economic policies to boost societal happiness (see for example Di Tella *et al.* (2001) and Campante and Yanagizawa-Drott (2015)). We could also apply *objective list theory*. This theory claims that what matters is whether individuals possess items on a list which are intrinsically valuable, such as health, education etc. This notion of welfare arguably comes closest to what policymakers have in mind. This approach, too, has been richly applied, particularly in the area of multidimensional poverty (see for example Alkire and Foster (2011)). None of these conceptions of welfare are unproblematic, but they show that our normal conception of welfare does not escape judgments about what is valuable in life.

Even if we agree on what welfare is and how to measure it, we still cannot escape value judgments. Some would argue that there is more to life than just welfare. As Atkinson (2011) puts it, *"One may be concerned with personal liberty and another with social justice. Where there are multiple [...] criteria, it makes no sense to talk about the welfare consequences, instead we have to apply multiple criteria and consider how conflicts can be resolved."*[2] In

[2]Sometimes welfare is used as a synonym to what matters. This appears to be the definition Atkinson

Figure 1.1, this takes us to the first level of assumptions – the arrow from *what matters* to *welfare*. Even if we observe all individuals' level of well-being, we may still not possess the necessary information to determine policy optimality. Conditional on the welfare levels of a society, we might care about individuals' ability to choose their own pathway in life; their agency, autonomy, capabilities, *freedom* and opportunities. To illustrate the point, consider the well-being (utilities, if you will) of two individuals, i and j, in two different policy worlds. Let the well-being numbers be interpersonally cardinally comparable.

	Status quo	New policy
i	10	7
j	5	7

In a standard textbook account, we could find the optimal policy by adding the utilities, possibly while giving a larger weight to the individual with a low utility. The argument I am making here is that the process matters. It matters whether j has 5 in the status quo because of lack of freedom, say, racial discrimination, or because of laziness. It has been known for a long time that there may be a trade-off between individuals' freedom and their level of well-being (see for example Sen (1970) and Kaplow and Shavell (2001)). This acknowledgment has spurred rich literatures, for example on capability theory. It has also spurred a literature trying to determine the value individuals place on freedom, such that it can be traded off with well-being (see for example Bartling *et al.* (2014)).[3]

We might also care not just about individuals' welfare or their freedom, but the joint distribution of these, which I – with the cost of severely simplifying centuries of debate on the topic – broadly will call *justice*. Some theories of justice, such as utilitarianism and strict egalitarianism, are easily compatible with only observing individuals' welfare levels. All we need to do is maximize the sum of welfare or equalize the levels of welfare, respectively. Other theories of justice, for example, the notion that individuals ought to have equal opportunities

(2011) has in mind, since the omitted part of the quote above – [...] – is "welfare." I use what I consider to be the concept of well-being applied in philosophy, which considers well-being "to describe what is non-instrumentally or ultimately good *for* a person" (Crisp, 2017, emphasis in original). Individuals may be concerned with, say, social justice independent of whether it is good for them or others.

[3]Some operationalizations of welfare make freedom one of the aspects of welfare; it may be one of the items on the objective list, freedom might give people happiness, and individuals may have a preference for freedom. Although certainly likely, one could claim that these would all be instrumental reasons for caring about freedom. We would care about freedom because it gives us well-being. I am sympathetic to this view, and if adopted, my claim here is that freedom is frequently neglected as a dimension of well-being, and that policy implications may differ substantially if the value of freedom is included. The argument for keeping freedom separate from welfare is that freedom gets an intrinsic value independent of welfare levels. This makes it easy to accommodate frequent policy ends, such as the notion that we ought to minimize state intervention in individuals' liberties.

in life, are non-welfarist. They are non-welfarist in the sense that we need to know more than individuals' level of well-being – particularly their opportunities and effort – before being able to select optimal policies. A vibrant literature building on the works of Roemer (1993) and Fleurbaey (1994), amongst others, has tried to accommodate this moral view in economic models.

In sum, there are many alternative moral objectives than the satisfaction of revealed preferences. Opening this box of alternative objectives of economic policies may appear to prohibit looking at economic policies from a value-free perspective. If we do not want researchers simply to select their own conception of what matters and blindly suggest policy proposals towards this end, then this is clearly problematic. The mere recognition of multiple moral ends, however, does not have to be in conflict with value neutrality. We can presuppose a normative value, and track how to design policies given our presumption. As Samuelson (1948) puts it, *"It is a legitimate exercise of economic analysis to examine the consequences of various value judgments."* In other words, one can easily analyze normative problems without taking a normative standpoint. If we want economics to be able to speak to the plurality of moral ends shared by policymakers and ordinary citizens, then examining the consequences of various value judgments is vital. This is precisely what this dissertation aims to do.

1.2 Tying the Pieces

Following this introductory chapter, this dissertation contains five self-contained chapters that aim to contribute to a richer understanding about the relationship between moral pluralism and economics. In particular, the chapters address three broad questions:

- Does it matter empirically which moral objective we target? (Chapter 2, 3, 5 & 6)
- Can the measurement of some of these objectives be improved? (Chapter 3 & 4)
- Are there empirical reasons to avoid using revealed preferences? (Chapter 5 & 6)

We can get at the first of the three questions I pose, whether it matters empirically which objective we target, by testing if a theory we know holds with one moral objective also holds with other moral objectives. This is the strategy of **Chapter 2**, which is written with Ohto Kanninen. In this chapter, we look at whether elements of prospect theory, which have been shown to apply when revealed preferences are used as the measure of welfare, also apply to mental state conceptions of welfare. The importance of prospect theory in economics is non-trivial. The original paper by Kahneman and Tversky (1979) has been cited nearly 50.000 times, it was one of the kick-starters of behavioral economics, and it was one of the reasons for Kahneman receiving the Nobel Prize in 2002. The theory confronts expected utility theory by suggesting that individuals receive value from their consumption bundle relative to a reference point, rather than based on diminishing marginal utility. The theory also postulates that individuals value losses from the reference point greater than gains, and that positive deviations from the reference point follow a concave pattern, while negative deviations follow a convex pattern. We test whether these elements of the theory apply with self-reported life satisfaction as the notion of welfare using 250,000 observations from the German Socio-Economic Panel. In line with experimental evidence on revealed preferences, we find evidence for all of these elements of prospect theory as well as for expected utility theory. Hence, this chapter suggests that the same theories – and consequently the same policy applications arising from these theories – apply to different notions of welfare.

Another way to test if it matters which measure of welfare we employ is to see if the different measures are highly correlated. **Chapter 3**, which is written jointly with Xavier Ramos, takes this approach. We investigate if estimates of equality of opportunity depend on which way we measure welfare. Equality of opportunity is a popular policy objective among politicians and citizens alike. However, it is not entirely clear *what* we ought to equalize opportunities for. Most existing attempts measure the extent to which there are unequal opportunities for income acquisition. From a philosophical point of view, an argument can be made that we ought to care about whether individuals have equal opportunities for welfare (Cohen, 1990). We take this approach and measure well-being in three different ways, corresponding to the three notions of well-being presented in Figure 1.1, as well as using income as a proxy for well-being. For each measure, we calculate inequality of opportunity and characterize the opportunity-deprived.

To measure individuals' opportunities, we first define a set of circumstance variables. These are variables individuals cannot be held accountable for, such as their gender, their parents' education, and their birth region. Next, we predict each individual's well-being from the circumstance variables and assign individuals "opportunity ranks" from these predictions. Individuals from fortunate backgrounds have higher opportunity ranks and vice versa. To measure inequality of opportunity, we correlate these opportunity ranks with individuals' welfare ranks. If individuals from the most fortunate backgrounds have the highest levels of welfare, then there is little equality of opportunity, and vice versa. This measure generalizes the frequently used Spearman's correlation in studies of intergenerational mobility (see for example Chetty *et al.* (2014)). It is a generalization since we rank individuals based on multiple circumstances rather than only parental income. We find that both the amount of inequality of opportunity and the characterization of the most opportunity-deprived largely is independent of how we measure well-being. Hence, akin to Chapter 2, this suggests that the way well-being is measured matters little empirically.

We developed the rank-based approach to measure inequality of opportunity because some of our well-being variables only had an ordinal interpretation. This is not an issue when income is used as the outcome variable, in which case most of the literature on equality of opportunity takes a different approach. Frequently, equality of opportunity is calculated by assigning each individual an opportunity set based on his/her circumstances, and calculating inequality of opportunity as the inequality in these opportunity set. One way of doing this is by segmenting the population into a number of types, such that within each type, individuals share the same circumstances. Each person is assigned the mean outcome of its type as its opportunity set. This implies that individuals of a good type (i.e. individuals from a fortunate background) are given high opportunity sets and vice versa. This approach suffers from a number of challenges. Particularly, the researcher has to decide which circumstances to use and how they relate to individuals' incomes. These are non-trivial decisions. If too few circumstances are taken into account, then the amount of inequality of opportunity will be underestimated, and inequality may be portrayed as arising dominantly from fair sources. Conversely, if too many circumstances are taken into account, part of the inequality that will be ascribed to factors beyond individual control will solely be due to sampling error, which does not reflect differences in opportunities.

In **Chapter 4**, written together with Paolo Brunori and Paul Hufe, we suggest utilizing machine learning methods to overcome these challenges. Machine learning methods can take all (measured) factors beyond individual control as inputs and use algorithms to endogenize the segmentation into types. Hence, the model of how unequal opportunities come about is not a judgement call of the researcher but a non-arbitrary outcome of data analysis. Specifically, we suggest using conditional inference trees, which segment the population into types based on sequential hypothesis tests. Whenever two groups of individuals are claimed to have different opportunities, this is based on statistical tests rather than researcher intuition. We show that these methods outperform the currently used methods in terms of predictive performance.

The chapters so far suggest that how we measure well-being has little policy implications *in general*. Yet, the two critiques of using preferences presented in Section 1.1 may suggest a different picture. Recall that the first critique is that individuals at times fail to reveal their preferences through choices due to lack of self-control and lack of information, and the second critique is that individuals are guided by other motivations than their own welfare. The final two chapters suggest that when either of these critiques apply, using revealed preferences can be problematic.

Chapter 5, written jointly with David Dreyer Lassen, looks at attitudes – or stated preferences, if you will – towards paternalistic policies. We look at policies trying to address smoking, obesity, lack of pension savings, and use of risky mortgages. These are all domains where there are good reasons to believe that many people are choosing inoptimally. We characterize what shapes attitudes towards paternalistic policies in these domains and investigate if individuals' stated preferences better capture their well-being in the presence of behavioral shortcomings. We find that individuals with self-declared self-control problems are more in favor of paternalism, suggesting that stated preferences may undo some of the problems of revealed preferences. In general, we find that individuals whose behavior is likely to be inoptimal are the ones most opposed to paternalism, even though these individuals likely would benefit most from such policies. Since paternalistic policies tend to restrict individuals' freedom, we investigate if their opposition towards paternalistic policies may be explained by individuals placing a separate value on freedom. Measuring the value of freedom is exceptionally difficult. We try to get at it indirectly by comparing targets of paternalism's attitudes towards nudges, which largely leave their freedom intact, with hard paternalism, which restrict individuals' freedom of choice. We find that targets of paternalism are relatively more against hard paternalistic policies but not relatively more against nudges, suggesting that they place a considerable value on freedom. In general, the paper points to many difficulties in using revealed preferences in areas where individuals suffer from behavioral biases and policies engage individuals' freedom. In these cases, one may plausibly rely on multiple accounts of welfare to try to establish the appropriate design of policies.

The other critique of revealed preferences suggests that individuals at times have greater objectives in life than satisfying their preferences. In many ways, this critique is the reverse of the behavioral critique. Whereas the behavioral critique investigated in Chapter 5 argues that it is too demanding to require preferences to be revealed through choices, this critique suggests that requiring preferences to be revealed through choices is too primitive. Although seemingly inconsistent, it is plausible that preferences are too demanding in some domains and too restrictive in others. The argument, loosely speaking, is that once we equate individual welfare with the satisfaction of revealed preferences, then, by construction, it is impossible for individuals to have non-selfish motivations for acting.[4] This is in contrast to everyday intuition that individuals often are guided by moral commitments, altruism or other non-selfish motivations, which drive *"a wedge between personal choice and personal welfare"* (Sen, 1977). This is not to say that individuals always are guided by such commitments: when individuals choose their labor supply, it seems safe to assume that they aim to maximize

[4]It is of course both possible and likely that individuals have a preference for the outcomes of other individuals (social preferences, Fehr and Schmidt (2006)) or for the act of helping other individuals (warm glow, Andreoni (1990)). As long as we equate revealed preferences with individual welfare, the motivation for these choices would still be selfish, as they would be done to increase the individual's own welfare.

their own well-being. However, when externalities or public goods are at play, or when the well-being costs to acting based on other motivations are minor, these commitments may be relevant.

Chapter 6 investigates if such altruistic motivations matter in a case where both of these conditions are met: voting in elections. A single vote is unlikely to change the outcome, so the cost to your own well-being of voting non-selfishly is negligible. On the other hand, it is plausible that individuals are encouraged to vote for what is best for society as a whole. Utilizing a representative sample of Danish voters, I ask respondents which party they would vote for if elections were held tomorrow, which party they would vote for if they were to consider what is best for themselves, and which party they would vote for if they were to consider what is best for society as a whole. This allows me to look at whether outcomes would be different if individuals' choices only reflect what they think is in line with their own welfare. I find that the election outcome would have looked very different if individuals voted based on what is best for themselves. Hence, non-selfish motivations are important in this domain, and equating satisfaction of revealed preferences with welfare gives wrong policy conclusions.

The conclusions of the different chapters may seem contradictory. Chapter 2 and 3 argue that it matters little how we measure welfare. Based on this, one may conclude that using revealed preferences is rather problem free. There are two important exceptions to this. Chapter 3, 4 and 5 suggest that policy implications change if we go beyond welfare and include justice, in the form of equality of opportunity, or freedom, when making policy suggestions. Chapter 5 and 6 point to two cases where using revealed preferences may be misleading, even within a welfarist framework. This concerns when we have reasons to believe that individuals lack self-control to act in their own interest (Chapter 5), and when individuals have moral reasons to act against their own interest (Chapter 6). The latter is likely to be an issue when considering public goods or goods with externalities. Hence, the overall conclusion drawn is that going beyond revealed preferences matters little in a welfarist framework and in the absence of internalities, externalities, and public goods. Are these criteria not met, I suggest that other measures of what matters should be utilized.

Bibliography

ALKIRE, S. and FOSTER, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, **95** (7), 476–487. (page 5)

ANDREONI, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, **100** (401), 464–477. (page 10)

ATKINSON, A. B. (2011). The restoration of welfare economics. *The American Economic Review*, **101** (3), 157–161. (page 5)

BARTLING, B., FEHR, E. and HERZ, H. (2014). The intrinsic value of decision rights. *Econometrica*, **82** (6), 2005–2039. (page 6)

BENJAMIN, D. J., HEFFETZ, O., KIMBALL, M. S. and SZEMBROT, N. (2014). Beyond happiness and satisfaction: Toward well-being indices based on stated preference. *The American Economic Review*, **104** (9), 2698–2735. (page 3)

BOULDING, K. E. (1969). Economics as a moral science. *The American Economic Review*, **59** (1), 1–12. (page 3)

CAMPANTE, F. and YANAGIZAWA-DROTT, D. (2015). Does religion affect economic growth and happiness? evidence from ramadan. *The Quarterly Journal of Economics*, **130** (2), 615–658. (page 5)

CAPPELEN, A. W., HOLE, A. D., SØRENSEN, E. Ø. and TUNGODDEN, B. (2007). The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, **97** (3), 818–827. (page 3)

CHETTY, R., HENDREN, N., KLINE, P. and 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. (page 8)

COHEN, G. A. (1990). Equality of what? On welfare, goods and capabilities. *Recherches Économiques de Louvain/Louvain Economic Review*, **56** (3-4), 357–382. (page 8)

Crisp, R. (2017). Well-being. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, fall 2017 edn., Metaphysics Research Lab, Stanford University. (page 6)

Di Tella, R., MacCulloch, R. J. and Oswald, A. J. (2001). Preferences over inflation and unemployment: Evidence from surveys of happiness. *The American Economic Review*, **91** (1), 335–341. (page 5)

Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. In S.-C. Kolm and J. M. Ythier (eds.), *Handbook of the economics of giving, altruism and reciprocity*, vol. 1, Elsevier, pp. 615–691. (page 10)

Fleurbaey, M. (1994). On fair compensation. *Theory and decision*, **36** (3), 277–307. (page 7)

— (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, **47** (4), 1029–1075. (page 2)

Goodin, R. E. (1986). Laundering preferences. In J. Elster and A. Hylland (eds.), *Foundations of Social Choice Theory*, vol. 75, Cambridge University Press, pp. 75–101. (page 5)

Griffin, J. (1986). Well-being: Its meaning, measurement, and moral importance. (page 5)

Hausman, D. and McPherson, M. S. (1997). Beware of economists bearing advice. *Policy Options*, **18** (7), 16–19. (page 5)

Hausman, D. M. and McPherson, M. S. (1993). Taking ethics seriously: Economics and contemporary moral philosophy. *Journal of Economic Literature*, **31** (2), 671–731. (page 2)

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47** (2), 263–292. (page 7)

Kaplow, L. and Shavell, S. (2001). Any non-welfarist method of policy assessment violates the pareto principle. *Journal of Political Economy*, **109** (2), 281–286. (page 6)

Parfit, D. (1984). *Reasons and persons*. OUP Oxford. (page 5)

Ravallion, M. (2015). *The economics of poverty: History, measurement, and policy*. Oxford University Press. (page 2)

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, pp. 146–166. (page 7)

Samuelson, P. (1948). Foundations of economic analysis. (page 7)

Sen, A. (1970). The impossibility of a paretian liberal. *Journal of Political Economy*, **78** (1), 152–157. (page 6)

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, pp. 317–344. (page 10)

Weinzierl, M. (2014). The promise of positive optimal taxation: Normative diversity and a role for equal sacrifice. *Journal of Public Economics*, **118**, 128–142. (page 3)

Chapter 2

The Value Function with Life Satisfaction Data

The Value Function with Life Satisfaction Data[1]

Ohto Kanninen
Labour Institute for Economic Research

Daniel Gerszon Mahler
University of Copenhagen

Abstract

Prospect theory and its empirical applications have shown that in some contexts people base their choices on a reference point. The resulting mapping from outcomes to utility is called the value function and it exhibits loss aversion and diminishing sensitivity. These properties make the value function an S-curve with a kink at the reference point. In this paper, we use the German Socio-Economic Panel ($n > 250,000$) to test whether the properties of the value function extend from narrow gambling choices in experiments to yearly changes in earnings evaluated with life satisfaction. We find that the mapping from changes in earnings to life satisfaction mimics the predicted S-curve remarkably well when the reference point is generated from individuals' past earnings. This finding is robust to a large set of alternative specifications. We emphasize that the S-curve we find need not be causal, since the changes in earnings were based on observational data. However, we can rule out that a wide array of other factors produced the observed relationship, including expected utility theory.

Keywords: Prospect theory, loss aversion, life satisfaction, subjective well-being
JEL codes: D03, D60, I31

[1]We are grateful for comments received from participants at presentations held at Helsinki Center of Economic Research, Norface 4Is project workshop, the University of Copenhagen, the Interdisciplinary Center for Economic Science at George Mason University, and the European Meeting of the Experimental Science Association, 2016. We are further thankful for comments from conversations with Abhijit Banerjee, John Gruber, David Dreyer Lassen, Michael Norton, Peter Norman Sørensen, and Andrei Shleifer. Any errors remain our own. Ohto Kanninen acknowledges financial support from NORFACE project number 465-13-186 and the Yrjö Jahnsson Foundation.

2.1 Introduction

The relationship between income and well-being is central to economic policy making. Most economic policies redistribute income between individuals or entail a trade-off between income and other goods. In order to evaluate the well-being consequences of such policies, the relationship between income and well-being must be pinned down. Until recent decades, expected utility theory combined with risk aversion provided the rationale for depicting well-being as a concave function of income.

To account for observed systematic departures from expected utility theory, Kahneman and Tversky (1979) proposed prospect theory as an alternative theory. At the core of prospect theory is the value function, which measures the subjective value derived from the argument in question, such as income. The value function encompasses three departures from expected utility theory. First, the argument of the value function is defined relative to a reference point. Second, it contains loss aversion, implying that it is steeper for losses from the reference point than gains from the reference point. Third, it exhibits diminishing sensitivity, meaning that it is concave for gains and convex for losses.

If prospect theory depicts an accurate relationship between income and well-being, then many policies may need to be revisited. Hence, obtaining empirical evidence on the basic elements of prospect theory is crucial. Starting from Kahneman and Tversky (1979), the properties of the value function have been well documented in experimental settings where subjects are asked to choose between various gambles. Two elements of the value function, reference dependence and loss aversion, have also been shown to influence behavior in many settings outside the lab (see Barberis (2013) for an excellent review). However, diminishing sensitivity, the feature of prospect theory that generates an S-curve, has remained elusive in non-experimental settings. Barberis (2013) concludes that diminishing sensitivity seems "much less important" in empirical applications of prospect theory. Shleifer (2012) notes that the value function occasionally is graphed as a simple piecewise linear function, thus fully neglecting diminishing sensitivity.

In this paper, we test for the existence of an S-curve outside of the lab. We find that the mapping from changes in earnings to life satisfaction follows the hypothesized S-curve remarkably well. In congruence with previous experimental evidence, we find that losses carry a larger weight than gains. We complement the findings of Kahneman and Tversky (1979) in three important ways. Firstly, we look at experienced utility, as measured by life satisfaction, rather than decision utility. Secondly, we consider changes in labor income rather than narrow gambling situations. Thirdly, we extend the previous laboratory findings to a non-experimental setting. For a wider applicability of any laboratory finding, this last step is pertinent, but it sets challenges to causal inference.[2]

We measure income by self-reported net monthly earnings. Our baseline reference point is last year's monthly earnings plus the average growth in earnings. Our data comes from 30 years of the German Socio-Economic Panel (GSOEP), which allows our baseline set-up to contain more than 250,000 observations.

We emphasize that we observe the S-curve in non-experimental data. In principle, an S-curve could arise spuriously from expected utility theory if there is a particular correlational structure between life satisfaction, income changes, and income levels. To assure that this is not the case, and inspired by the reference-dependent utility function in Kőszegi and Rabin (2006), we test for the presence of the value function while simultaneously controlling for income levels. We find evidence for both the S-curve and expected utility theory, suggesting that both theories map to experienced utility. This is in line with prior experimental evidence (Harrison and Rutström, 2009). In particular, we find that if the median earner experiences a normal change in earnings, the value function has roughly half the impact on life satisfaction relative to the standard utility function. For individuals with higher earnings, diminishing marginal utility kicks in and prospect theory plays about the same role as expected utility theory.

To minimize the possibility that omitted variable bias is behind our observed S-curve, we also control for age, health status, work hours, marital status, household composition, and job changes. In our main specification, we use individuals' current net monthly earnings. To address potential concerns about reverse causality, we conduct robustness checks where our income measure is the annual income in the year prior to the survey. The S-curve remains.

Another potential concern is that diminishing sensitivity arises mechanically due to the boundedness of the life satisfaction scale, which goes from 0 to 10. Individuals who – following an income increase – report 10 out of 10 in life satisfaction could not have increased their life satisfaction further had they achieved an even higher income increase. The reverse applies to individuals who report 0 out of 10. Hence, for large deviations from the reference point, one may expect the marginal impact of income on life satisfaction to level off. Three points argue against this generating diminishing sensitivity. Firstly, rather few individuals find themselves at the boundary of the life satisfaction scale (4.5% report being 10/10, 0.2% report being 0/10 and only 1.25% being below 3). If we exclude individuals reporting 0, 1, 9, or 10 our results still hold. Secondly, if diminishing sensitivity is purely mechanical, any variable positively associated with life satisfaction should generate S-curves as well. We show that this is not the case. Thirdly, as a robustness check we transform the life satisfaction variable such that

[2] A fourth departure from Kahneman and Tversky (1979) is that we look at realized outcomes. Consequently, we are not concerned with uncertainty and probability weighting.

responses close to the boundary carry a larger weight. This does not change the results.

In sum, we fail to find an alternative explanation for the S-curve and infer that it is likely that the value function plays a causal role in converting income changes into utility. Experimental or quasi-experimental downward income changes of varying magnitudes remain to be studied to confirm the causality of our observed relationship.

In further robustness checks, we vary the dependent variable, the income variable, the dataset (with the British Household Panel Survey), and the reference point. We also test the results with various subsamples. The robustness checks suggest that the relationship is not spurious. However, not all the properties of the S-curve hold with all specifications. Generally, diminishing sensitivity is more robust to changes in the specification than loss aversion, which fluctuates depending on the reference point and the econometric specification.

To our knowledge, we are the first to find that the value function follows the predictions from prospect theory outside of lab settings. The S-curve has previously been found using experienced utility in experimental settings (see for example Galanter (1990) and Carter and McBride (2013)). The most extensive study using experienced utility in a non-experimental context is by Vendrik and Woltjer (2007). They look for the S-curve, also using the German Socio-Economic Panel, but where the reference point is defined as the mean income of a reference group and the independent variable is relative income changes, not absolute income changes. They observe a globally concave value function and thus rule out an S-curve with their reference point. Lamela (2015) likewise tests if an S-curve is present in Uruguayan data using the mean income of a reference group as the reference point. No significant S-curve is found. We complement these findings by showing that generating a reference point based on past earnings using incomes in levels does generate an S-curve. The S-curve we find holds also with log changes and weakly and non-robustly with relative changes.

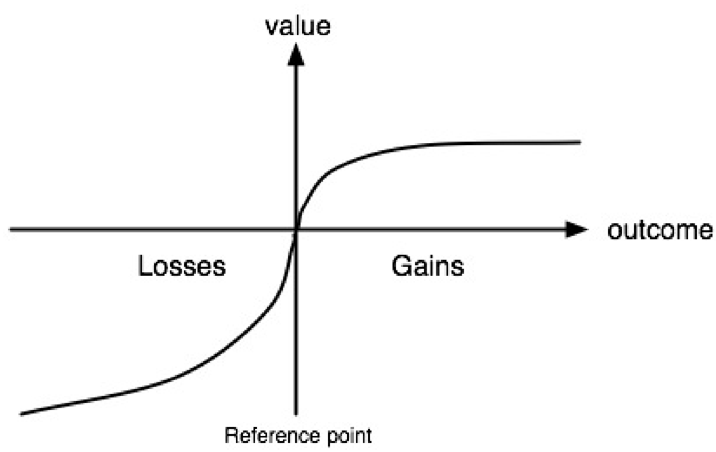
A number of papers have documented elements of the value function by using experienced utility outside of experiments. For example, Luttmer (2005) and Ferrer-i Carbonell (2005) find that individuals' well-being depend on the distance between their own income and the income of a reference group. Boyce *et al.* (2013) study loss aversion using panel data. They observe bigger impacts of losses than gains in income. De Neve *et al.* (forthcoming) exploit macroeconomic variation in incomes to study loss aversion, which they find support for.

The rest of the paper is organized as follows. Section 2.2 describes the theory and outlines our main hypotheses. Section 2.3 describes the data and our empirical specification. Section 2.4 presents the results. Section 2.5 concludes.

2.2 Theory

At the core of prospect theory is the value function, which has three fundamental properties: loss aversion, diminishing sensitivity, and reference dependence.[3] The argument in the value function is changes in income or consumption rather than levels of income or consumption as in the neoclassical approach. Kahneman and Tversky (1979) argued that people tend to be more sensitive to differences between small than large changes and more sensitive to losses than gains. Thus, the value function is concave for gains, convex for losses, and steeper for losses than gains, i.e. it exhibits diminishing sensitivity and loss aversion. Together with differentiability everywhere except at the reference point, this generates a value function with an S-shape as shown in Figure 2.1. The value function is not expected to be predictive of losses so large that "ruin or near ruin is a possible outcome" (Kahneman, 2003).

Figure 2.1: Textbook Version of the Value Function



Notes: From the Wikipedia entry on prospect theory.

Let $V(\Delta y)$ be the value function, where Δy is changes in y, which in the value function can be income, wealth or consumption. Income can be the argument of the value function formulation in its own right or serve as a proxy for consumption. In principle, differences in consumption and income relative to a reference point could have independent effects in the value function. In expected utility theory, the argument is always consumption and income only acts as a proxy for it.

We assume that $V(\Delta y)$ is continuous for all Δy, twice differentiable for all $\Delta y \neq 0$, and that $V(0) = 0$. This notation implicitly assumes that the reference point is no increase in

[3] Prospect theory also hypothesizes that people tend to overweight small probabilities and underweight large ones. Our focus is on realized incomes for which there is no uncertainty.

y. Given that the average increase in real income typically is larger than zero, this may seem like an overly pessimistic reference point. In our empirical specification we will allow individuals to have a larger than zero increase as the reference point. Hence, strictly speaking our value function will take the form $V(y_{it} - r_{it})$, where r_{it} is the idiosyncratic reference point for individual i at time t. For notational convenience we denote $y_{it} - r_{it} = \Delta y$ in this section.

Based on Kahneman and Tversky (1979), Bowman *et al.* (1999), and Kőszegi and Rabin (2006), we make three sets of testable hypotheses regarding the properties of the value function:

H1a: $V'(\Delta y) \geq 0$ and $V''(\Delta y) < 0$, for $\Delta y > 0$,

H1b: $V'(\Delta y) \geq 0$ and $V''(\Delta y) > 0$, for $\Delta y < 0$.

H1 defines diminishing sensitivity in the positive (**H1a**) and negative (**H1b**) domain.

H2: $\frac{|V(-\Delta y)|}{V(\Delta y)} > 1$, for $\Delta y > 0$,

H2': $\lim\limits_{\Delta y \to 0} \frac{V'(-\Delta y)}{V'(\Delta y)} = \delta > 1$, for $\Delta y > 0$,

H2": $\frac{|V(-\Delta y)|}{V(\Delta y)} = \delta^c > 1$, for $\Delta y > 0$.

H2 assumes that for any given distance from the reference point, a loss carries a larger weight than an equivalent gain. **H2'** defines loss aversion around the reference point, where $\delta > 1$ is the loss aversion parameter. It implies that very small losses have a larger impact than very small gains. **H2"** is the strongest version, as it assumes loss aversion is constant in the whole domain of the value function (Tversky and Kahneman, 1991). We denote the parameter of constant loss aversion δ^c, defined by Tversky and Kahneman (1991) and estimated to be around 2.25 over the range of a few hundred dollars in Tversky and Kahneman (1992).

Loss aversion for large changes is ambiguously defined and has taken many definitions in the literature. Kahneman (2003) and Köbberling and Wakker (2005) only look at loss aversion for small changes (**H2'**), whereas Kahneman and Tversky (1979) and Tversky and Kahneman (1992) assume it holds at changes of all sizes (**H2**). As we will explain later, our empirical approach only allows us to look at **H2**.

Inspired by Kőszegi and Rabin (2006), throughout the empirical work, we formulate a reference dependent utility function that combines the standard utility function and the value function:[4]

[4] Kőszegi and Rabin (2006) use the terms "consumption utility" and "gain-loss utility." Since we make some important departures from their model, we will instead use the terms standard utility function and the value function.

H3: $u(c|r) = m(c) + n(c|r) \approx m(y) + V(\Delta y)$,

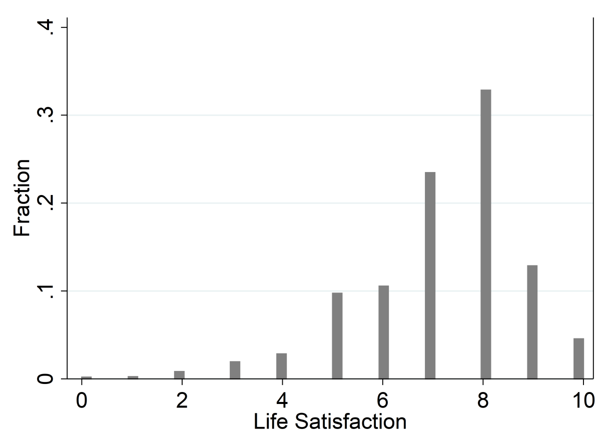
where r is the reference consumption, $m(c)$ is the standard utility function and $n(c|r)$ is a gain-loss utility function. We approximate the consumption in $m(c)$ with income, y. Kőszegi and Rabin (2006) assume that the decision-maker assesses gain-loss utility in each dimension separately. We focus on just one general dimension of consumption as proxied by earnings. In all specifications, we control for the standard utility function. We will test whether the two components of **H3** translate into experienced utility.

2.3 Data & Empirical Specification

2.3.1 Data

We use the German Socio-Economic Panel (GSOEP), which is a nationally representative household survey conducted yearly since 1984. We use data from 1984–2015. The total sample meeting our baseline empirical specification contains 42,002 individuals and 259,384 observations. As the measure of value we use life satisfaction, which is the answer to the question *"How satisfied are you with your life, all things considered?"* The answer categories range from 0 (completely dissatisfied) to 10 (completely satisfied). A histogram over the answers to the question is given in Figure 2.2.

Figure 2.2: Histogram of Life Satisfaction



Notes: Histogram of answers to the question, *"How satisfied are you with your life, all things considered?"* 0 means completely dissatisfied and 10 completely satisfied.

We use self-reported net monthly earnings as the income variable. Only employed individuals can report any earnings. Thus, unemployed and people outside the labor market are excluded from the main analysis. The income variable has been deflated and is expressed in constant 2010 EUR. Income variables that include all respondents and all forms of income exist at the household level. Most of these other income measures reflect the income for the year prior to the interview, and as such need not reflect the income available to the respondent at the time of the interview. We therefore prefer to use contemporaneous labor earnings. Later on, we show that the results also hold when using income variables from the year before the interview using the entire sample.

2.3.2 Empirical Specification

In our baseline specification, we run the following regression:

$$LifeSat_{it} = \alpha_i + m(y_{it}) + V^{-}(y_{it} - r_{it})N_{it} + V^{+}(y_{it} - r_{it})P_{it} + \sum_{k \in K} \beta_k x_{itk} + \epsilon_{it},$$

where y_{it} is earned income and r_{it} is the reference level of earnings. In our baseline case, we set $r_{it} = (1 + \mu_t)y_{it-1}$, where $\mu_t = \frac{1}{n_t} \sum_{j \in n_t} \frac{y_{jt} - y_{jt-1}}{y_{jt-1}}$ and n_t is the number of respondents with income data in year t and year $t-1$. That is, we take the reference point to be last year's earnings plus the average growth in earnings from the year before. To contain the influence of outliers on the average growth rate, we top-code the growth rates at 100%. μ_t ranges from 3% to 9% over the 30 years of data. The mean over the entire sample is 6%. We will use several other reference points, including the median growth rate, as robustness checks.

P_{it} and N_{it} are indicators of positive and negative difference from the reference point. $m(y_{it})$ and $V^{-}(y_{it} - r_{it})$ and $V^{+}(y_{it} - r_{it})$ are estimated with a piecewise second-order polynomial or a restricted cubic spline. Splines use a number of pre-selected knots to fit piecewise third-order polynomials with continuous first-order and second-order derivatives. This allows for a flexible, localized structure. However, one needs to make choices regarding the number of knots, and where these should be located. The more knots, the more flexible the functional form can be. We place the knots based on guidelines by Harrell (2015) and judge 4 knots to be the most reasonable compromise between precision and bias. We show results when using 3, 5 and 6 knots in the Appendix (Figure 2A.8). We prefer splines to power functions such that we can be agnostic about the true functional form while at the same time easily accommodate individual fixed effects in the regression.[5]

Since our income changes are not randomized, it is pertinent to add relevant control variables, x_k, to the regression. We control for work hours, voluntary/involuntary job changes[6],

[5] Splines are known to be imprecise at the tails of distributions. Restricted cubic splines try to deal with this by making linear predictions beyond the first and last knot, which in our case will be below the 5th percentile and above the 95th percentile. All the splines we show contain only the 80% most central part of the data where this linearity is not in effect.

[6] Voluntary: own resignation, employee requested transfer within company, end of self-employment. In-

health status (proxied by yearly doctor visits), marital status, number of kids in household, number of adults in household, 10-year age cohorts, and year fixed effects.

The treatment of age groups and year fixed effects is particular relevant, as both will matter directly for life satisfaction and potentially also for shaping the reference point. Age matters directly as life satisfaction is known to follow an inverse U in age (Blanchflower and Oswald, 2008) and indirectly since individuals' expectations and reference points plausibly depend on their age (Schwandt, 2016). Years matters directly since life satisfaction is lower in times of economic hardship, and vice versa, and indirectly since individuals plausibly adjust their reference point based on average growth rates. In our preferred specification, we account for years both through the reference point and by including year dummies. We control for age directly in the regression but not through the reference point. We run several robustness checks altering these assumptions. Notice that if individuals do not form their reference points based on annual growth rates or based on the age cohort to which they belong, but simply use last year's income as the reference, then controlling for either will remove some of the variation we want the value function to pick up.

All estimates include individual fixed effects and cluster standard errors at the individual level. We deliberately interpret the life satisfaction answers as cardinal. Using an ordered logit model or a related model would take out information that is useful for the present purposes. The distribution of $y_{it} - y_{it-1}$ is given in Figure 2.3. 95.6% of the observations are contained in the window. The histogram of changes from the reference point, $y_{it} - r_{it}$, looks almost identical but shifted slightly to the left. To assure that our results are not driven by outliers, we remove the 2% most extreme changes in incomes (1% in each tail). In addition, we top-code the income variable at the 99.5th percentile.

Carter and McBride (2013) and Kahneman and Tversky (1979) also study the value function as a function of absolute income changes, while Vendrik and Woltjer (2007) use the relative income gap. We will show results later using relative differences and log differences.

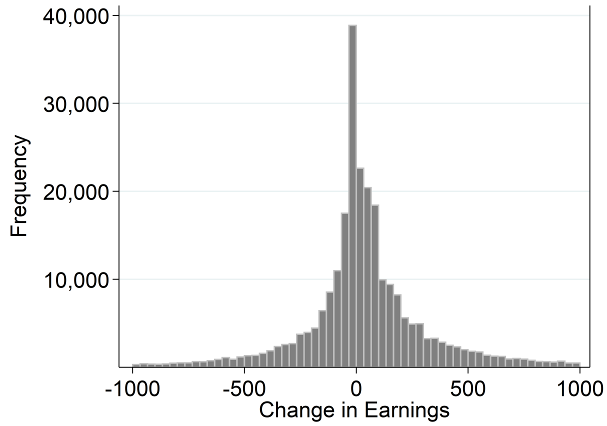
2.4 Results

2.4.1 Preliminary Graphical Results

To start with, we graphically inspect if the relationship between changes from the reference point and life satisfaction follows the pattern presented in Figure 2.1. In Figure 2.4 we regress

voluntary: terminated by employer, temporary contract expired, company transferred employee, company closed down.

Figure 2.3: Histogram of Changes in Earnings



Notes: Histogram of changes in monthly net earnings (2010 EUR).

life satisfaction on a restricted cubic spline of changes from the reference point with controls. In panel (a) we use the mean income growth to form the reference point, while in panel (b) we use the median income growth to form the reference point. In the figures, and throughout the paper, we use bootstrapping to derive uncertainty. We resample at individual level clusters and use the percentile method to generate confidence intervals. In Figure 2.4, and all the figures that follow, we restrict the x-axis to contain 80% of the sample to assure that our conclusions are not driven by extreme changes from the reference point.

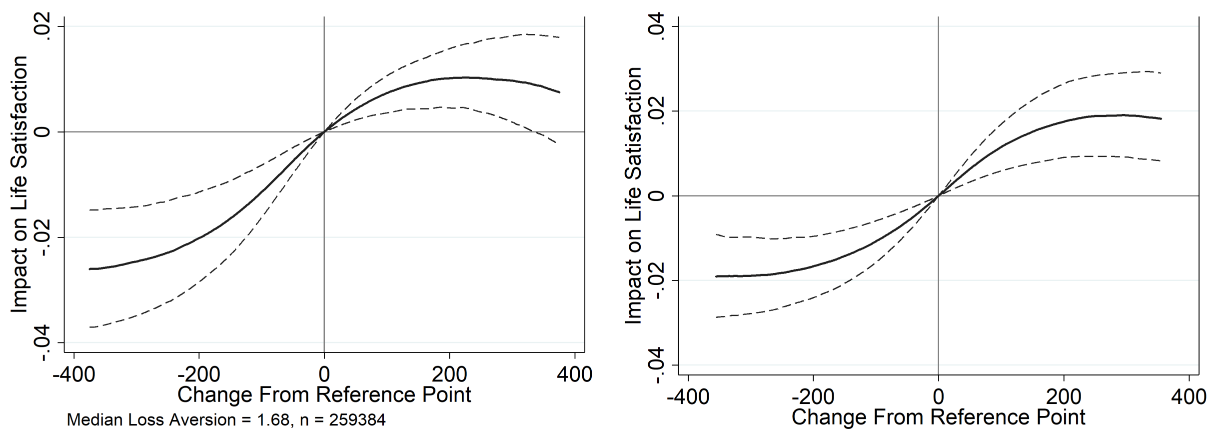
The splines suggest that we have evidence for both diminishing sensitivity and loss aversion in the form of **H2** (positive changes from the reference point have a smaller impact than equivalent negative changes) when the mean is used to form the reference point. When the median is used to form the reference point, we have evidence for diminishing sensitivity but not for loss aversion.

2.4.2 The Reference Point

The previous discussion imposed very specific reference points on the respondents, the reason being that there is no obvious way to define the reference point for the value function (see Barberis (2013) for a discussion of this). Kahneman (2003) states that the reference point is "usually the status quo." In our context, it is not obvious what each individual considers to be the status quo. Kőszegi and Rabin (2006) assume that the reference point is defined as rational expectations about outcomes, which Abeler *et al.* (2011) find convincing support for. It is unclear how best to operationalize these expectations, but it is plausible that individuals rationally expect to have an average increase in earnings.

Figure 2.4: Restricted Cubic Splines

(a) $r = inc_{t-1}$(1+mean growth) (b) $r = inc_{t-1}$(1+median growth)



Notes: Both panels show the predicted values from a fixed effects regression of life satisfaction on changes from the reference point. Both regressions control for a spline in income levels, marital status, work hours, job changes, household composition, health, age groups, and year dummies. Panel (a) uses the mean income growth to form the reference point, while panel (b) uses the median income growth to form the reference point. We use a restricted cubic spline with 4 knots. The window captures 80% of the sample. Dashed lines indicate bootstrapped 95 pct. confidence interval.

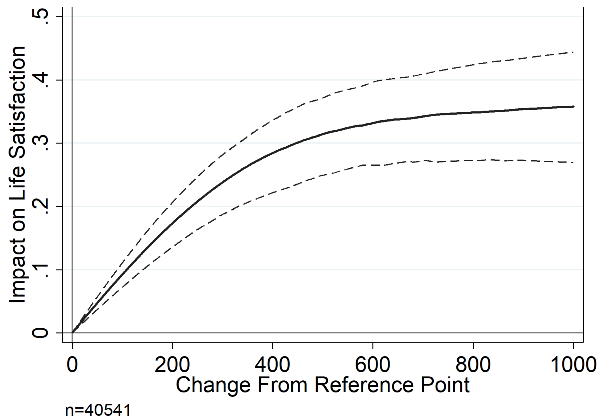
Our favored reference point is last year's earnings plus the average growth in earnings from the year before. It is important to note that our choice of reference point is mechanically related to the amount of loss aversion we find. If we choose a higher reference point than the mean growth, then, essentially, the vertical line in Figure 2.4 shifts to the right, and loss aversion increases. A reference point larger than the mean growth might be accurate if individuals suffer from over-optimism bias (Sharot, 2011), and think that they will experience better conditions than average. They may expect, for example, to get an earnings increase around the 75th percentile.

The GSOEP provides an explicit question that tries to get to the reference point in the surveys of 1992, 1997, and 2007. The respondents are asked to state what would be a *somewhat inadequate income*, a *barely adequate income*, a *good income*, and a *very good income* in net monthly terms for the household (see Van Praag and Frijters (1999) for more information on these questions). In addition, in 2002, 2007, and 2012 individuals are asked what would be *the minimum net household income needed to get by.* In Figure 2.5, we plot a similar graph as before, but now the reference points are some of these self-reported, subjective income levels. Hence, $r_{it} = y_{it,good}$ etc. We use a measure of households' monthly net income as the income variable to align the income measure with these particular questions.

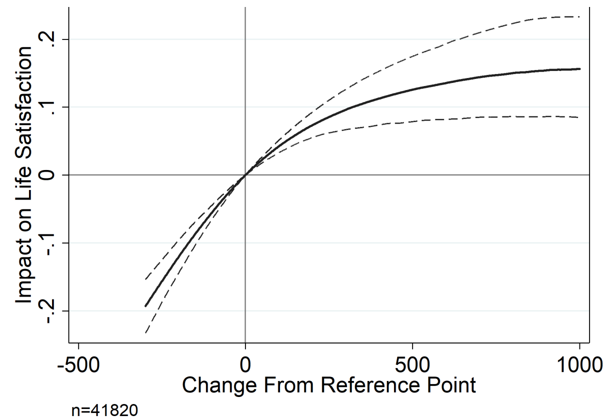
In panels (a)-(c), we have evidence for the top half of the S-curve, but since few individuals have incomes lower than what they consider to be *a somewhat inadequate income*, *a barely*

Figure 2.5: Using Subjective Reference Points

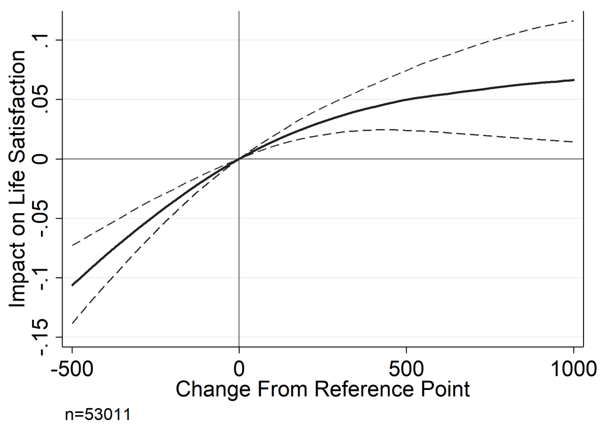
(a) Somewhat Inadequate Income



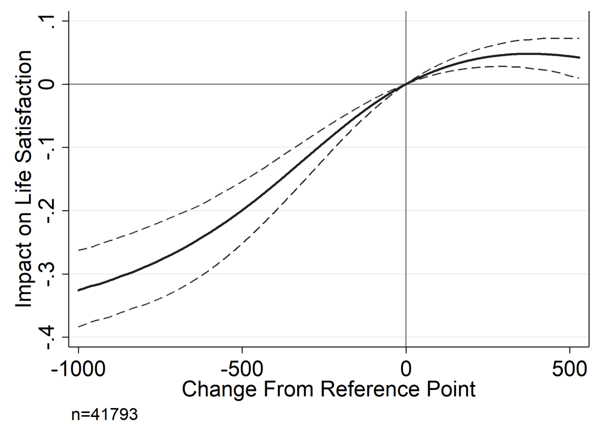
(b) Barely Adequate Income



(c) Minimum Needed Income



(d) Good Income



Notes: The figure shows the predicted values from regressing life satisfaction on the distance between net monthly household income and the net monthly household income deemed necessary by the respondent to have "a good income" etc. All splines are based on fixed effects regressions, which control for year dummies, age groups, income levels, marital status, work hours, job changes, household composition, and health. The graph windows capture individuals with a change from the reference point larger than the 10th percentile and smaller than the 90th percentile.

adequate income, and *a minimum needed income*, we cannot confirm the lower half of the S-curve. If we use a *good income*, we get a beautiful S, but with extreme loss aversion. This indicates that a reference point just below *a good income* generates an S-curve with the amount of loss aversion found in experimental settings. The median person thinks a good income is 256 EUR more than his or her own household income. This suggests that individuals use a reference point slightly above what they possess, which is consistent with our preferred choice of reference point.

Note that the self-reported numbers seem to be a more precise measure of the reference point. The effects are about five times as large with the subjective reference points compared to the reference points based on lagged earnings. Thus, lagged earnings could be considered a noisy measure of the reference point. One potential reason for this is that lagged earnings are subject to some degree of hedonic adaptation (see Frederick and Loewenstein, 1999; Loewenstein and Ubel, 2008; Kimball *et al.*, 2015). People expect life events to have a larger effect on their well-being than actually is the case (see Riis *et al.*, 2005; Ubel *et al.*, 2005). The respondents are usually not interviewed the moment they hear about their raise in salary. By the time they are surveyed about their life satisfaction, they may have partly adapted to the higher or lower earnings compared to the previous year. As they are asked to state subjective reference earnings at the same time as they report their income, no such adaptation takes place and the measured effect is larger.

Fundamentally, it is problematic to use subjective measures of the reference point, since they are likely to have correlated measurement errors with life satisfaction. An advantage of generating reference points based on lagged earnings is that we have more than five times the sample size. Alternatively, we could try to infer possible rational expectations of future growth through a time-series fit of the income path of the individual. Such a model seems to give a more precise estimate of the S-curve (see Section 2.4.4 and Figure 2A.4 for a discussion of this). We use past earnings plus average growth for a more transparent main specification.

One important limitation of having a noisy measure of the reference point is that a potential kink at the reference point is severely attenuated. As such, we do not expect any discrete changes to occur at the reference point, and focus on S-curves with a continuous slope at the reference point. Using an S-curve with a continuous slope at the reference point implies that the impact of a loss relative to a gain at the reference point is constrained to equal one. We are therefore unable to test if the slope changes at the reference point (**H2'**) and cannot expect to find constant loss aversion (**H2"**). We are, however, able to study whether the loss aversion ratio approaches levels found in experimental settings as we move further away from the reference point.

2.4.3 Testing the Hypotheses

Diminishing Sensitivity (H1)

To test the relationship between earnings changes and life satisfaction more formally, we explore which of our hypotheses are consistent with the data. We want to emphasize that our setting is not experimental. Thus, the tests we present here are not definite and further research is needed. However, we diligently study whether confounding variables or reversed causality could drive the results.

First, we test **H1**, diminishing sensitivity. To test for diminishing sensitivity, we run the same regressions presented in Figure 2.4. Instead of a spline, we fit quadratic terms on the positive and negative domain separately but with a continuous linear term. Next to using the mean growth rate to form the reference point, we also show results using the median growth rate, and the 60th percentile growth rate. The results are presented in Table 2.1.

Table 2.1: Regression Results

	(1) $r = inc_{t-1}\times$ (1 + mean growth)		(2) $r = inc_{t-1}\times$ (1 + median growth)		(3) $r = inc_{t-1}\times$ (1 + 60th pctl growth)	
	Coef	SE	Coef	SE	Coef	SE
Exp. Utility Theory:						
y	.260***	(.020)	.255***	(.020)	.257***	(.020)
y^2	−.020***	(.003)	−.020***	(.003)	−.020***	(.003)
Prospect Theory:						
$y-r$	.061***	(.018)	.076***	(.019)	.068***	(.018)
$(y-r)^2$ [gains]	−.073***	(.025)	−.088***	(.023)	−.083***	(.024)
$(y-r)^2$ [losses]	.042**	(.020)	.066**	(.026)	.052**	(.022)
Controls:						
Married (baseline)						
Single	−.088***	(.022)	−.088***	(.022)	−.088***	(.022)
Widowed	−.310***	(.076)	−.311***	(.076)	−.310***	(.076)
Divorced	−.045	(.031)	−.048	(.031)	−.047	(.031)
Separated	−.397***	(.035)	−.398***	(.035)	−.398***	(.035)
Work hours	−.003***	(.001)	−.003***	(.001)	−.003***	(.001)
# of kids in hh	−.008	(.008)	−.008	(.008)	−.008	(.008)
# of adults in hh	.003	(.007)	.003	(.007)	.003	(.007)
Pos. job change	.062***	(.017)	.062***	(.017)	.062***	(.017)
Neg. job change	−.097***	(.024)	−.093***	(.024)	−.096***	(.024)
Annual doctor visits	−.014***	(.000)	−.014***	(.000)	.014***	(.000)
Age: <30 (baseline)						
Age: 31-40	−.014	(.017)	−.015	(.017)	−.014	(.017)
Age: 41-50	−.032	(.024)	−.034	(.024)	−.033	(.024)
Age: 51-60	.001	(.030)	−.000	(.030)	.001	(.030)
Age: >60	0.172***	(.040)	.170***	(.040)	.171***	(.040)
Year fixed effects	Yes		Yes		Yes	

Notes: * p<0.10, ** p<0.05, *** p<0.01. All models have life satisfaction as the dependent variable, contain individual fixed effects, and individual-level clustered standard errors. Incomes are expressed in 1,000 EUR (2010 prices). $n = 259,384$.

Model (1), our preferred model, shows significant evidence for diminishing sensitivity both in the loss domain and in the gain domain. This does not change if we use the median growth rate instead of the mean growth rate (model (2)), or if we use the 60th percentile as the growth rate (model (3)). The coefficients on the other variables are largely as expected.

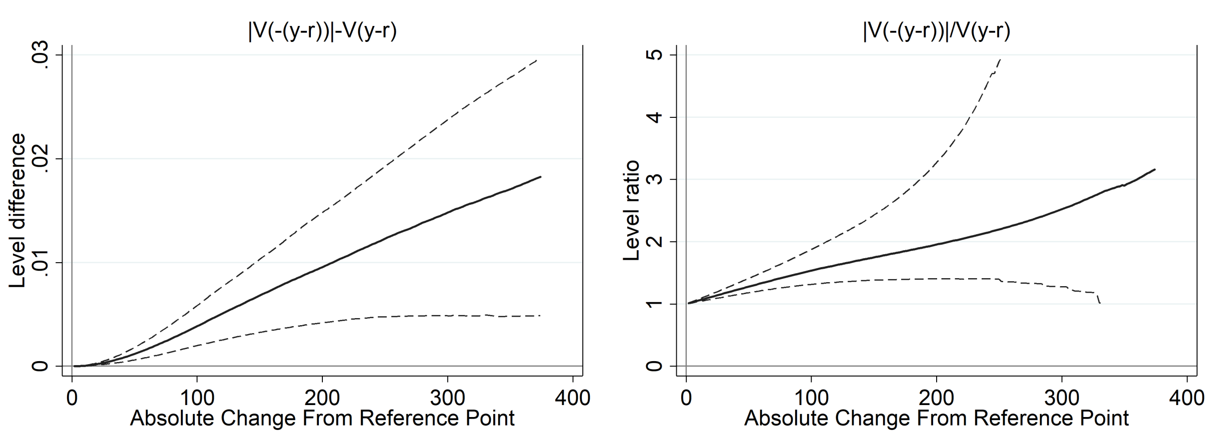
The predictions from expected utility theory show a positive but declining role of additional income. Everything else equal, being widowed or separated has a negative impact on life satisfaction. The same applies to increasing work hours and having a worse health. Positive job changes are associated with a higher life satisfaction while negative job changes are associated with lower life satisfaction.

Loss Aversion (H2)

We next turn to testing whether loss aversion is present. Figure 2.6 tests whether loss aversion in the form of **H2** is present. The figure is derived from the spline presented in Figure 2.4(a), which is equivalent to model (1) in Table 2.1 using splines instead of polynomials. It shows whether a decrease of a given size from the reference point has a larger impact on life satisfaction than a similar increase. This is the case for all changes from the reference point, which is consistent with **H2**. For example, according to Figure 2.6(a), we predict that losing 200 EUR is 0.010 [0.004-0.015] life satisfaction units more impactful than gaining 200 EUR.

Figure 2.6: Loss Aversion

(a) Level Difference (b) Level Ratio



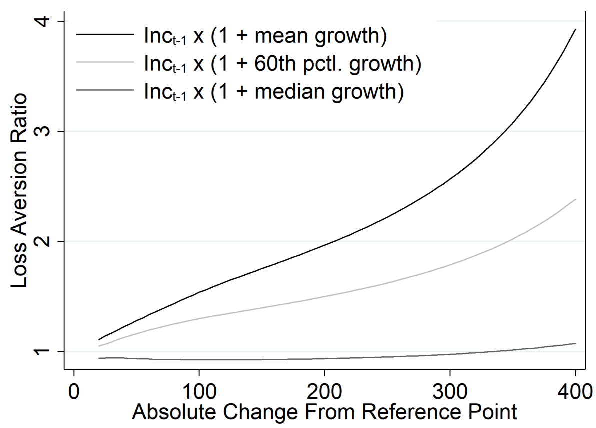
Notes: The graphs display the nature of loss aversion by computing differences and ratios in the impact of losses and gains. The graphs are based on model (1) from Table 2.1 using splines instead of polynomials. Sample size in both graphs is 259,384. Confidence bands are generated by 500 bootstraps resamples at individual level clusters using the percentile method.

Figure 2.6(b) shows the corresponding loss aversion ratio; the factor by which a loss of a given size is worse than an equivalent gain. Recall that the continuous slope implies that a loss and a gain will have the same impact at the reference point. Hence, the figure is constrained to be 1 at the reference point. As we move away from the reference point, losses carry a larger impact than gains. At distances from the reference point of around 200 EUR, we find a loss aversion factor of around 2, which is consistent with Tversky and Kahneman (1992).

The amount of loss aversion depends crucially on where we place our reference point. In Figure 2.7, we display the amount of loss aversion for different reference points. If we use a median growth to construct the reference point, we have no evidence of loss aversion. We need to assume that individuals expect earnings increases equivalent to the 60th percentile or higher before replicating the amount of loss aversion from experiments. Again, this is consistent with evidence on people suffering from over-optimism bias.

Figure 2.7: Loss Aversion with Different Reference Points

(a)



Notes: The figure shows the impact of a loss relative to a gain for different changes from the reference point. Three different models are compared based on which reference point they adopt.

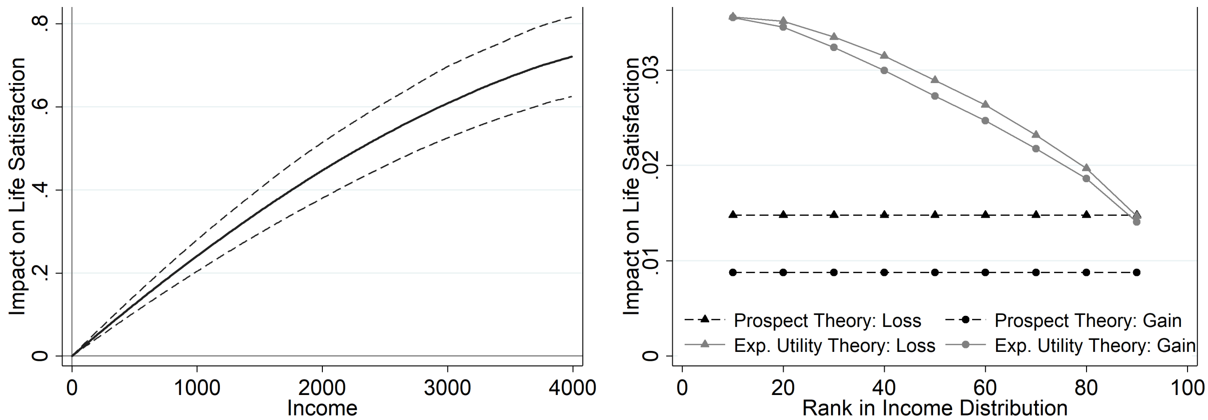
Prospect Theory vs Standard Utility (H3)

Finally, we show whether individuals' life satisfaction is increasing both in the positive distance from their reference point, and in the absolute amount of earnings that they have in line with (**H3**). The previous splines all controlled for a spline of absolute income levels as well. We used a spline to be agnostic about how the level of earnings is transmitted into life satisfaction. The corresponding predictions for the relationship between absolute income levels and life satisfaction are graphed in Figure 2.8(a).

In addition to the previously shown evidence for an S-curve, we also appear to have evidence for a standard utility function in line with **H3**. This finding suggests that expected utility theory and prospect theory both play an independent role in transmitting income into life satisfaction.

Figure 2.8: Testing H3

(a) Expected Utility Theory (b) Impact of Gaining/Losing 133 EUR



Notes: Panel (a) shows the predicted impact of income on life satisfaction from our main specification. Panel (b) shows the predicted impact on life satisfaction from expected utility theory and prospect theory, respectively, when losing/gaining 133 EUR, which is the median distance from the reference point. The expected utility theory curves converge at the top and at the bottom of the distribution since the spline makes a linear prediction as the tails of distributions.

In order to put the magnitudes in perspective, Figure 2.8(b) shows the impact of losing/gaining 133 EUR – the median distance from the reference point – channeled through the standard utility function and the value function, respectively. We show the impact losing/gaining this amount has for individuals at varying income levels. For poorer individuals, diminishing marginal utility kicks in and expected utility theory has a much larger impact on life satisfaction. For the wealthiest 10 percent, the value function plays about the same role as expected utility theory in shaping individuals' life satisfaction. At the median income, the effect is around twice as large for expected utility theory compared to prospect theory.

Thus far we have implicitly assumed that all individuals, wealthy and poor alike, share the same S-curve. This means that a specific deviation from the reference point has the same impact on individuals regardless of their income level. One could imagine that the appropriate specification scales the S-curves according to individuals' income levels. This would imply that the horizontal lines in Figure 2.8(b) would be downwards sloping – a certain change from a reference point will have a larger impact on poorer individuals. Instead of specifying the difference from the reference point in our regressions in terms of levels of income ($y_{it} - r_{it}$), we try specifying differences from the reference point as, respectively, the percentage gap (y_{it}/r_{it}), and the log difference ($ln(y_{it}) - ln(r_{it})$). Although individuals then will have different S-curves, the S-curves will have the same shape but be on different scales. For any given reference point, we can revert the x-axis back to displaying absolute distances from the reference point. In panel (a) and (c) of Figure 2.9 we illustrate the corresponding S-curves for individuals who have the median income as the reference point (which is around 1500 EUR), and in panel (b) and (d) the resulting relative importance of prospect theory and expected utility theory.

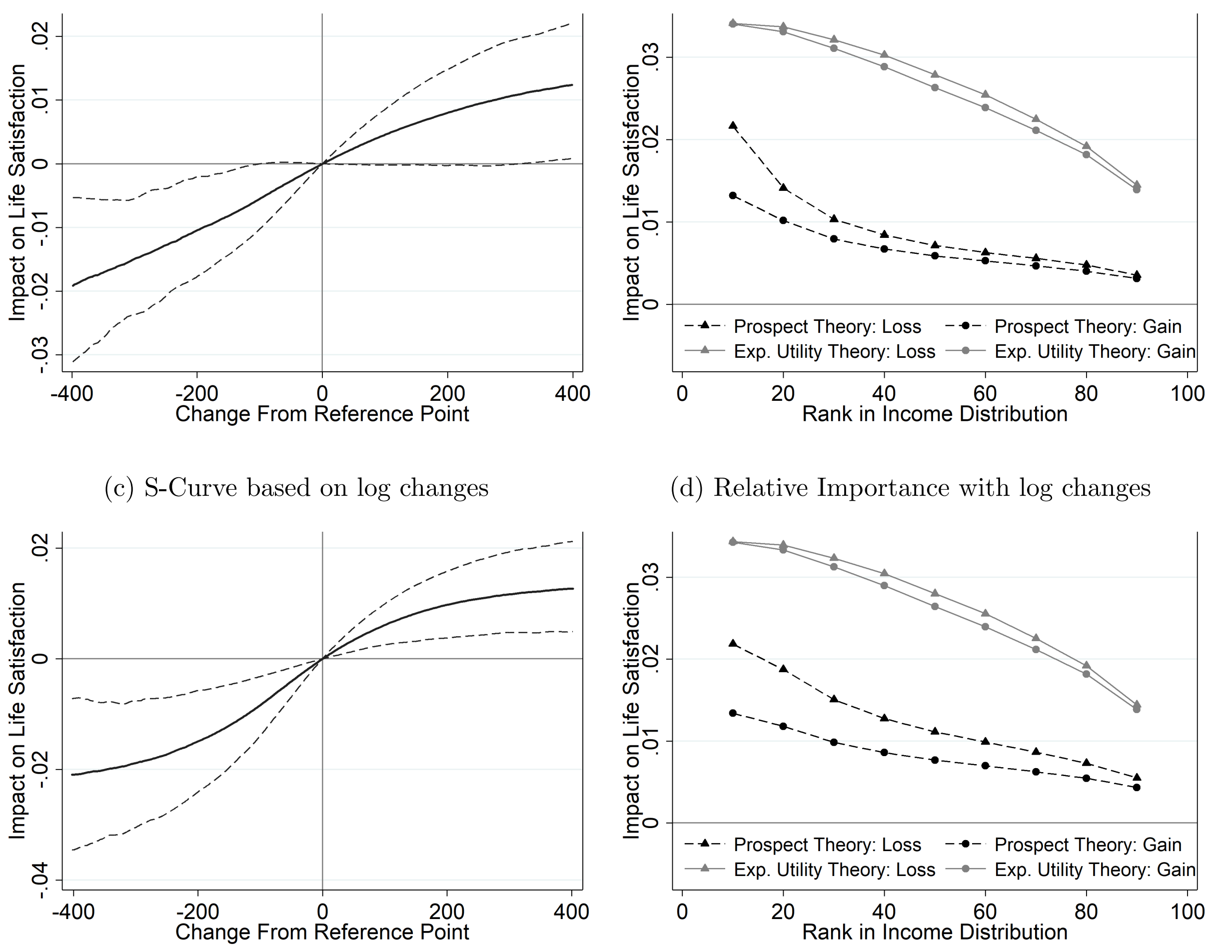
Figure 2.9: Relative Prospect Theory vs. Expected Utility Theory

(a) S-Curve based on pct. changes

(b) Relative Importance with pct. changes

(c) S-Curve based on log changes

(d) Relative Importance with log changes



Notes: Predictions when the distance from the reference point is measured in relative terms (panel (a)) or in log terms (panel (c)) for individuals with a reference point of 1500 EUR (the median income). Panel (b) and panel (d) show the resulting relative importance of prospect theory and expected utility theory.

Notice first that the predictions based on relative changes only show very weak evidence for diminishing sensitivity.[7] With log changes we get an S-curve very similar looking to our main specifications.[8] If each individual has his own S-curve, the impact from prospect theory

[7] This, however, does not hold up in alternative specifications; if we use five knots instead of four knots we get a clear S-curve. We continue using four knots here for the sake of consistency.

[8] Interestingly, if we display predictions without converting the x-variable back to absolute changes from the reference point, and thus use log changes from the reference point as the x-variable, we still get an S-curve.

of gaining/losing a specific amount will depend on the income level of the individual. We show this in panel (b) and (d), where the prospect theory curves now are declining. In these specifications, expected utility theory has about twice the impact of prospect theory at all income levels.

We are unable to distinguish between which of these three models is a best manifestation of the data. It would be interesting for future research to try to investigate if individuals share the same S-curves or if the S-curves should be scaled according to income levels.

2.4.4 Robustness Checks

The results until now were based on a number of assumptions concerning the dependent variable, the income measure, the reference point etc. In this section, we perform a number of alternative specifications to clarify when the main results are robust. Table 2.2 shows an overview of the robustness checks we make. Figures supporting all of these results are presented in the Appendix. We list whether we have evidence for diminishing sensitivity (DS) and loss aversion (LA) separately. We measure loss aversion as the loss aversion ratio for the median distance from the reference point. That is, we find the individual j for whom $F(|y_j - r_j|) = F(d_j) = 0.5$, and report loss aversion as $\frac{-V(-d_j)}{V(d_j)}$. This measure is comparable across different income variables and allows us to study loss aversion even though there is no loss aversion at the reference point. In all cases, unless otherwise stated, we use restricted cubic splines with four knots, use last year's income times the mean income growth as the reference point, and control for our usual set of covariates.

In general, our measure of loss aversion is smaller than 2, which is the level usually found in experimental settings. Aside from the possibility that loss aversion simply is smaller in our set-up, we believe three factors may explain our lower level of loss aversion. Firstly, since we do not have the exact reference point, our loss aversion will be partly attenuated. Secondly, evidence presented later suggests that loss aversion is larger if we use a more flexible spline. Thirdly, individuals may use a more optimistic reference point than the mean income growth. If true, this would generate higher loss aversion. In sum, our empirical specification may result in a conservative estimate of loss aversion.

Dependent Variable

First, we test if the S-curve holds with other subjective well-being variables than life satisfaction. To this end, we utilize a battery of variables in GSOEP where individuals are asked how satisfied they are with certain domains of their life on a scale from 0–10. We use questions on satisfaction with job, personal income, household income, and leisure. In addition, we

Table 2.2: Robustness Checks

Robustness check	**Main specification**	**Alternative specification**	**DS**	**LA**
Dependent variable	Life satisfaction	Personal income satisfaction	Yes	1.50
		Household income satisfaction	Yes	1.62
		Work satisfaction	Yes	1.37
		Happiness	No	1.81
		Leisure satisfaction	No	0.92
Income variable	Monthly net earnings	Monthly gross earnings	Yes	2.13
		Monthly net HH income	Yes	1.61
		Annual gross earnings	Yes	2.16
		Annual net HH income	No	2.08
		Annual gross HH earnings	Yes	1.29
		Annual gross HH income	Yes	1.25
Survey	GSOEP	BHPS, monthly net earnings	Yes	1.24
		BHPS, monthly gross earnings	Yes	1.32
		BHPS, annual gross earnings	Yes	1.19
		BHPS, weekly net HH income	No	0.97
Reference point	Lagged earnings plus average growth in earnings	Predicted earnings from AR(1)	Yes	1.47
		Predicted earnings from AR(2)	Yes	1.28
		Peer mean by region and educ	No	1.19
		Peer mean by region, age, sex, and educ	No	1.45
Accounting for age	Control for age groups in the regression	Control for age through reference point	Yes	1.56
		Control for age both ways	Yes	1.61
		No age controls	Yes	1.62
Transformation of dependent variable	Cardinal from 0-10	log(lifesat/(10-lifesat)) transformation	Yes	1.64
		Remove individuals at boundary	Yes	1.88
Outliers	Excl. 2% in $\lvert y-r \rvert$, top-code 0.5% in y	Excl. 5% in $\lvert y-r \rvert$	Yes	2.00
		Top-code 2% in y	Yes	1.68
Functional form	Spline with 4 knots	Spline with 3 knots	No	1.65
		Spline with 5 knots	Yes	3.25
		Spline with 6 knots	Yes	-
Sample	All except for outliers	Full-time employed	Yes	1.81
		Bottom 50% in income	Yes	1.58
		Top 50% in income	No	6.27
		Men	Yes	1.53
		Women	Yes	1.72
		West Germany	Yes	1.83
		East Germany	Yes	1.30
		1984-1999	Yes	1.57
		2000-2014	Yes	1.62

Notes: All results are based on fixed effects regressions controlling for a spline of income levels, marital status, age groups (unless otherwise stated), work hours, job changes, household composition, and year fixed effects. A restricted cubic spline with 4 knots is used in all cases (unless otherwise stated). *DS* indicates whether there is evidence for diminishing sensitivity, *LA* reports the median loss aversion.

use a question on how often individuals have felt happy in the past four weeks. The answer categories to this question are very rarely, rarely, occasionally, often, and very often. The results when using these measures as the dependent variable are given in Figure 2A.1.

The results using personal income satisfaction, household income satisfaction, and work satisfaction are quite similar to using life satisfaction. In fact, the work satisfaction results generate more narrow confidence bans. This is not surprising since earnings changes are more important when people evaluate satisfaction with their job rather than with life as a whole. When we use leisure satisfaction we find no evidence for prospect theory. In fact, there seems to be a negative relationship between leisure satisfaction and distance from the reference point. This suggests that incomes above the reference point primarily stem from increased work intensity.

When the question on happiness is used, we get quite wide confidence bans indicating that this variable is noisier. Although the point estimates suggest a weak S-curve, we do not have the statistical power to support loss aversion nor diminishing sensitivity.

Income Variable

Next, we test the robustness of our result to using other income measures. The main results were based on monthly self-reported net earnings expressed in constant 2010 EUR. We try six other income measures. First, we use self-reported *gross* earnings and self-reported monthly net *household* income rather than net individual earnings. The latter variable includes all income sources so now we also include unemployed and people outside the labor market. As shown in Figure 2A.2, we have evidence for both diminishing sensitivity and loss aversion.

Next, we use income variables from the Cross National Equivalent File, which contains yearly income measures, mostly at the household level. We use gross individual/household earnings and gross/net total household income. Since the incomes are from the year prior to the self-reported life satisfaction, these variables make a great test for whether our main results were driven by reversed causality. The vast majority of individuals are interviewed in the first months of the year implying that the time gap from the income being earned to the life satisfaction answer is not too large. Results using these income measures are also given in Figure 2A.2. We find evidence for loss aversion in all cases, and evidence for diminishing sensitivity in all measures but annual net household income.

Survey

We also try to see if our results replicate using the British Household Panel Survey (BHPS). The BHPS is an annual survey that ran from 1991 to 2008. The sample we use contains around 143,000 observations spread across 24,000 individuals. The life satisfaction question is phrased slightly differently in BHPS. Respondents are asked *"How dissatisfied or satisfied are you with your life overall?"* on a scale from 1 (not satisfied at all) to 7 (completely satisfied). The BHPS has several income variables. Here we use four different measures. Results are displayed in Figure 2A.3.

First, we use monthly net earnings, which is similar to our baseline specification with the German Socio-Economic Panel. We have evidence for both diminishing sensitivity and loss aversion. This does not change if we use monthly gross earnings or annual gross household earnings rather than net monthly earnings. Only when we use weekly net household income, which is a derived variable, we do not have evidence for an S-curve. This may be because what really matters is individuals' perceptions of their income and not their actual income. In general, the results using the BHPS suggest rather small loss aversion. One explanation for this may be that Brits have higher expectations than Germans. If Brits expect their earnings increase to be around the 75th percentile (such that 3 in 4 get lower earnings increase) then loss aversion is at the German level. It may also be, of course, that loss aversion simply is smaller in the British sample.

Reference Point

We also vary the reference point. First we predict each individual's income using AR(1) and AR(2) models. We consider the predicted income as the reference point. Hence, incomes larger than predicted are considered positive deviations from the reference point and vice versa. Findings are given in Figure 2A.4. In both models, we find evidence for both diminishing sensitivity and loss aversion. These results are consistent with the idea that the reference point is the rational expectation of future earnings.

We also try using mean peer earnings as a reference point, which is what Vendrik and Woltjer (2007) used in their analysis. We construct two different mean peer earnings measures: i) by region and education level (three categories), and ii) by region, education, gender, and 10-year age groups. Although life satisfaction is increasing in the difference from mean peer earnings, the relationship between life satisfaction and these reference points shows little evidence for diminishing sensitivity. Similar to Vendrik and Woltjer (2007), we therefore do not find evidence for the predictions from prospect theory when using peer earnings as the reference point. This could be because prospect theory does not apply to this reference point, or because the reference groups we can create based on the survey do not capture whom individuals actually compare themselves with. It is plausible that individuals use specific colleagues, friends or family members as their reference group, which these broad measures have a hard time capturing.

Accounting for age

We also try to account for age differently. Instead of controlling for age groups in the regression, we can account for age groups directly in the reference point, by assuming that individuals' reference point are given by last year's income times the average growth rate *within age groups*. As shown in 2A.5, results are unchanged. We can also try to account for age both ways; both through the reference point and in the regression. Again, results are broadly unchanged. Finally, it may be the case that individuals do not specify reference points based on the their age, in which case accounting for age one way or the another will remove variation that the value function ideally should pick up. If we do not control for age in any way, we still get an S-curve.

Transformation of Dependent Variable

A possible concern is that diminishing sensitivity arises mechanically due to floor and ceiling effects. No matter the income change, individuals cannot report life satisfaction levels below 0 or greater than 10. We test whether this is driving the results in three ways. Firstly, we transform the dependent variable such that our new dependent variable equals $log(\frac{lifesat}{10-lifesat})$. With this transformation there is further between life satisfaction levels close to 0 and 10 and closer between life satisfaction levels close to 5. Hence, more weight is attached to changes close to the boundaries. As shown in Figure 2A.6, this does not change the results.

Secondly, we try to deal with the boundedness concern more directly by deleting observations where life satisfaction is reported to be 0, 1, 9, or 10. As also shown in in Figure 2A.6, this too does not change our findings.

A third way in which we could see if our results are driven by floor and ceiling effects is to regress life satisfaction on other variables, which we know from prior research are positively correlated with life satisfaction. If the boundedness is a concern, we should see S-curves also in these regressions. We have already demonstrated that this is not the case when we used peer income as the reference point, and when we showed the spline relating income levels to life satisfaction. Hence, we find it unlikely that the S-curve is driven by floor and ceiling effects.

Outliers

Next, we try to change the rules governing outliers that we applied in the main analysis. In the main analysis, we top-coded the income variable at the 99.5th percentile and removed individuals who had the 2% largest deviations from the reference point (1% in each direction). First, we try to remove the 5% with the largest deviation instead of 2%. Second, we top-code the income variable at 2% instead of 0.5%. Results are displayed in Fig 2A.7. In both cases, our S-curve remains largely unchanged.

Functional form

We also try to test if the functional form we chose to represent the S-curve matters. We try using 3, 5, or 6 knots in the spline instead of 4 knots. Recall that the more knots we use the more flexibility we allow at the cost of greater imprecision. Results are shown in Figure 2A.8. Results look broadly similar with 5 and 6 knots, albeit with greater confidence intervals and with larger loss aversion. With 3 knots, we get what looks like a linear function. This illustrates that we need to allow for a relatively large degree of flexibility before the S-curve becomes visible.

Sample

Finally, we try to divide the sample according to a number of characteristics to see if any particular subsample is driving the results. We first look at only full-time employed. Although we have already excluded unemployed and individuals outside the labor market, only looking at full-time employed may reassure that results are not driven by individuals who were only marginally attached to the labor market. Figure 2A.9 shows that the point estimates at the center of the distribution are largely unchanged when only using full-time employed. Due to large confidence bands, however, we are unable to nail down a precise S.

Next, we divide the sample into two depending on whether they are in the top half or bottom half of the income distribution, by gender, region (east/west), and survey year (before/after 2000). As shown in Figures 2A.9 and 2A.10, we find S-curves in all cases except for the wealthiest 50%, where confidence bands are too wide to say anything conclusive. We find strongest loss aversion for women and West Germany.

In sum, we find that our results our robust to most specifications. However, the amount of loss aversion changes quite a bit depending on the exact specification.

2.5 Conclusion

Prospect theory is one of the most canonical results in behavioral economics in the past century. The theory holds that individuals derive value from changes in their income with respect to a reference point. The theory further postulates that individuals display loss aversion, meaning that losses are valued more heavily than gains, and diminishing sensitivity, meaning that large changes from the reference point have diminishing marginal impacts. Together, these properties generate an S-curve.

In this paper, we provided a comprehensive test for the S-curve using data on life satisfaction from the German Socio-Economic Panel. Rather than considering choices under risk, we looked at non-experimental realized outcomes. We used experienced utility rather than decision utility as the measure of value and data on last year's earnings to generate a reference point. Our findings revealed an S-curve similar to experimental evidence, with support for both diminishing sensitivity and loss aversion. Our main results are robust to a number of alterations, including using other subjective well-being variables, applying different definitions of income, and using other reference points.

Throughout the analysis, we controlled for a number of covariates as well as for the level of income. This allowed to us test the relative importance of prospect theory next to expected utility theory. Our baseline estimates suggest that for the median earner, expected utility theory is about twice as important as prospect theory. For wealthy individuals, however, diminishing marginal utility kicks in, and prospect theory plays as large a role as expected utility theory.

To our knowledge, we are the first to find that the predictions from prospect theory with respect to diminishing sensitivity hold with life satisfaction outside of an experiment. This result gives support to applying the value function in policy analyses as a significant source of well-being.

Bibliography

ABELER, J., FALK, A., GOETTE, L. and HUFFMAN, D. (2011). Reference points and effort provision. *The American Economic Review*, **101** (2), 470–492. (page 25)

BARBERIS, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *The Journal of Economic Perspectives*, **27** (1), 173–195. (page 17, 25)

BLANCHFLOWER, D. G. and OSWALD, A. J. (2008). Is well-being u-shaped over the life cycle? *Social Science & Medicine*, **66**, 1733–1749. (page 24)

BOWMAN, D., MINEHART, D. and RABIN, M. (1999). Loss aversion in a consumption–savings model. *Journal of Economic Behavior & Organization*, **38** (2), 155–178. (page 21)

BOYCE, C. J., WOOD, A. M., BANKS, J., CLARK, A. E. and BROWN, G. D. (2013). Money, well-being, and loss aversion: Does an income loss have a greater effect on well-being than an equivalent income gain? *Psychological science*, pp. 2557–2562. (page 19)

CARTER, S. and MCBRIDE, M. (2013). Experienced utility versus decision utility: Putting the 's' in satisfaction. *The Journal of Socio-Economics*, **42**, 13–23. (page 19, 24)

DE NEVE, J.-E., WARD, G. W., DE KEULENAER, F., VAN LANDEGHEM, B., KAVETSOS, G. and NORTON, M. I. (forthcoming). Individual experience of positive and negative growth: Global evidence using subjective well-being data. *The Review of Economics & Statistics*. (page 19)

FERRER-I CARBONELL, A. (2005). Income and well-being: An empirical analysis of the comparison income effect. *Journal of Public Economics*, **89** (5), 997–1019. (page 19)

FREDERICK, S. and LOEWENSTEIN, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener and N. Schwarz (eds.), *Well-being: Foundations of hedonic psychology*, Russell Sage Foundation. (page 28)

GALANTER, E. (1990). Utility functions for nonmonetary events. *The American Journal of Psychology*, pp. 449–470. (page 19)

HARRELL, F. E. J. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer. (page 23)

HARRISON, G. W. and RUTSTRÖM, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, **12** (2), 133–158. (page 18)

KAHNEMAN, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American psychologist*, **58** (9), 697. (page 20, 21, 25)

— and TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, pp. 263–291. (page 17, 18, 20, 21, 24)

KIMBALL, M., NUNN, R. and SILVERMAN, D. (2015). Accounting for adaptation in the economics of happiness. *NBER Working Paper*. (page 28)

KÖBBERLING, V. and WAKKER, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, **122** (1), 119–131. (page 21)

KŐSZEGI, B. and RABIN, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, pp. 1133–1165. (page 18, 21, 22, 25)

LAMELA, M. L. (2015). *Essays on the role of social status and beliefs on intergenerational mobility*. Ph.D. thesis. (page 19)

LOEWENSTEIN, G. and UBEL, P. A. (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*, **92** (8), 1795–1810. (page 28)

LUTTMER, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly Journal of Economics*, pp. 963–1002. (page 19)

RIIS, J., LOEWENSTEIN, G., BARON, J., JEPSON, C., FAGERLIN, A. and UBEL, P. A. (2005). Ignorance of hedonic adaptation to hemodialysis: A study using ecological momentary assessment. *Journal of Experimental Psychology: General*, **134** (1), 3. (page 28)

SCHWANDT, H. (2016). Unmet aspirations as an explanation for the age u-shape in wellbeing. *Journal of Economic Behavior & Organization*, **122**, 75–87. (page 24)

SHAROT, T. (2011). The optimism bias. *Current biology*, **21** (23), R941–R945. (page 26)

SHLEIFER, A. (2012). Psychologists at the gate: A review of Daniel Kahneman's thinking, fast and slow. *Journal of Economic Literature*, **50** (4), 1080–1091. (page 17)

TVERSKY, A. and KAHNEMAN, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, pp. 1039–1061. (page 21)

— and — (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, **5** (4), 297–323. (page 21, 31)

Ubel, P. A., Loewenstein, G. and Jepson, C. (2005). Disability and sunshine: Can hedonic predictions be improved by drawing attention to focusing illusions or emotional adaptation? *Journal of Experimental Psychology: Applied*, **11** (2), 111–123. (page 28)

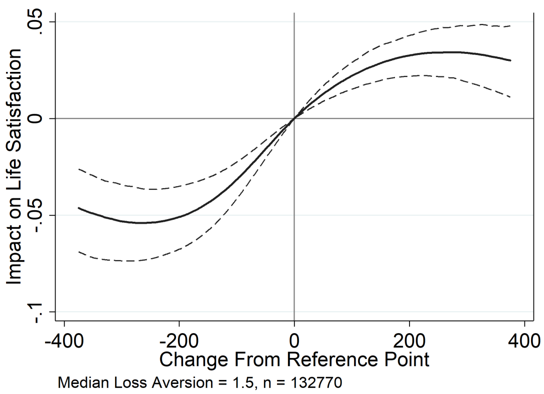
Van Praag, B. M. and Frijters, P. (1999). The measurement of welfare and well-being: The leyden approach. *Well-Being: Foundations of Hedonic Psychology*, p. 413. (page 26)

Vendrik, M. C. and Woltjer, G. B. (2007). Happiness and loss aversion: Is utility concave or convex in relative income? *Journal of Public Economics*, **91** (7), 1423–1448. (page 19, 24, 37)

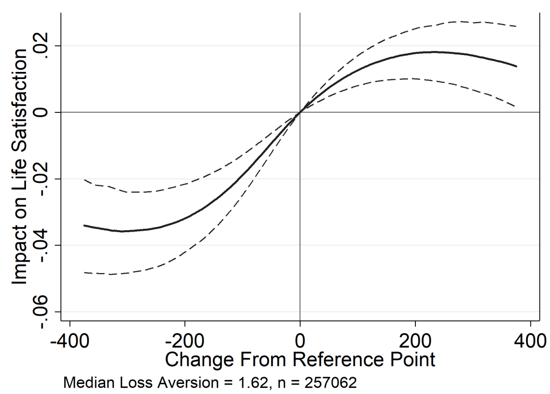
2.A Appendix

Figure 2A.1: Changing the Dependent Variable

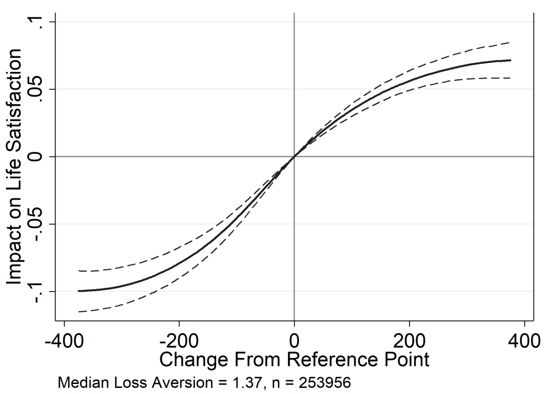
(a) Personal Income Satisfaction



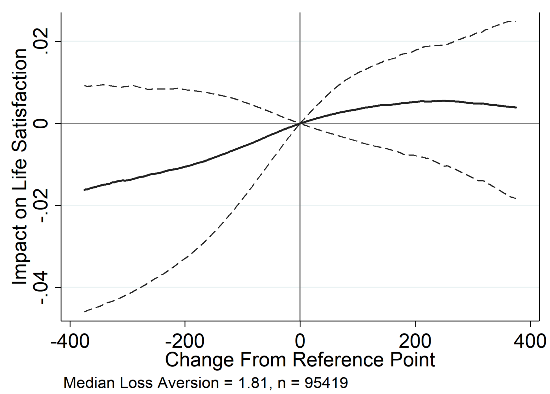
(b) Household Income Satisfaction



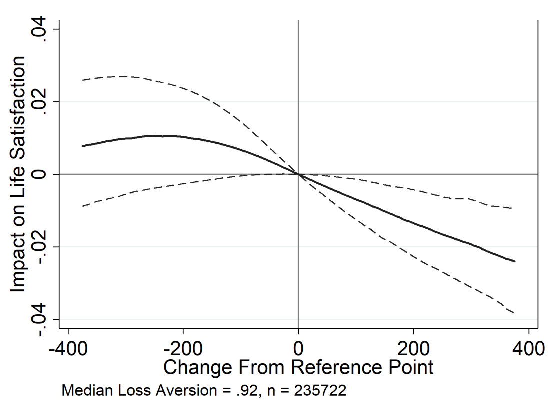
(c) Work Satisfaction



(d) Happiness

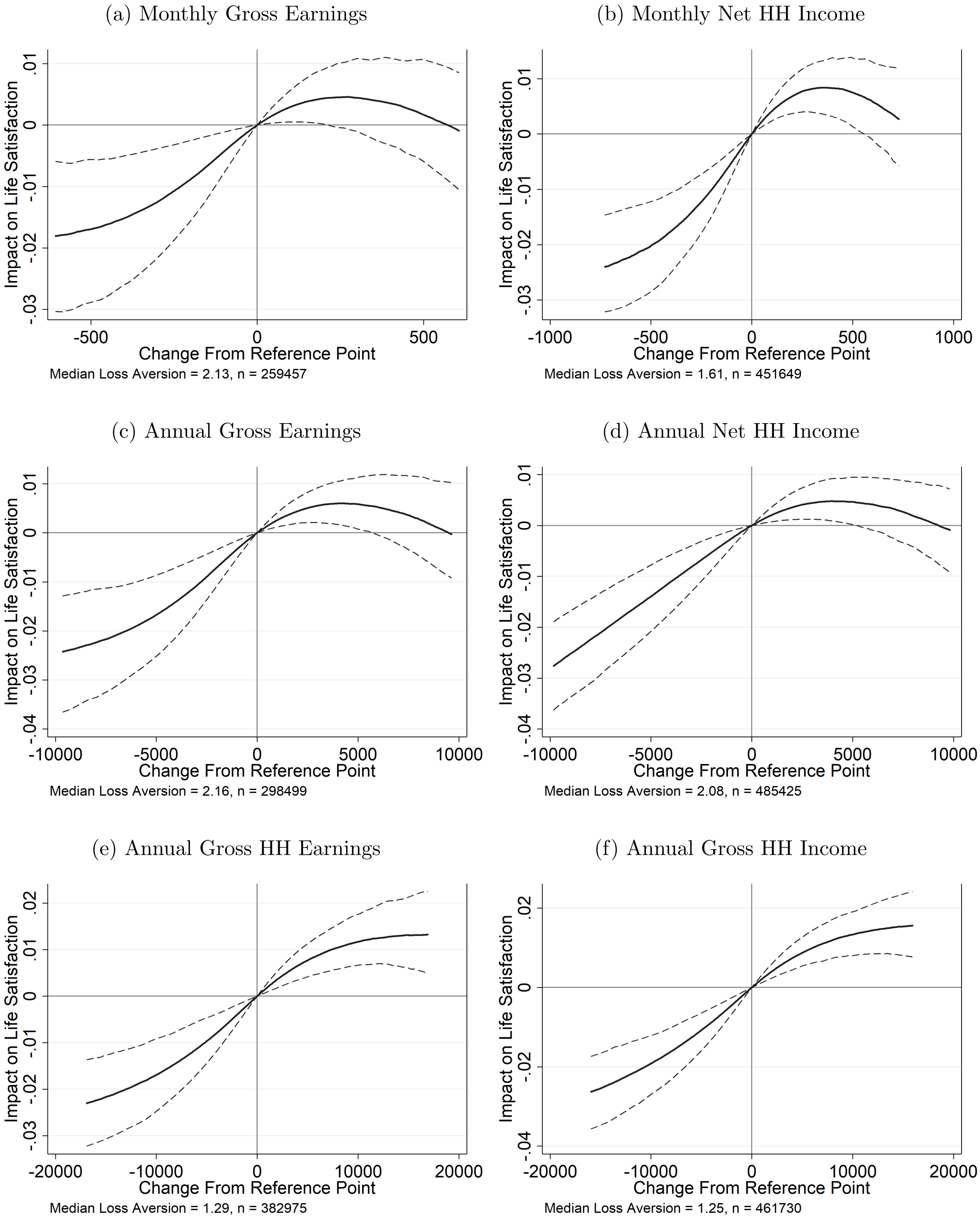


(e) Leisure



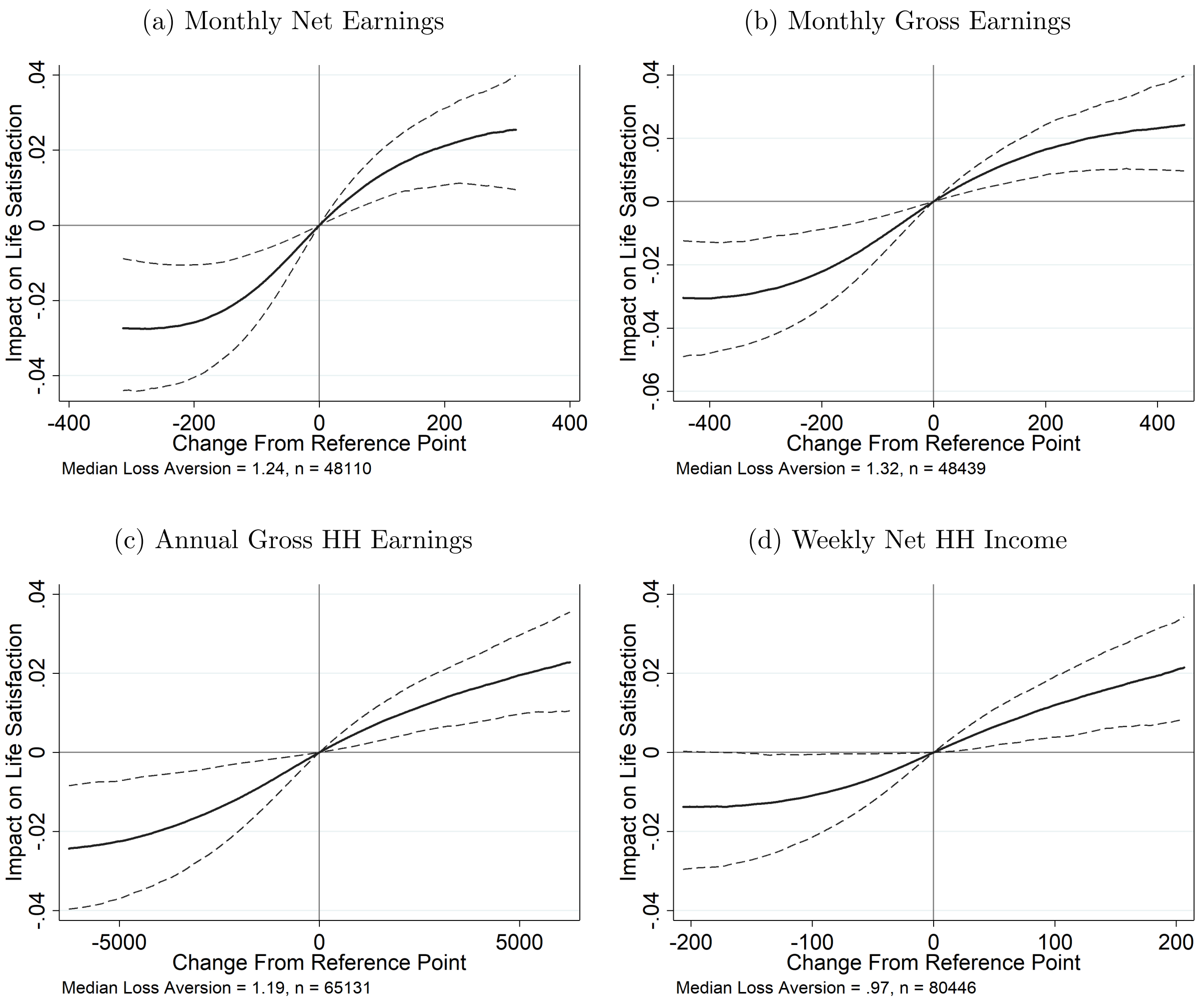
Notes: All variables range from 0-10. Happiness is the answer to how often individuals have felt happy in the past four weeks. The answer categories to this question are very rarely, rarely, occasionally, often, and very often (coded 0, 2.5, 5, 7.5, 10, such that the range is comparable to the domain satisfaction questions).

Figure 2A.2: Changing the Income Measure



Notes: Predicted values from fixed effects regressions of life satisfaction on changes from the reference point. The bottom four figures use income in the year prior to the survey. Panel (b), (d), and (f) use the entire GSOEP sample including unemployed and people outside the labor market.

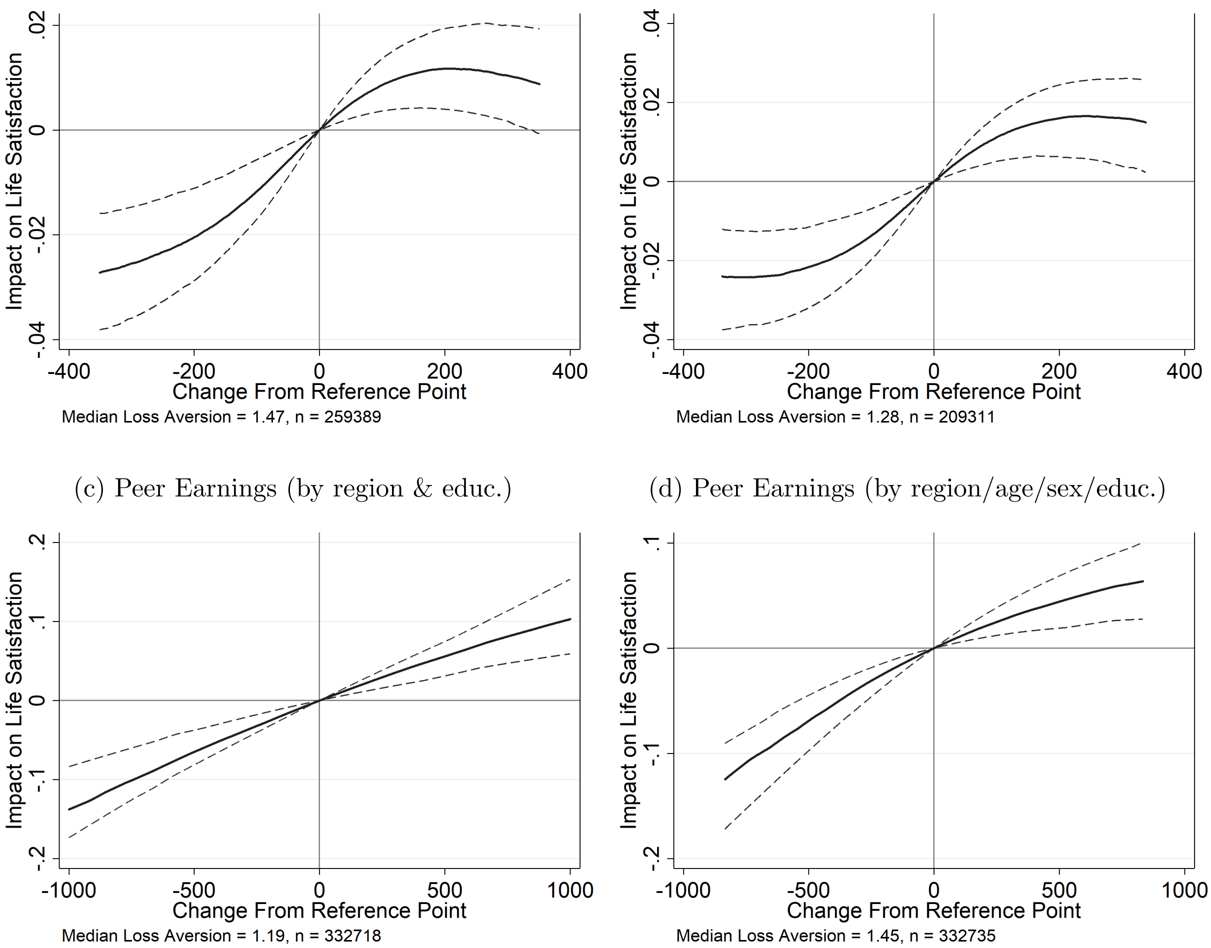
Figure 2A.3: Changing the Survey: BHPS



Notes: Predicted values from fixed effects regression of life satisfaction on changes from the reference point. The four income variables are from the British Household Panel Survey.

Figure 2A.4: Changing the Reference Point

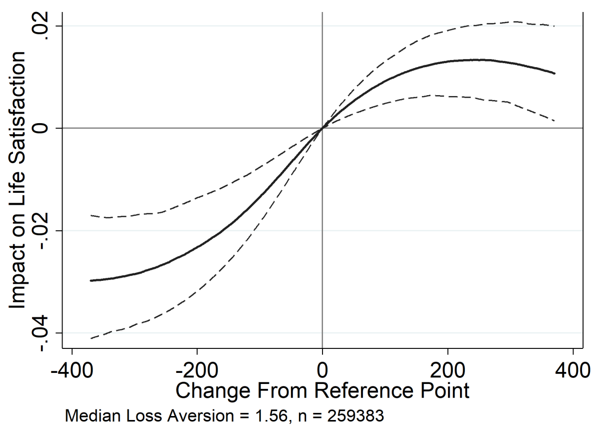
(a) AR(1) | (b) AR(2) | (c) Peer Earnings (by region & educ.) | (d) Peer Earnings (by region/age/sex/educ.)



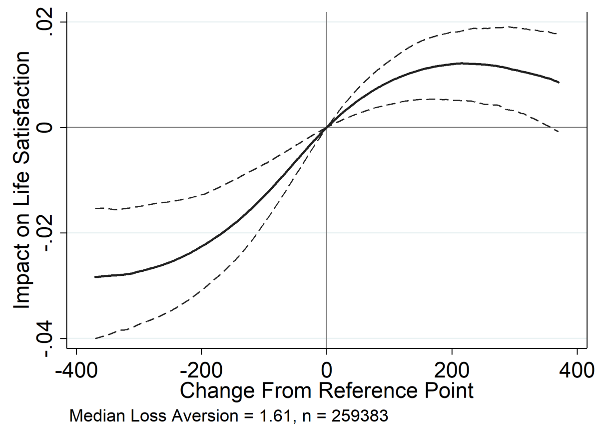
Notes: Predicted values from regressing life satisfaction on changes from the reference point using a variety of different reference points. The bottom two panels use the mean earnings of a peer group as a reference point. Education is split into three categories and age groups are 10-year intervals.

Figure 2A.5: Accounting for Age

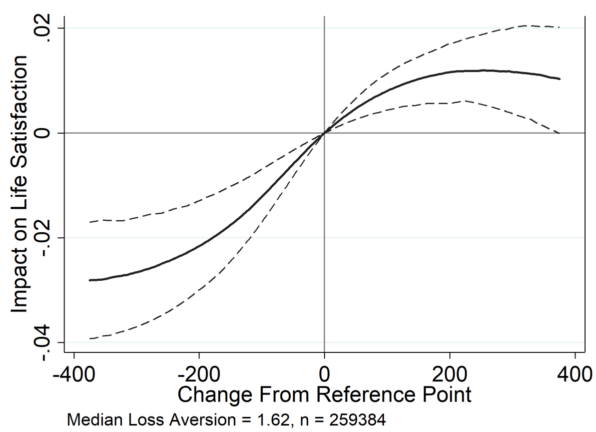
(a) Through Reference Point



(b) Through Ref. Point and in Regression



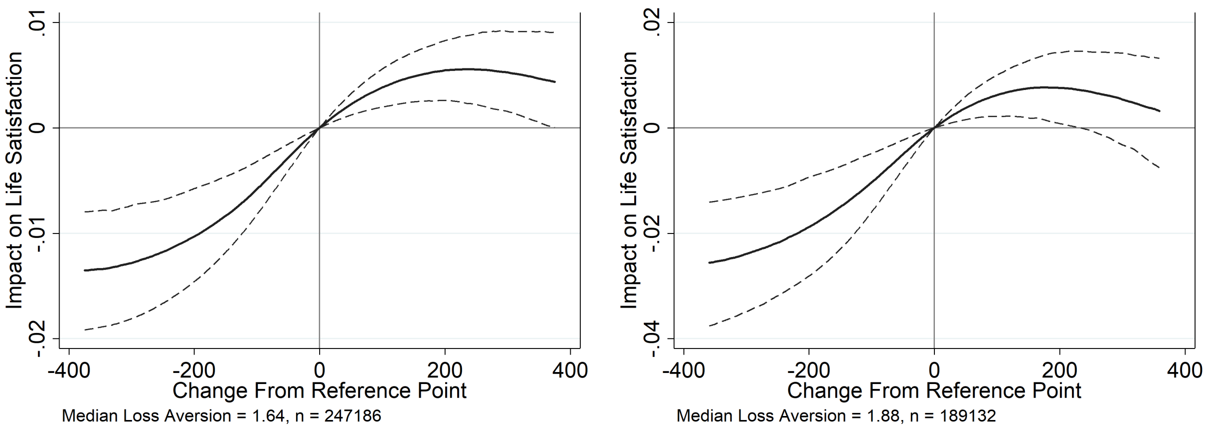
(c) No Age Control



Notes: Predicted values from regressing life satisfaction on changes from the reference point using different ways of accounting for age. Panel (a) assumes that the reference point is last year's income times the average growth rate *within age groups*. Panel (b) mimics the specification in panel (a) but in addition also controls for age groups. Panel (c) does not take age into account in any way.

Figure 2A.6: Transformations of the Dependent Variable

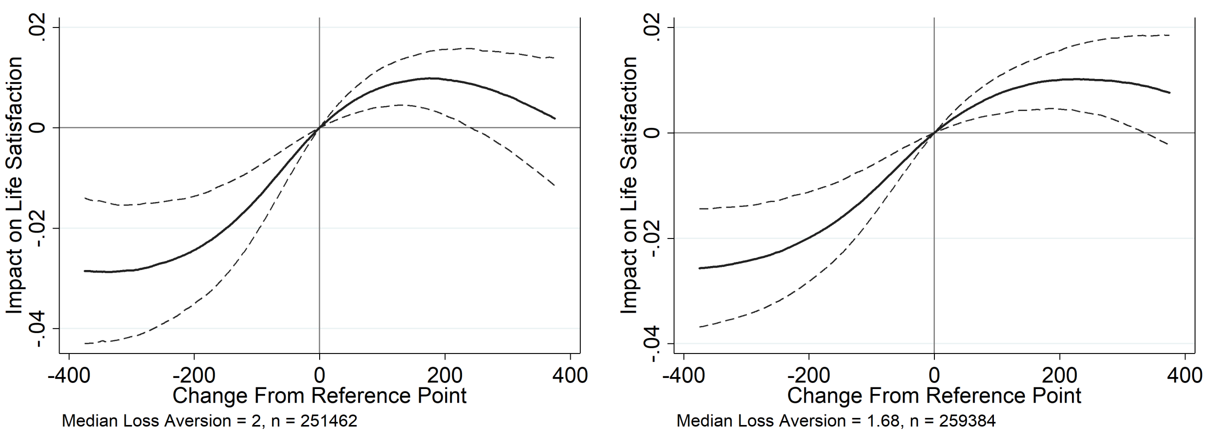
(a) y=log(lifesat/(10-lifesat)) (b) Removing 0-1 & 9-10 in Life Sat.



Notes: Panel (a) transforms the left-hand side variable to equal $log(\frac{lifesat}{10-lifesat})$ (individuals with 0 or 10 in life satisfaction are discarded). Panel (b) removes individuals who report 0, 1, 9, or 10 in life satisfaction.

Figure 2A.7: Outliers

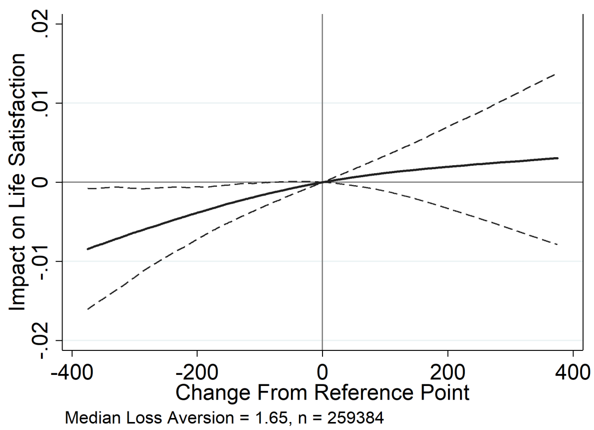
(a) Drops 5% Largest Changes from Ref. Point (b) Top-codes 2% Largest Incomes



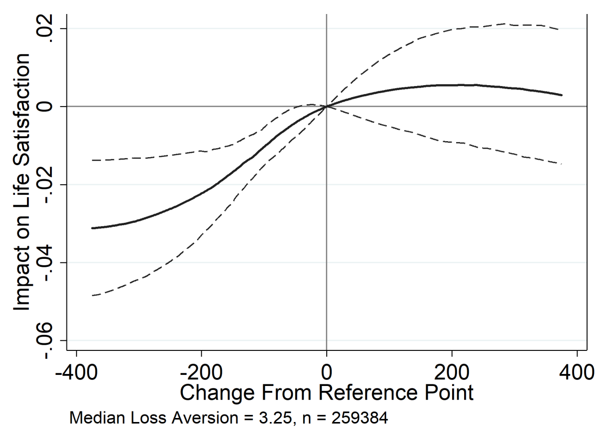
Notes: Predicted values using different rules for how to deal with outliers.

Figure 2A.8: Functional Form

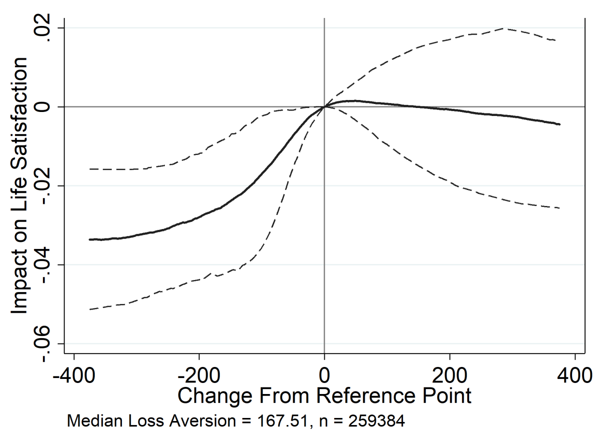
(a) Spline with 3 Knots



(b) Spline with 5 Knots



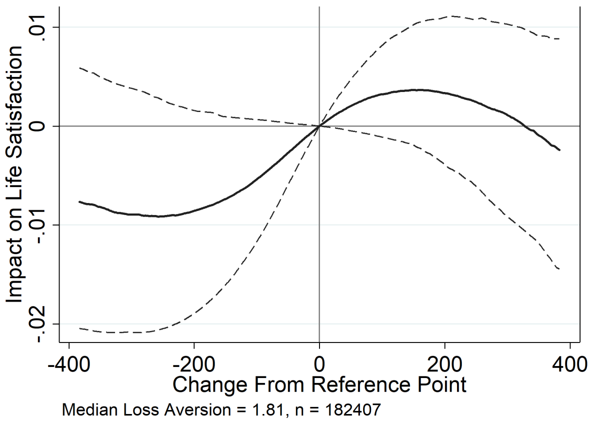
(c) Spline with 6 Knots



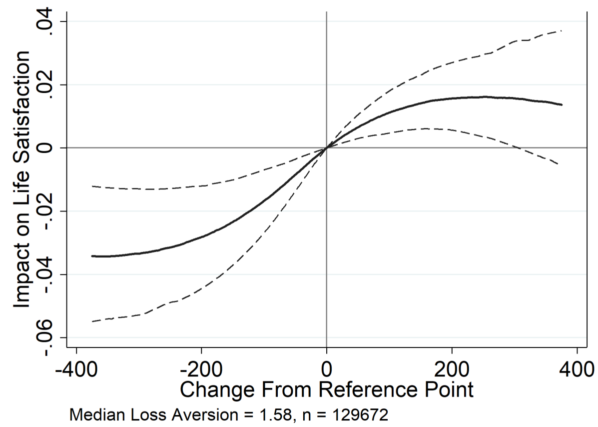
Notes: Predicted values using a different functional form than our 4-knotted spline.

Figure 2A.9: Changing the Sample 1/2

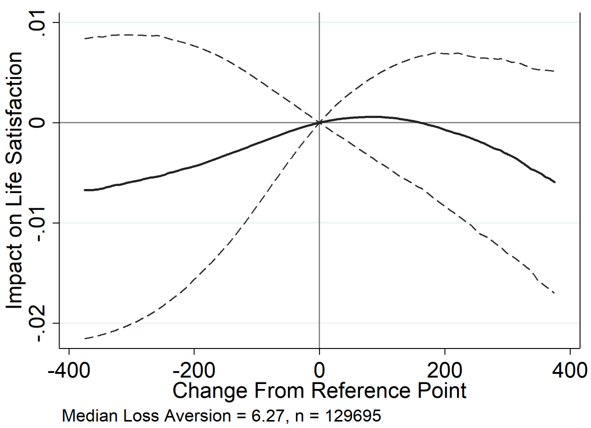
(a) Full time workers



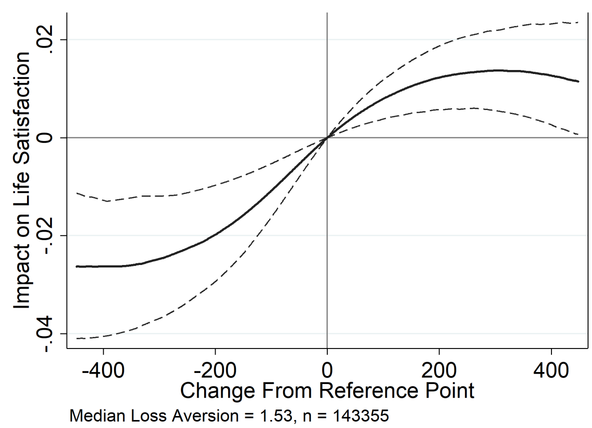
(b) Bottom 50%



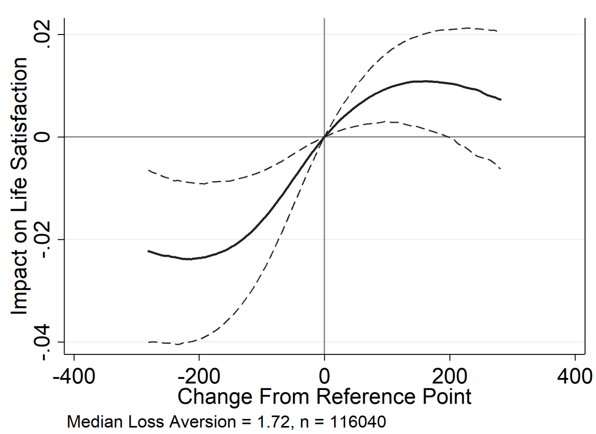
(c) Top 50%



(d) Men

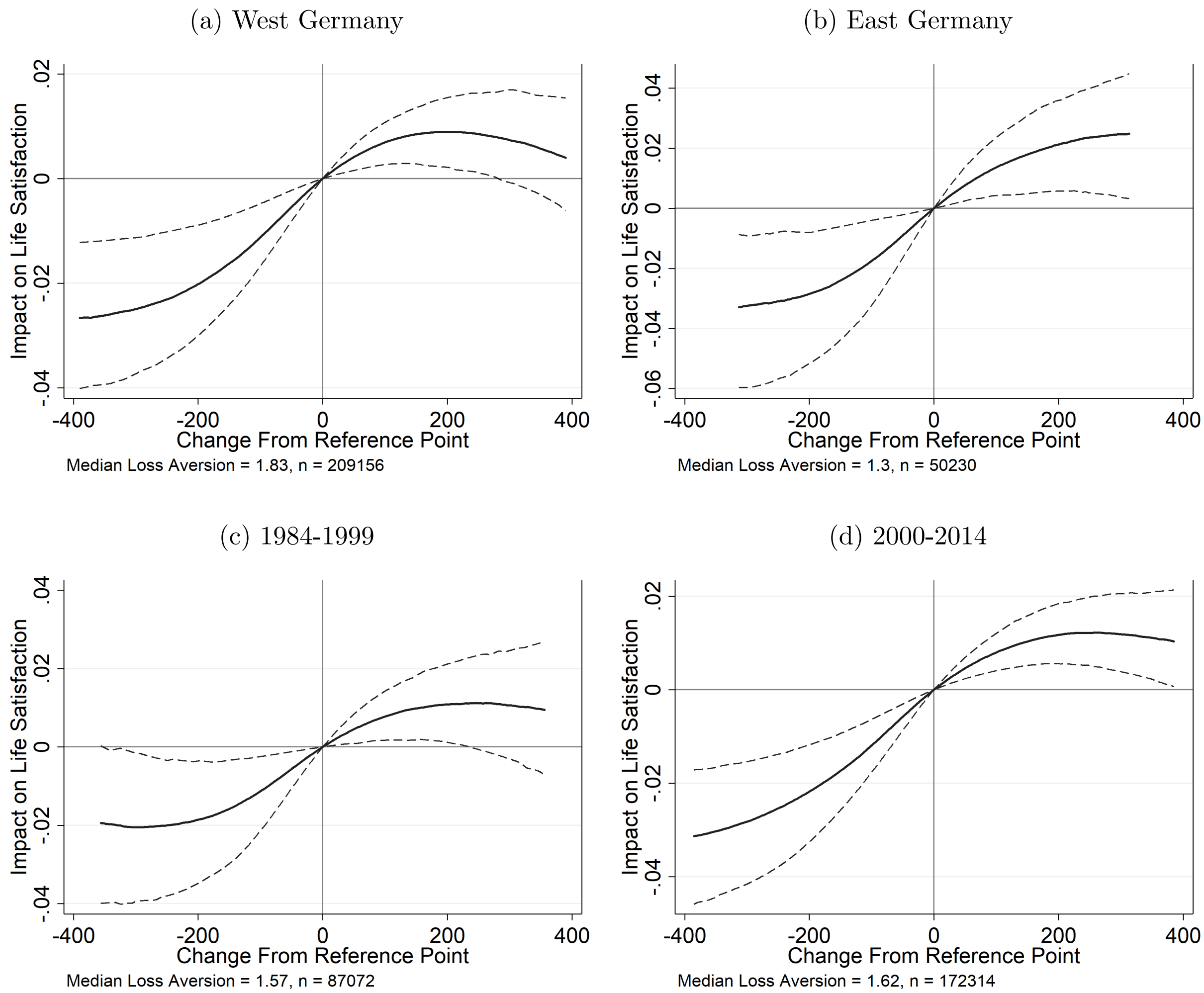


(e) Women



Notes: Predicted values using only parts of the sample.

Figure 2A.10: Changing the Sample 2/2



Notes: Predicted values using only parts of the sample.

Chapter 3

Equality of Opportunity for Well-Being

Equality of Opportunity for Well-Being[1]

Daniel Gerszon Mahler
University of Copenhagen

Xavier Ramos
Universitat Autònoma de Barcelona

Abstract

A growing literature has tried to measure the extent to which individuals have equal opportunities to acquire income. At the same time, policy makers have doubled down on efforts to go beyond income when measuring well-being. We attempt to bridge these two areas by measuring the extent to which individuals have equal opportunities to achieve a high level of well-being. We use the German Socio-Economic Panel to measure well-being in four different ways including incomes. This makes it possible to determine if the way well-being is measured matters for identifying who the opportunity-deprived are and for tracking inequality of opportunity over time. We find that, regardless of how well-being is measured, the same people are opportunity-deprived and equality of opportunity has improved over the past 20 years. This suggests that going beyond income has little relevance if the objective is to provide equal opportunities.

Keywords: Equality of opportunity, measurement, responsibility, well-being
JEL codes: D3, D63, I31

[1]We are very grateful to Brice Magdalou for comments and advice on the use of norm-based inequality metrics, to Dirk Van de gaer and Paul Hufe for fruitful discussions and comments on several parts of the paper, to participants of the Seventh ECINEQ meeting in New York, the 2017 IARIW-Bank of Korea Conference "Beyond GDP: Past Experiences and Future Challenges in the Measurement of Economic Well-being", and the 2017 Winter School on Inequality and Social Welfare Theory.

3.1 Introduction

The notion that individuals ought to have equal opportunities in life is popular among politicians, the general public, and philosophers alike. A sizable number of empirical studies have been carried out analyzing the extent to which individuals have equal opportunities for income acquisition (see Ramos and Van de gaer (2016), Roemer and Trannoy (2015), and Ferreira and Peragine (2016) for recent reviews). These studies are based on the idea that when evaluating the progress of societies, looking at the level and distribution of incomes provides an incomplete picture. A distinction has to be made between income differences arising from factors individuals ought to be held personally responsible for and income differences arising from factors outside the realm of personal responsibility.

In recent years, there has been a growing interest in going beyond income to measure individual well-being (Fitoussi *et al.*, 2010). Well-being (or welfare, we use the two interchangeably) is inherently multidimensional, and growth and income statistics fail to capture this multiplicity. Given the growing interest in going beyond income, it seems pertinent to apply this discussion to the equality of opportunity framework. If individuals ought to have equal opportunities for well-being, then using income as the acquisition variable in equality of opportunity studies could be problematic. Incomes ignore the disutility of effort, the well-being individuals receive from other dimensions of life, and the differences in preferences over income and these other dimensions. Indeed, the philosophers who advocate for equality of opportunity do not advocate for equality of opportunity for income acquisition, but rather something broader than income such as welfare or advantage (Arneson, 1989; Cohen, 1990).

Once it is acknowledged that incomes are not sufficient for measuring well-being, the door opens for many alternatives. Which other well-being dimensions are necessary? Should we measure these dimensions separately or somehow aggregate them to a single number? How can we incorporate the fact that individuals have different preferences over these various dimensions? Should we try to measure well-being directly by alluding to self-reported happiness levels?

We use 30 years of data from the German Socio-Economic Panel (GSOEP) to measure welfare in four ways; with incomes, life satisfaction, a multidimensional index, and equivalent incomes. We use incomes to facilitate comparisons with the generic way of measuring equality of opportunity. The other three measures have roots in different philosophical theories about well-being (Parfit, 1984; Griffin, 1986). Life satisfaction explicitly tries to measure mental states, the multidimensional index defines and aggregates an objective list of dimensions of importance for well-being, and equivalent incomes incorporate preference heterogeneity.

We will investigate if the measurement of welfare matters for 1) characterizing the opportunity-deprived and 2) tracking inequality of opportunity over time. In both cases, we first convert our welfare measures into welfare ranks. This minimizes our reliance on cardinality and assures that the different distributions the welfare measures follow do not drive the results. Next, we regress the welfare ranks on a set of effort and circumstance variables, or equivalently, variables we hold individuals responsible for and variables we do not hold individuals responsible for. Based on these regressions, we assign each individual a "fair" rank and an "unfair" rank. The fair ranks order individuals according to how much their effort variables contribute to their welfare, while the unfair ranks order individuals according to how much their circumstances contribute to their welfare. The latter ranks can be interpreted as an ordinal measure of individuals' opportunities, where the highest ranked possesses the best combination of circumstances and vice versa. With this procedure, for any given welfare measure, an individual will have a welfare rank, a fair rank, and an unfair rank.

In order to answer our first research question, if the measure of welfare matters for characterizing the opportunity-deprived, we compare the unfairness ranks across the four well-being measures. Broadly speaking, if individuals have similar unfairness ranks across the well-being measures, then a characterization of the opportunity-deprived will not depend on how we measure welfare.

In order to answer our second research question, if the measure of welfare matters for tracking inequality of opportunity over time, we use the norm-based approach (see Ramos and Van de gaer (2016) for more information on this approach). In our set-up, this implies using divergence measures to compare the divergence between the welfare ranks and either the fair ranks or the unfair ranks. If the welfare ranks and the fair ranks are highly related, then the individuals exerting the most effort also have the highest well-being, and there is a low presence of inequality of opportunity. If the welfare ranks and unfair ranks are highly related, then the individuals with the best circumstances also have the best outcomes, and there is a high presence of inequality of opportunity. In this connection, we introduce a new simple summary statistic of equality of opportunity; the correlation between the welfare ranks and the unfairness ranks. If we consider income to be the outcome variable of interest and consider parental income as the only circumstance variable, then this summary statistic boils down to the Spearman's correlation between parents' and childrens' incomes. Hence, it generalizes a frequently used measure of intergenerational mobility.

We find that the measure of well-being matters little, both with regards to characterizing the opportunity-deprived and with regards to tracking inequality of opportunity over time. In particular, we find that regardless of how welfare is measured, inequality of opportunity has decreased in the last two decades. These results are robust to using different divergence measures and, for the most part, to changing what we hold individuals responsible for. This is encouraging news for policymakers interested in providing equal opportunities while going beyond income, as they may broadly get things right if they proxy well-being with income.

To our knowledge, this is the first paper to explicitly address the Beyond GDP agenda in a Roemerian equality of opportunity framework. We are certainly not the first, however, to relate notions of fairness with welfare measurement. Fleurbaey and Maniquet (2011) summarize extensive work on this topic. This prior work generally incorporates concerns about fairness directly into the well-being measure. Our approach, in contrast, first computes measures of welfare and then analyzes the extent to which factors beyond individual control are driving the welfare differences. A particularly relevant paper for our approach is Ravallion (2017), which incorporates the disutility of effort into estimates of inequality of opportunity. We go in a different direction by analyzing whether the concept of welfare matters for estimates of inequality of opportunity. In previous studies, the measurement of welfare has been shown to matter for assessments of how average welfare has developed over time (Blanchflower and Oswald, 2004), for how inequality in welfare has developed over time (Stevenson and Wolfers, 2008), and for identifying who the most welfare-deprived are (Decancq and Neumann, 2016).

Roemer (2012) explicitly argues against using welfare as the outcome variable in equality of opportunity estimations. He does so on the grounds that policymakers are interested in dimensions of well-being separately, such as health, income, or education, rather than well-being itself. This may certainly be the case, but if the ultimate objective is to equalize opportunities for well-being, then equalizing opportunities for only one dimension of well-being might actually bring about the opposite result (Calsamiglia, 2009). To see this, consider a policy that targets people born in a certain regions of a country because they have fewer opportunities to acquire a high income. If these people simultaneously have better health, more leisure, or different preferences over the importance of income, they need not have less opportunities to acquire a high level of well-being. Our framework helps clarify if such examples have empirical leverage.

The rest of the paper is organized as follows. Section 3.2 explains both the philosophical and axiomatic theory behind measuring equality of opportunity for well-being. Section 3.3 details our data and measurement approach. Section 3.4 outlines the results and provides several robustness checks. Section 3.5 concludes.

3.2 Theory

3.2.1 Well-Being

Three overarching theories of well-being exist in the philosophical literature; objective list theory, preference satisfaction theory, and mental state theory (Parfit, 1984; Griffin, 1986). *Preference satisfaction theory* is the most commonly assumed in economics. It claims that an individual's welfare depends on the degree to which his preferences are satisfied. Often preference orderings are assumed to be revealed through choice behavior. The underlying tenet behind these revealed preferences is that if an agent chooses bundle A over bundle B, then the agent must prefer A over B, and the agent must be better off with A rather than B. *Mental state theory* takes its starting point in what goes on inside the mind of individuals rather than their observed choices. According to this theory, well-being is the degree to which individuals are happy or the extent to which they experience pleasure over pain. *Objective list theory* argues that individuals' lives go well to the degree that they are in possession of certain items on a list, which could be income, education, health, safety, etc.

In short, mental state theory cares about what individuals feel, (revealed) preferences about what individuals choose, and objective list theory about external factors which could be independent of the choices and feelings of individuals. Each theory has its advantages and shortcomings. Preference satisfaction theory, at least in the revealed form, can be criticized when individuals' decision-making is subject to imperfect knowledge and behavioral biases. If individuals have mistaken beliefs about what is best for themselves or lack willpower to choose what is best for themselves, then there is little reason to believe that their choices are a good manifestation of their well-being. Mental state theory can be criticized for its "physical condition neglect" (Sen, 1985), whereby individuals might feel well only because they have adapted to horrible conditions. In these scenarios, Sen argues, mental states are not an appropriate yardstick. Objective list theory can be criticized for being elitist in the sense that a set of indicators and weights are chosen somewhat independent of the preferences of individuals.[2]

This three-part division of well-being concepts is still very much in use today in both theoretical and empirical literature about well-being (see for example the chapter division in the Oxford Handbook of Well-Being and Public Policy (Adler and Fleurbaey, 2016) and the Stanford Encyclopedia of Philosophy on Well-Being (Crisp, 2016)). We will operationalize a measure of well-being with roots in each of these theories and see if they lead to different conclusions about equality of opportunity. We are not attempting to argue in favor of one of

[2] All of these critiques can of course be counteracted, but doing so would be outside the scope of this paper.

these welfare concepts. Rather, we will take some of the operationalizations of these concepts at face value and investigate if equality of opportunity estimations depend on which measure is used.

3.2.2 Distributive Justice

Until Rawls published his Theory of Justice (Rawls, 1971), the predominant view of justice was defined in utilitarian terms. Under this view, the just outcome maximizes total welfare or, equivalently, equalizes marginal utilities. This view is welfarist in the sense that if all individuals' welfare levels are known, then no additional information is needed to decide whether one scenario is more desirable than another. Rawls argued against this welfarist view, emphasizing that we should not seek to equalize marginal utilities, but rather primary goods, which is a broader notion that also encompasses rights and liberties.

A number of subsequent scholars proposed variations of what the right equalizandum ought to be, building on the work of Rawls. Sen (1980) argued that neither utilities nor primary goods were enough to judge outcomes. He concluded that we need to look at what individuals are capable of achieving with these goods, thus advocating for basic capability equality. Dworkin (1981) contended that resources are the right equalizandum, while Arneson (1989) and Cohen (1990) argued that the right equalizandum is, respectively, equality of opportunity for welfare and equal access to advantage (see Roemer and Trannoy (2016) for a more complete account on the developments in distributive justice since Rawls).

Although these philosophers differ in their preferred equalizandum, they all adhere to the point of view that knowing all individuals' welfare levels is not sufficient. We need to know how that welfare came about, whether it came from fortunate backgrounds or from factors we can hold individuals accountable for. As such, they agree on the need to go beyond welfarism and accept some degree of individual responsibility, and thereby some degree of just inequalities. Notably, none of the philosophers defined the equalizandum in terms of income. Rather, they considered broader notions than income such as welfare, advantage, or functionings. Our approach attempts to get a bit closer to these frameworks. In particular, our approach is closely related to that of Arneson (1989), who precisely argued for equalization of opportunities for welfare.[3]

[3] That being said, Arneson (1989) considered welfare to be preference satisfaction, thus differing from our take, where we will look at different theories of welfare.

3.2.3 Equality of Opportunity

The philosophical theories of distributive justice have been operationalized in economics through the works of Roemer (1993), Van de gaer (1993), and Fleurbaey (1994) amongst others. The starting point in many of these operationalizations is to consider a population, $\mathcal{N} = \{1, 2, \ldots, n\}$, and a distribution of an outcome variable for this population, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$. Often $\boldsymbol{y}$ is considered to be income, but here we will take welfare/well-being as the outcome, such that y_i is the welfare of individual $i \in \mathcal{N}$. An individual's outcome is assumed to be a product of two sets of variables: circumstances, $\boldsymbol{a^C}$, and effort, $\boldsymbol{a^E}$. Circumstances are the factors outside the realm of control for the individual, the factors one ought not to hold an individual responsible for. These are often taken to be gender, region of birth, parental education, parental income etc. Effort variables are the factors one ought to hold an individual responsible for. The well-being of individual i is thus assumed to be given by $y_i = f(\boldsymbol{a_i^C}, \boldsymbol{a_i^E})$.

Inequality of opportunity is frequently measured using either the direct approach or the indirect approach. The *direct approach* obtains a counterfactual distribution of the outcome variable, which only depends on circumstances, and therefore only incorporates unfair variation. This can be obtained by fixing the effort variables at a certain level, $\boldsymbol{\tilde{a}^E}$. Inequality of opportunity is then measured as the inequality in the resulting counterfactual distribution, $I\big(f(\boldsymbol{a^C}, \boldsymbol{\tilde{a}^E})\big)$. The *indirect approach* obtains a counterfactual outcome distribution that only depends on effort, and therefore only incorporates fair variation. This can be done by fixing the circumstance variables at a given level, $\boldsymbol{\tilde{a}^C}$. Inequality of opportunity is then given as the difference between overall inequality and the inequality in this counterfactual distribution, $I(\boldsymbol{y}) - I\big(f(\boldsymbol{\tilde{a}^C}, \boldsymbol{a^E})\big)$.[4]

The axiomatic literature on fair allocations has put forward two criteria inequality of opportunity estimates ideally ought to reflect, these being the compensation principle and the reward principle. The *compensation principle* states that differences in well-being due to differences in circumstances should be eliminated. The *reward principle* is concerned with the proper reward of effort for individuals with the same circumstances. Unfortunately, certain versions of these two criteria are mutually incompatible (Bossert, 1995; Fleurbaey, 1995) unless one assumes that the outcome is linearly separable in circumstance and effort, such that $y_i = g(\boldsymbol{a_i^C}) + h(\boldsymbol{a_i^E})$. We will follow much of the literature and make that assumption.

We want to minimize our reliance on cardinality assumptions, and hence instead of using the welfare levels as the outcome variable, we convert them into welfare ranks, $r_i^y = F(y_i)$, where F is the cumulative distribution function. We likewise convert our counterfactual

[4] See Ramos and Van de gaer (2015) for more methods to obtain counterfactual distributions using the direct and the indirect approach.

distributions into ranks, such that $r_i^{unfair} = G(\boldsymbol{a_i^C})$ and $r_i^{fair} = H(\boldsymbol{a_i^E})$ where G and H are the cumulative distributions of $g(\boldsymbol{a^C})$ and $h(\boldsymbol{a^E})$, respectively. The fair ranks order individuals according to how much their effort variables contribute to their welfare, while the unfair ranks order individuals according to how much their circumstances contribute to their welfare. Within this set-up, neither the direct approach nor the indirect approach is feasible, as the inequalities in rank variables are identical. In other words, $I(\boldsymbol{r^y}) = I(\boldsymbol{r^{unfair}}) = I(\boldsymbol{r^{unfair}})$. Instead, we will use the norm-based approach.

3.2.4 Norm-Based Inequality Metrics

The norm-based approach (Ramos and Van de gaer, 2016), also called the *fairness gap* (Fleurbaey and Schokkaert, 2009), evaluates equality of opportunity by measuring the divergence between the actual outcomes and fair outcomes. The fair outcomes only depend on effort variables and reflect how much each individual ideally is entitled to under some allocation principles. The more aligned these two distributions are, the lower inequality of opportunity. In our context, the fairness gap can be calculated by employing a divergence measure, $D(\boldsymbol{r^y} \| \boldsymbol{r^{fair}})$, which evaluates the divergence between the two distributions, $\boldsymbol{r^y}$ and $\boldsymbol{r^{fair}}$.

An intuitive measure of the (absence of) divergence between $\boldsymbol{r^y}$ and $\boldsymbol{r^{fair}}$ is simply their correlation. A correlation of 1 suggests complete equality of opportunity with respect to rewarding effort; individuals with higher effort levels also have higher levels of welfare. A correlation of 0, in contrast, suggests no equality of opportunity with respect to the rewarding effort.

We can construct a parallel measure which uses $\boldsymbol{r^{unfair}}$ instead of $\boldsymbol{r^{fair}}$. A high correlation between $\boldsymbol{r^y}$ and $\boldsymbol{r^{unfair}}$ suggests a high degree of inequality of opportunity with respect to the compensation principle; individuals with the best circumstances also have the highest welfare levels. A correlation of 0, in turn, reflects that the compensation principle is completely fulfilled; the quality of individuals' circumstances is not correlated with their welfare.

If we use income as the outcome variable and consider parental income as the only circumstance variable, then the correlation between $\boldsymbol{r^y}$ and $\boldsymbol{r^{unfair}}$ boils down to a frequently used measure of mobility; Spearman's correlation between parents' and children's income level (see for example Chetty *et al.* (2014)). We generalize this measure since we consider more than one circumstance when constructing the unfairness ranks and apply other outcome variables than income.

Magdalou and Nock (2011) (MN) provide a framework for more axiomatically grounded divergence measures, which we will use as robustness checks.[5] They put forth a class of divergence measures between an outcome distribution, $\boldsymbol{y}$, and a reference distribution, $\boldsymbol{z}$:

$$D_{MN}(\boldsymbol{y}\|\boldsymbol{z}) = \frac{1}{n}\sum_{i\in\mathcal{N}} \left[\phi(y_i) - \phi(z_i) - (y_i - z_i)\phi'(z_i)\right], \tag{3.1}$$

where for all $c \in \mathbb{R}_{++}$, the function $\phi(c)$ is given by:

$$\phi(c) \coloneqq \begin{cases} \frac{1}{s(s-1)}\, c^s\,, & \text{if} \quad s \neq 0, 1\,, \\ c\,\ln c\,, & \text{if} \quad s = 1\,, \\ -\,\ln c\,, & \text{if} \quad s = 0\,. \end{cases} \tag{3.2}$$

The class $D_{MN}(\boldsymbol{y}\|\boldsymbol{z})$ satisfies partial symmetry (i.e. it is invariant to permutations of (y_i, z_i) pairs) along with other relevant properties. It is suitable only for distributions with equal means. This is not a problem in our set-up since the rank-based measures that we will use in place of $\boldsymbol{y}$ and $\boldsymbol{z}$ by construction have equal means.[6] By using the function ϕ in (3.2), one obtains

$$D_{MN}(\boldsymbol{y}\|\boldsymbol{z}) = \begin{cases} \frac{1}{n}\frac{1}{s(s-1)}\, \sum_{i\in\mathcal{N}} \left[y_i^s + (s-1)z_i^s - s\, y_i\, z_i^{s-1}\right]\,, & \text{if} \quad s \neq 0, 1\,, \\ \frac{1}{n}\sum_{i\in\mathcal{N}} \left[y_i\, \ln\left({}^{y_i}\!/_{z_i}\right)\right]\,, & \text{if} \quad s = 1\,, \\ \frac{1}{n}\sum_{i\in\mathcal{N}} \left[{}^{y_i}\!/_{z_i} - \ln\left({}^{y_i}\!/_{z_i}\right) - 1\right]\,, & \text{if} \quad s = 0\,. \end{cases} \tag{3.4}$$

Cowell (1985) suggests a different class of divergence measures, which he calls measures of distributional change, that satisfies different properties. The population invariant measure equivalent to (3.1) is:[7]

$$D_C(\boldsymbol{y}\|\boldsymbol{z}) = \frac{1}{n}\sum_{i\in\mathcal{N}} \left[z_i\, \phi(y_i/z_i)\right]. \tag{3.5}$$

[5] We are heavily indebted for comments and advice from Brice Magdalou on the use and interpretation of appropriate divergence measures.

[6] Had not our actual and norm distributions had the same mean we could have normalized our divergence class further to obtain a (strong) scale invariant class:

$$D^m_{MN}(\boldsymbol{y}\|\boldsymbol{z}) = \frac{1}{n}\sum_{i\in\mathcal{N}} \left[\phi(\hat{y}_i) - \phi(\hat{z}_i) - (\hat{y}_i - \hat{z}_i)\phi'(\hat{z}_i)\right], \tag{3.3}$$

where $\hat{y}_i = y_i/\mu(\boldsymbol{y})$, $\hat{z}_i = z_i/\mu(\boldsymbol{z})$, and $\mu(\boldsymbol{z}) = \sum_{i=1}^n z_i/n$. It is worth noting that this class boils down to the generalized entropy class of standard inequality measures if the reference distribution is assumed to be the mean of the actual distribution.

[7] A scale invariant measure can be obtained by replacing y and z by $\hat{y}$ and $\hat{z}$ in (3.5). Devooght (2008) provides an empirical application of this measure to equality of opportunity in Belgium.

By using the function ϕ in (3.2), $D_C(\boldsymbol{y}\|\boldsymbol{z})$ can be written as

$$D_C(\boldsymbol{y}\|\boldsymbol{z}) = \begin{cases} \frac{1}{n}\frac{1}{s(s-1)} \sum_{i\in\mathcal{N}} \left[y_i^s z_i^{1-s} - 1\right] , & \text{if} \quad s \neq 0, 1 , \\ \frac{1}{n} \sum_{i\in\mathcal{N}} y_i \, \ln\left[{}^{y_i}/_{z_i}\right] , & \text{if} \quad s = 1 , \\ \frac{1}{n} \sum_{i\in\mathcal{N}} z_i \, \ln\left[{}^{z_i}/_{y_i}\right] , & \text{if} \quad s = 0 . \end{cases} \tag{3.6}$$

The two different classes of divergence measures, $D_{MN}(\boldsymbol{y}\|\boldsymbol{z})$ and $D_C(\boldsymbol{y}\|\boldsymbol{z})$, coincide for one – and unique – parameter value, $s = 1$. For this reason we are going to use $s = 1$ as one of our robustness checks. Parameters $s = 0$ and $s = 2$ with $D_{MN}(\boldsymbol{y}\|\boldsymbol{z})$ will be used as other robustness checks. Why these two values? One of the features of D_{MN} is that a progressive transfer in the actual distribution, $\boldsymbol{y}$, reduces the divergence between $\boldsymbol{y}$ and $\boldsymbol{z}$ as long as the individuals involved in the transfer have the same reference, $\boldsymbol{z}$. It is a kind of priority given to the worse-off individuals when the individuals involved in the transfer share the same $\boldsymbol{z}$. Moreover, if (and only if) $s < 2$, the further down the distribution $\boldsymbol{y}$ such transfer takes place, the more the divergence between $\boldsymbol{z}$ and $\boldsymbol{y}$ is reduced. This property resembles the principle of diminishing transfers in the context of inequality measurement, which holds for the class of entropy indices when $s < 2$. When $s = 2$ the measure is ordinally equivalent to the Euclidian distance, and it is thus insensitive to the position on the distribution where the progressive transfer (among individuals with the same reference) takes place. Thus, the parameter value $s = 2$ can be seen as a threshold. Contrary to this, the parameter value $s = 0$ yields a measure that is more sensitive to transfers lower down the distribution than our baseline measure with $s = 1$.

As another robustness check we will also use a divergence measure which is a generalization of the standard Gini coefficient developed by Almås *et al.* (2011), called the Unfairness Gini, $D_{Gini}(\boldsymbol{y}\|\boldsymbol{z})$:

$$D_{Gini}(\boldsymbol{y}\|\boldsymbol{z}) = \frac{1}{2n(n-1)\mu(\boldsymbol{y})} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} |(y_i - z_i) - (y_j - z_j)| \tag{3.7}$$

3.3 Data & Measurement

We use data from the German Socio-Economic Panel (GSOEP), which is a yearly panel that started in 1984. The panel contains detailed questions on household income, life satisfaction, other well-being dimensions, as well as biographical and historical data that we use to construct circumstance variables. We use data from 1984-2014 and include all working and unemployed individuals but drop individuals outside the labor market and observations with missing values. In total, we have 176,196 person-year observations meeting our baseline specification. These are spread around 21,038 individuals in 15,067 different households.

Our baseline analysis will use the following circumstance variables: gender, father's education (3 categories), mother's education (3 categories), father's occupation (6 categories), polynomial of age, height, place of birth (West Germany, East Germany, abroad), degree of urbanization at place of birth (4 categories), and number of siblings. As baseline effort variables we use years of education, work hours, a dummy for whether the respondent is self-employed, and a dummy for whether the respondent works in the public sector. Effort may be a slightly misleading term here. The point of these four variables is that they plausibly lie within individual control and hence constitute factors that we may hold individuals accountable for. We will provide robustness checks where we move these four variables to the other side of the responsibility cut. Summary statistics of the circumstance and effort variables are given in Table 3.1.

3.3.1 Constructing Welfare Variables

We use four different welfare variables in the analysis. First, we use incomes. This is the most frequently used outcome variable in equality of opportunity studies. We use it as a baseline for comparison to the other well-being measures. We use annual net household income expressed in 2010 constant EUR. The other three welfare variables are rooted in the three concepts of well-being that Parfit (1984) and Griffin (1986) put forward.

The second welfare measure we use is life satisfaction, which has roots in mental state theory. Life satisfaction is the answer to the question, "How satisfied are you with your life, all things considered?" The answer categories range from 0 (completely dissatisfied) to 10 (completely satisfied). For the purpose of this study, we consider the answers interpersonally comparable. This is not meant as an endorsement of this particular account of well-being but rather as an inquiry into how inequality of opportunity estimates would look if one accepted these assumptions.

The third welfare measure we use is a multidimensional welfare measure, which has roots in objective list theory. To construct the measure of multidimensional welfare we partly follow Decancq and Neumann (2016). We consider four dimensions; income, health, leisure, and unemployment.[8] Income is measured the same way as above but converted into logs. Unemployment is a binary variable taking the value of 1 if the respondent had a job at the time of the survey. Leisure is measured as the amount of daily hours spent on leisure (capped at 6 hours). Health is itself a composite index composed of 1) an indicator for whether the

[8]Although we would like to include more dimensions, such as education, we run into estimation problems since this also is considered an effort variable. As we will regress the welfare variable on circumstance and effort variables, and since we do not want to have the same variables on each side of the regression, we omit this dimension.

Table 3.1: Summary Statistics

	mean	sd	min	max
Circumstance Variables				
Father's Education: Primary school	0.68	0.47	0	1
Father's Education: Secondary school	0.19	0.40	0	1
Father's Education: Tertiary school	0.13	0.33	0	1
Mother's Education: Primary school	0.71	0.45	0	1
Mother's Education: Secondary school	0.22	0.42	0	1
Mother's Education: Tertiary school	0.07	0.25	0	1
Father's Occupation: Blue-collar (untrained)	0.14	0.35	0	1
Father's Occupation: Blue-collar (trained)	0.34	0.47	0	1
Father's Occupation: Not employed	0.06	0.24	0	1
Father's Occupation: White-collar	0.26	0.44	0	1
Father's Occupation: Self-employed	0.12	0.32	0	1
Father's Occupation: Civil servant	0.08	0.28	0	1
Place of Upbringing: Large city	0.22	0.41	0	1
Place of Upbringing: Medium city	0.18	0.38	0	1
Place of Upbringing: Small city	0.23	0.42	0	1
Place of Upbringing: Countryside	0.37	0.48	0	1
Place of Birth: West Germany	0.66	0.47	0	1
Place of Birth: East Germany	0.27	0.44	0	1
Place of Birth: Abroad	0.07	0.26	0	1
Height (cm)	173	9.13	80	210
Female	0.47	0.50	0	1
Number of siblings	1.94	1.68	0	17
Age	42.0	12.1	17	91
Effort Variables				
Years of education	12.6	2.72	7	18
Weekly working time	35.9	15.9	0	80
Self-employed	0.09	0.29	0	1
Works in public sector	0.18	0.39	0	1

Notes: Summary statistics of circumstance and effort variables. $n = 176,196$.

individual is disabled, 2) the number of doctor appointments the respondent had last year and 3) the number of inpatient nights in hospitals the respondent had last year. To aggregate these sub-dimensions into one health dimension we regress a health satisfaction question on the three variables and use the coefficients as weights. The health satisfaction variable is composed of answers to how satisfied individuals are with their health on a scale from 0 (not at all satisfied) to 10 (completely satisfied). For the income, leisure, and health dimension, we standardize the values such that the highest possible level is 1 and the lowest possible level is 0. Now we have four dimensions each bounded between 0 and 1. To arrive at the final multidimensional index, we simply add these four together.

The fourth welfare measure, equivalent incomes, is based on preference satisfaction theory. In short, equivalent incomes are the incomes individuals need together with a reference bundle, to make them indifferent with their actual bundle. Although preferences are often estimated from choice behavior, this is difficult when the arguments cover a wide array of dimensions of well-being. An alternative method to recover preferences used by Decancq *et al.* (2015), is to regress life satisfaction on the dimensions of well-being and interpret the weights as marginal rates of substitution. The resulting utility functions seem to be highly correlated with the utility functions one would recover from choice behavior (Akay *et al.*, 2015). This approach easily accommodates preference heterogeneity by simply allowing for interactions between sociodemographic characteristics, w, and the various dimensions, dim. We follow this approach and use the following subset of the circumstance and effort variables as preference heterogeneity parameters, w: birth location, sex, age, age^2, education, work hours, self-employed, public sector worker. As the non-income dimensions, dim, we consider health, employment, and leisure, like in the multidimensional index. Our life satisfaction regression looks as follows:

$$lifesat_{it} = [\beta^{inc} + \gamma^{inc} w_{it}]ln(inc_{it}) + [\beta^{dim} + \gamma^{dim} w_{it}]dim_{it} + \mu_t + \alpha_i + \varepsilon_{it} \tag{3.8}$$

In order to calculate the equivalent incomes, we first select a reference vector, $\tilde{dim}$, of all other dimensions than income. Here we choose the mean outcome (mode for categorical variables), since this avoids favoring any extreme marginal rates of substitution.[9] Then we calculate the income needed together with $\tilde{dim}$ for this joint bundle to make individuals indifferent with their actual bundle. That is, we isolate inc_{it}^{eq} below:

$$\begin{aligned} &lifesat(inc_{it}, dim_{it}) = lifesat(inc_{it}^{eq}, \tilde{dim}) \\ &\qquad \Leftrightarrow inc_{it}^{eq} = exp\Big(ln(inc_{it}) + \frac{\beta^{dim} + \gamma^{dim} w_{it}}{\beta^{inc} + \gamma^{inc} w_{it}}(dim_{it} - \tilde{dim})\Big) \end{aligned} \tag{3.9}$$

The final result is an interpersonally comparable measure of individual welfare that takes differences in preferences into account.

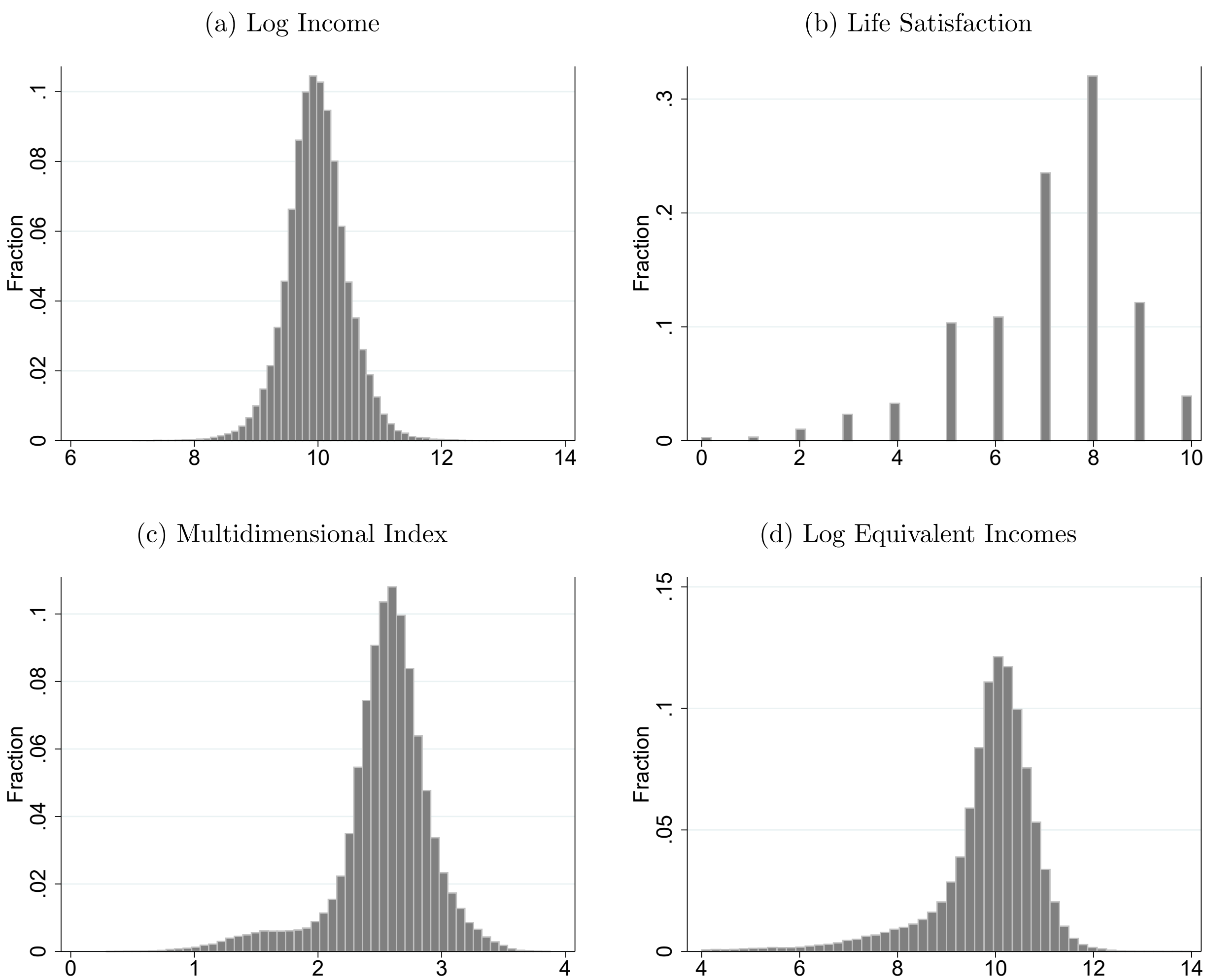
By employing equivalent incomes as a welfare measure in our analysis, we are implicitly taking sides in a rich philosophical debate about whether individuals should be held responsible for their preferences. Our empirical approach implicitly deems differences in well-being arising from preference heterogeneity unfair if they stem from circumstance variables, but

[9] To see this, suppose we were to choose the maximum value rather than the mean outcome. Then we would favor individuals with a high preference for income, as these individuals need a relatively high income together with this reference bundle to be indifferent to their actual bundle.

fair if they stem from effort variables. This is in contrast to most applications of equivalent incomes, which hold that individuals should be responsible for all their preferences. We can amend our approach such that individuals are held responsible for all their preferences by decomposing the equivalent income measure into a part that is due to preference heterogeneity and a part that does not incorporate preferences. We will do so as a robustness check. A detailed discussion of this method along with an illustration is given in Appendix 3.A.1.

Histograms of the four final welfare measures (with the incomes and equivalent incomes expressed in logs) are presented in Figure 3.1. We turn these welfare measures into annual welfare ranks, $\boldsymbol{r}_t^{\boldsymbol{y}}$, and break ties by adding a small amount of random noise to the well-being levels.

Figure 3.1: Histograms



Notes: Histograms of the four welfare measures.

3.3.2 Estimating Equality of Opportunity

For our empirical specification, we consider the well-being ranks to be linear in effort and circumstance variables:

$$r_{it}^{y} = \beta^{C} a_{it}^{C} + \beta^{E} a_{it}^{E} + \epsilon_{it} \tag{3.10}$$

Two important issues remain unsettled. First is the issue of how to interpret the error term, ϵ_{it}. The error contains omitted effort variables, omitted circumstance variables, measurement error, and general uncertainty. It is unclear whether this should be considered within individual control. This is an important decision as it accounts for most of the variation in the welfare ranks. In our baseline specification, we follow the inequality of opportunity literature and consider it an effort variable, but as a robustness check we shift it to the other side of the responsibility cut.

The other unsettled issue is what to do with the correlation between the effort and circumstance variables. Individuals' effort levels are partly determined by their own choices and partly by their circumstances. Years of education, for example, is partly influenced by individuals' social background and partly by an individual's own choices. We follow Roemer's approach and consider this correlation to be outside the realm of individual responsibility (Roemer, 1998).[10] In practice this means that prior to estimating the impact of circumstances and efforts on well-being we perform an auxiliary regression of the following form:

$$a_{it}^{E} = \gamma a_{it}^{C} + \eta_{it} \tag{3.11}$$

We perform such a regression for each effort variable and use the residuals from these regressions as our effort variables in our main regression, which then becomes:

$$r_{it}^{y} = (\beta^{C} + \gamma\beta^{E}) a_{it}^{C} + \beta^{E} \eta_{it} + \epsilon_{it} \tag{3.12}$$

Due to the Frisch-Waugh-Lowell theorem, the coefficients on the effort variables will be the same in (3.10) and (3.12). The coefficients on the circumstance variables will be different as they in (3.12) also incorporate the indirect effect of circumstances on effort. We will later report specifications where we omit this auxiliary regression.

To compare who the opportunity-deprived are across the four well-being measures, we rank individuals according to their opportunity profile, that is, according to the quality of

[10] Jusot *et al.* (2013) likewise call this Roemer's view, while not correcting for this correlation is termed Barry's view (Barry, 2005). A final possibility, where the correlation between effort and circumstances is considered effort, is called Swift's view (Swift, 2005).

their circumstances. We calculate each person's yearly unfairness rank as

$$r_{it}^{unfair} = G_t\big[(\beta^C + \gamma\beta^E)a_{it}^C\big], \tag{3.13}$$

where G_t is the yearly cumulative distribution of individuals' unfair advantage. This allows us to compare the opportunity-deprived across the four well-being measures.

To estimate equality of opportunity over time using the norm-based approach, we also compute fair well-being ranks. These are given by

$$r_{it}^{fair} = H_t\big[\beta^E\eta_{it} + \epsilon_{it}\big], \tag{3.14}$$

where H_t is the yearly cumulative distribution of individuals' fair advantage. Using a divergence measure, we are now able to compute the level of inequality of opportunity for each of the welfare variables for each year of the survey.

3.4 Results

3.4.1 Who are the Opportunity-Deprived?

Table 3A.1 in Appendix 3.A.2 shows the results of the regressions from equation (3.12). Based on this output and equation (3.13), we calculate each person's yearly unfairness rank. Table 3.2 shows the correlations between these unfairness ranks for the four welfare measures. The correlations reveal the extent to which the same people are opportunity-deprived across the four measures. In the table – and throughout the paper – we bootstrap confidence intervals in order to take all derived uncertainty into account, including the uncertainty when constructing the welfare measures. We bootstrap 500 resamples at individual-level clusters.

The unfairness ranks display rather high correlations, suggesting that the same people are opportunity-deprived regardless of how we measure welfare. The welfare measures we have constructed are of course partly contained within each other; the income variable is included in both the multidimensional index and the equivalent income measure, and the latter two use the same four dimensions but aggregate them differently. We can analyze the extent to which this is driving the high unfairness correlations by comparing the unfairness correlations with the correlation between the welfare ranks, $\boldsymbol{r^y}$, which are shown in Table 3.3. In all cases, the correlations are higher when we look at $\boldsymbol{r^{unfair}}$. This indicates that the high unfairness correlations are not driven solely by the measures' interrelatedness. This also suggests that the way welfare is measured matters more if we target the welfare-deprived

Table 3.2: Correlation Between Unfairness Ranks

	Income	Life Sat.	Multidim. Index	Equivalent Inc.
Income	-	-	-	-
Life Satisfaction	0.60	-	-	-
	(0.55, 0.66)			
Multidim. Index	0.64	0.90	-	-
	(0.61, 0.68)	(0.87, 0.93)		
Equivalent Inc.	0.88	0.72	0.82	-
	(0.79, 0.92)	(0.66, 0.79)	(0.76, 0.88)	

Notes: Correlations between $\boldsymbol{r^{unfair}}$ for the four welfare measures. Bootstrapped 95^{th} percentile confidence intervals in parenthesis.

than if we target the opportunity-deprived.

Table 3.3: Correlation Between Welfare Ranks

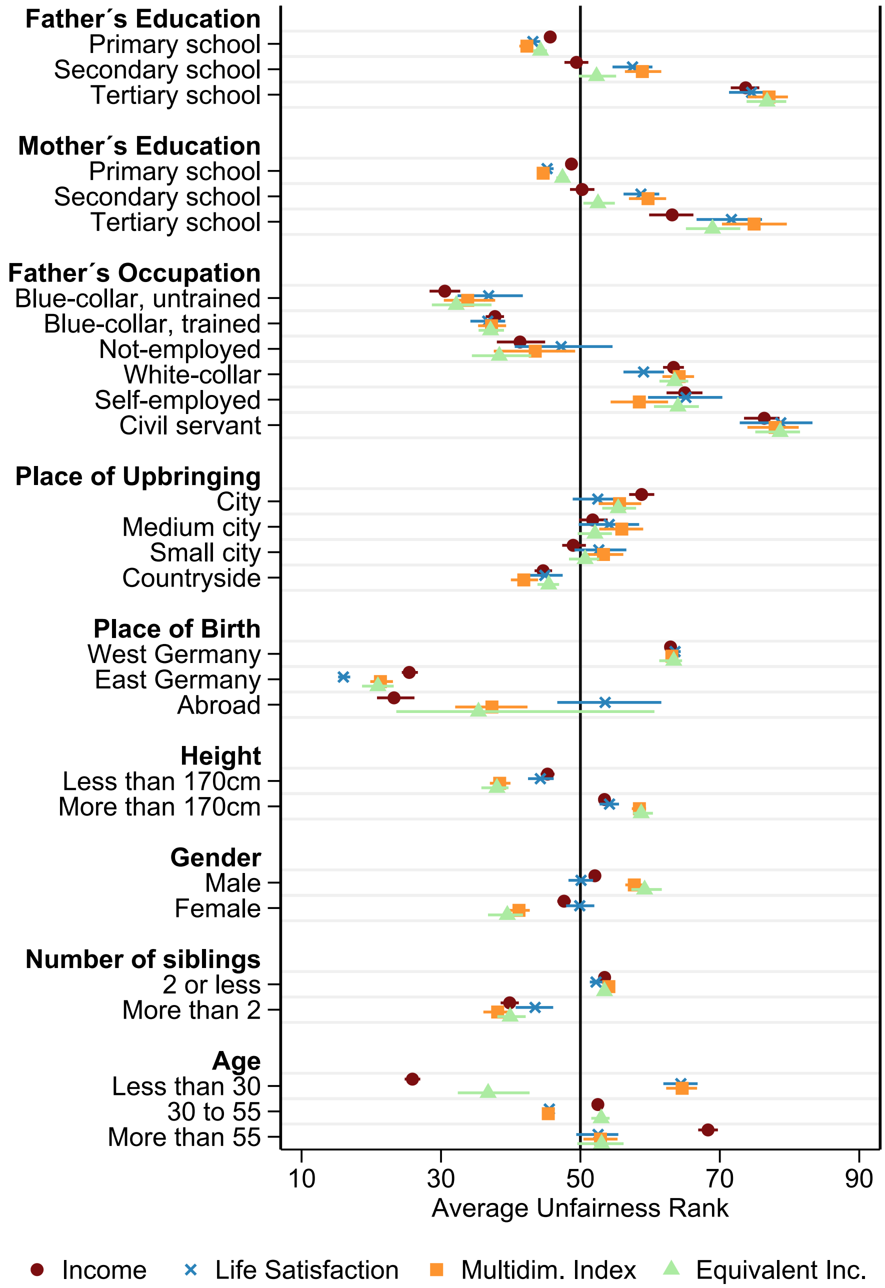
	Income	Life Sat.	Multidim. Index	Equivalent Inc.
Income	-	-	-	-
Life Satisfaction	0.21	-	-	-
	(0.20, 0.22)			
Multidim. Index	0.41	0.23	-	-
	(0.40, 0.43)	(0.22, 0.24)		
Equivalent Inc.	0.66	0.27	0.77	-
	(0.61, 0.71)	(0.26, 0.28)	(0.74, 0.80)	

Notes: Correlation between welfare ranks, $\boldsymbol{r^{y}}$. Bootstrapped 95^{th} percentile confidence intervals in parenthesis.

Since individuals' opportunities are unobservable, policymakers may have to assist the opportunity-deprived indirectly. One way of doing so is by targeting individuals with circumstance profiles that are highly correlated with having low opportunities. We can use the unfairness ranks to test if the characteristics of the opportunity-deprived are similar across the four welfare measures. To do so, we calculate the average unfairness rank for individuals with a given circumstance. Results are shown in Figure 3.2. The lower the average unfairness rank, the less opportunities individuals with the given circumstance have, and the more this circumstance is a potential factor policymakers can use to target the opportunity-deprived. If the average rank is less than 50, then the particular group has less than average opportunities.

There are many similarities across the welfare measures. Individuals with low educated parents and individuals whose father was a blue-collar worker or not employed have low

Figure 3.2: Who are the Opportunity-deprived?



Notes: The figure shows the average unfairness rank for individuals that share a given circumstance. If the points are to the left of the line at 50, then individuals with this circumstance are more than average opportunity-deprived and vice versa. Bars indicate bootstrapped 95^{th} percentile confidence bans.

opportunities. The same applies to individuals who grew up in the countryside, individuals born in East Germany, short individuals, females, and individuals with many siblings.

Important differences emerge only in two places, for people born abroad and for different age groups. People born abroad are more opportunity-deprived in all measures but life satisfaction. A possible explanation for this is that people born abroad understand the life satisfaction scale more optimistically than Germans. Alternatively, it may be because people born abroad tend to have other circumstances which are particularly good for life satisfaction. We can indirectly check which effect is driving the result by excluding place of birth from the regression, calculating new unfairness ranks, and re-computing the average unfairness rank of individuals born abroad. With this approach, the direct channel from being born abroad on life satisfaction is omitted. Now the average unfairness rank of people born abroad falls below 50 (not shown in figure). This suggests that the previous higher rank was solely driven by the direct positive effect of being born abroad on life satisfaction and that individuals born abroad fare worse than people born in Germany with respect to the remaining circumstances. Hence, we cannot exclude that the high unfairness rank for people being born abroad is solely due to scaling effects.

With respect to age, young people are opportunity-deprived in income but not in the multidimensional index. This is hardly surprising as the multidimensional index includes health, and young people on average are healthier. It is questionable whether resources should be allocated such that individuals have equal opportunities in every part of their life. For this reason, we will later on place age on the other side of the responsibility cut. This may seem counterintuitive but it amounts to saying that individuals should have equal opportunities on expectation over their lifecycle rather than in every point of their life (see Almås *et al.* (2011) for a similar approach).

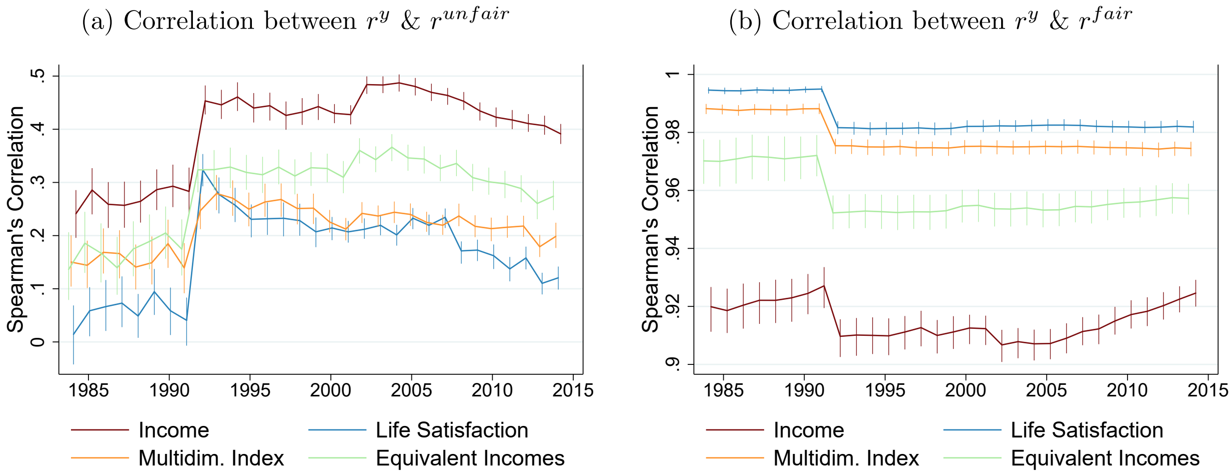
In sum, there seems to be relatively large agreement about whom the opportunity-deprived are across the four measures. Hence, if a policymaker strives to target the individuals with low opportunities, it matters relatively little how welfare is measured.

3.4.2 Equality of Opportunity over Time

To compare the development in equality of opportunity over time across the four welfare measures, we calculate the yearly correlations between $\boldsymbol{r_t^y}$ and, respectively, $\boldsymbol{r_t^{unfair}}$ and $\boldsymbol{r_t^{fair}}$. Results are displayed in Figure 3.3

Before commenting on the trends over time, one remark is necessary. The *level* of the correlations cannot be compared across the different welfare measures. Some of the welfare ranks, particularly the life satisfaction ranks, are by construction noisier. This means that the residuals of the regression take up most of the variation. In our baseline set-up, the residual is considered an effort variable that gets incorporated into the fairness ranks. The fairness ranks will therefore be very highly correlated with the actual ranks for a measure such as life satisfaction. It cannot be concluded that there is more equality of opportunity

Figure 3.3: Equality of Opportunity over Time



Notes: Development in inequality of opportunity in each well-being variable from 1984-2014. In panel (a) large correlations suggest a high presence of inequality of opportunity, while in panel (b) high correlations suggest a low presence of inequality of opportunity. Bars indicate bootstrapped 95^{th} percentile confidence bans.

with life satisfaction than, say, income, as the difference in levels may be entirely due to the degree of random noise inherent in the measures. This explains why the correlations in panel (b) of Figure 3.3 are much greater than in panel (a). To circumvent this problem, we will compare the trends over time.

With regards to the trends over time, all measures undergo a sharp increase in inequality of opportunity in the early '90s. This is caused by the introduction of East Germany into the sample. In panel (a), there is a gradual improvement in inequality of opportunity over the past 15 years. In other words, the quality of individuals' circumstances is becoming less correlated with their welfare levels – regardless of how welfare is measured. At the same time panel (b) shows that for both the income and the equivalent income measure, the effort ranks are becoming more associated with the welfare ranks, suggesting that effort is being more rewarded. Only for life satisfaction and the multidimensional index do we not see improvements in this aspect over the past 15 years.

The figures suggest that regardless of what concept of welfare we use to measure inequality of opportunity, the trends are broadly similar; inequality of opportunity increased with the introduction of East Germany, and overall inequality of opportunity – understood as some combination of the trends of panel (a) and panel (b) – has improved over the past 15 years.

3.4.3 Altering the Responsibility Cut

The analysis thus far was based on important normative assumptions regarding what individuals were held responsible for. We assumed that individuals were responsible for four variables (*4var*); their education, work hours, and whether they are self-employed or work in the public sector. We also assumed that individuals should not be held responsible for the part of these variables that could be accounted for by circumstance variables. That is, the correlation (*cor*) between circumstances and effort was itself considered outside the control of individuals. We further assumed that the part of individual well-being that was unaccounted for by circumstance or effort variables (*residual*) was within individual control. Next, we implicitly considered well-being differences across different age groups (*age*) as unfair. Finally, for the equivalent income measure, we assumed that individuals were not responsible for well-being differences due to preference heterogeneity arising from circumstances (*pref*).

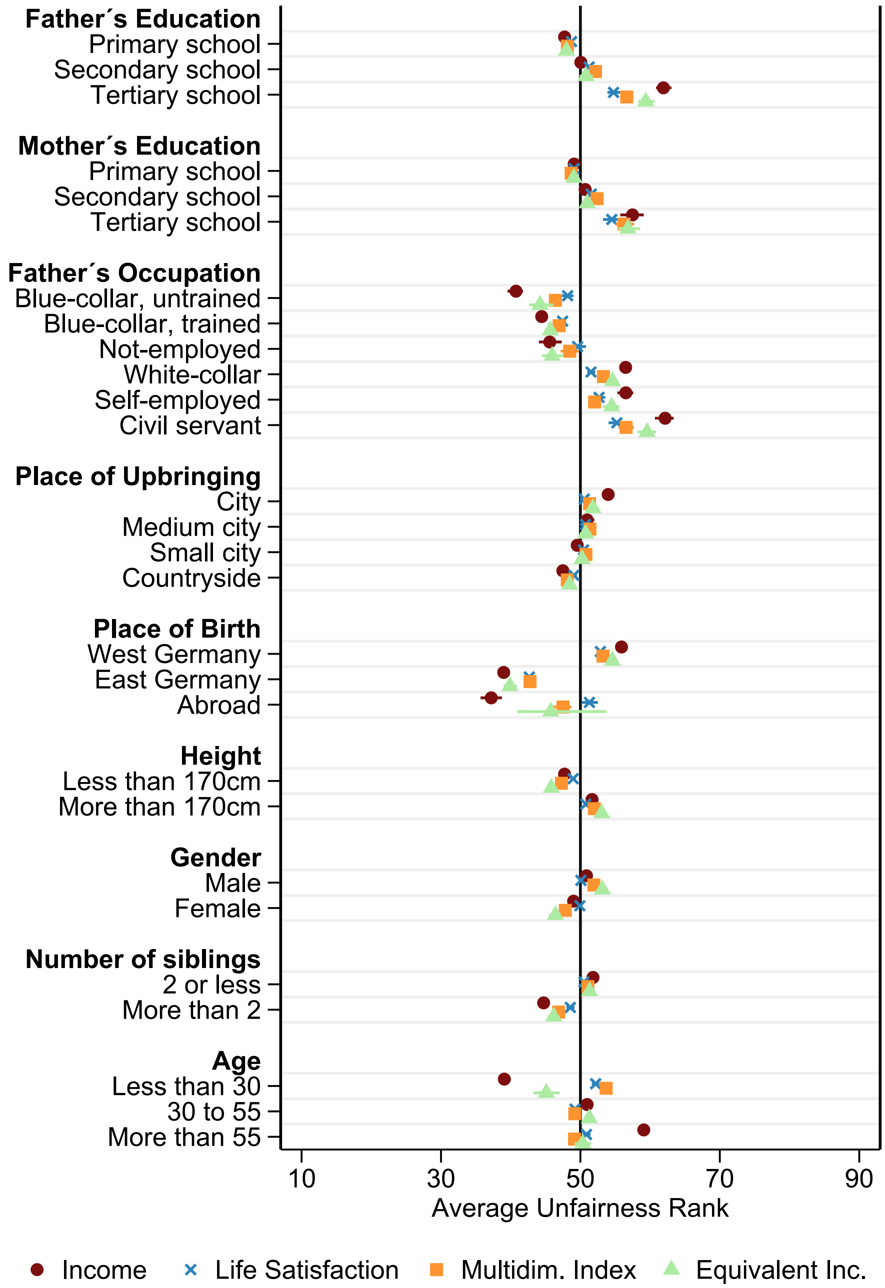
Our baseline effort set contained *4var* and *residual*. In this section, we try to shift the responsibility cut by altering what goes into the effort set. First, we look at whether the characteristics of the opportunity-deprived change, as we change the effort set. This is analyzed in Figures 3.4 and 3.5. Figure 3.4 uses our most narrow effort set, only *4var*, while Figure 3.5 uses the widest possible effort set, *4var, residual, cor, age, pref*.

The figures show the same overall pattern as our main results. For the most part the opportunity-deprived share the same characteristics across all four measures. The only disputes are, once again, for individuals born abroad, and for different age groups. The responsibility cut does matter, however, for quantifying the degree to which a particular group is opportunity-deprived. For example, individuals born in East Germany have an average opportunity rank of about 35 with the smallest effort set and 15 with the largest effort set.

Next, we look at whether the developments over time depend on where we place the responsibility cut. We try six different specifications, which gradually expand the effort set. Results are displayed in Figure 3.6. Our primary interest is not whether the trends change as we change the responsibility cut, but rather whether the four well-being measures follow similar trends regardless of where we place the cut.

Our baseline results are displayed in panel (c). If we add more to the responsibility side (panel (d) to (f)), then the pattern remains the same; inequality of opportunity increased with the inclusion of East Germany but since then gradually fell. This applies to all four welfare variables. Hence, broadening our effort set does not influence our baseline results. The picture looks similar but less pronounced if we only consider the residual in the effort set (panel (b)). There are slightly diverging patterns in the '90s, but over the last fifteen years all four well-being display improvements in equality of opportunity. If we only hold individuals responsible for the four choice variables and not the residual (panel (a)), then

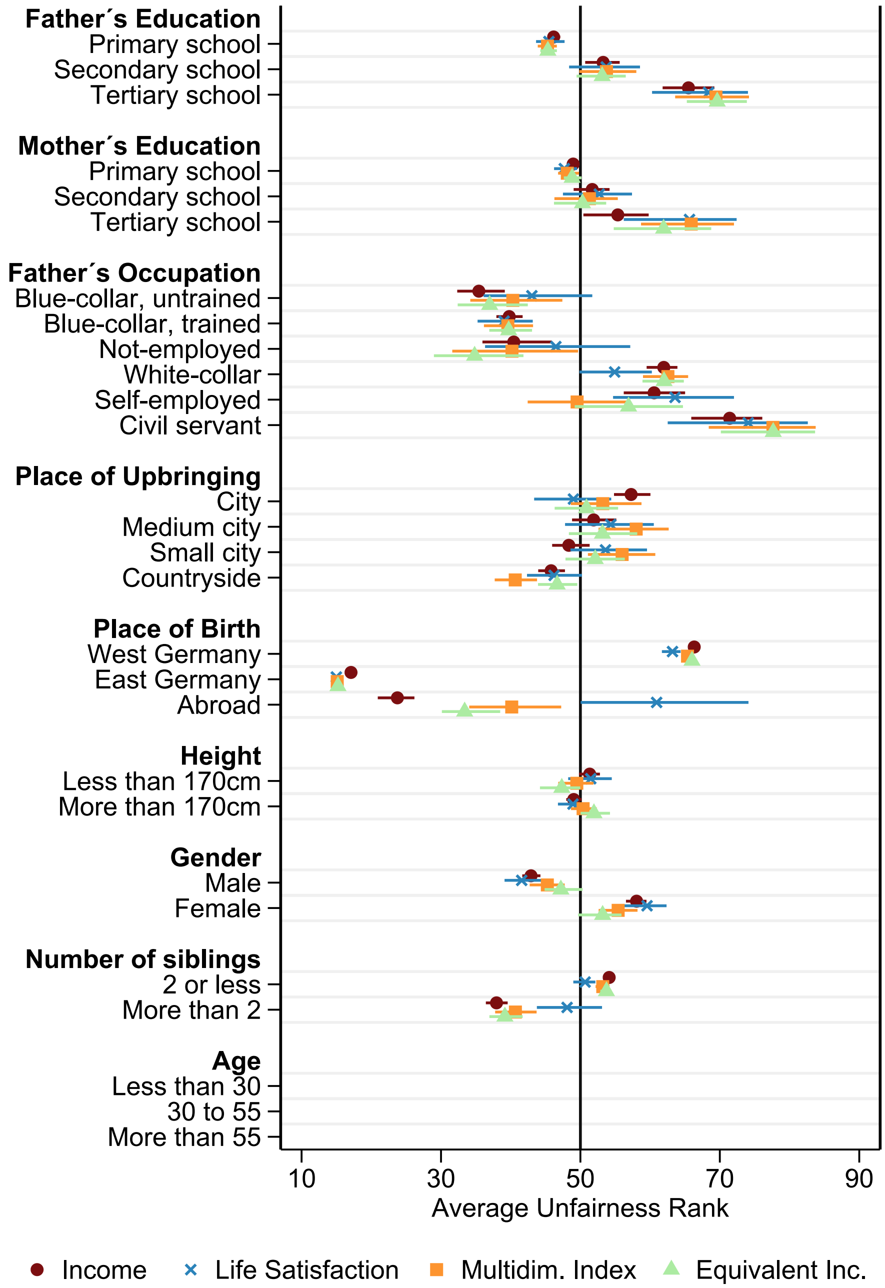
Figure 3.4: Changing the Responsibility Cut: Effort = {4var}



Notes: The figure shows the average unfairness rank for people sharing a particular circumstance using a narrow effort set. Bars indicate bootstrapped 95^{th} percentile confidence bans.

the picture looks very different, particularly for life satisfaction. As already argued, the life satisfaction ranks carry a lot of noise, implying that the four choice variables only are able to explain a very little part of the variance. When we rank individuals according to their unfairness and let the residual be part of the unfairness ranks, these will be almost perfectly

Figure 3.5: Changing the Responsibility Cut: Effort = {4var, residual, cor, age, pref}



Notes: The figure shows the average unfairness rank for people sharing a particular circumstance using a wide effort set. We no longer report differences in the average unfairness rank by age groups, as age is not considered a circumstance in this specification. Bars indicate bootstrapped 95^{th} percentile confidence bans.

aligned with the welfare ranks. Other studies that have tried to switch the residual to the other side of the responsibility cut likewise found this to have a great impact on the results (see for example Almås *et al.* (2011) and Devooght (2008)). Still, the trends over time are broadly the same across the other three well-being measures.

In Figure 3A.2 in Appendix 3.A.3, we show the impact of changing the responsibility cut for the correlations between $\boldsymbol{r^y}$ and $\boldsymbol{r^{fair}}$. As long as we do not change the residual to the other side, the figures display a modest decrease in inequality of opportunity for incomes and equivalent incomes since 2000 and a flat trend for life satisfaction and the multidimensional index. The picture looks different, but the trends are somewhat similar, if we switch the residual to the other side of the responsibility cut. This is not too different from our baseline results.

In sum, we find that when characterizing whom the opportunity-deprived are, neither the measure of welfare nor the precise location of the responsibility cut is of great importance. When analyzing developments in equality of opportunity over time, where we place the residual matters quite a bit, while enlarging the effort set further has few implications. For all well-being variables, we find that over the past fifteen years, the correlation between welfare ranks and unfairness ranks have decreased, while the correlation between welfare ranks and fairness ranks in no case has become worse.

3.4.4 Robustness Checks

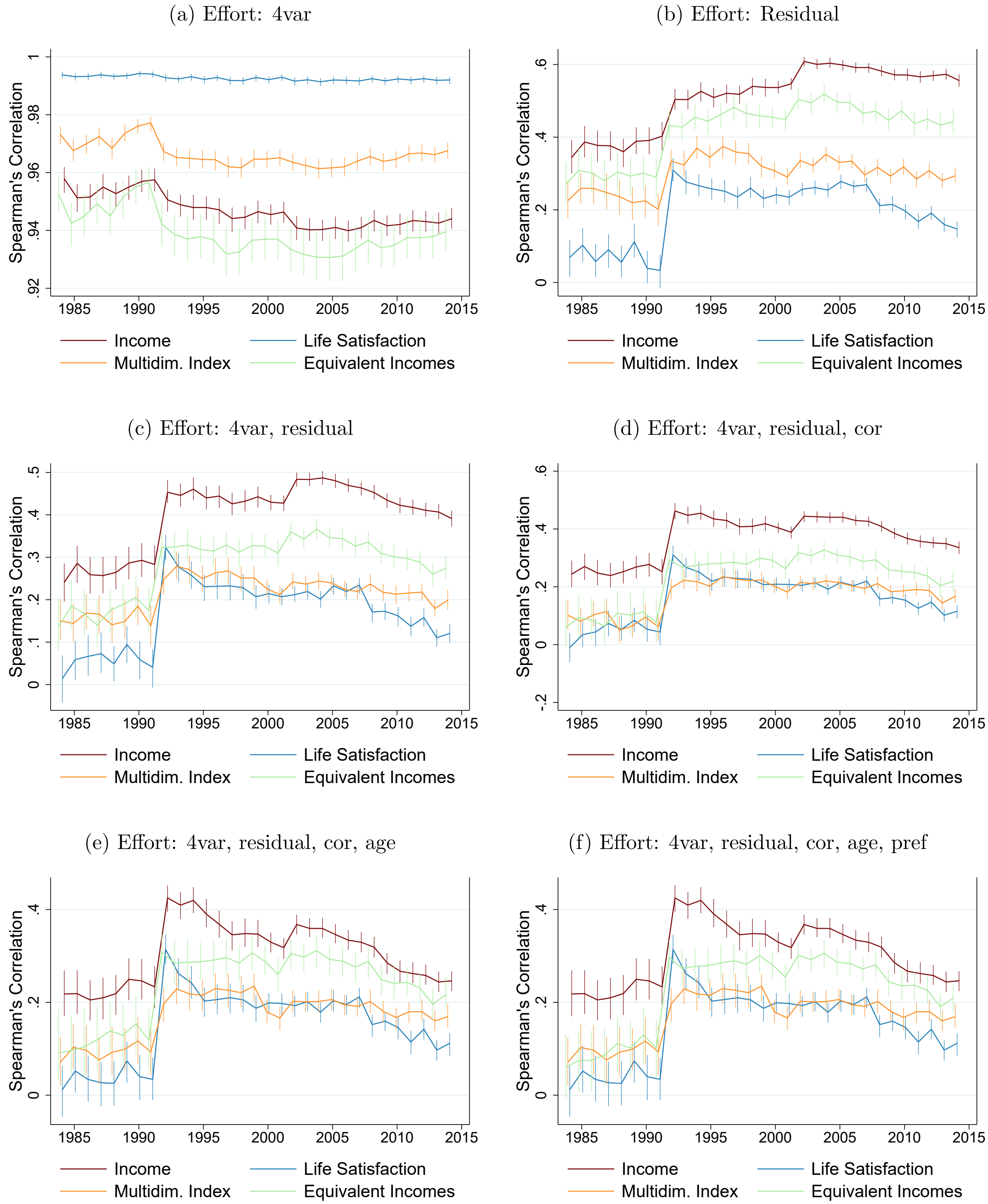
The main analysis was based on a specific choice of a divergence measure. In this section, we test whether our findings are sensitive to using other divergence measures. We use the Magdalou-Nock divergence measures with $s = 0$, $s = 1$, and $s = 2$, as well as the fairness Gini. We display the results both when calculating the divergence between $\boldsymbol{r^y}$ and $\boldsymbol{r^{fair}}$ (Figure 3A.3 in Appendix 3.A.3) and when calculating the divergence between $\boldsymbol{r^y}$ and $\boldsymbol{r^{unfair}}$ (Figure 3A.4 in Appendix 3.A.3). For these measures, a large divergence between $\boldsymbol{r^y}$ and $\boldsymbol{r^{unfair}}$ suggests *greater* equality of opportunity. The reverse applies to the divergence between $\boldsymbol{r^y}$ and $\boldsymbol{r^{fair}}$.

The results are qualitatively unchanged when we use the Magdalou-Nock divergence measures with $s = 1$, $s = 2$, or the fairness Gini. In all three cases, the unfairness ranks and the welfare ranks are growing more apart in the past fifteen years. In contrast, the fairness ranks and the welfare ranks are growing more alike for equivalent incomes and incomes while no trend is visible for life satisfaction and the multidimensional index. This mimics exactly our baseline results, suggesting that they are not driven by our particular choice of divergence measure. Only the Magdalou-Nock divergence measure with $s = 0$ displays different developments. In particular, the developments are quite erratic with large confidence bands making it hard to extract any meaningful changes. This divergence measure puts great emphasis on divergences at the bottom of the distribution and is particularly sensitive to values close to zero. If we add 1 to all ranks (such that the ranks go from 1 to 101), then the Magdalou-Nock index with $s = 0$ shows the same trends as the other divergence measures.

3.5 Conclusion

We have investigated if equality of opportunity estimates depend on what, precisely, we seek to equalize opportunities for. Based on philosophical literature on well-being, we constructed four measures of welfare that are candidates for what we ought to equalize opportunities for. Upon constructing these, we analyzed if the way welfare is measured matters for 1) characterizing the opportunity-deprived and 2) tracking inequality of opportunity over time. We found that, for the most part, neither depend greatly on what measure of well-being we use. These results are robust to most alternative measurement assumptions and changes to the responsibility cut. This is encouraging news for researchers and policymakers interested in going beyond GDP. Whereas previous research has shown that going beyond GDP matters greatly for defining the welfare-deprived and for tracking growth in welfare over time, our findings suggest that going beyond income is less important if the object of interest is to equalize opportunities. Circumstances beyond individual control influence welfare in a similar fashion regardless of how welfare is measured. Hence, for matters of distributive justice, alternative measures of GDP seem to have less importance, and a good picture may be achieved by simply using incomes as a proxy variable for welfare.

Figure 3.6: Altering the Responsibility Cut, $D(\boldsymbol{r^y} \| \boldsymbol{r^{unfair}})$



Notes: Development in inequality of opportunity in each well-being variable from 1984-2014 for different responsibility cuts. *4var:* The four variables work hours, education, self-employed, and works in public sector are considered effort. *Residual:* The residuals from the regressions of the well-being variables on circumstance and effort variables are considered effort. *Age:* Age is considered effort (implying we are equalizing lifetime opportunities). *Cor:* The correlation between effort and circumstance variables is not considered a circumstance. *Pref:* Individuals are fully held responsible for their preferences (only applies to the equivalent income measure). Our baseline specification used *4var* and *residual* as effort.

Bibliography

Adler, M. D. and Fleurbaey, M. (2016). *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press. (page 58)

Akay, A., Bargain, O. and Jara, H. X. (2015). Back to Bentham: Should we? Large-scale comparison of decision versus experienced utility for income-leisure preferences. *Working paper*. (page 66)

Almås, I., Cappelen, A. W., Lind, J. T., Sørensen, E. Ø. and Tungodden, B. (2011). Measuring unfair (in)equality. *Journal of Public Economics*, **95** (7), 488–499. (page 63, 72, 76)

Arneson, R. J. (1989). Equality and equal opportunity for welfare. *Philosophical studies*, **56** (1), 77–93. (page 55, 59)

Barry, B. (2005). *Why social justice matters*. Polity. (page 68)

Blanchflower, D. G. and Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, **88** (7), 1359–1386. (page 57)

Bossert, W. (1995). Redistribution mechanisms based on individual characteristics. *Mathematical Social Sciences*, **1** (31), 51. (page 60)

Calsamiglia, C. (2009). Decentralizing equality of opportunity. *International Economic Review*, **50** (1), 273–290. (page 57)

Chetty, R., Hendren, N., Kline, P., Saez, E. and Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *The American Economic Review*, **104** (5), 141–147. (page 61)

Cohen, G. A. (1990). Equality of what? On welfare, goods and capabilities. *Recherches Économiques de Louvain/Louvain Economic Review*, pp. 357–382. (page 55, 59)

Cowell, F. A. (1985). Measures of distributional change: An axiomatic approach. *The Review of Economic Studies*, **52** (1), 135–151. (page 62)

Crisp, R. (2016). Well-being. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, summer 2016 edn., Stanford University. (page 58)

Decancq, K., Fleurbaey, M. and Schokkaert, E. (2015). Happiness, equivalent incomes and respect for individual preferences. *Economica*, **82** (s1), 1082–1106. (page 66)

— and Neumann, D. (2016). Does the choice of well-being measure matter empirically? In M. D. Adler and M. Fleurbaey (eds.), *Oxford Handbook of Well-Being and Public Policy*, Oxford University Press, pp. 553–587. (page 57, 64)

Devooght, K. (2008). To each the same and to each his own: A proposal to measure responsibility-sensitive income inequality. *Economica*, **75**, 280–295. (page 62, 76)

Dworkin, R. (1981). What is equality? Part 2: Equality of resources. *Philosophy & Public Affairs*, pp. 283–345. (page 59)

Ferreira, F. H. and Peragine, V. (2016). Individual responsibility and equality of opportunity. In M. D. Adler and M. Fleurbaey (eds.), *Oxford Handbook of Well-Being and Public Policy*, Oxford University Press, pp. 746–784. (page 55)

Fitoussi, J.-P., Sen, A. and Stiglitz, J. E. (2010). *Mismeasuring our lives: Why GDP doesn't add up*. The New Press. (page 55)

Fleurbaey, M. (1994). On fair compensation. *Theory and decision*, **36** (3), 277–307. (page 60)

— (1995). The requisites of equal opportunity. In W. Barnett, H. Moulin and M. Salles (eds.), *Social Choice, Welfare and Ethics*, Cambridge University Press, pp. 37–53. (page 60)

— and Maniquet, F. (2011). *A Theory of Fairness and Social Welfare*. Cambridge University Press. (page 57)

— and Schokkaert, E. (2009). Unfair inequalities in health and health care. *Journal of health economics*, **28** (1), 73–90. (page 61)

Griffin, J. (1986). *Well-being: Its meaning, measurement, and moral importance*. Clarendon Press. (page 55, 58, 64)

Jusot, F., Tubeuf, S. and Trannoy, A. (2013). Circumstances and efforts: How important is their correlation for the measurement of inequality of opportunity in health? *Health economics*, **22** (12), 1470–1495. (page 68)

Magdalou, B. and Nock, R. (2011). Income distributions and decomposable divergence measures. *Journal of Economic Theory*, **146** (6), 2440–2454. (page 61)

Parfit, D. (1984). *Reasons and persons*. Oxford University Press. (page 55, 58, 64)

Ramos, X. and Van de gaer, D. (2015). Measurement of inequality of opportunity based on counterfactuals. (page 60)

— and — (2016). Empirical approaches to inequality of opportunity: Principles, measures, and evidence. *Journal of Economic Surveys*, **30** (5), 855–883. (page 55, 56, 61)

Ravallion, M. (2017). Inequality and poverty when effort matters. *Econometrics*, **5** (4). (page 57)

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press. (page 59)

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, pp. 146–166. (page 60)

— (1998). *Equality of Opportunity*. Harvard University Press. (page 68)

— (2012). On several approaches to equality of opportunity. *Economics and Philosophy*, **28** (02), 165–200. (page 57)

— and Trannoy, A. (2015). Equality of opportunity. In A. B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution, Volume 2A-2B*, North Holland, pp. 217–300. (page 55)

— and — (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, **54**, 1288–1332. (page 59)

Sen, A. (1980). Equality of what? *The Tanner Lecture on Human Values*, pp. 197–220. (page 59)

— (1985). *Commodities and Capabilities*. North-Holland. (page 58)

Stevenson, B. and Wolfers, J. (2008). Happiness inequality in the United States. *Journal of Legal Studies*, **37**. (page 57)

Swift, A. (2005). *Justice, luck and the family*. Princeton University Press, Princeton. (page 68)

Van de gaer, D. (1993). *Equality of opportunity and investment in human capital*. Ph.D. thesis, Leuven: Katholieke Universiteit Leuven. (page 60)

3.A Appendix

3.A.1 Decomposition of Equivalent Incomes

The equivalent incomes are calculated from the following equation:

$$inc_{it}^{eq} = exp\Big(ln(inc_{it}) + \frac{\beta^{dim} + \gamma^{dim} w_{it}}{\beta^{inc} + \gamma^{inc} w_{it}}(dim_{it} - \tilde{dim})\Big) \tag{15}$$

We want to decompose this into a part that reflects the welfare individuals derive from their specific preferences, and a part that reflects the welfare individuals receive independent of their particular preferences. To do so, we define a set of norm preferences by fixing the preference heterogeneity variables at a given level, $\tilde{w}$, and computing individuals' equivalent incomes with these norm preferences:

$$\tilde{inc}_{it}^{eq} = exp\Big(ln(inc_{it}) + \frac{\beta^{dim} + \gamma^{dim} \tilde{w}}{\beta^{inc} + \gamma^{inc} \tilde{w}}(dim_{it} - \tilde{dim})\Big) \tag{16}$$

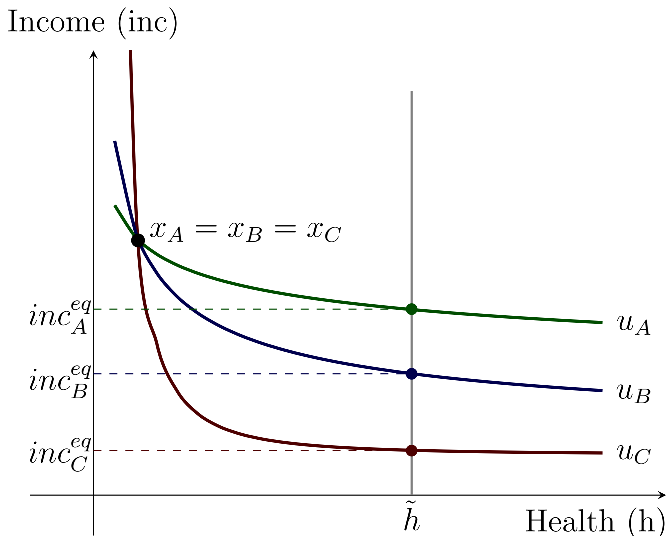
We choose the mean (mode for categorical variables) of the preference heterogeneity variables as the norm preferences. $\tilde{inc}_{it}^{eq}$ does not depend on an individual's preferences, and has thus taken the preference heterogeneity out of the original equivalent incomes. We can now express an individual's equivalent income as:

$$inc_{it}^{eq} = \tilde{inc}_{it}^{eq} + \phi_{it}, \tag{17}$$

where ϕ_{it} is the contribution to equivalent incomes individuals receive from the match between their particular bundle and their preferences. An illustrative example is given in Figure 3A.1. We consider three individuals, A, B, and C, and two dimensions of well-being, income and health. A, B, and C all consume the same bundle, which contains a lot of income but only little health. They have different preferences, though, with A putting the largest preference on income relative to health compared to B and C. It seems fitting that A should derive the most welfare from this income-heavy bundle, which also is the case when we calculate the three individuals' equivalent incomes.

Suppose that A, B, and C have different preferences because of variation in a circumstance variable, for example their parents' level of education. If we use the equivalent income ranks as the welfare variable that is regressed on effort and circumstance variables, which include parental education, then the coefficient on parents' education will eat up the differences in equivalent incomes. When we construct the unfairness ranks, A will have a higher unfairness rank than B and C, because A's circumstances – through his preferences – are yielding a

Figure 3A.1: Equivalent Income Illustration



higher well-being rank. This means that well-being differences due to preference heterogeneity from circumstances are not considered fair, and individuals are not held fully responsible for their preferences.

Our decomposition can circumvent this problem. In this example, we choose individual B's preferences as the norm preferences. Our decomposition asks what level of equivalent incomes each individual would have had with B's preferences. In this case, all three individuals would have $\tilde{inc}^{eq}_i = inc^{eq}_B$ for $i = A, B, C$. This means that $\phi_A = inc^{eq}_A - inc^{eq}_B > 0$ and $\phi_C = inc^{eq}_C - inc^{eq}_B < 0$, implying that A gets a positive boost from the match between his bundle of goods and his preferences, while the reverse applies to C.

In essence, we want to hold individuals responsible for their ϕ-term. Suppose we use ranks of $\tilde{inc}^{eq}_{it}$ instead of inc^{eq}_{it} to regress on circumstance and effort variables. Since the ϕ-terms are not part of these ranks, the circumstance variables can no longer pick up well-being differences due to preference heterogeneity. Hence, when we calculate the unfair ranks, preference heterogeneity does not enter, and individuals are held responsible for their preferences. We still use the baseline equivalent incomes, inc^{eq}_{it}, to calculate the welfare ranks we use for the divergence measures, but the unfairness ranks will be based on a regression with $\tilde{inc}^{eq}$.

The above method applies to unfairness ranks. Unfortunately, the method does not allow us to calculate fairness ranks that hold individuals responsible for their preferences. If we follow the same approach as above, and hence regress the ranks of $\tilde{inc}^{eq}_{it}$ on circumstance and effort variables, we face the problem that the fairness ranks should be composed of both the contribution from the effort variables and the ϕ_{it}-terms. As these two terms are on different scales, it is not entirely clear how to add them and turn them into a ranked-based measure. Hence, we will only use the decomposition of equivalent incomes for divergences between $\boldsymbol{r^y}$ and $\boldsymbol{r^{unfair}}$.

3.A.2 Regression Output

Table 3A.1: Regressing Welfare Ranks on Circumstances and Effort

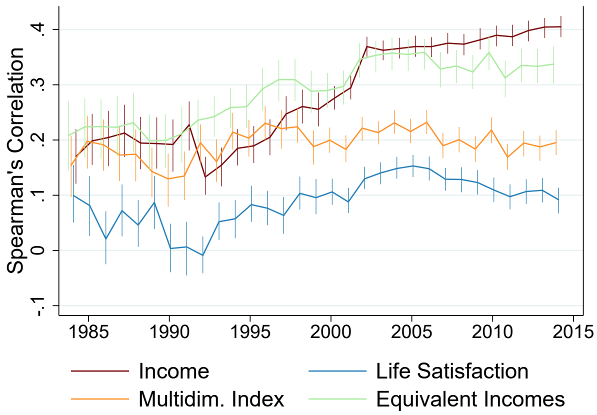
	Income		Life Satisfaction		Multidim. Index		Equivalent Income	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Circumstance Variables								
Father's education (ref: primary school)								
Secondary school	4.37***	(0.48)	1.79***	(0.51)	2.56***	(0.48)	3.44***	(0.46)
Tertiary school	7.81***	(0.64)	2.98***	(0.66)	4.14***	(0.63)	6.76***	(0.62)
Mother's education (ref: primary school)								
Secondary school	2.38***	(0.46)	1.10**	(0.49)	1.13**	(0.46)	1.15**	(0.45)
Tertiary school	2.86***	(0.76)	2.11***	(0.76)	2.56***	(0.77)	2.33***	(0.73)
Father's occup. (ref: blue collar, untrained)								
Blue-collar, trained	3.08***	(0.53)	0.73	(0.57)	1.20**	(0.53)	2.30***	(0.49)
Not employed	2.44***	(0.76)	1.15	(0.81)	0.68	(0.76)	1.02	(0.70)
White-collar	8.98***	(0.59)	2.50***	(0.63)	4.17***	(0.58)	6.71***	(0.55)
Self-employed	7.81***	(0.68)	2.65***	(0.70)	2.58***	(0.66)	5.73***	(0.65)
Civil servant	9.13***	(0.76)	3.43***	(0.81)	4.51***	(0.74)	7.19***	(0.73)
Place of upbringing (ref: large city)								
Medium city	-0.90*	(0.52)	0.79	(0.54)	0.90*	(0.51)	0.35	(0.50)
Small city	-1.11**	(0.49)	0.93*	(0.50)	0.87*	(0.48)	0.64	(0.47)
Countryside	-1.16***	(0.45)	0.32	(0.47)	-0.50	(0.44)	0.55	(0.43)
Place of Birth (ref: West Germany)								
East Germany	-15.53***	(0.38)	-10.08***	(0.39)	-10.04***	(0.37)	-13.45***	(0.33)
Abroad	-12.56***	(0.65)	-0.60	(0.69)	-4.10***	(0.64)	-5.07***	(0.61)
Height (cm)	0.29***	(0.02)	0.18***	(0.03)	0.16***	(0.02)	0.24***	(0.02)
Female	2.75***	(0.45)	2.49***	(0.47)	-1.33***	(0.45)	-2.67***	(0.43)
Number of siblings	-1.71***	(0.10)	-0.26**	(0.11)	-0.75***	(0.10)	-1.18***	(0.09)
Age	1.31***	(0.07)	-0.64***	(0.07)	-0.85***	(0.07)	1.07***	(0.06)
Age squared	-0.01***	(0.00)	0.01***	(0.00)	0.01***	(0.00)	-0.01***	(0.00)
Effort Variables								
Years of education	2.77***	(0.07)	0.79***	(0.07)	1.07***	(0.07)	2.01***	(0.07)
Weekly working time	0.42***	(0.01)	0.18***	(0.01)	0.46***	(0.01)	0.57***	(0.01)
Self-employed	5.55***	(0.56)	-0.21	(0.53)	1.72***	(0.53)	5.38***	(0.51)
Works in public sector	2.01***	(0.32)	2.71***	(0.35)	1.88***	(0.34)	1.77***	(0.35)
r^2	0.30		0.05		0.12		0.23	

Notes: * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$. Based on a pooled OLS with individual level clustered standard errors. The dependent variables range from 0-100 with 100 representing the largest yearly welfare and vice versa. $n = 176,196$.

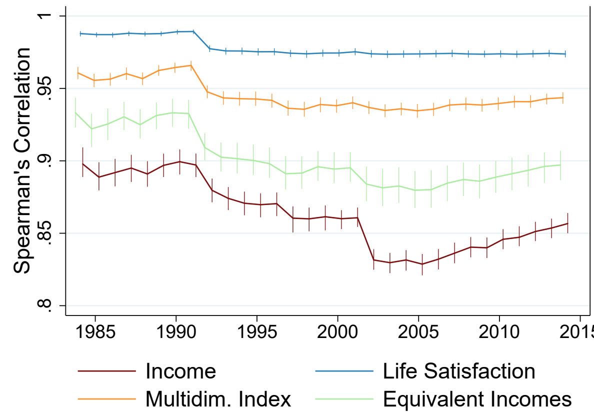
3.A.3 Robustness Checks

Figure 3A.2: Altering the Responsibility Cut, $D(\boldsymbol{r^y} \| \boldsymbol{r^{fair}})$

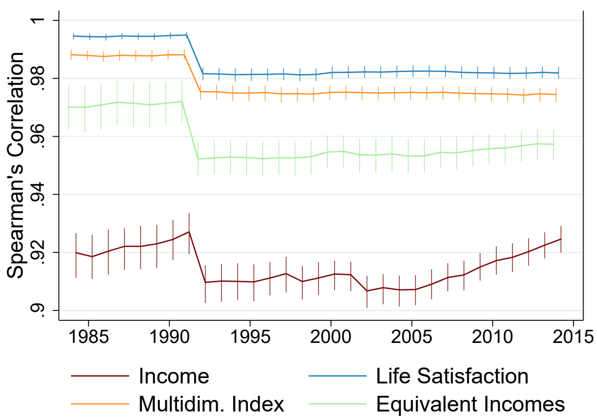
(a) Effort: 4var



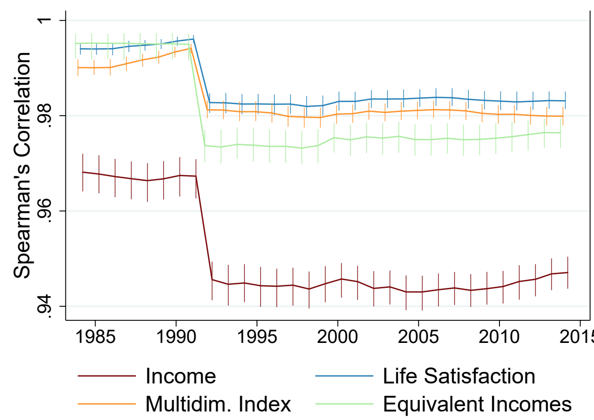
(b) Effort: Residual



(c) Effort: 4var, residual



(d) Effort: 4var, residual, cor



(e) Effort: 4var, residual, cor, age

Notes: Inequality of opportunity from 1984-2014 for different responsibility cuts. *4var:* Work hours, education, self-employed, and works in public sector are considered effort. *Residual:* The residuals from the regressions of well-being on circumstance and effort variables are considered effort. *Age:* Age is considered effort (implying we are equalizing lifetime opportunities). *Cor:* The correlation between effort and circumstance variables is not considered a circumstance. Our method for decomposing equivalent incomes cannot be used for generating fairness ranks, hence we do not show a figure where individuals are held responsible for their preferences (see App. 3.A.1 for details).

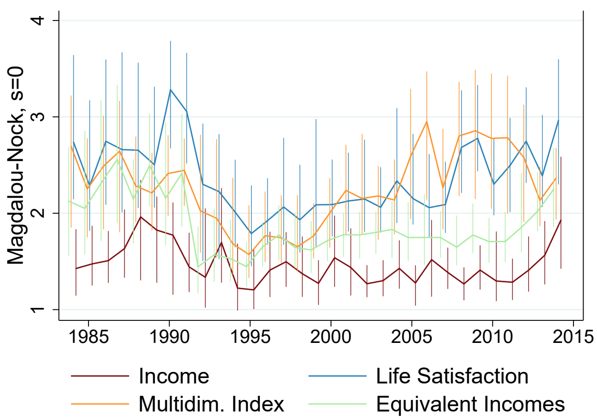
Figure 3A.3: Changing the Divergence Measure, $D(\boldsymbol{r^y} \| \boldsymbol{r^{fair}})$

(a) Magdalou-Nock, $s = 0$ | (b) Magdalou-Nock, $s = 1$ | (c) Magdalou-Nock, $s = 2$ | (d) Fairness Gini

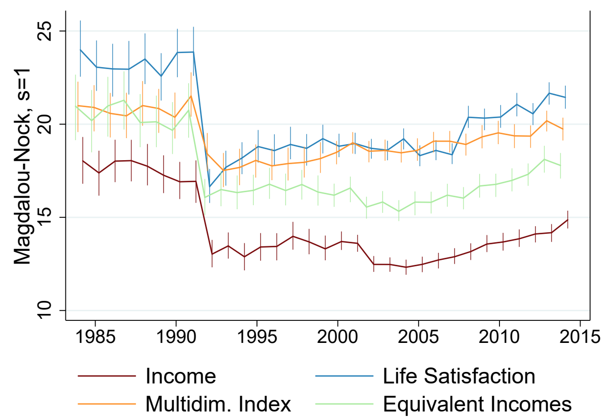
Notes: Development in inequality of opportunity in each well-being variable from 1984-2014 using different divergence measures.

Figure 3A.4: Changing the Divergence Measure, $D(\boldsymbol{r}^{\boldsymbol{y}} \| \boldsymbol{r}^{\boldsymbol{unfair}})$

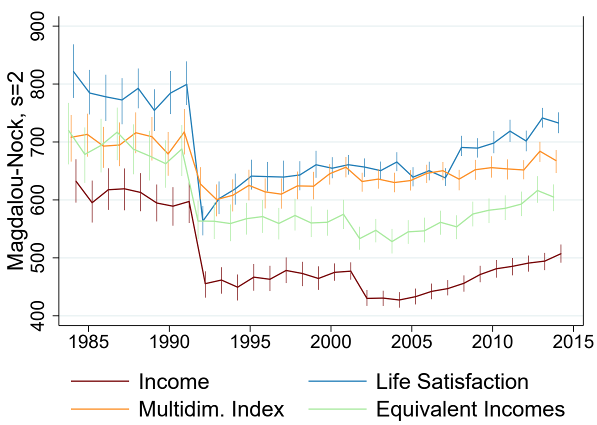
(a) Magdalou-Nock, $s = 0$



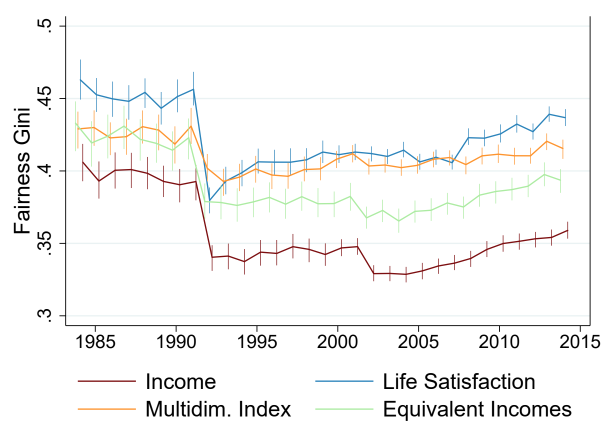
(b) Magdalou-Nock, $s = 1$



(c) Magdalou-Nock, $s = 2$



(d) Fairness Gini



Notes: Development in inequality of opportunity in each well-being variable from 1984-2014 using different divergence measures.

Chapter 4

The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees

The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees[1]

Paolo Brunori
University of Florence

Paul Hufe
ifo Munich and LMU Munich

Daniel Gerszon Mahler
University of Copenhagen

Abstract

We propose a set of new methods to estimate inequality of opportunity based on conditional inference regression trees. We illustrate how these methods represent a substantial improvement over existing empirical approaches to measure inequality of opportunity. First, they minimize the risk of arbitrary and ad hoc model selection. Second, they provide a standardized way to trade off upward and downward biases in inequality of opportunity estimations. Finally, regression trees can be graphically represented; their structure is immediate to read and easy to understand. This makes the measurement of inequality of opportunity more easily comprehensible to a large audience. These advantages are illustrated by an empirical application based on the 2011 wave of the European Union Statistics on Income and Living Conditions.

Keywords: Inequality of opportunity, machine learning, random forests
JEL codes: C51, C52, D63

[1]We are grateful for comments received from participants at presentations held at the University of Copenhagen, the University of Essex, the World Bank, and the 13th Winter School on Inequality and Social Welfare Theory.

4.1 Introduction

John Roemer's (1998) seminal contribution, *Equality of Opportunity*, has incited a flourishing empirical literature on the measurement of unequal opportunities. At the heart of Roemer's formulation is the idea that factors that determine individual outcomes can be divided into two categories: factors over which individuals have control, which he calls *effort*, and factors for which individuals cannot be held responsible, which he calls *circumstances*. Individuals characterized by exactly the same exogenous circumstances are assigned to a circumstance *type*. Members of a type have the same background conditions to transform resources into outcomes. Therefore, while within-type inequality, as caused by the differential exertion of effort, is morally irrelevant, between-type differences in achievements are inequitable and call for compensation. Thus, opportunity-equalizing policies have the objective of neutralizing the impact of circumstances on the distribution of the desirable outcome.

Following Roemer's approach, a battery of methods to measure inequality of opportunity have been proposed (see Van de gaer and Ramos, 2016; Roemer and Trannoy, 2015, for recent overviews).[2] Today, well established empirical methods include summary indexes that quantify the extent of unequal opportunities (Bourguignon *et al.*, 2007; Checchi and Peragine, 2010; Almås *et al.*, 2011) as well as statistical tests that detect the mere existence thereof (Lefranc *et al.*, 2009; Kanbur and Snell, 2017). In either case, empirical results are sensitive to critical choices of model selection, which are under complete discretion of the researcher.

First, researchers have to make a decision on which circumstance variables to consider for estimation.[3] Observable circumstances beyond individual control are typically a subset of the real number of exogenous variables affecting individual outcomes. This issue has been largely discussed by the literature, and the prevailing view is that partial observability implies downward-biased inequality of opportunity estimates (Ferreira and Gignoux, 2011). To counteract this downward bias, one strategy is to resort to high-quality datasets that provide very detailed information with respect to individual circumstances (Hufe *et al.*, 2017). Naturally, the scope of improvement of this approach is limited by sample sizes. Consider for example the increasing availability of genetic datasets with billions of polymorphisms per person (Altshuler *et al.*, 2015). While the genetic make-up of individuals clearly is beyond individual control and must be considered a circumstance, the number of circumstances exceeds the available degrees of freedom, which forces the researcher to choose selectively

[2] Note that a number of contributions from the social choice literature on fair allocation had previously proposed similar methods (Van de gaer, 1993; Fleurbaey, 1995, 2008).

[3] Roemer does not provide a fixed list of variables that are to be considered as circumstances. Rather he suggests that the set of circumstances should evolve from a political process (Roemer and Trannoy, 2015). In empirical implementations, typical circumstances are biological sex, socioeconomic background, race, or the area of birth.

from the available set of circumstances.

Second, the influence of circumstances may be dependent on the expression of other circumstance characteristics. For example, it is a well-established finding that the influence of similar child-care arrangements on various life outcomes varies strongly by biological sex (García *et al.*, 2017). In contrast to such evidence, many empirical applications presume that the effect of circumstances on individual outcome is fixed and additive (Bourguignon *et al.*, 2007; Ferreira and Gignoux, 2011). On the one hand, analogous to partial observability, this functional form assumption forces a downward bias on inequality of opportunity estimates. On the other hand, limitations in the available degrees of freedom may prove the estimation of fully saturated models impractical. Again the researcher is left to her own devices in selecting the best model for estimating inequality of opportunity.

While the downward bias of inequality of opportunity estimates is prominently discussed in the extant literature, the reliability of estimates has been largely disregarded. Holding the sample size constant, increasing the type partition by including additional circumstances or relaxing the linearity assumption directly translates into reduced variation for estimating the relevant parameters. In fact, Brunori *et al.* (2016) show that overfitting the model may instill an upward bias on inequality of opportunity estimates.

This discussion highlights the non-trivial challenge in selecting the appropriate model for estimating inequality of opportunity. Scholars must balance between different sources of bias while trying to avoid ad-hoc solutions.

In this paper, we propose the use of classification and regression tree methods to address the outlined shortcomings of current approaches. Introduced by Morgan and Sonquist (1963) and popularized by Breiman *et al.* (1984), classification and regression trees belong to a family of statistical methods that are commonly summarized under the labels of "machine learning" or "statistical learning" (Friedman *et al.*, 2009). Originating from the fields of computer science and statistics, these methods are being increasingly adopted by economists (Varian, 2014; Mullainathan and Spiess, 2017; Athey, 2017). Classification and regression tree methods were developed to make out-of-sample predictions of a dependent variable based on a number of observable predictors. They let algorithms automatically segment the predictor space into non-overlapping regions to find the best model for predicting the outcome of interest. In the context of estimating equality of opportunity, this means that we let an automated algorithm decide how to partition the population into mutually exclusive types for the purpose of calculating measures of inequality of opportunity in the spirit of Roemer's theory. To be precise, within the class of classification and regression tree methods we focus on *conditional inference trees* and *conditional inference forests*, both of which bear a number of substantial advantages (Hothorn *et al.*, 2006).

First, by drawing on a clear-cut algorithm one minimizes the degree of arbitrariness in model selection. In both trees and forests, types are obtained in the attempt to explain outcome variability without assuming anything about which circumstances play a role in shaping individual opportunities and how they interact. Thus, the partition of the population into Roemerian types is no longer a judgment call of the researcher but a non-arbitrary outcome of data analysis. Second, the conditional inference algorithm branches trees (and constructs forests) by using a sequence of hypothesis tests that prevent model overfitting. Therefore, by using the conditional inference algorithm we can both derive a test for the null hypothesis of equal opportunity and avoid the potential upward bias of inequality of opportunity estimates. As a consequence of avoiding upwards and downwards biases, our estimates are better suited for comparisons across time and between countries when sample sizes differ.

Aside from those shared merits, trees and forests bear some distinct advantages, which the researcher needs to trade off when selecting her preferred approach within the class of conditional inference methods. Trees, on the one hand, have intuitive appeal and their graphical illustrations are instructive tools for longitudinal or cross-sectional comparisons of opportunity structures. Forests, on the other hand, perform better in trading off the different sources of bias outlined above. In fact, we will show that conditional inference forests outperform other prevalent estimation techniques in terms of out-of-sample prediction accuracy.

To be sure, just as in the literature on intergenerational mobility (Black and Devereux, 2011), scholars of equality of opportunity are reluctant to give their estimates a causal interpretation. The ambition of the literature is to understand how much variation in outcomes can be attributed to root causes that are commonly perceived as unfair. It is precisely the *prediction* character of these empirical exercises that makes this branch of the literature a useful field to leverage the advantages of machine learning algorithms.

The remainder of this paper is organized as follows: section 4.2 gives a brief introduction with respect to current empirical approaches in the literature. Section 4.3 introduces regression trees and illustrates how to use them in the context of inequality of opportunity estimations. An empirical illustration based on the EU Survey of Income and Living Conditions is contained in section 4.4, in which we will also highlight the particular advantages of using tree-based estimation methods. Section 4.5 concludes.

4.2 Empirical Approaches to Equality of Opportunity

Consider a population of size N indexed by $i \in \{1, ..., N\}$ and an associated vector of incomes $Y = \{y_1, ..., y_i, ..., y_N\}$. Individual i's outcome, y_i, is the result of two sets of factors. First, a set of *circumstances* beyond her control of size P: $\Omega_i = \{C_i^1, ..., C_i^p, ..., C_i^P\}$. Second, a set of *efforts* of size Q: $\Theta_i = \{E_i^1, ..., E_i^q, ..., E_i^Q\}$. In general, the outcome generating function $g : \Omega \times \Theta \rightarrow \mathbb{R}_+$ can therefore be written as

$$y_i = g(\Omega_i, \Theta_i). \tag{4.1}$$

Each circumstance $C^p \in \Omega$ is characterized by a total of X^p possible realizations, where each realization is denoted as x^p. Based on the realizations, x^p, we can partition the population into a set of non-overlapping *types*, $T = \{t_1, ..., t_m, ..., t_M\}$. A type is a subgroup of the original population uniform in terms of circumstances, i.e. individuals i and j belong to the same type $t_m \in T$ if $x_i^p = x_j^p \; \forall C^p \in \Omega$. They belong to different types $t_m \in T$ if $\exists C^p \in \Omega : x_i^p \neq x_j^p$. The number of types in the population is given by $M = \prod_{p=1}^{P} X^p$. Following Roemer (1998) we assume that the joint realizations of the effort variables $E^q \in \Theta$ can be summarized by a scalar $\pi \in [0, 1]$. Individuals sharing the same expression of effort are called a *tranche*. Hence, types and tranches define two particular ways of partitioning the population into subgroups, where group membership either indicates uniformity in circumstances (types) or effort (tranches).

In the literature we can distinguish two broad classes of equality of opportunity definitions.[4] First, the *ex-ante* view focuses on between-type differences in the value of opportunity sets without paying attention to the specific effort realizations of individual type members. According to this perspective, equality of opportunity is satisfied if the value of opportunity sets is equalized across types. One example is the *ex-ante utilitarian* perspective according to which the value of opportunity sets is indicated by the average outcome within the specific type. Thus, equality of opportunity would be realized if the mean outcome of each type was equal to the population mean. Second, the *ex-post* view focuses on individual outcomes conditional on effort exertion. According to this perspective, equality of opportunity would be satisfied if individual outcomes were equalized within each tranche, i.e. individuals with equal levels of effort exertion realize the same outcomes. A comprehensive discussion of the ex-ante and ex-post principles of equality of opportunity can be found in Fleurbaey and Peragine (2013). In the context of this paper, we will restrict ourselves to the ex-ante utilitarian approach only.

[4]Measures different from the ones illustrated here have been proposed in the literature. The interested reader is referred to Van de gaer and Ramos (2016) for a comprehensive overview.

Tests and Measures The extant literature has witnessed the development of empirical *tests* and *measures* for ex-ante utilitarian inequality of opportunity. A prominent example for the former category is provided by Lefranc *et al.* (2009), who show that rejecting the null hypothesis of no first-order stochastic dominance in type-specific outcome distributions is sufficient to reject the existence of equal opportunities in the population from an ex-ante utilitarian perspective. Furthermore, in a recent contribution Kanbur and Snell (2017) develop likelihood ratio tests that can serve to test for ex-ante utilitarian equality of opportunity.

A widely adopted example of the latter category, is the measure developed by Van de gaer (1993) and Checchi and Peragine (2010). They propose to measure inequality in a counterfactual distribution $Y^{EA} = \{y_1^{EA}, ..., y_i^{EA}, ..., y_N^{EA}\}$ obtained by removing inequality within types from the original distribution. To be precise, individual outcomes are re-scaled to match their respective type mean:

$$y_i^{EA} = \frac{1}{N_m} \sum_{i \in t_m} y_i = \mu_m, \ \forall i \in t_m, \forall t_m \in T, \tag{4.2}$$

where N_m is the size and μ_m the average outcome of type t_m. Therefore, any remaining inequality in Y^{EA} reflects inequality between types and inequality of opportunity can now be summarized by applying any standard scalar measure of inequality I$(\cdot)$, like the Gini index or a member of the generalized entropy class (Cowell, 2016), to the counterfactual distribution Y^{EA}. Any such measure obtains its minimal value in the case of equality of all type means, i.e. if $\mu_m = \mu_l = \mu \ \forall t_m, t_l \in T$.

Estimation In practice we do not observe the full set of circumstances Ω. Rather, we observe the subset $\check{\Omega} \subseteq \Omega$ of size $\check{P}$. For example, in most datasets we do not have full information on the genetic make-up of individuals. Neither do we have a gapless documentation of the socioeconomic conditions in which individuals grew up. Analogously, for most $C^p \in \check{\Omega}$ we only observe the subset $\check{X}^p$ of the true number of realizations X^p. For example, in many datasets information on parental education and occupation is coded in categorical variables of varying detail, which may mask more nuanced socioeconomic differences among households.

Depending on the strength of their distributional assumptions, estimations of inequality of opportunity are typically classified as either non-parametric or parametric. A point in case for the former approach is the abovementioned measure put forward by Van de gaer (1993) and Checchi and Peragine (2010). The researcher partitions the sample into mutually exclusive cells based on the realizations of all circumstance variables under consideration. Hence, the researcher makes no assumption on the interaction of circumstance variables in the determination of individual outcomes. This comes at a high cost. To avert overfitting, the partition must be constructed such that a sufficient number of observations belongs to each cell. Conditional on the dataset being rich enough in information on circumstances, this forces the researcher to make a discretionary choice on the *relevant* partition. Consider for instance a continuous circumstance variable like parental income. Employing the non-parametric estimation approach, the researcher must split the parental income distribution into quantiles for constructing the type partition. The potential granularity of this split obviously depends on the sample sizes of the ensuing cells. Additionally, the researcher must balance the informational content of a finer partition of parental income against the opportunity cost of being forced to exclude another circumstance variable from the investigation. To put it in formal terms: the researcher must select a subset $\hat{\Omega} \subseteq \check{\Omega} \subseteq \Omega$ from the set of observed circumstances. Furthermore, within the confines of limited degrees of freedom the researcher must also decide for each $C^p \in \hat{\Omega}$, how to restrict the number of realizations $\hat{X}^p \subseteq \check{X}^p \subseteq X^p$ in order to construct a statistically meaningful type partition.

To address this problem, the literature commonly resorts to parametric estimation approaches. Here, the researcher obtains the counterfactual distribution by estimating a Mincerian regression with circumstances as the sole right-hand side variables (Bourguignon *et al.*, 2007; Ferreira and Gignoux, 2011):

$$\ln(y_i) = \beta_0 + \sum_{p=1}^{\check{P}} \beta_p C_i^p + \epsilon_i. \tag{4.3}$$

The counterfactual distribution, Y^{EA} can then be constructed from the predicted values

$$y_i^{EA} = \exp\left[\sum_{p=1}^{\check{P}} \hat{\beta}_p C_i^p\right]. \tag{4.4}$$

Although the parametric approach solves some of the shortcomings of the non-parametric approach, it is not a panacea. The standard version of the parametric approach assumes a linear impact of all circumstances and therefore neglects the existence of interdependencies and non-linearities in the impact of circumstances. To pick up the example from the introduction, the researcher cannot allow for a differential impact of the same child-rearing arrangement on male and female children. Of course, to alleviate this shortcoming the researcher may integrate interaction terms and higher order polynomials into equation (4.3). At the extreme the researcher may even estimate a fully saturated model, in which case parametric and non-parametric estimation coincide. This congruence, however, reiterates the fundamental problem of current approaches towards the estimation of inequality of opportunity. In view of restrictions on the available degrees of freedom, the researcher is forced to make a discretionary choice on the model she estimates, which in itself is a strong determinant of the ensuing results when testing and measuring equality of opportunity. Furthermore, just as the non-parametric approach, the parametric estimation is at risk of overfitting the data when the set of circumstances is large.

In analogy to this paper, Li Donni *et al.* (2015) have discussed the issue of ad-hoc model selections in the empirical literature on equality of opportunity. To resolve this issue, they propose a data-driven type partition by estimating a latent class model. In this approach, observable circumstances are considered indicators of membership in an unobservable latent type, t_m. For each possible number of latent types, M, the model obtains the partition into types by minimizing the within-type correlation of observable circumstances, $C^p \in \check{\Omega}$. The optimal number of types is selected by minimizing an appropriate model selection criterion such as Schwarz's Bayesian Information Criterion (BIC). The latent class approach therefore partly solves the issue of arbitrary model selection. However, it cannot solve the problem of model selection once the potential number of type characteristics exceeds the available degrees of freedom. In these cases the latent class approach replicates the limitations of other prevalent approaches towards estimating inequality of opportunity: the researcher must pre-select the relevant set of circumstances, their subpartition as well as the respective interactions. To the contrary, our approach embodies a method to select circumstances from the set of all observed variables in a non-arbitrary fashion. Furthermore, latent types are constructed in the attempt to explain circumstances' correlation. The partition is therefore insensitive to the degree of association between circumstances and outcome. One may consider explaining outcome variability as function of circumstances to be precisely the purpose of inequality of opportunity measurement.[5] Lastly, we prefer the conditional inference approach as it provides the particular advantages of being econometrically more tractable while providing a stronger economic meaning of the identified types.

4.3 Estimating Inequality of Opportunity from Regression Trees

Originally, tree-based methods were developed to make out-of-sample predictions of a dependent variable based on a number of observable predictors. As we will outline in the following, they can be straightforwardly applied to equality of opportunity estimations and solve many

[5] This issue is common to any two-stage analysis in which latent classes serve as controls for a distal outcome. The effect of latent class membership on the distal outcome is attenuated and the explained variability is reduced (Lanza *et al.*, 2013).

of the issues associated with the prevalent estimations approaches outlined in section 4.2. While we put a particular emphasis on *regression* trees, our main arguments also hold for *classification* trees. Thus, the proposed estimation methods are not restricted to continuous variables like income, but can also be fruitfully employed with respect to non-continuous outcomes, such as health (Trannoy *et al.*, 2010) or education (Oppedisano and Turati, 2015).

In what follows we will present two tree-based estimation procedures both of which solve the model selection problem outlined in section 4.2. First, we will introduce conditional inference regression trees. Their simple graphical illustration is particularly instructive for longitudinal or cross-sectional comparisons of opportunity structures. Second, we will introduce conditional inference forests, which are – loosely speaking – a collection of many conditional inference trees. Forests do not have the intuitive appeal of regression trees. However, they perform significantly better in terms of out-of-sample predictions. In fact, we will show in section 4.4.5 that they outperform all other considered estimation techniques along this dimension.

4.3.1 Conditional Inference Trees

Tree-based methods obtain predictions for outcome y as a function of the input variables $I = \{I^1, ..., I^p, ...I^P\}$. Specifically, they use the set I to partition the population into a set of non-overlapping groups, $G = \{g_1, ..., g_m, ..., g_M\}$, where each group g_m is homogeneous in the expression of each input variable. These groups are also called *terminal nodes* or *leafs* in a regression tree context. The predicted value for outcome y of observation i is calculated from the mean outcome μ_m of the group g_m to which the individual is assigned. Hence, in addition to the observed income vector $Y = \{y_1, ..., y_i, ..., y_N\}$ one obtains a vector of predicted values $\hat{Y} = \{\hat{y}_1, ..., \hat{y}_i, ..., \hat{y}_N\}$, where

$$\hat{y}_i = \mu_m = \frac{1}{N_m} \sum_{i \in g_m} y_i, \ \forall i \in g_m, \forall g_m \in G. \tag{4.5}$$

The mapping from regression trees to equality of opportunity estimation is straightforward. Conditional on the input variables being circumstances only, i.e. $I \subseteq \check{\Omega} \subseteq \Omega$, it is evident that each resulting group $g_m \in G$ can be interpreted as a circumstance type $t_m \in T$. Furthermore, $\hat{Y}$ is analogous to the smoothed distribution Y^{EA}, the construction of which we have outlined in section 4.2 to illustrate ex-ante utilitarian measures of inequality of opportunity. In view of the fact that our predictor space is confined to circumstance variables only, we use the terms "input variables" and "circumstances" as well as "groups" and "types" interchangeably in the following. Input variables will be denoted by C^p and groups by t_m. In line with equation

(4.5), we will refer to individual predictions $\hat{y}_i$ as μ_m.

Algorithm Considering all possible ways in which the population can be split into groups is a daunting task when the set of input variables is large. In conventional estimation approaches the researcher is left to her own devices in (i) selecting $\hat{\Omega}$ from $\check{\Omega}$, (ii) to restrict the number of realizations of each $C^p \in \hat{\Omega}$, and (iii) to determine the relevant interactions among all $C^p \in \hat{\Omega}$. The magnitude of this choice set oftentimes leads to arbitrary model selection. To the contrary, with regression trees the researcher does not need to make these choices herself. The researcher only submits the full and unrestricted set of observed variables that qualify as circumstances, $\check{\Omega}$, while the algorithm chooses the relevant circumstances, their subpartition and the respective interactions. To be precise, the observations are divided into M groups (or types) by what is known as *recursive binary splitting*. Recursive binary splitting starts by dividing the full sample into two distinct groups according to the value they take in one input variable C^p. If C^p is a continuous or ordered variable, then $i \in t_m$ if $C_i^p < x^p$ and $i \in t_l$ if $C_i^p \geq x^p$. If C^p is a categorical variable then the categories can be split into any two arbitrary groups. The process is continued such that one of the two groups is divided into further subgroups (potentially based on another $C^p \in \check{\Omega}$), and so on. Graphically, this division into groups can be presented like an upside-down tree (Figure 4.1).

The exact manner in which the split is conducted depends on the type of regression tree that is used. In this paper we follow the methodology proposed by Hothorn *et al.* (2006), leading to what they call conditional inference trees.[6]

Conditional inference trees are grown by a series of permutation tests according to the following 4-step procedure:

1. Test the null hypothesis of independence, $H_0^{C^p} : D(Y|C^p) = D(Y)$, for each input variable $C^p \in \check{\Omega}$, and obtain a p-value associated with each test, p^{C^p}.

   ⇒ Adjust the p-values for multiple hypothesis testing, such that $p_{adj.}^{C^p} = 1-(1-p^{C^p})^P$ (Bonferroni Correction).

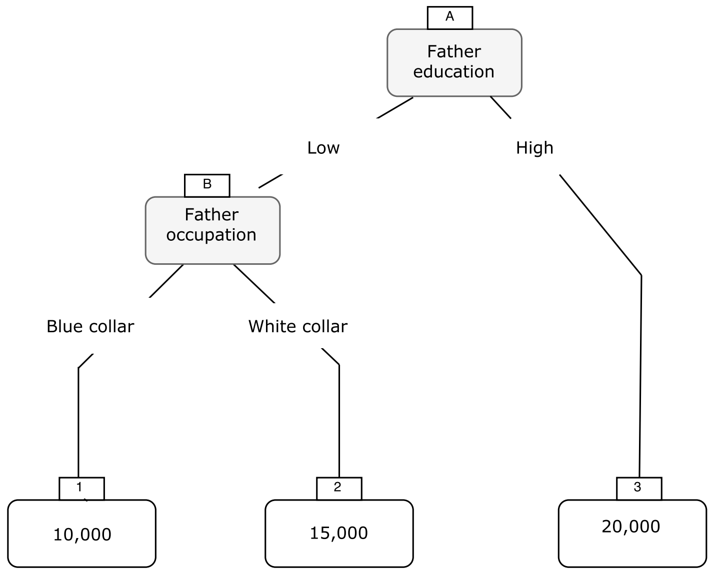
2. Select the variable, C^*, with the lowest adjusted p-value, i.e. $C^* = \{C^p : \text{argmin}\, p_{adj.}^{C^p}\}$.

   ⇒ If $p_{adj.}^{C^*} > \alpha$: Exit the algorithm.

   ⇒ If $p_{adj.}^{C^*} \leq \alpha$: Continue, and select C^* as the splitting variable.

[6] An alternative would be Classification and Regression Trees (CART) as introduced by Breiman *et al.* (1984). CART chooses splits so as to minimize the mean squared error, $\text{MSE} = \frac{1}{N}\sum_m \sum_{i \in t_m}(y_i - \mu_m)^2$. We prefer conditional inference trees since CART are biased towards variables with many possible splitting points (Hothorn *et al.*, 2006). Furthermore, we avoid the intricacies of tree *pruning* (Friedman *et al.*, 2009) by establishing a test criterion that considers the bias-variance trade-off before making an additional split.

Figure 4.1: Exemplary Tree Representation



Note: Artificial example of a regression tree. The grey boxes indicate splitting points, while the white boxes indicate terminal nodes. The values inside the white boxes show predicted values associated with each terminal node (μ_m).

3. Test the discrepancy between the subsamples for each possible binary partition, s, based on C^*, i.e. $Y_s = \{Y_i : C_i^* < x^p\}$ and $Y_{-s} = \{Y_i : C_i^* \geq x^p\}$, and obtain a p-value associated with each test, $p^{C_s^*}$.

   ⇒ Split the sample based on $C_{s^*}^*$, by choosing the split point s that yields the lowest p-value, i.e. $C_{s^*}^* = \{C_s^* : \text{argmin}\, p^{C_s^*}\}$.

4. Repeat the algorithm for each of the resulting subsamples.

In words, conditional inference trees start by testing the relationship between income and each circumstance separately. The circumstance that is most related to incomes is chosen as the splitting variable, if the dependence between the two is sufficiently strong. If it is not, no split is made. Whenever a circumstance can be split in several ways, the sample is split into two such that the two resulting groups' dependence with the income variable is maximized. This procedure is repeated in each of the two subsamples.

Conditional inference trees offer a particularly relevant structure in the context of inequality of opportunity. Each hypothesis test is essentially a test for whether equal opportunities exist within a particular (sub)sample. If the algorithm results in no splits at all, then we cannot reject the null hypothesis of equality of opportunity. The deeper the tree is grown, the more types are necessary to fully account for the inherent inequality of opportunities in the society under consideration. Each split tells us that the resulting types have significantly different opportunities under an ex-ante utilitarian interpretation. In all of the resulting types (i.e. the terminal nodes of the tree), we cannot reject the null of equal opportunities.

Tuning Note that the structure and depth of the resulting opportunity tree hinges crucially on the level of α. The less stringent the α-requirement, the more we allow for false positives, i.e. the more splits will be detected as significant and the deeper the tree will be grown. So how should α be chosen? On the one hand, α can be chosen a priori in line with the disciplinary convention to require significance levels of at least 5% or even 1%. On the other hand, we can let the data speak on the optimal specification of α, i.e. we can *tune* the α-parameter to find a model that performs optimally according to a pre-specified testing criterion.

If opting for the latter option, α is typically chosen by K-fold cross-validation (CV). To perform cross validation, one starts by splitting the sample into K subsamples, also called folds. Then, one implements the conditional inference algorithm on the union of $K-1$ folds for varying levels of α, while leaving out the kth subsample. This makes it possible to compare the predictions emanating from the $K-1$ folds with the real, unused data points observed in the kth fold. The mean squared prediction error serves as an evaluation criterion:

$$\text{MSE}_k^{CV}(\alpha) = \sum_m \frac{N_m^k}{N^k} \sum_{i \in t_m} \frac{1}{N_m^k} (y_i^k - \mu_m(\alpha))^2. \tag{4.6}$$

This exercise is repeated for all K folds, so that $\text{MSE}^{CV}(\alpha) = \frac{1}{K}\sum_k \text{MSE}_k^{CV}(\alpha)$. One then chooses the α^* that delivers the lowest $\text{MSE}^{CV}(\alpha)$:

$$\alpha^* = \{\alpha \in A : \text{argmin } \text{MSE}^{CV}(\alpha)\}.^{7} \tag{4.7}$$

In our empirical application we fix $\alpha^* = 0.01$, which is in line with the disciplinary convention for hypothesis tests. However, we provide a sensitivity check using cross-validated α's in Figure 4A.1 of Appendix 4.A.3.

[7] One may argue that a criterion that evaluates models according to their capacity to predict *individual* outcomes is misplaced for ex-ante utilitarian inequality of opportunity estimations. Afterall, we have demonstrated above that we are mainly concerned with estimating type means rather than individual outcomes. In Appendix 4.A.1 we show that the MSE-criterion and its handling of the variance-bias trade-off straightforwardly extends to balancing upward and downward biases in inequality of opportunity estimations.

4.3.2 Conditional Inference Forests

Regression trees solve the model selection problem outlined in section 4.2 and provide a simple and non-arbitrary way of dividing the population into types. Furthermore, trees are easily mapped and thus lay bare the opportunity structure of a given society for a larger audience. However, constructing the counterfactual distribution Y^{EA} from conditional inference trees suffers from two shortcomings: first, they only make limited use of the information inherent in the set of observed circumstances since not all $C^p \in \check{\Omega}$ are used for the construction of the tree. Yet, the omitted circumstances may possess some informational content that can increase predictive power even though they are not significantly associated with Y at level α^*. This is a particular issue if two important circumstances are highly correlated. Once a split is done using either of the two, the other will unlikely yield enough information to cause another split. Second, the predictions and thus the values of opportunity sets, μ_m, emanating from trees have a high variance. The structure of trees - and therefore the ensuing distribution Y^{EA} - is fairly sensitive to alternations in the respective data samples. This is a particular issue if there are various circumstances that are close competitors for defining the first split (Friedman *et al.*, 2009). In what follows we will introduce conditional inference forests, which build methodologically on conditional inference trees and are able to deal with both of these shortcomings (Breiman, 2001).

Algorithm In short, random forests create many trees and average over all of these when making predictions. Trees are constructed according to the same 4-step procedure outlined in the previous subsection. However, two tweaks are made. First, each tree is estimated on a random subsample b of the original data.[8] In total B such trees are estimated. Second, a random subset of circumstances $\bar{\Omega} \subseteq \check{\Omega}$ of size $\bar{P}$ is allowed to be used at each splitting point. Together these two tweaks remedy the shortcomings of single conditional inference trees. Drawing only on subsets $\bar{\Omega} \subseteq \check{\Omega}$ increases the likelihood that all circumstances with informational content at some point will be identified as the splitting variable C^* and thus addresses the limited information use of regression trees. Furthermore, averaging over the B predictions cushions the variance of individual predictions μ_m and thus addresses the second shortcoming identified in relation with single regression trees. Therefore, predictions are formed as follows:

$$\hat{y}_i(\alpha, \bar{P}, B) = \frac{1}{B} \sum_{b=1}^{B} \mu_m^b(\alpha, \bar{P}). \tag{4.8}$$

[8] Alternatively, one can draw bootstrapped samples, i.e. sample with replacement until a dataset with the same size as the original data is reached. We use the subsampling technique since it has been shown that using bootstrapping leads to biased variable selection (Strobl *et al.*, 2007).

Tuning From equation (4.8) it is evident that individual predictions are a function of α – the significance level governing the implementation of splits –, $\bar{P}$ – the number of circumstances to be considered at each splitting point –, and B – the number of subsamples to be drawn from the data. Again, these parameters can be imposed a priori by the researcher or they can be determined by tuning the three-dimensional grid $(\alpha, \bar{P}, B)$ to optimize the out-of-sample fit of the model. In our empirical illustration we proceed as follows. First, to reduce computational costs we fix B at a level at which the marginal gain of drawing an additional subsample in terms of out-of-sample prediction accuracy becomes negligible.[9]

Second, we determine α^* and $\bar{P}^*$ by minimizing the *out-of-bag* error. This entails the following four steps for a grid of values of α and $\bar{P}$:

1. Run a random forest with B subsamples, where $\bar{P}$ circumstances are randomly chosen to be considered at each splitting point, and α is used as the value for the hypothesis tests.

2. Calculate the average predicted value of observation i using each of the subsamples b_{-i} (the so called *bags*) in which i does not enter: $\hat{y}_i^{OOB}(\alpha, \bar{P}) = \frac{1}{B_{-i}} \sum_{b_{-i}} \mu_m^b(\alpha, \bar{P})$.

3. Calculate the out-of-bag mean squared error: $\text{MSE}^{OOB}(\alpha, m) = \frac{1}{N} \sum_i [y_i - \hat{y}_i^{OOB}(\alpha, \bar{P})]^2$.

4. Choose $(\alpha^*, \bar{P}^*) = \{(\{\alpha \in A\}, \{\bar{P} \in \check{P}\}) : \text{argmin}\, \text{MSE}^{OOB}\}$.

The logic behind this tuning exercise is similar to cross-validation. However, instead of leaving out the kth fraction of the dataset to make out-of-sample predictions, we leverage the fact that each tree of a forest is grown on a subsample b_{-i} that excludes all observations i. Hence, for each tree we can use the out-of-bag data points to evaluate the predictive accuracy of the respective model.[10]

The improved predictive quality of random forests comes at a cost. It is no longer possible to identify a fixed set of types T into which we can partition the population. For example, depending on the subset $\bar{\Omega} \subseteq \check{\Omega}$ used for a particular tree as well as the the particular subsample b drawn from the data, it may be that $i, j \in t_m^b$ but $i \in t_m^{b+1}$ while $j \notin t_m^{b+1}$. As a consequence, the individual prediction and hence the valuation of the individual opportunity set is an average over the value of opportunity sets μ_m^b associated with each tree of the

[9] Empirical tests show that this is the case with $B^* = 200$ for most countries in our sample (see Figure 4A.2 of Appendix 4.A.3).

[10] In principle tuning can be conducted analogously to regression trees by means of k-fold cross validation. This, however, is computationally expensive. Cross-validation would require to repeat the entire estimation exercise for a total of K folds. This is not necessary when using the out-of-bag error since out-of-sample points are already delivered by leaving out observations i when using bag b_{-i}. Hence, in the case of forests using the out-of-bag error is K times more computationally efficient than cross-validation.

forest. Therefore, the valuation of opportunity sets is less straightforward and opportunity structures are hard to illustrate in a graphical manner. It is nevertheless possible to describe opportunity structures by calculating the relative variable importance of each $C^p \in \check{\Omega}$ in constructing the forest (see section 4.4.3 for an illustration).

4.4 Empirical Application

In this section we provide an illustration of our methodology using harmonized survey data from 31 European countries. As outlined above, conditional inference trees and random forests solve the issue of model selection associated with the prevalent approaches to equality of opportunity estimations. Conditional inference trees are easily tractable and lend themselves to cross-sectional (and longitudinal) comparisons of opportunity structures. Conditional inference forests are less tractable but outperform the former approach in terms of predictive accuracy. In the following, we will illustrate the merits of both approaches. Furthermore, we will compare the results from both versions of our method with prevalent measurement approaches in the extant literature; namely parametric, non-parametric and latent class models. Comparisons will be made along two dimensions. First, the estimates themselves, and second, the respective out-of-sample accuracy. The latter criterion should be interpreted as an indicator of how well the respective method balances upward and downward biases in inequality of opportunity estimations. A formal argument for why this is the case, is provided in Appendix 4.A.1.

4.4.1 Data

The empirical illustration is based on the 2011 wave of the European Union Statistics on Income and Living Conditions (EU-SILC). EU-SILC provides harmonized survey data with respect to incomes, poverty, and living conditions on an annual basis and covers a cross-section of 31 European countries in the 2011 wave.[11] We draw on the 2011 wave since it contains an ad-hoc module about the intergenerational transmission of (dis)advantages, which allows us to construct finely-grained type partitions. The set of observed circumstances $\check{\Omega}$ and their respective expressions x^p are listed in Table 4.1. Descriptive statistics concerning circumstances are reported in Appendix 4.A.2. As an additional advantage, EU-SILC has been extensively studied by the empirical literature on inequality of opportunity and thus

[11]The sample consists of Austria (AT), Belgium (BE), Bulgaria (BG), Switzerland (CH), Cyprus (CY), Czech Republic (CZ), Germany (DE), Denmark (DK), Estonia (EE), Greece (EL), Spain (ES), Finland (FI), France (FR), Croatia (HR), Hungary (HU), Ireland (IE), Iceland (IS), Italy (IT), Malta (MT), Lithuania (LT), Luxembourg (LU), Latvia (LV), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Sweden (SE), Slovenia (SI), Slovak Republic (SK), and Great Britain (UK).

provides appropriate benchmarks against which we can compare our method (Marrero and Rodríguez, 2012; Palomino *et al.*, 2016; Checchi *et al.*, 2016).

The unit of observation is the individual, whereas the outcome of interest is equivalent disposable household income. Aware that inequality statistics tend to be heavily influenced by outliers (Cowell and Victoria-Feser, 1996), we adopt a standard winsorization method according to which we set all non-positive incomes to 1 and scale back all incomes exceeding the 99.5th percentile of the country-specific income distribution to this lower threshold. Our analysis is focused on the working age population. Therefore, we restrict the sample to respondents aged between 30 and 59. To assure the representativeness of our country samples, all results are calculated by using appropriate individual cross-sectional weights.

Table 4.2 shows considerable heterogeneity in the income distributions of our country sample. While households in Norway (NO) and Switzerland (CH) on average obtained incomes above €40,000 in 2010, the average households in Bulgaria (BG), Romania (RO) and Lithuania (LT) did not exceed the €5,000 mark. The lowest inequality prevails in the Nordic countries of Norway (NO), Sweden (SE) and Iceland (IS), all of which are characterized by Gini coefficients below 0.22. At the other end of the spectrum we find the Eastern European countries of Latvia (LV), Lithuania (LT) and Romania (EL) with Gini coefficients well above 0.33.

4.4.2 Benchmark Methods

We compare our estimates from trees and forests against three benchmark methods that have been proposed in the extant literature.

First, we draw on the parametric approach as proposed by Bourguignon *et al.* (2007) and Ferreira and Gignoux (2011). In line with equation (4.3), estimates are obtained by a Mincerian regression of equivalent household income on the following circumstances: father's occupation (10 categories), father and mother's education (five categories), area of birth (three categories), and tenancy status of the household (two categories). The model specification therefore includes 20 binary variables and resembles the specification used in Palomino *et al.* (2016).[12]

Second, we draw on the non-parametric approach as proposed by Checchi and Peragine

[12] We have estimated the predicted outcomes both as the exponential of the predicted log outcome, $y_i^{EA} = \exp\left[\sum_{p=1}^{\check{P}} \hat{\beta}_p C_i^p\right]$, and by introducing, assuming a normally distributed error term, the correction $y_i^{EA} = \exp\left[\sum_{p=1}^{\check{P}} \hat{\beta}_p C_i^p + \sigma^2/2\right]$, where σ^2 is the estimated variance of the error term. We do not find any significant differences in the level of estimated inequality of opportunity when introducing the correction. This may explain why the need of such correction has never been explicitly discussed in previous contributions.

Table 4.1: List of Circumstances

- Respondent's sex:
    - Male
    - Female
- Respondent's country of birth:
    - Respondent's present country of residence
    - European country
    - Non-European country
- Presence of parents during childhood:
    - Both present
    - Only mother
    - Only father
    - Without parents
    - Lived in a private household without any parent
- Number of adults (aged 18 or more) in respondent's childhood home
- Number of working adults (aged 18 or more) in respondent's childhood home
- Number of children (under 18) in respondent's childhood home
- Father/mother country of birth and citizenship:
    - Born/citizen of the respondent's present country of residence
    - Born/citizen of another EU-27 country
    - Born/citizen of another European country
    - Born/citizen of a country outside Europe
- Father/mother education (based on international Standard Classification of Education 1997 (ISCED-97)):
    - Unknown father/mother
    - Illiterate
    - Low (0-2 ISCED-97)
    - Medium (3-4 ISCED-97)
    - High (5-6 ISCED-97)
- Father/mother occupational status:
    - Unknown or dead father/mother
    - Employed
    - Self employed
    - Unemployed
    - Retired
    - House worker
    - Other inactive
- Father/mother main occupation (based on International Standard Classification of Occupations, published by the International Labour Office ISCO-08):
    - Managers (I-01)
    - Professionals (I-02)
    - Technicians (I-03)
    - Clerical support workers (I-04)
    - Service and sales workers (including also armed force) (I-05 and 10)
    - Skilled agricultural, forestry and fishery workers (I-06)
    - Craft and related trades workers (I-07)
    - Plant and machine operators, and assemblers (I-08)
    - Elementary occupations (I-09)
    - Father/mother did not work, was unknown or was dead (I-0)
- Managerial position of the father/mother:
    - Supervisory
    - Non-supervisory
- Tenancy status of the respondent's childhood home:
    - Owned
    - Not owned

Table 4.2: Summary Statistics

Country	Sample size	Avg. eq. income	Std. dev.	Gini
AT	6,220	25,451	13,971	0.268
BE	6,011	23,291	10,948	0.249
BG	7,154	3,714	2,491	0.333
CH	7,583	42,208	24,486	0.279
CY	4,589	21,058	11,454	0.279
CZ	8,711	9,006	4,320	0.250
DE	12,683	22,221	12,273	0.276
DK	5,897	32,027	13,836	0.232
EE	5,338	6,922	3,912	0.330
EL	6,184	13,184	8,651	0.334
ES	15,481	17,088	10,597	0.329
FI	9,743	27,517	13,891	0.246
FR	11,078	24,299	14,583	0.288
HR	6,969	6,627	3,819	0.306
HU	13,330	5,327	2,863	0.276
IE	4,318	24,867	14,307	0.296
IS	3,684	22,190	9,232	0.210
IT	21,070	18,786	11,730	0.309
LT	5,403	4,774	3,150	0.344
LU	6,765	37,911	19,977	0.271
LV	6,423	5,334	3,618	0.363
MT	4,701	13,006	6,747	0.277
NL	11,411	25,210	11,414	0.235
NO	5,026	43,260	16,971	0.202
PL	15,545	6,103	3,690	0.316
PT	5,899	10,781	7,296	0.334
RO	7,867	2,562	1,646	0.337
SE	6,599	26,346	10,700	0.215
SI	13,183	13,772	5,994	0.225
SK	6,779	7,304	3,416	0.257
UK	7,391	25,936	16,815	0.320

Note: Summary statistics for the 31 countries in the 2011 wave of EU-SILC. Income variables are measured in Euros.

(2010). Non-parametric estimates are obtained by partitioning the sample into 40 types. Individuals in type t_m have parents of equivalent education (five categories), share their migration status (a binary variable whether the respondent is a first or second generation immigrant), and have fathers working in the same occupation. To minimize the frequency of sparsely populated types we divert from the occupational list given in Table 4.1 by recoding occupations into three categories: highly skilled non-manual (I-01–I-03), lower-skilled non-manual (I-04–I-05 and I-10), skilled manual and elementary occupation (I-06–I-09 and father/mother unknown or dead). This partition is similar but more parsimonious than the one used in Checchi *et al.* (2016), who base their analysis on a total of 96 types.

Lastly, we compare our estimates against the latent class approach as proposed by Li Donni *et al.* (2015). The eligible set of circumstances is the full set of observable circumstances, $\check{\Omega}$. We follow Li Donni *et al.* (2015) in using Schwartz's Bayesian Information

Criterion (BIC) to select the most adequate number of latent types.

4.4.3 Estimates of Inequality of Opportunity

Table 4.3 shows inequality of opportunity estimates for our country sample according to five different estimation procedures. Columns 2-4 list results using the parametric, the non-parametric, and the latent class approach, all of which have been proposed in the extant literature (see section 4.4.2). Columns 5 and 6 list results from conditional inference trees and conditional inference forests, respectively. For all methods, inequality of opportunity estimates are obtained by calculating the Gini index in the counterfactual distribution Y^{EA}.

Of all methods under consideration, the parametric approach delivers the highest estimates. For 29 out of 31 countries the inequality of opportunity estimates are higher than the results from both conditional inference trees and forests. Analogously, the unweighted average Gini over all countries equals 0.103 for the parametric approach as compared to 0.079 and 0.078 for trees and forests, respectively. Also in terms of country rankings, the parametric approach delivers markedly different results in comparison to our preferred methods. While the parametric approach identifies Romania (RO), Bulgaria (BG) and Greece (EL) as the countries in which opportunities are most unequally distributed, these countries rank 6th, 1st and 5th (6th, 2nd and 7th) in the case of trees (forests).

Non-parametric measures of inequality of opportunity take a middle ground between the parametric approach and our preferred methods. For 16 (19) out of 31 countries the non-parametric estimate exceeds the estimate coming from trees (forests), while the unweighted cross-country average Gini amounts to 0.084. In terms of country rankings, the non-parametric approach shows much closer resemblance to our preferred methods than the parametric approach. For example, the three most unequal countries from an opportunity perspective as identified by the non-parametric approach are Bulgaria (BG), Portugal (PT) and Luxembourg (LU), which is congruent with the top three countries identified by trees and forests.

Lastly, the latent class model tends to furnish much lower estimates than all other methods, including trees and forests. This is not very surprising if one considers how latent types are constructed. Latent classes are obtained in the attempt to maximize local independence, that is, to minimize the within-type correlation of circumstances. The algorithm constructs types (and selects their most appropriate number) ignoring the covariance of circumstances and outcome. Conditional inference trees instead construct types by maximizing the outcome variability that can be explained by circumstances. For 8 (9) out of 31 countries the latent class estimate falls short of the estimate coming from trees (forests), while the unweighted cross-country average Gini amounts to 0.069. Also in terms of country rankings, the latent class approach differs markedly from our preferred methods. It identifies Romania (RO), Greece (EL) and Portugal (PT) as the countries in which opportunities are most unequally distributed, whereas these countries rank 6th, 5th and 1st (6th, 7th and 3rd) in the case of trees (forests).

Table 4.3: Inequality of Opportunity Estimates

Country	Parametric	Non-Parametric	Latent Class	Cond. Inf. Tree	Cond. Inf. Forest
AT	0.089	0.075	0.080	0.087	0.088
BE	0.111	0.087	0.053	0.087	0.091
BG	0.154	0.136	0.115	0.136	0.134
CH	0.092	0.083	0.063	0.080	0.090
CY	0.094	0.083	0.074	0.080	0.080
CZ	0.072	0.066	0.060	0.057	0.051
DE	0.070	0.059	0.047	0.070	0.079
DK	0.077	0.041	0.029	0.021	0.020
EE	0.111	0.102	0.074	0.097	0.101
EL	0.148	0.121	0.117	0.126	0.109
ES	0.142	0.120	0.089	0.128	0.120
FI	0.069	0.052	0.048	0.020	0.028
FR	0.086	0.086	0.072	0.090	0.098
HR	0.131	0.088	0.076	0.082	0.076
HU	0.110	0.103	0.095	0.113	0.108
IE	0.105	0.097	0.048	0.084	0.078
IS	0.067	0.032	0.030	0.012	0.016
IT	0.121	0.091	0.080	0.108	0.097
LT	0.095	0.067	0.059	0.070	0.067
LU	0.134	0.121	0.090	0.133	0.136
LV	0.134	0.110	0.095	0.110	0.111
MT	0.087	0.080	0.057	0.071	0.072
NL	0.066	0.053	0.041	0.028	0.019
NO	0.048	0.041	0.030	0.020	0.023
PL	0.111	0.097	0.095	0.1020	0.099
PT	0.138	0.124	0.116	0.136	0.127
RO	0.170	0.104	0.119	0.120	0.111
SE	0.118	0.060	0.025	0.025	0.031
SI	0.077	0.073	0.059	0.032	0.036
SK	0.063	0.051	0.042	0.050	0.046
UK	0.101	0.090	0.062	0.071	0.079
Avg	0.103	0.084	0.069	0.079	0.078

Note: Estimates of inequality of opportunity using five different estimation methods. Inequality of opportunity is measured as the Gini coefficient in the counterfactual distribution Y^{EA}.

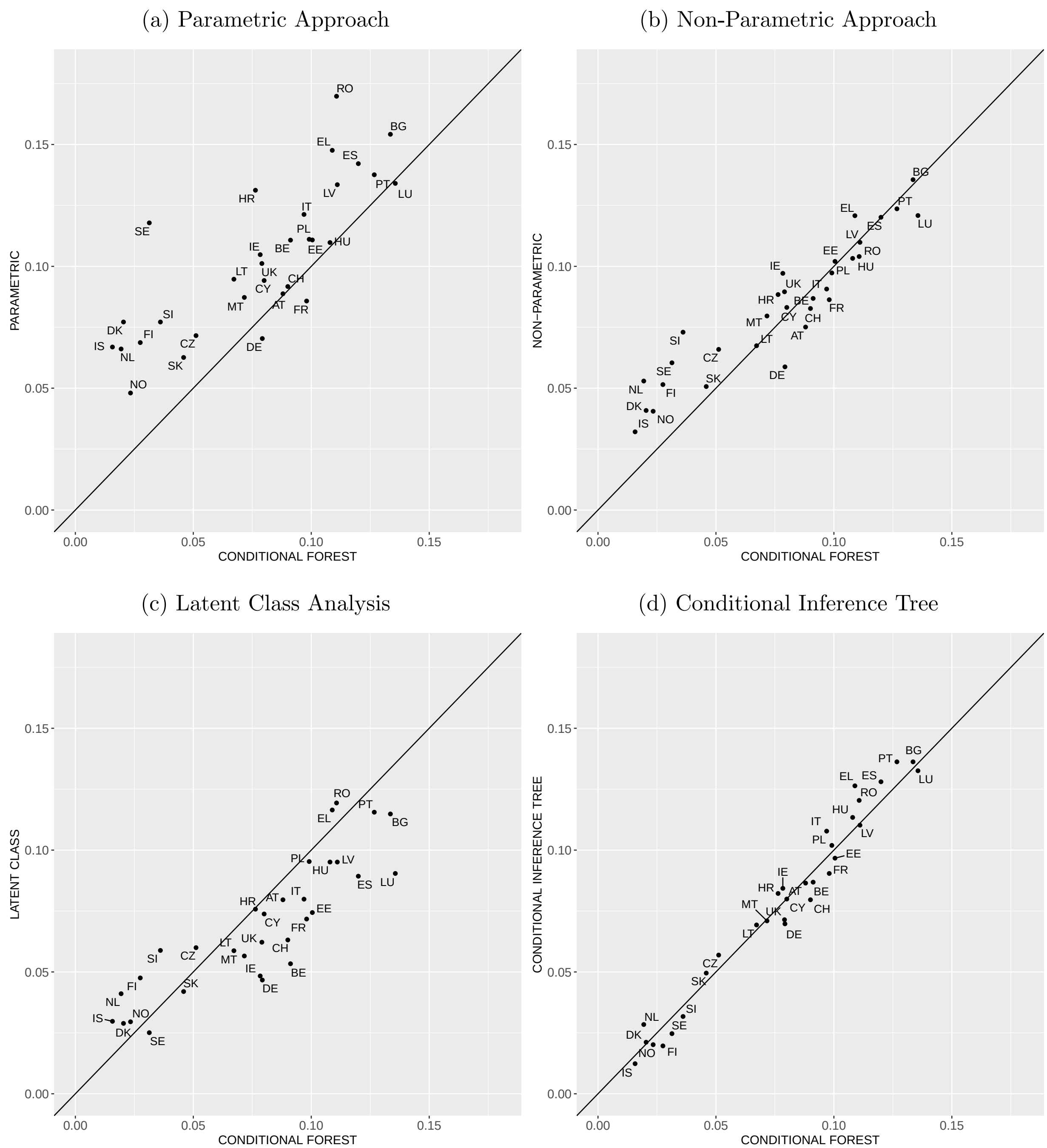
To gain further understanding as regards the relation of existing measurement approaches to our preferred methods, Figure 4.2 plots the estimates from each method against the estimates from conditional inference forests. The black diagonal indicates the 45 degree line, on which all data points should align if the different methods were perfectly congruent. The upper left panel plots the estimates from the parametric approach against the forest estimates. We can confirm the previous diagnosis that the parametric approach delivers higher estimates than forests (and trees). The difference is particularly pronounced for countries that are characterized by relatively low levels of inequality of opportunity, like the Nordic countries.

The upper right panel shows the same plot for the non-parametric approach. We again find relatively high upward divergences in comparison to conditional forest estimates for countries in which inequality of opportunity is low. However, the differences are less pronounced. Interestingly, this pattern is reversed when looking at the correlation plot for the latent class

Figure 4.2: Correlation of Estimates by Method

(a) Parametric Approach (b) Non-Parametric Approach

(c) Latent Class Analysis (d) Conditional Inference Tree



Note: Comparison of inequality of opportunity estimates based on random forests with estimates based on four other methods. Along the solid line inequality of opportunity is the same for the two methods.

approach in the lower left panel. Instead of overestimating the impact of circumstances in societies of low inequality of opportunity, it underestimates the impact of circumstances in societies that are characterized by high inequality in opportunities. Finally, as expected, trees and forests tend to produce very similar results. The correlation between estimates is high (0.98) and in contrast to all other approaches, the sign of the difference is uncorrelated with the level of the estimate.

4.4.4 Opportunity Structure

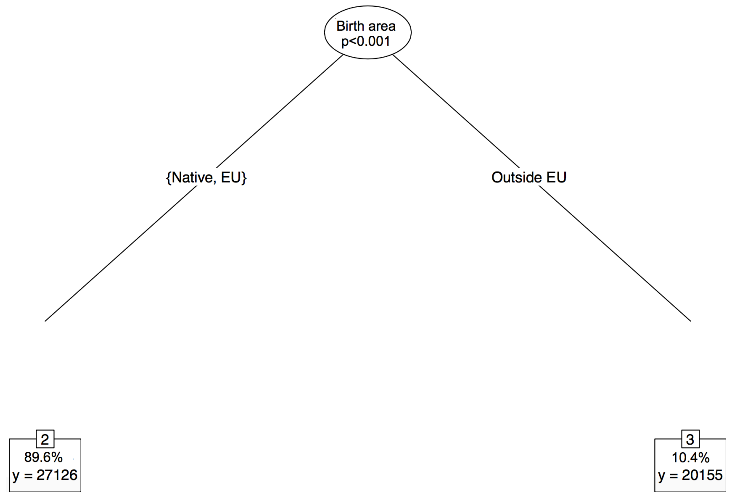
Endowed with an estimate of inequality of opportunity, adequate policy responses must be informed by the particular opportunity structure of a society. That is, policymakers want to learn about the particular circumstances, which cause the existence of inequality of opportunity. In this section we illustrate such analyses for both trees and forests. To keep the analysis intelligible we restrict ourselves to two interesting cases: Sweden and Germany. Readers interested in the opportunity structures of the remaining 29 countries are referred to Appendix 4.A.4.

Trees As outlined in section 4.3.1, the analysis of opportunity structures is particularly intuitive in the case of trees as the relevant information can be directly read off their graphical illustration.

Figure 4.3 illustrates the opportunity structure of Sweden that can be summarized by a tree with two terminal nodes. Inequality of opportunity in Sweden arises from marked differences between first-generation immigrants born outside Europe and the collective group of native residents and European immigrants. The former type accounts for about 10% of the population and on average obtains an equivalent household income that is 35% lower than the corresponding income of the latter group. Recall that each split is based on a statistical test for the existence of equality of opportunity within the respective internal node. Thus, in Sweden we can reject the null hypothesis of equal opportunities for first-generation immigrants born outside Europe and the remainder of the population. However, within these sub-groups the null hypothesis of equality of opportunity cannot be rejected.

A different picture arises when considering Germany. Parental occupation, parental education, migration status, the number of working adults in the respondent's childhood home, and parental tenancy status interact in creating a complex tree made of 14 splits and 15 terminal nodes. The null hypothesis of equality of opportunity is most firmly rejected for individuals whose fathers work in different occupations. If a respondent's father worked in one of the higher ranked occupations (I-01–I-05), the individual belongs to a more advantaged circumstance type than otherwise (Terminal nodes 5-10). These types together account for

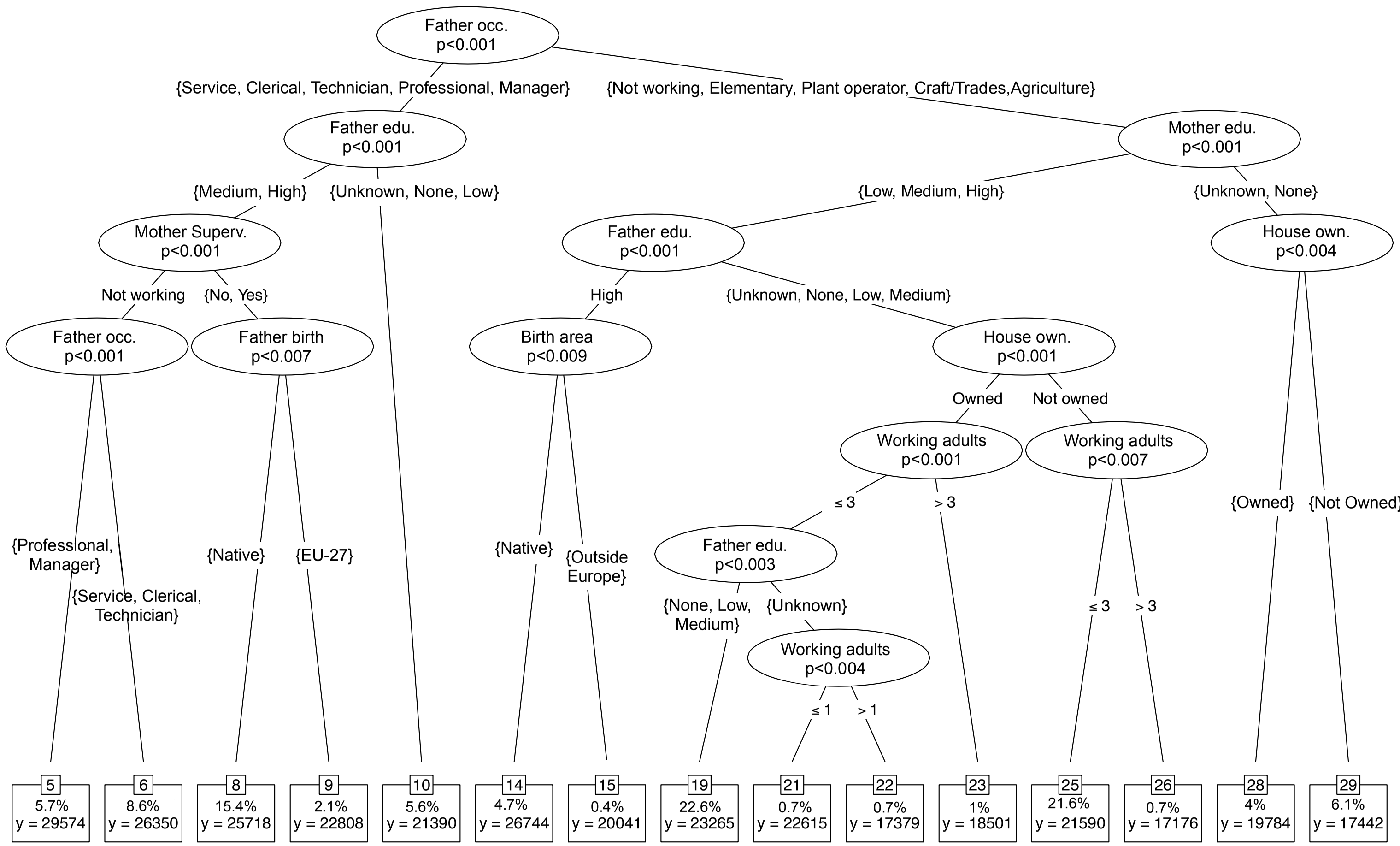
Figure 4.3: Opportunity Tree: Sweden



Note: Opportunity tree for Sweden. White rectangular boxes indicate terminal nodes. The first number inside the rectangular boxes indicates the share of the population belonging to this group, while the second number indicates the predicted income.

37.4% of the population and have an average outcome of €26,380 – far above the population average of €22,221. However, the advantage of this circumstance characteristic is contingent on the educational status of the father. If fathers of respondents had no or low education, the offspring earned less (€21,390) than the country average in spite of the fact that fathers made a career in a high-rank occupation. Conditional on the father both being highly educated and working in a high-rank occupation, the intra-household division of labor plays a strong role. On the one hand, those individuals coming from single earner households in which the mother stayed at home are the most advantaged circumstance types of Germany in 2010, especially if their father worked as a manager or professional (Terminal nodes 5 and 6). On the other hand, offspring of double-earner households tend to be differentiated by their migration status. Comparing terminal nodes 8 and 9 we learn that the advantage of coming from a highly educated double-earner household is substantially diminished from €25,718 to €22,808 if the respondent's father was born outside Germany. A similar distinction based on migration status can be observed on the right-hand side of the tree, in which individuals were born to fathers with a lower occupational status (I-05–I-0). Individuals in this group lived in above average income households if both of their parents were fairly educated *and* their father had no migration background (Terminal node 14). This advantage again vanishes substantially if the respondent's father was born outside Europe (Terminal node 15). Overall, when analyzing the right-hand side of the tree, it is clear that circumstances interact in a very different way in determining individuals' outcomes. In addition to parental education

Figure 4.4: Opportunity Tree: Germany



Note: Opportunity tree for Germany. White rectangular boxes indicate terminal nodes. The first number inside the rectangular boxes indicates the share of the population belonging to this group, while the second number indicates the predicted income. Occupation refers to ISCO-08 one digit codes. All variables describing household characteristics refer to the period in which the respondent was about 14 years old. See Table 4.1 for details.

and the migration status of individuals, the tenancy status during childhood as well as the number of working adults in the respondents' childhood home play an important role.

There is marked heterogeneity in tree structures across countries (Appendix 4.A.4). For the remaining countries in our sample, terminal nodes range from three (Denmark, Iceland and Norway) to 27 (Italy). It is noteworthy that the rank-rank correlation between the number of terminal nodes and the inequality of opportunity estimates presented in section 4.4.3 is positive but not perfect (Appendix 4.A.5).

Forests Forests cannot be analyzed in the straightforward graphical manner of trees. However, we can use variable importance measures to assess the importance of circumstance variables for the construction of forests. One measure of variable importance, as proposed by Strobl *et al.* (2007), is obtained by permuting input variable $C^p \in \check{\Omega}$ such that its dependence with y is lost. After this, the out-of-bag error rate, MSE^{OOB}, is re-computed. The increase of MSE^{OOB} in comparison to the baseline out-of-bag error indicates the importance of $C^p \in \check{\Omega}$ for prediction accuracy. Repeating this procedure for all $C^p \in \check{\Omega}$ affords a relative comparison of all circumstances.

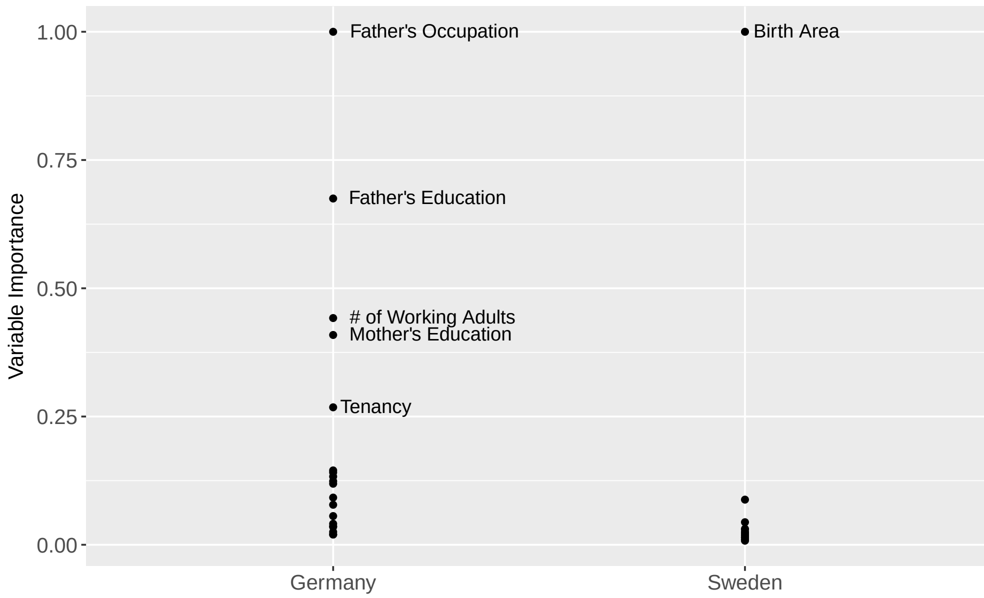
Figure 4.5 shows the results from this procedure for our example cases of Germany and Sweden. Each black dot is the importance of one of the $\check{P}$ variables in the set of observed circumstances $\check{\Omega}$. We standardize the ensuing results such that the variable importance measure for the circumstance with the greatest impact in each country equals one. For the case of Sweden, birth area is the only circumstance that has a meaningful predictive value. In Germany, father's occupation and father's education are most important, followed by the number of working adults in the household and mother's education.

It is reassuring that these findings are in line with the graphical analysis of opportunity trees. In Figure 4A.30 of Appendix 4.A.4 we show variable importance plots for all countries in our sample. Broadly, we can divide the countries into three groups according to the circumstances that determine their opportunity structure. First, there is a handful of primarily Nordic countries where the respondent's birth area is the most important circumstance. Second, there is a large group of primarily Western and Southern European countries for which father's occupation or father's education is most important. Third, there is a group of Eastern European countries for which mother's education or mother's occupation is most important.

4.4.5 Out-of-Sample Performance

Recall that current approaches towards estimating inequality of opportunity are subject to different biases. Models are downward biased to the extent that the full set of circumstances

Figure 4.5: Variable Importance for Germany and Sweden



Note: Each dot shows the importance of a particular circumstance for the predictions from our random forest. The importance of a circumstance is measured by permuting the circumstance, calculating a new MSE^{OOB}, and computing the difference in the MSE^{OOB} between the original model and the model with the permuted circumstance. The importance measure is standardized such that the circumstance with the greatest importance in each country equals one. Occupation refers to ISCO-08 one digit codes. All variables describing household characteristics refer to the period in which the respondent was about 14 years old. See Table 4.1 for details.

Ω is unobserved. Models are upward biased to the extent that they over-utilize the set of observed circumstances $\check{\Omega}$ leading to overfitted estimates that do not replicate out-of-sample (see Appendix 4.A.1 for the formal argument).

In order to assess how well different estimations approaches trade off these biases, we follow the machine-learning practice of splitting our sample into a *training set* with $i_{train} \in \{1, ..., N_{train}\}$ and a *test set* with $i_{test} \in \{1, ..., N_{test}\}$. For each country in our sample, $N_{train} = \frac{2}{3}N$ while $N_{test} = \frac{1}{3}N$. We fit our models on the training set and compare their performance on the test set according to the following procedure:

1. Run the chosen models on the training data (for the specific estimation procedures, see section 4.3.1 for trees, section 4.3.2 for forests, and section 4.4.2 for our benchmark methods).

2. Store the prediction functions $\hat{f}_{train}(\check{\Omega})$.

3. Predict the outcomes of observations in the test set: $\hat{y}_{i_{test}} = \hat{f}_{train}(\check{\Omega}_{i_{test}})$.

4. Calculate the out-of-sample error: $\text{MSE}^{test} = \frac{1}{N_{test}} \sum_{i_{test}} [y_{i_{test}} - \hat{y}_{i_{test}}]^2$.

Figure 4.6 compares the resulting MSE^{test} of the different models. For each country, the MSE^{test} of random forests is standardized to equal 1, such that if a model has an MSE^{test} larger than 1, then this represents a worse fit out-of-sample than forests. This implies that the model in question performs worse than forests in trading off upward and downward biases, either by neglecting the use of circumstances or overfitting. We derive 95% confidence intervals based on 200 bootstrapped re-samples of the test data using the normal approximation method.

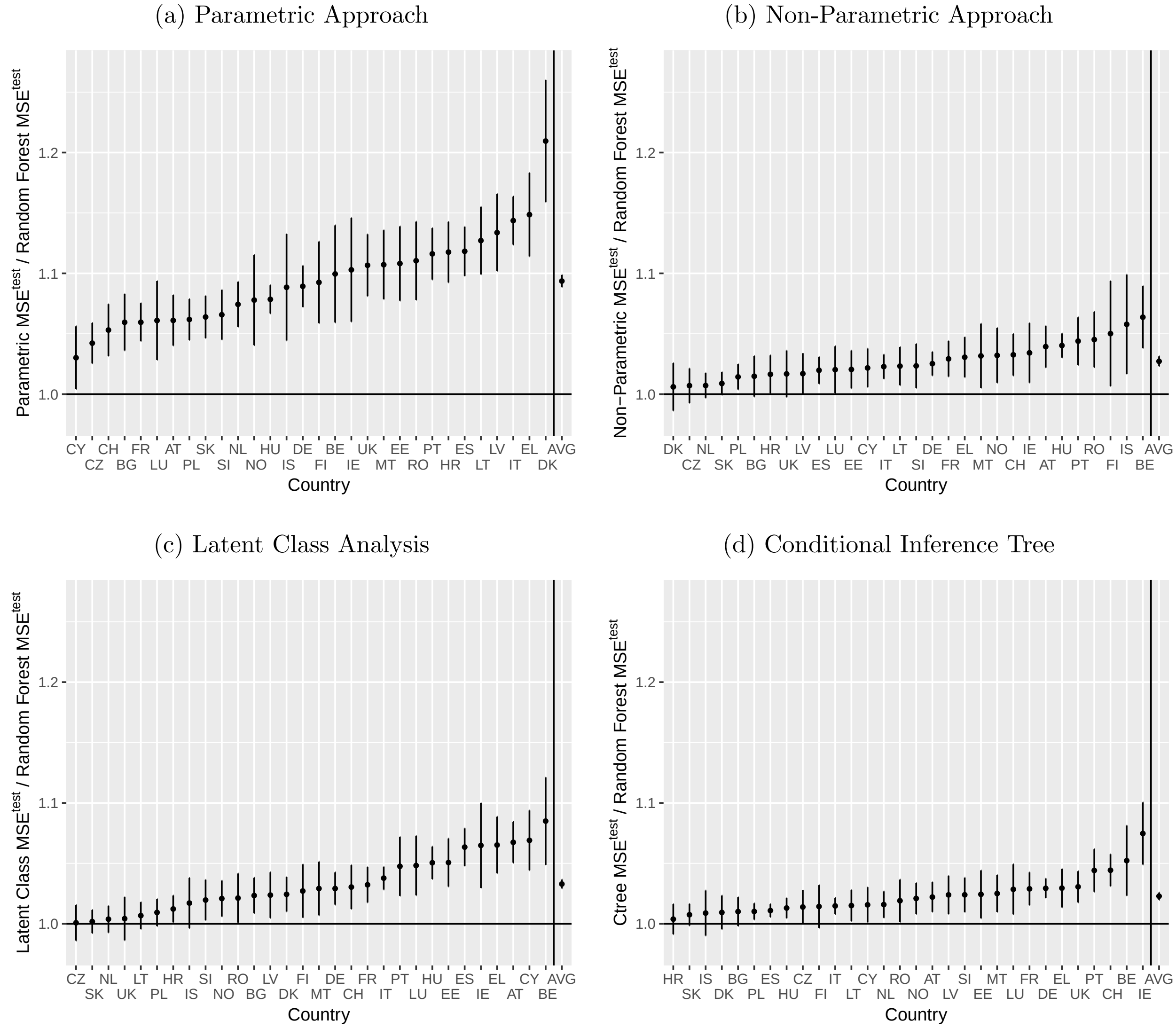
As expected, random forests outperform all other methods in nearly all cases. On average, the parametric approach gives a fit 9.4% worse than forests. With average shortfalls of around 3%, out-of-sample prediction errors are less pronounced for non-parametric models and latent class analysis. Yet both methods perform worse than conditional inference forests for the vast majority of countries in our sample. Hence, relative to random forests, our benchmark methods either underutilize or overutilize the information contained in $\check{\Omega}$ and are therefore biased in their inequality of opportunity estimates. The estimates presented in section 4.4.3 suggest that the parametric and the non-parametric partitions are overfitting the data, while the type partition delivered by latent class analysis is too coarse.

On average conditional inference trees are closest to the test error rate of forests. Yet they also fall short of the performance of forests due to their poorer utilization of the information given in $\check{\Omega}$.

4.5 Conclusion

In this paper we have proposed two novel approaches towards estimating inequality of opportunity based on regression trees. Both conditional inference trees and forests minimize arbitrary model selection by the researcher, while trading off downward and upward biases in inequality of opportunity estimates. On the one hand, conditional inference forests outperform all methods considered in this paper in terms of their out-of-sample performance. Hence, they deliver the best estimates of inequality of opportunity. On the other hand, conditional inference trees are econometrically less complex and provide a handy graphical illustration that can be used for the straightforward analysis of opportunity structures. The fact that trees are very close to forests in terms of their inequality of opportunity estimates (section 4.4.3), the importance they assign to specific circumstances (4.4.4) and their out-of-sample performance (4.4.5) makes us confident that they are a useful tool for communicating issues related to inequality of opportunity to a larger audience.

Figure 4.6: Comparison of Models' Test Error



Notes: The figure compares the test error of the different models. The test error of random forests is standardized to 1, such that a test error larger than 1 represent worse fits than random forests. 95% confidence intervals are derived based on 200 bootstrapped re-samples of the test data using the normal approximation method. Sweden is excluded from the figure since it is an outlier. The test errors for Sweden are 1.43 $[1.21, 1.66]$ for the parametric approach, 1.11 $[1.01, 1.21]$ for the non-parametric approach, 1.06 $[1.02, 1.11]$ for latent class analysis, and 1.06 $[1.01, 1.11]$ for conditional inference trees.

To be sure, the development of machine learning algorithms and their integration into the analytical toolkit of economists is a highly dynamic process. We are well aware that finding the best machine learning algorithm for inequality of opportunity estimations is a methodological horse race with frequent entry of new competitors that eventually will lead to some method outperforming the ones we proposed in this work. Therefore, the main contribution of this work should be understood as paving the way for new methods that are able to handle the intricacies of model selection for inequality of opportunity estimations. While we restricted ourselves to ex-ante utilitarian measures of inequality of opportunity, the exploration of these algorithms for other methods in the inequality of opportunity literature, such as ex-post measures à la Pistolesi (2009) or ex-ante and ex-post tests à la Lefranc *et al.* (2009), provides an interesting avenue for future research.

Bibliography

Almås, I., Cappelen, A. W., Lind, J. T., Sørensen, E. Ø. and Tungodden, B. (2011). Measuring unfair (in)equality. *Journal of Public Economics*, **95** (7–8), 488–499. (page 93)

Altshuler, D., Durbin, R. M., Donnelly, P., Green, E. D., Nickerson, D. A., Boerwinkle, E. and Doddapaneni, H. (2015). A global reference for human genetic variation. *Nature*, **526** (7571), 68–74. (page 93)

Athey, S. (2017). The Impact of Machine Learning on Economics. *mimeo*. (page 94)

Black, S. E. and Devereux, P. J. (2011). Recent Developments in Intergenerational Mobility. In D. Card and O. Ashenfelter (eds.), *Handbook on Labor Economics*, vol. 4, *16*, Amsterdam: Elsevier, pp. 1487–1541. (page 95)

Bourguignon, F., Ferreira, F. H. G. and Menéndez, M. (2007). Inequality of Opportunity in Brazil. *Review of Income and Wealth*, **53** (4), 585–618. (page 93, 94, 98, 107)

Breiman, L. (2001). Random forests. *Machine Learning*, **45** (1), 5–32. (page 104)

—, Friedman, J., Stone, C. and Olshen, R. (1984). *Classification and Regression Trees*. Belmont: Taylor & Francis. (page 94, 101)

Brunori, P., Peragine, V. and Serlenga, L. (2016). Upward and downward bias when measuring inequality of opportunity. *ECINEQ Working Paper Series*, **2016-406**. (page 94, 125)

Checchi, D. and Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality*, **8** (4), 429–450. (page 93, 97, 107)

—, — and Serlenga, L. (2016). Inequality of Opportunity in Europe: Is There a Role for Institutions? In L. Cappellari, S. Polachek and K. Tatsiramos (eds.), *Inequality: Causes and Consequences*, *Research in Labor Economics*, vol. 43, Bingley: Emerald, pp. 1–44. (page 107, 109)

Cowell (2016). Inequality and Poverty Measures. In M. D. Adler and M. Fleurbaey (eds.), *Oxford Handbook of Well-Being and Public Policy*, *4*, Oxford: Oxford University Press, pp. 82–125. (page 97)

Cowell, F. A. and Victoria-Feser, M.-P. (1996). Robustness Properties of Inequality Measures. *Econometrica*, **64** (1), 77–101. (page 107)

Ferreira, F. H. G. and Gignoux, J. (2011). The Measurement of Inequality of Opportunity: Theory and an Application to Latin America. *Review of Income and Wealth*, **57** (4), 622–657. (page 93, 94, 98, 107)

Fleurbaey, M. (1995). Three solutions for the compensation problem. *Journal of Economic Theory*, **65** (2), 505–521. (page 93)

— (2008). *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press. (page 93)

— and Peragine, V. (2013). Ex Ante Versus Ex Post Equality of Opportunity. *Economica*, **80** (317), 118–130. (page 96)

Friedman, J., Hastie, T. and Tibshirani, R. (2009). *The elements of statistical learning*. New York: Springer. (page 94, 101, 104)

García, J. L., Heckman, J. J. and Ziff, A. L. (2017). Gender Differences in the Benefits of an Influential Early Childhood Program. *European Economic Review*, **Forthcoming**. (page 94)

Hothorn, T., Hornik, K. and Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, **15** (3), 651–674. (page 94, 101)

Hufe, P., Peichl, A., Roemer, J. E. and Ungerer, M. (2017). Inequality of income acquisition: the role of childhood circumstances. *Social Choice and Welfare*, **Forthcoming**. (page 93)

Kanbur, R. and Snell, A. (2017). Inequality Measures as Tests of Fairness. *mimeo*. (page 93, 97)

Lanza, S., Xianming, T. and Bethany, B. (2013). Latent class analysis with distal outcomes: A flexible model- based approach. *Structural Equation Modeling*, **20** (1), 1–26. (page 99)

Lefranc, A., Pistolesi, N. and Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France. *Journal of Public Economics*, **93** (11–12), 1189–1207. (page 93, 97, 120)

Li Donni, P., Rodríguez, J. G. and Rosa Dias, P. (2015). Empirical definition of social types in the analysis of inequality of opportunity: A latent classes approach. *Social Choice and Welfare*, **44** (3), 673–701. (page 99, 109)

Marrero, G. A. and Rodríguez, J. G. (2012). Inequality of Opportunity in Europe. *Review of Income and Wealth*, **58** (4), 597–621. (page 107)

Morgan, J. N. and Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, **58** (302), 415–434. (page 94)

Mullainathan, S. and Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, **31** (2), 87–106. (page 94)

Oppedisano, V. and Turati, G. (2015). What are the causes of educational inequality and of its evolution over time in Europe? Evidence from PISA. *Education Economics*, **23** (1), 3–24. (page 100)

Palomino, J. C., Marrero, G. A. and Rodríguez, J. G. (2016). Channels of inequality of opportunity: The role of education and occupation in Europe. *ECINEQ Working Paper Series*, **2016-411**. (page 107)

Pistolesi, N. (2009). Inequality of opportunity in the land of opportunities, 1968–2001. *The Journal of Economic Inequality*, **7** (4), 411–433. (page 120)

Roemer, J. E. (1998). *Equality of Opportunity*. Cambridge: Harvard University Press. (page 93, 96)

— and Trannoy, A. (2015). Equality of Opportunity. In A. B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution*, vol. 2, *4*, Amsterdam: Elsevier, pp. 217–300. (page 93)

Strobl, C., Boulesteix, A.-L., Zeileis, A. and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, **8** (1), 25. (page 104, 116)

Trannoy, A., Tubeuf, S., Jusot, F. and Devaux, M. (2010). Inequality of opportunities in health in France: A first pass. *Health Economics*, **19** (8), 921–938. (page 100)

Van de gaer, D. (1993). *Equality of Opportunity and Investment in Human Capital*. Ph.D. thesis, University of Leuven. (page 93, 97)

Van de gaer, D. and Ramos, X. (2016). Empirical Approaches to Inequality of Opportunity: Principles, Measures, and Evidence. *Journal of Economic Surveys*, **30** (5), 855–883. (page 93, 96)

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, **28** (2), 3–27. (page 94)

4.A Appendix

4.A.1 Model Evaluation by the MSE

We use the MSE as a model evaluation criterion when cross-validating α in the case of trees (Section 4.3.1) and when determining the values of α and $\bar{P}$ using the out-of-bag error rate in the case of forests (Section 4.3.2). Analogously, when comparing the predictive performance of different estimation approaches in the test sample N^{test}, we prefer the estimation approach that yields a lower MSE (Section 4.4.5). The following discussion applies to all of these applications. For the sake of conciseness, superscript h always indicates observations in the hold-out sample regardless of the specific application.

The general MSE evaluation criterion can be written as follows:

$$\frac{1}{N^h}\sum_h (y_i^h - \hat{y}_i)^2. \tag{9}$$

In the case where observed circumstances $\check{\Omega}$ are the sole input variables, individual predictions $\hat{y}_i$ are given by the mean outcomes of the type to which individuals are allocated and we can write:

$$\frac{1}{N^h}\sum_h (y_i^h - \mu_m)^2, \tag{10}$$

where $\mu_m = \frac{1}{N}\sum_{i\in t_m} y_i$ and t_m denotes a specific type in the model we want to evaluate. It is instructive to rewrite the MSE as a weighted average over types as follows:

$$\sum_m \frac{N_m^h}{N^h}\sum_{i\in t_m}\frac{1}{N_m^h}(y_i^h - \mu_m)^2. \tag{11}$$

We can expand the previous expression and spell out the binomial formula:

$$\sum_m \frac{N_m^h}{N^h}\sum_{i\in t_m}\frac{1}{N_m^h}\left[(y_i^h - \mu_m^h) + (\mu_m^h - \mu_m)\right]^2 \tag{12}$$

$$= \sum_m \frac{N_m^h}{N^h}\sum_{i\in t_m}\frac{1}{N_m^h}\left[(y_i^h - \mu_m^h)^2 + (\mu_m^h - \mu_m)^2\right] + 2\sum_m \frac{N_m^h}{N^h}(\mu_m^h - \mu_m)\sum_{i\in t_m}\frac{1}{N_m^h}(y_i^h - \mu_m^h). \tag{13}$$

Evidently, $\sum_{i \in t_m} \frac{1}{N_m^h}(y_i^h - \mu_m^h) = 0$ and the formula simplifies to:

$$\sum_m \frac{N_m^h}{N^h} \sum_{i \in t_m} \frac{1}{N_m^h} \left[\underbrace{(y_i^h - \mu_m^h)^2}_{(1)} + \underbrace{(\mu_m^h - \mu_m)^2}_{(2)} \right], \tag{14}$$

where (1) is the intra-type variance of outcomes in the hold-out sample and (2) is the variance of type-means between the hold-out sample and the training sample. Recall that we prefer models that minimize formula (14). For the sake of exposition, we generalize the previous expression by introducing the weighting parameter $\delta \in [0, 1]$. Note that the standard MSE yields equivalent rankings to the special case in which $\delta = 0.5$, i.e. the case in which we give equal weight to both (1) and (2):

$$\sum_m \frac{N_m^h}{N^h} \sum_{i \in t_m} \frac{1}{N_m^h} \left[\delta \underbrace{(y_i^h - \mu_m^h)^2}_{(1)} + (1 - \delta) \underbrace{(\mu_m^h - \mu_m)^2}_{(2)} \right]. \tag{15}$$

Now assume two extreme cases:

1. $\delta = 1$: In this case we give full priority to (1), i.e. we would always prefer a model that decreased the intra-type variance in the hold-out sample the most. Naturally, one reduces intra-type variance by increasing the granularity of the type partition. Hence, we would always prefer the model that used more circumstances and interactions. Thus, (1) addresses the downward bias of equality of opportunity estimates as induced by using only a subset $\hat{\Omega}$ of the full set of observed circumstances $\check{\Omega}$.

2. $\delta = 0$: In this case we give full priority to (2), i.e. we would always prefer a model that decreased the variance between type means in the hold-out sample and the type means in the training sample. Invoking the law of large numbers it is evident that the ideal model from this perspective is the model with no partition at all, i.e. the model in which individual predictions μ_m are given by the sample mean μ. Thus, (2) addresses the upward bias identified by Brunori *et al.* (2016) that originates from overfitting the model to the training data.

To conclude, the more weight we put on (1), the less the downward bias in our estimation, since we allow circumstances to have explanatory scope for observed outcomes. Intuitively, if we set $\delta = 0$, our estimates would be deeply downward biased because we would effectively say that inequality of opportunity was non-existent. The more weight we put on (2), the more accurate our estimates of type means, i.e. the less the out of sample variance in our estimates of the type means. Intuitively, with $\delta = 1$ we would say that we did not care about the precision of our estimates at all, which is the standard practice in today's inequality of opportunity estimations. This instills overfitting and an upward bias in inequality of opportunity estimates. Hence by giving equal weight to both components, the MSE balances upwards and downward biases in inequality of opportunity estimations and thus is a sensible criterion for model evaluation in this context.

4.A.2 Descriptive Statistics

Table 4A.1: Descriptive Statistics (Individual and Household)

		Sex		Birth area		Presence parents		Household members			Tenancy
Country	Eq. income	Male	Female	Native	EU	Both	One	Adults	Working adults	Children	Owner
AT	25,451	0.499	0.501	0.790	0.070	0.856	0.017	2.73	1.76	2.60	0.585
BE	23,291	0.502	0.498	0.824	0.076	0.855	0.019	2.38	1.59	2.78	0.750
BG	3,714	0.500	0.500	0.994	0.001	0.904	0.012	2.44	2.01	2.07	0.910
CH	42,208	0.495	0.505	0.684	0.197	0.837	0.017	2.55	1.90	2.53	0.546
CY	21,058	0.475	0.525	0.787	0.096	0.900	0.015	2.64	1.67	2.70	0.784
CZ	9,006	0.492	0.508	0.964	0.026	0.851	0.013	2.09	1.92	2.24	0.597
DE	22,221	0.504	0.496	0.868	0.000	0.830	0.020	2.24	1.68	2.32	0.499
DK	32,027	0.495	0.505	0.923	0.026	0.809	0.027	2.22	2.31	2.24	0.736
EE	6,922	0.475	0.525	0.847	0.000	0.756	0.011	2.10	1.80	2.09	0.859
EL	13,184	0.502	0.498	0.890	0.025	0.931	0.019	2.31	1.56	2.33	0.834
ES	17,088	0.505	0.495	0.834	0.051	0.893	0.012	2.88	2.11	2.43	0.819
FI	27,517	0.501	0.499	0.954	0.018	0.829	0.016	2.36	1.75	2.30	0.772
FR	24,299	0.491	0.509	0.885	0.036	0.820	0.022	2.47	1.66	1.75	0.630
HR	6,627	0.499	0.501	0.875	0.017	0.874	0.020	2.56	1.35	2.31	0.902
HU	5,327	0.483	0.517	0.988	0.008	0.844	0.041	2.14	1.75	2.27	0.830
IE	24,867	0.476	0.524	0.783	0.149	0.893	0.078	3.17	3.20	3.20	0.727
IS	22,190	0.493	0.507	0.920	0.042	0.899	0.012	2.42	1.90	2.63	0.893
IT	18,786	0.498	0.502	0.880	0.040	0.901	0.011	2.59	1.62	2.41	0.685
LT	4,774	0.479	0.521	0.939	0.004	0.846	0.016	2.32	2.02	2.46	0.698
LU	37,911	0.501	0.499	0.480	0.401	0.868	0.020	2.53	1.64	2.71	0.734
LV	5,334	0.480	0.520	0.865	0.000	0.763	0.012	1.97	1.76	2.28	0.455
MT	13,006	0.503	0.497	0.944	0.000	0.932	0.020	3.02	1.84	2.68	0.576
NL	25,210	0.491	0.509	0.903	0.020	0.882	0.016	2.10	1.54	3.25	0.575
NO	43,260	0.489	0.511	0.907	0.041	0.913	0.014	2.02	1.76	1.87	0.922
PL	6,103	0.496	0.504	0.999	0.000	0.889	0.015	2.70	1.96	2.44	0.644
PT	10,781	0.494	0.506	0.906	0.022	0.854	0.017	2.68	2.23	2.68	0.544
RO	2,562	0.494	0.506	0.999	0.000	0.919	0.009	2.77	1.90	2.27	0.861
SE	26,346	0.507	0.493	0.846	0.050	0.820	0.035	2.07	1.78	2.35	0.757
SI	13,772	0.504	0.496	0.876	0.000	0.855	0.019	2.53	1.77	2.20	0.746
SK	7,304	0.481	0.519	0.987	0.010	0.920	0.010	2.52	2.08	2.34	0.694
UK	25,936	0.493	0.507	0.848	0.042	0.825	0.024	2.34	2.24	2.41	0.649

Note: Omitted categories are: "Outside Europe" for birth area and "None/Collective house" for the presence of parents, and "Not owned" for the tenancy variable.

Table 4A.2: Descriptive Statistics (Fathers)

	Birth area		Citizenship		Education			Activity					Main occupation ISCO-08 1-digit								Superv.
Country	Native	EU	Resid.	EU	Prim.	Sec.	Tert.	Empl.	Self-empl.	Unempl.	Retired	House work	1	2	3	4	5	6	7	8	Yes
AT	0.743	0.093	0.777	0.068	0.007	0.398	0.421	0.714	0.215	0.003	0.010	0.072	0.085	0.063	0.284	0.145	0.147	0.051	0.064	0.046	0.338
BE	0.748	0.100	0.762	0.093	0.016	0.491	0.199	0.699	0.179	0.007	0.011	0.130	0.041	0.127	0.209	0.057	0.054	0.084	0.104	0.126	0.278
BG	0.933	0.004	0.936	0.001	0.029	0.466	0.333	0.899	0.028	0.005	0.004	0.078	0.142	0.207	0.216	0.135	0.058	0.029	0.047	0.065	0.093
CH	0.588	0.286	0.603	0.280	0.051	0.227	0.487	0.653	0.292	0.001	0.003	0.055	0.054	0.077	0.223	0.111	0.065	0.057	0.140	0.131	0.397
CY	0.803	0.082	0.808	0.094	0.045	0.667	0.178	0.566	0.381	0.004	0.009	0.053	0.125	0.122	0.245	0.161	0.109	0.029	0.074	0.071	0.229
CZ	0.878	0.065	0.910	0.036	0.003	0.602	0.216	0.891	0.017	0.001	0.006	0.094	0.053	0.195	0.305	0.039	0.051	0.036	0.125	0.070	0.233
DE	0.800	0.200	0.855	0.145	0.004	0.125	0.496	0.819	0.123	0.008	0.013	0.062	0.040	0.154	0.266	0.059	0.061	0.051	0.158	0.104	0.299
DK	0.935	0.025	0.970	0.020	0.000	0.368	0.418	0.708	0.272	0.004	0.014	0.021	0.009	0.072	0.288	0.160	0.103	0.043	0.070	0.122	0.447
EE	0.603	0.270	0.637	0.233	0.000	0.300	0.338	0.823	0.006	0.003	0.006	0.177	0.053	0.253	0.221	0.034	0.027	0.014	0.053	0.092	0.153
EL	0.887	0.016	0.911	0.015	0.042	0.587	0.135	0.449	0.517	0.002	0.004	0.034	0.055	0.099	0.210	0.308	0.060	0.087	0.026	0.047	0.182
ES	0.836	0.047	0.846	0.046	0.052	0.762	0.064	0.702	0.219	0.006	0.016	0.081	0.137	0.113	0.191	0.145	0.101	0.055	0.076	0.045	0.191
FI	0.827	0.007	0.827	0.007	0.019	0.491	0.182	0.592	0.209	0.016	0.009	0.253	0.044	0.138	0.146	0.135	0.053	0.016	0.085	0.089	
FR	0.789	0.078	0.857	0.057	0.040	0.695	0.073	0.753	0.170	0.003	0.006	0.079	0.223	0.055	0.155	0.103	0.050	0.072	0.111	0.068	0.335
HR	0.822	0.006	0.834	0.004	0.006	0.464	0.312	0.763	0.103	0.037	0.019	0.137	0.228	0.103	0.214	0.049	0.079	0.036	0.088	0.041	0.129
HU	0.962	0.017	0.969	0.012	0.017	0.599	0.241	0.892	0.043	0.001	0.011	0.064	0.137	0.193	0.279	0.094	0.067	0.017	0.052	0.060	0.117
IE	0.792	0.107	0.758	0.094	0.014	0.574	0.258	0.659	0.221	0.049	0.009	0.120	0.158	0.065	0.149	0.155	0.092	0.022	0.042	0.092	0.344
IS	0.918	0.050	0.923	0.044	0.001	0.334	0.486	0.638	0.332	0.001	0.001	0.031	0.042	0.094	0.220	0.180	0.096	0.024	0.076	0.121	0.570
IT	0.823	0.022	0.827	0.020	0.030	0.708	0.136	0.614	0.244	0.016	0.016	0.143	0.118	0.105	0.227	0.099	0.082	0.057	0.074	0.040	0.199
LT	0.899	0.004	0.926	0.004	0.014	0.538	0.228	0.916	0.011	0.000	0.004	0.076	0.214	0.179	0.241	0.080	0.030	0.017	0.038	0.074	0.110
LU	0.387	0.467	0.400	0.466	0.037	0.484	0.316	0.757	0.174	0.001	0.009	0.070	0.039	0.183	0.228	0.112	0.046	0.048	0.118	0.093	0.251
LV	0.572	0.248	0.642	0.165	0.002	0.381	0.297	0.767	0.005	0.002	0.008	0.229	0.083	0.218	0.199	0.069	0.036	0.010	0.037	0.083	0.070
MT	0.952	0.041	0.953	0.040	0.164	0.561	0.180	0.717	0.214	0.013	0.011	0.073	0.106	0.099	0.244	0.050	0.169	0.045	0.106	0.046	0.225
NL	0.829	0.028	0.888	0.022	0.008	0.376	0.285	0.726	0.173	0.006	0.003	0.103	0.031	0.079	0.200	0.086	0.084	0.051	0.155	0.124	0.310
NO	0.897	0.046	0.908	0.041	0.005	0.328	0.390	0.712	0.255	0.002	0.014	0.032	0.032	0.100	0.227	0.111	0.075	0.029	0.167	0.110	0.285
PL	0.955	0.012	0.980	0.003	0.004	0.462	0.448	0.701	0.238	0.002	0.005	0.064	0.078	0.157	0.254	0.237	0.053	0.025	0.053	0.044	0.111
PT	0.932	0.006	0.945	0.006	0.193	0.700	0.031	0.650	0.248	0.002	0.014	0.102	0.077	0.114	0.264	0.185	0.082	0.038	0.060	0.032	0.190
RO	0.938	0.001	0.939	0.001	0.017	0.726	0.088	0.642	0.237	0.004	0.014	0.139	0.104	0.121	0.249	0.253	0.040	0.016	0.034	0.040	0.045
SE	0.945	0.022	0.851	0.061	0.000	0.422	0.350	0.745	0.211	0.002	0.014	0.192	0.019	0.108	0.230	0.086	0.105	0.031	0.067	0.118	0.337
SI	0.769	0.200	0.000	0.000	0.001	0.684	0.166	0.773	0.099	0.013	0.020	0.128	0.173	0.080	0.257	0.089	0.059	0.037	0.100	0.052	0.242
SK	0.935	0.020	0.945	0.011	0.001	0.362	0.497	0.921	0.011	0.002	0.005	0.071	0.128	0.209	0.285	0.030	0.052	0.028	0.095	0.060	0.145
UK	0.800	0.064	0.869	0.039	0.033	0.508	0.228	0.795	0.147	0.025	0.009	0.059	0.083	0.133	0.236	0.036	0.091	0.040	0.085	0.142	0.398

Note: Omitted categories are: “Outside EU” for birth area, “Outside EU” for citizenship, “Illiterate” for education, “Unknown/Dead” and “Other inactive” for activity. ISCO-08 occupation definitions are: 1 “Elementary”, 2 “Plant Operator” 3, “Craft/Trades” 4, “Agriculture” 5, “Service and Army” 6, “Clerical”, 7 “Technician”, 8 “Professional”, 9 “Manager”; ”Dead/Unknown/Not working” is not shown. Omitted categories for supervisory are: “No.”

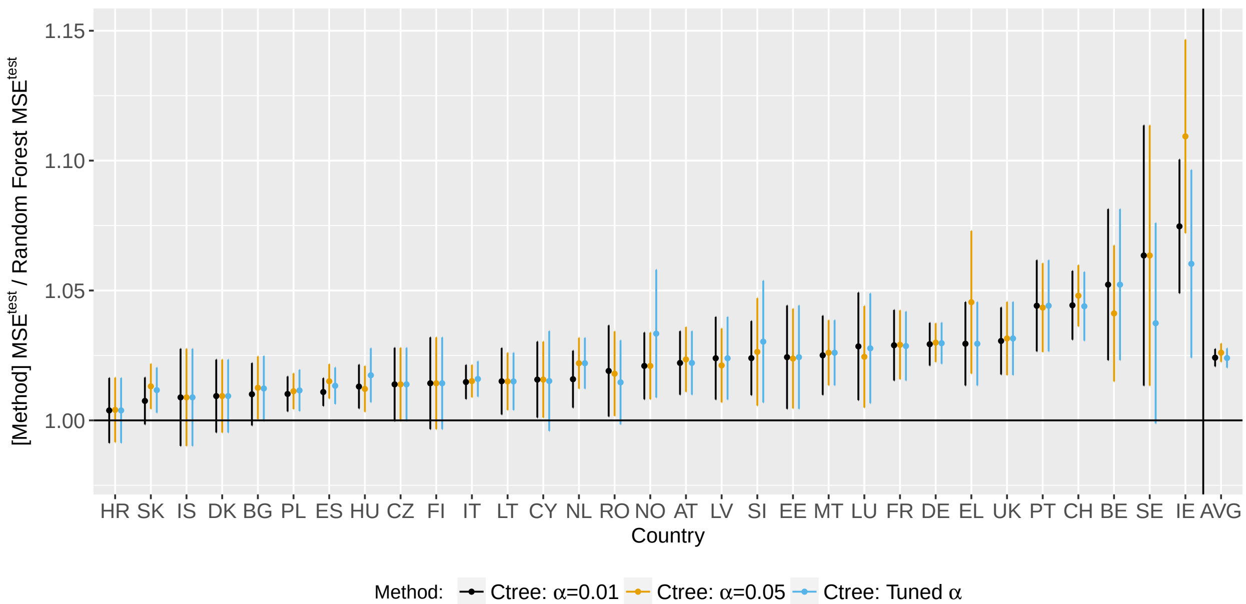
Table 4A.3: Descriptive Statistics (Mothers)

	Birth area		Citizenship		Education			Activity					Main occupation ISCO-08 1-digit								Superv.
Country	Native	EU	Resid.	EU	Prim.	Sec.	Tert.	Empl.	Self-empl.	Unempl.	Retired	House work	1	2	3	4	5	6	7	8	Yes
AT	0.740	0.096	0.789	0.065	0.026	0.587	0.328	0.369	0.169	0.002	0.005	0.463	0.087	0.010	0.045	0.128	0.155	0.071	0.009	0.024	0.092
BE	0.755	0.097	0.790	0.092	0.030	0.564	0.201	0.320	0.117	0.006	0.002	0.651	0.069	0.024	0.016	0.002	0.046	0.058	0.045	0.081	0.034
BG	0.931	0.003	0.981	0.002	0.039	0.464	0.357	0.878	0.026	0.007	0.003	0.101	0.152	0.064	0.099	0.181	0.140	0.092	0.040	0.123	0.030
CH	0.567	0.307	0.599	0.286	0.078	0.410	0.399	0.382	0.152	0.001	0.001	0.466	0.068	0.025	0.039	0.055	0.125	0.069	0.069	0.056	0.064
CY	0.804	0.080	0.812	0.091	0.088	0.684	0.162	0.325	0.166	0.001	0.001	0.509	0.220	0.042	0.022	0.036	0.067	0.037	0.020	0.045	0.048
CZ	0.882	0.061	0.946	0.037	0.005	0.670	0.261	0.898	0.007	0.003	0.002	0.096	0.139	0.080	0.105	0.074	0.160	0.149	0.104	0.080	0.088
DE	0.811	0.189	0.862	0.138	0.010	0.284	0.475	0.482	0.050	0.009	0.004	0.493	0.033	0.087	0.015	0.025	0.116	0.089	0.079	0.051	0.059
DK	0.922	0.029	0.935	0.023	0.000	0.531	0.283	0.630	0.069	0.006	0.012	0.321	0.001	0.026	0.052	0.035	0.225	0.123	0.095	0.103	0.122
EE	0.601	0.272	0.726	0.250	0.001	0.334	0.391	0.906	0.004	0.001	0.004	0.092	0.113	0.124	0.051	0.084	0.110	0.097	0.109	0.169	0.085
EL	0.888	0.016	0.916	0.016	0.078	0.592	0.133	0.193	0.277	0.001	0.004	0.532	0.049	0.021	0.034	0.223	0.048	0.039	0.004	0.027	0.026
ES	0.836	0.046	0.849	0.046	0.082	0.802	0.048	0.186	0.069	0.001	0.003	0.748	0.071	0.009	0.021	0.028	0.059	0.021	0.010	0.025	0.029
FI	0.826	0.007	0.933	0.006	0.019	0.559	0.238	0.658	0.204	0.019	0.006	0.151	0.202	0.057	0.046	0.048	0.145	0.122	0.091	0.126	
FR	0.806	0.067	0.880	0.047	0.063	0.724	0.079	0.454	0.085	0.001	0.001	0.463	0.109	0.005	0.049	0.059	0.108	0.111	0.050	0.036	0.072
HR	0.823	0.008	0.848	0.003	0.017	0.634	0.189	0.352	0.053	0.027	0.011	0.596	0.122	0.013	0.034	0.022	0.070	0.046	0.036	0.058	0.033
HU	0.964	0.016	0.980	0.012	0.025	0.655	0.243	0.729	0.022	0.001	0.007	0.252	0.167	0.087	0.075	0.061	0.118	0.113	0.063	0.049	0.044
IE	0.787	0.114	0.761	0.103	0.011	0.546	0.324	0.253	0.048	0.007	0.000	0.700	0.060	0.007	0.014	0.017	0.059	0.052	0.007	0.061	0.082
IS	0.905	0.059	0.924	0.046	0.002	0.626	0.275	0.598	0.102	0.001	0.000	0.305	0.130	0.013	0.028	0.064	0.180	0.109	0.045	0.095	0.149
IT	0.820	0.024	0.862	0.024	0.042	0.779	0.112	0.224	0.080	0.005	0.005	0.698	0.062	0.022	0.031	0.035	0.051	0.029	0.022	0.038	0.041
LT	0.902	0.002	0.959	0.003	0.014	0.519	0.316	0.867	0.014	0.001	0.001	0.124	0.293	0.034	0.112	0.067	0.110	0.049	0.046	0.129	0.068
LU	0.374	0.483	0.393	0.485	0.074	0.587	0.245	0.318	0.106	0.000	0.004	0.579	0.108	0.024	0.015	0.054	0.061	0.036	0.046	0.049	0.047
LV	0.585	0.234	0.793	0.182	0.006	0.414	0.399	0.891	0.003	0.002	0.007	0.106	0.221	0.023	0.093	0.085	0.122	0.098	0.084	0.138	0.074
MT	0.950	0.043	0.957	0.038	0.150	0.652	0.145	0.073	0.015	0.001	0.002	0.919	0.010	0.009	0.004	0.002	0.018	0.009	0.007	0.019	0.011
NL	0.829	0.027	0.907	0.023	0.012	0.532	0.288	0.282	0.056	0.003	0.000	0.665	0.060	0.008	0.011	0.016	0.089	0.052	0.038	0.050	0.037
NO	0.877	0.048	0.891	0.043	0.014	0.368	0.437	0.623	0.106	0.008	0.016	0.270	0.091	0.026	0.017	0.053	0.214	0.114	0.142	0.041	0.065
PL	0.957	0.010	0.990	0.004	0.004	0.524	0.410	0.518	0.261	0.008	0.002	0.226	0.118	0.018	0.080	0.262	0.097	0.071	0.053	0.057	0.050
PT	0.928	0.008	0.950	0.007	0.283	0.631	0.029	0.359	0.197	0.003	0.010	0.444	0.145	0.032	0.059	0.158	0.075	0.025	0.017	0.031	0.048
RO	0.936	0.001	0.939	0.001	0.023	0.728	0.112	0.370	0.219	0.005	0.010	0.440	0.080	0.040	0.076	0.218	0.060	0.026	0.024	0.034	0.010
SE	0.942	0.024	0.855	0.058	0.000	0.409	0.369	0.731	0.058	0.002	0.007	0.582	0.035	0.021	0.009	0.016	0.152	0.057	0.033	0.087	0.095
SI	0.791	0.178	0.000	0.000	0.004	0.752	0.148	0.578	0.071	0.005	0.010	0.351	0.193	0.006	0.066	0.061	0.091	0.085	0.093	0.047	0.089
SK	0.932	0.023	0.980	0.010	0.001	0.451	0.482	0.846	0.006	0.004	0.002	0.153	0.203	0.052	0.096	0.034	0.161	0.107	0.110	0.075	0.048
UK	0.808	0.064	0.877	0.036	0.042	0.679	0.099	0.577	0.051	0.087	0.003	0.375	0.127	0.044	0.028	0.005	0.152	0.078	0.068	0.097	0.104

Note: Omitted categories are: “Outside EU” for birth area, “Outside EU” for citizenship, “Illiterate” for education, “Unknown/Dead” and “Other inactive” for activity. ISCO-08 occupation definitions are: 1 “Elementary”, 2 “Plant Operator” 3, “Craft/Trades” 4, “Agriculture” 5, “Service and Army” 6, “Clerical”, 7 “Technician”, 8 “Professional”, 9 “Manager”; ”Dead/Unknown/Not working” is not shown. Omitted categories for supervisory are: “No.”

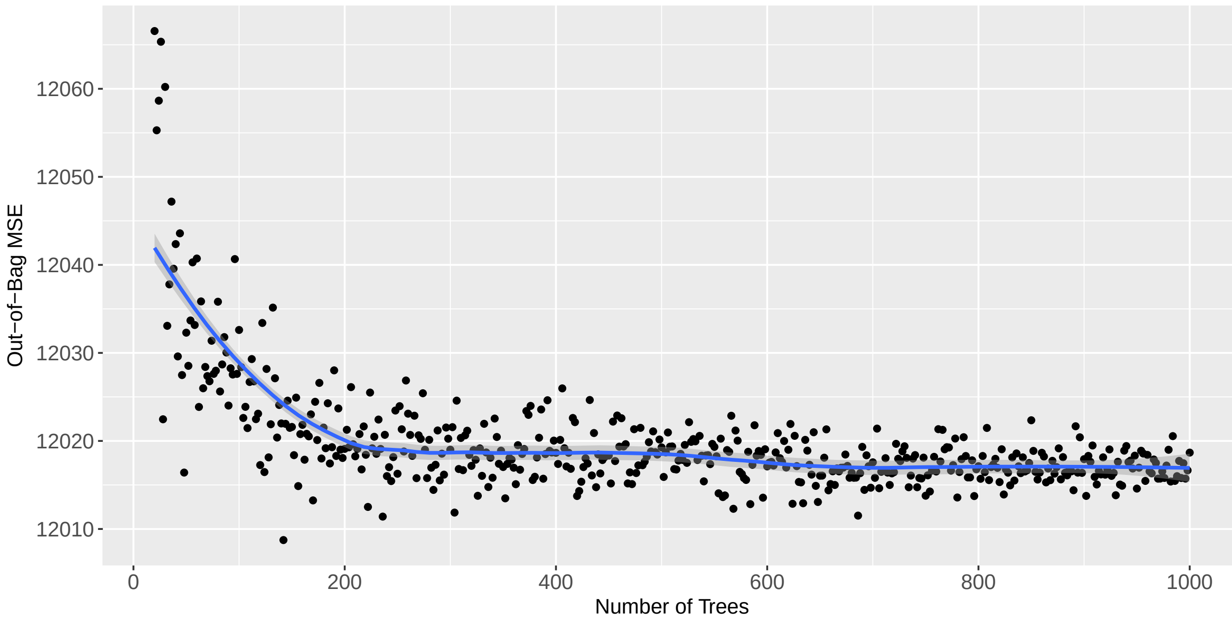
4.A.3 Empicial Robustness Checks

Figure 4A.1: Tuning Conditional Inference Trees



Note: The figure compares the test error of random forests with different conditional inference trees. The test error of random forests is standardized to equal 1, such that a test error larger than 1 represent worse fits than random forests. "Ctree: tuned α" uses cross-validation to tune α. 95% confidence intervals are derived based on 200 bootstrapped re-samples of the test data using the normal approximation method.

Figure 4A.2: Optimal Size of Forests



Note: The figure compares the out-of-bag mean squared error, MSE^{OOB}, for Germany using varying forest sizes (different levels of B). We allow for 6 circumstances to be considered at each splitting point ($\bar{P} = 6$). The blue line indicates a fitted line. After around 200 trees, improvements in the error tend to be negligible. Similar patterns were found with other countries and other levels of $\bar{P}$. For this reason, we set $B^* = 200$ in our random forests.

4.A.4 Opportunity Structures

Trees

Figure 4A.3: Opportunity Tree: Austria

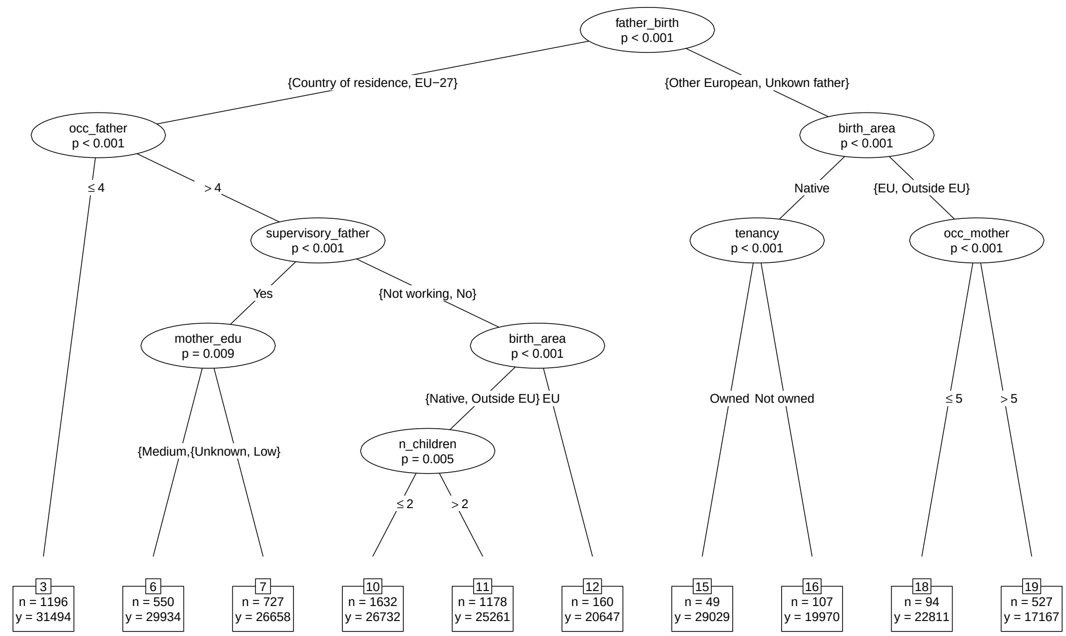


Figure 4A.4: Opportunity Tree: Belgium

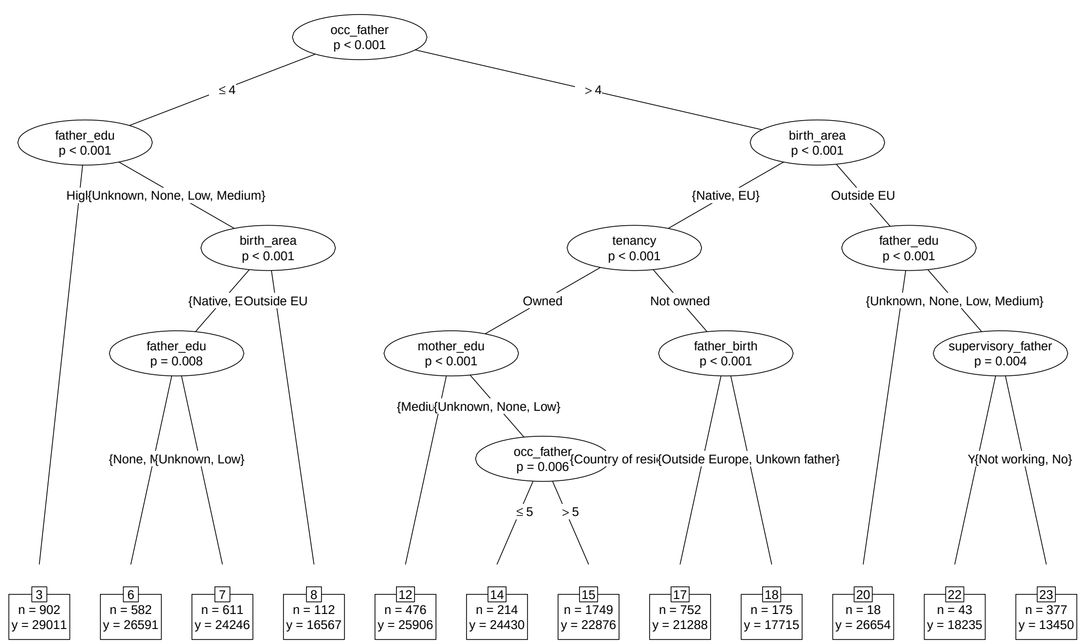


Figure 4A.5: Opportunity Tree: Switzerland

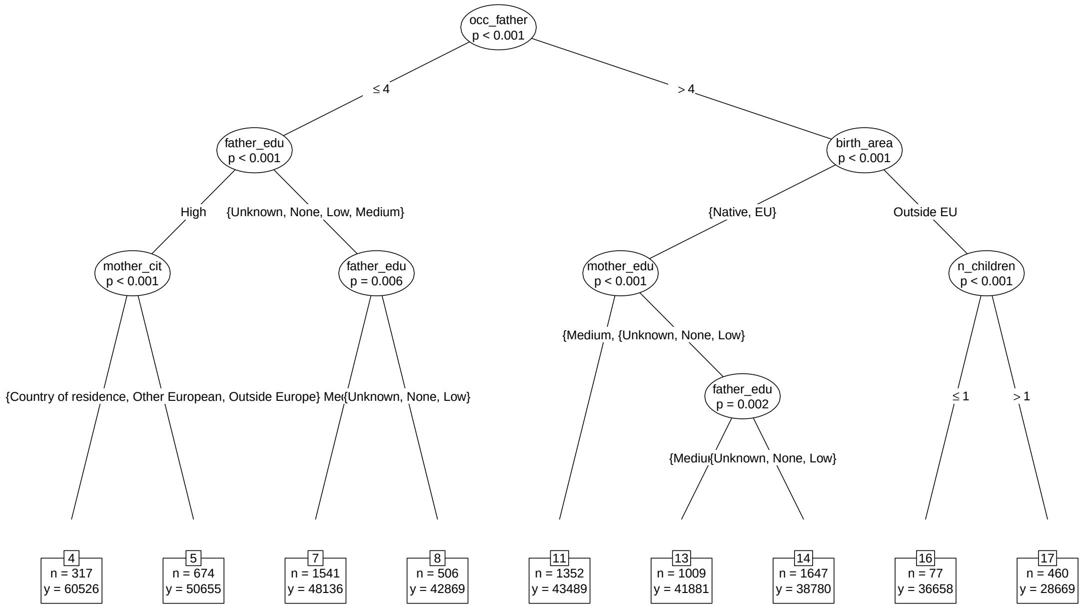


Figure 4A.6: Opportunity Tree: Cyprus

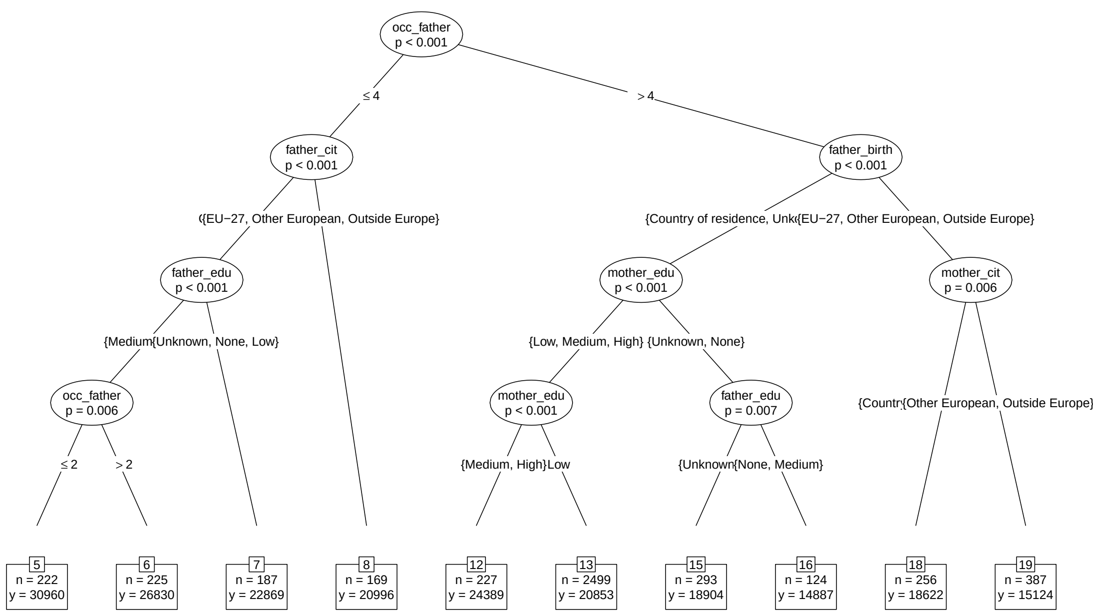


Figure 4A.7: Opportunity Tree: Czech Republic

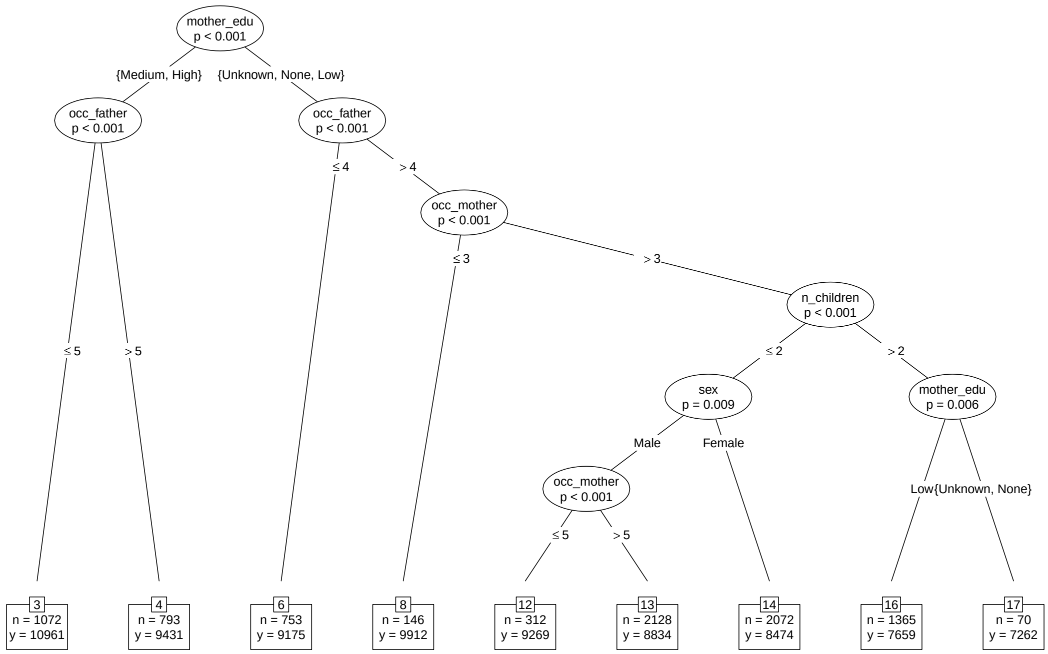


Figure 4A.8: Opportunity Tree: Denmark

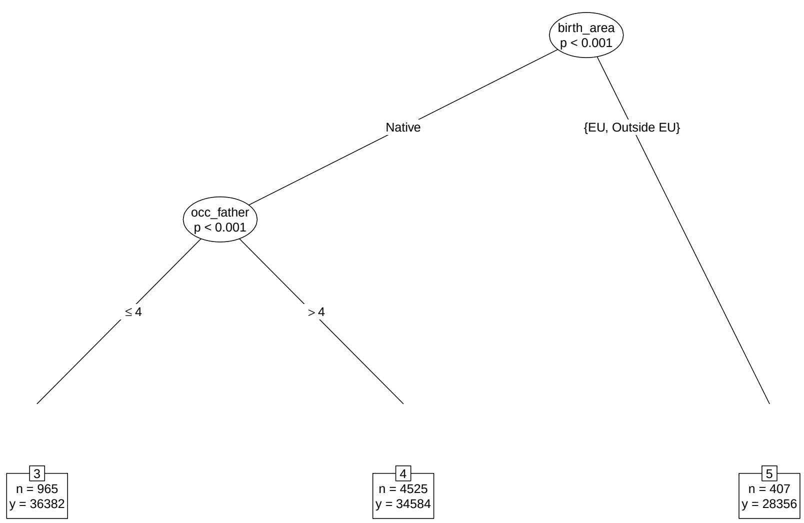


Figure 4A.9: Opportunity Tree: Estonia

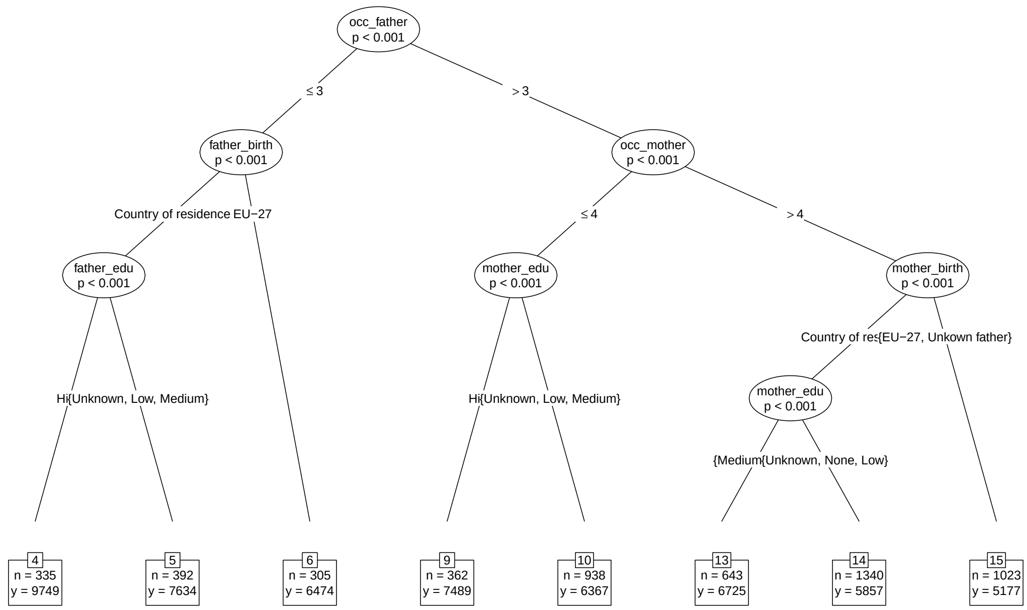


Figure 4A.10: Opportunity Tree: Greece

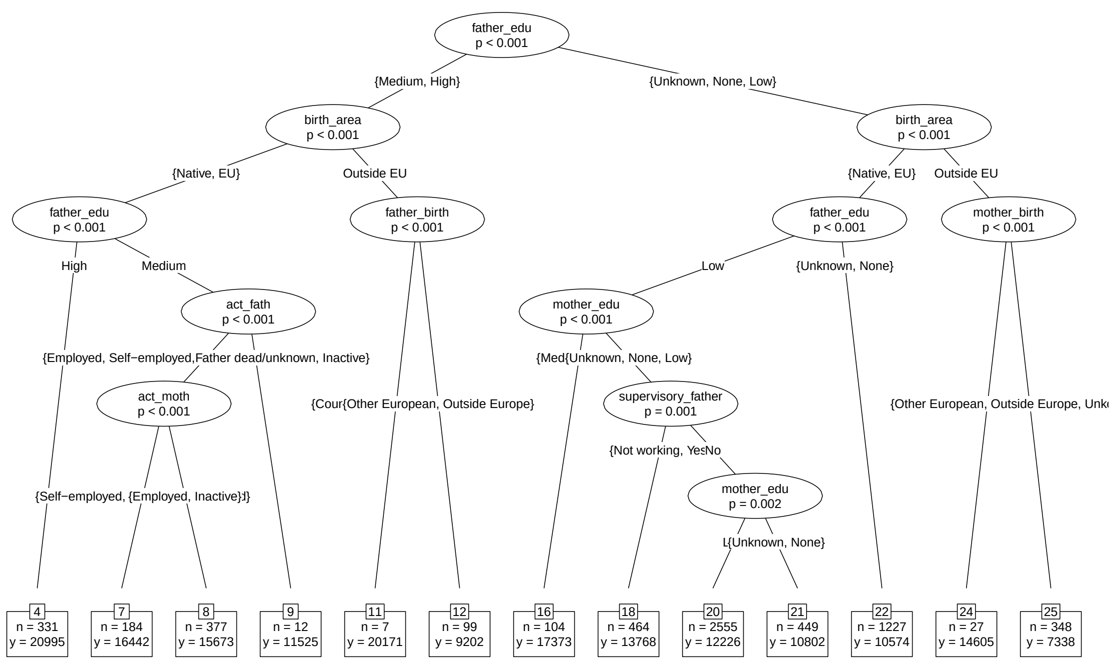


Figure 4A.11: Opportunity Tree: Spain

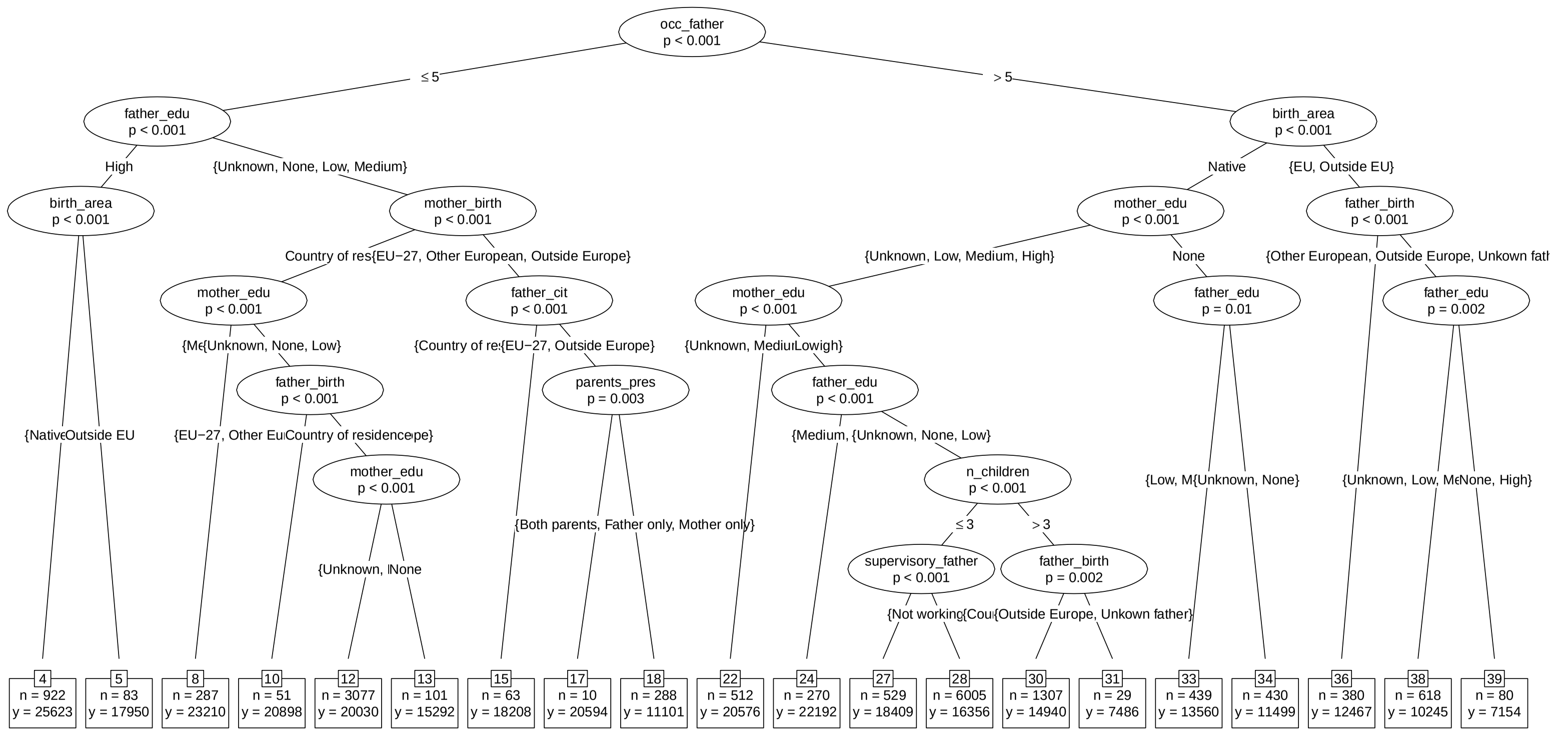


Figure 4A.12: Opportunity Tree: Finland

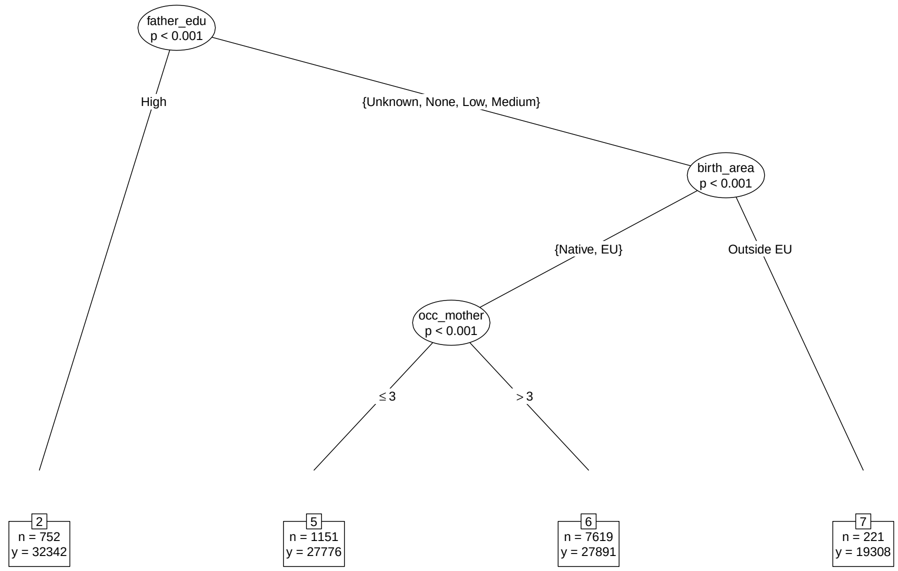


Figure 4A.13: Opportunity Tree: France

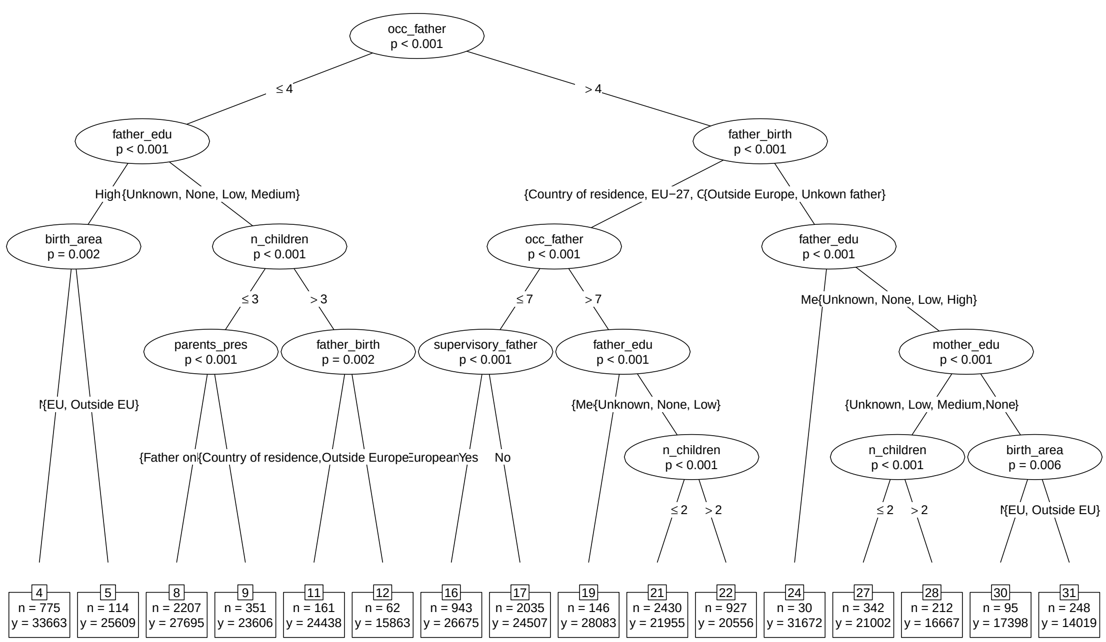


Figure 4A.14: Opportunity Tree: Croatia

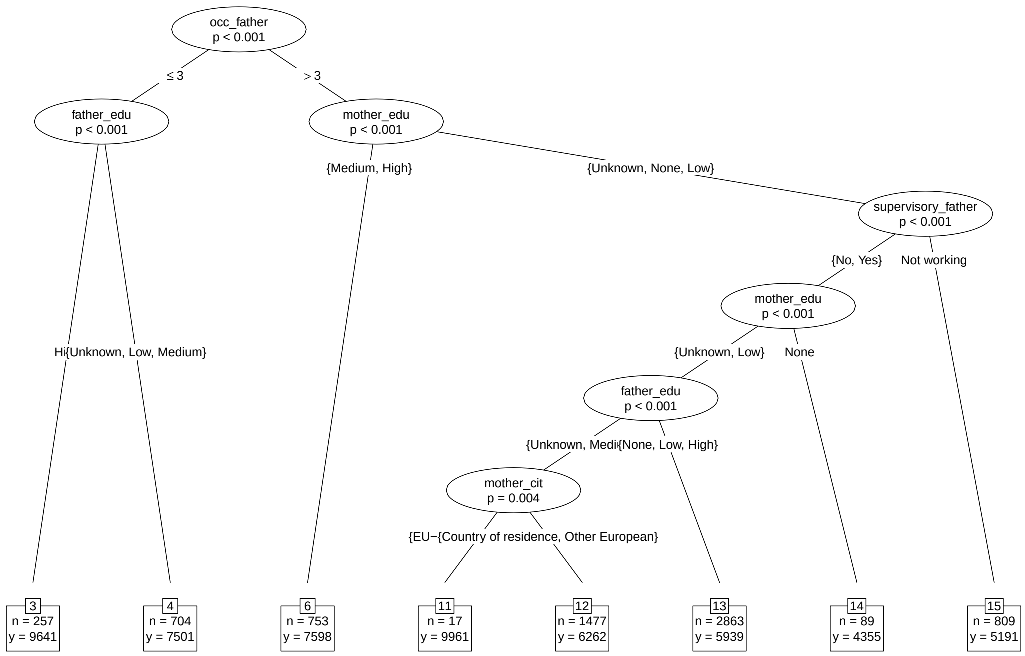


Figure 4A.15: Opportunity Tree: Ireland

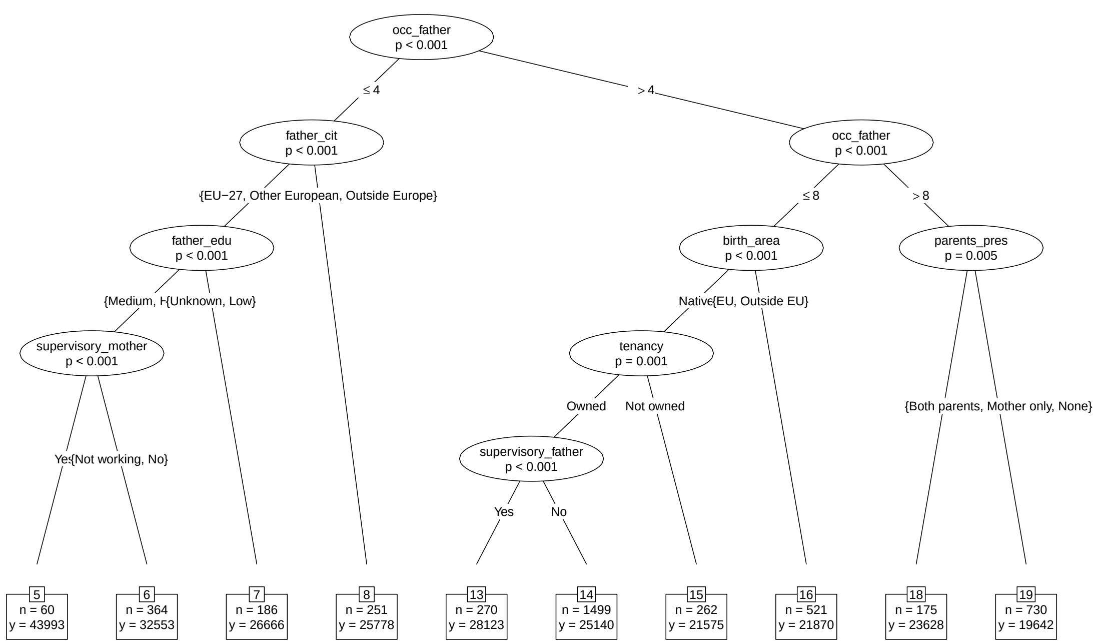


Figure 4A.16: Opportunity Tree: Hungary

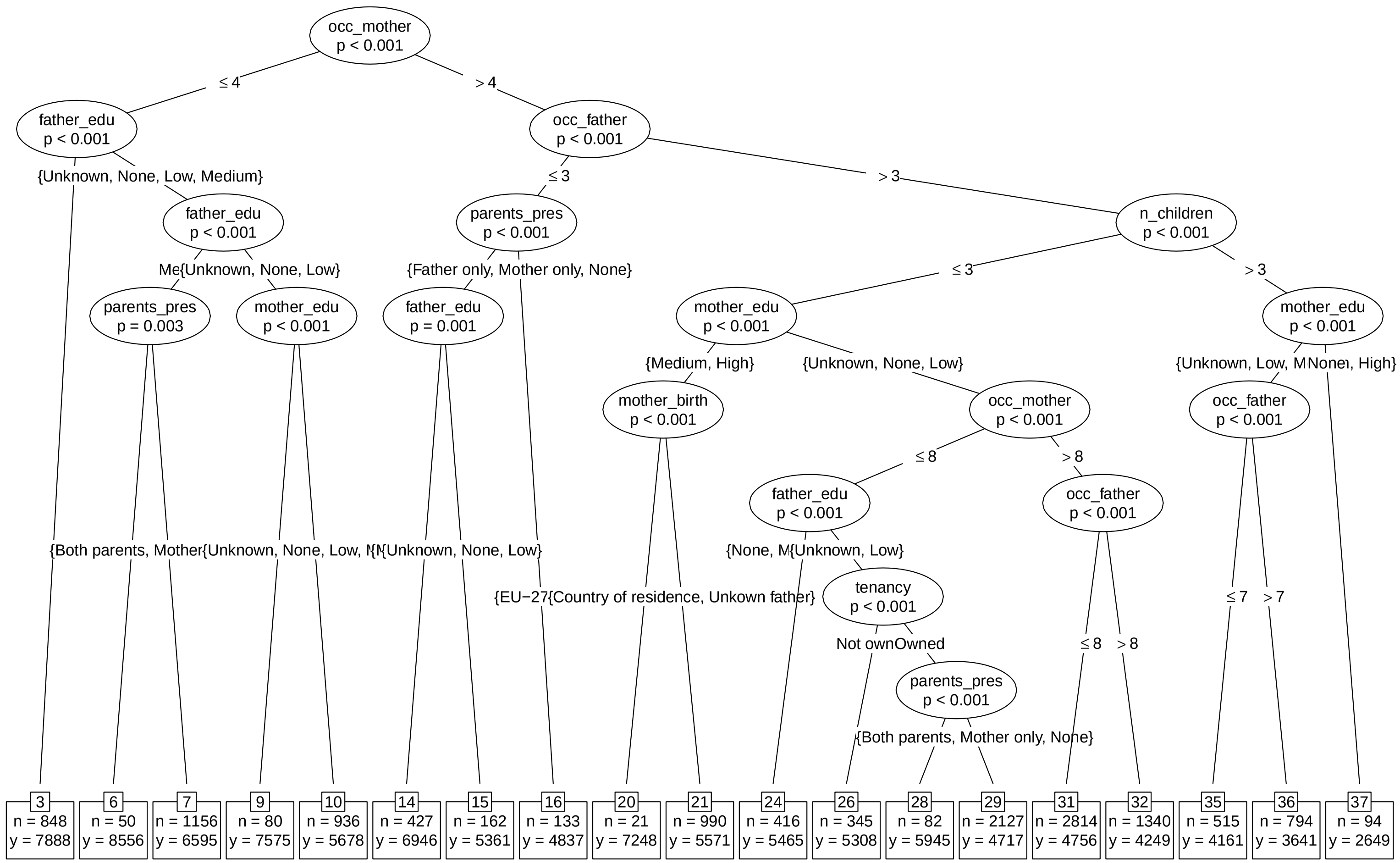


Figure 4A.17: Opportunity Tree: Italy

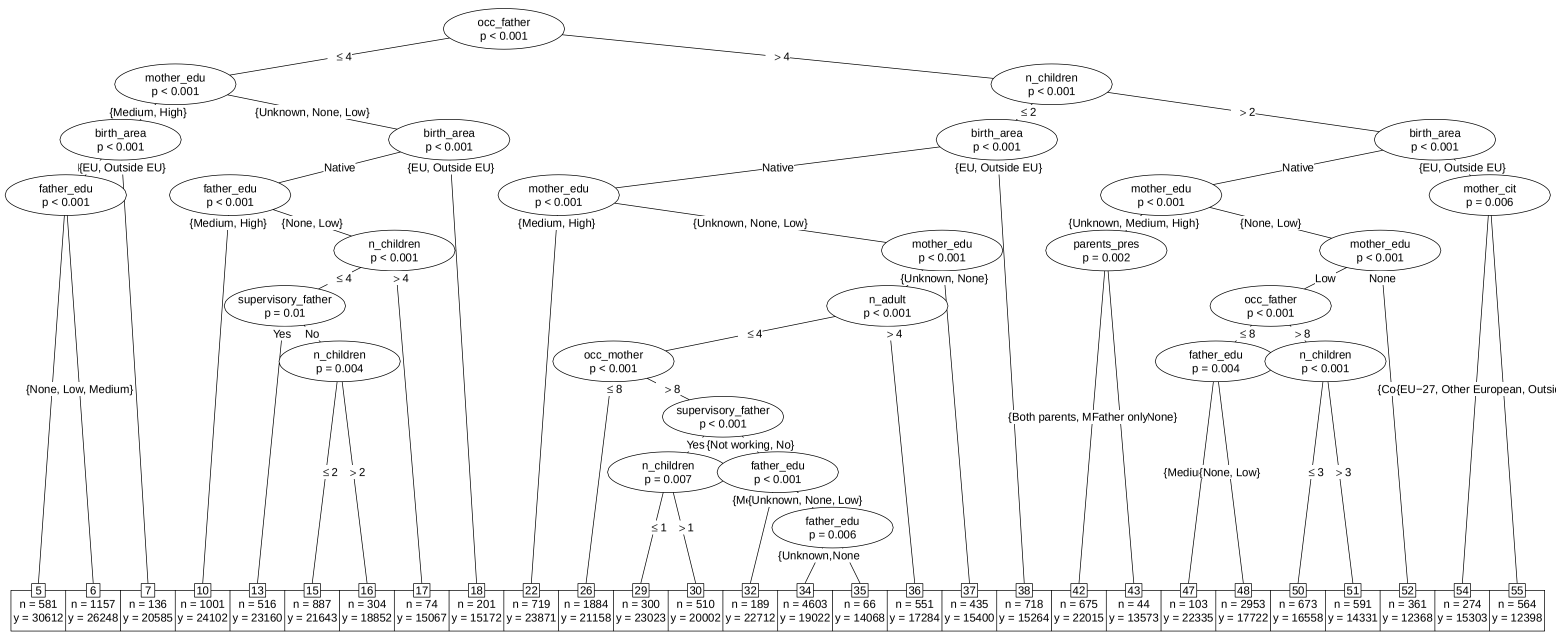


Figure 4A.18: Opportunity Tree: Luxembourg

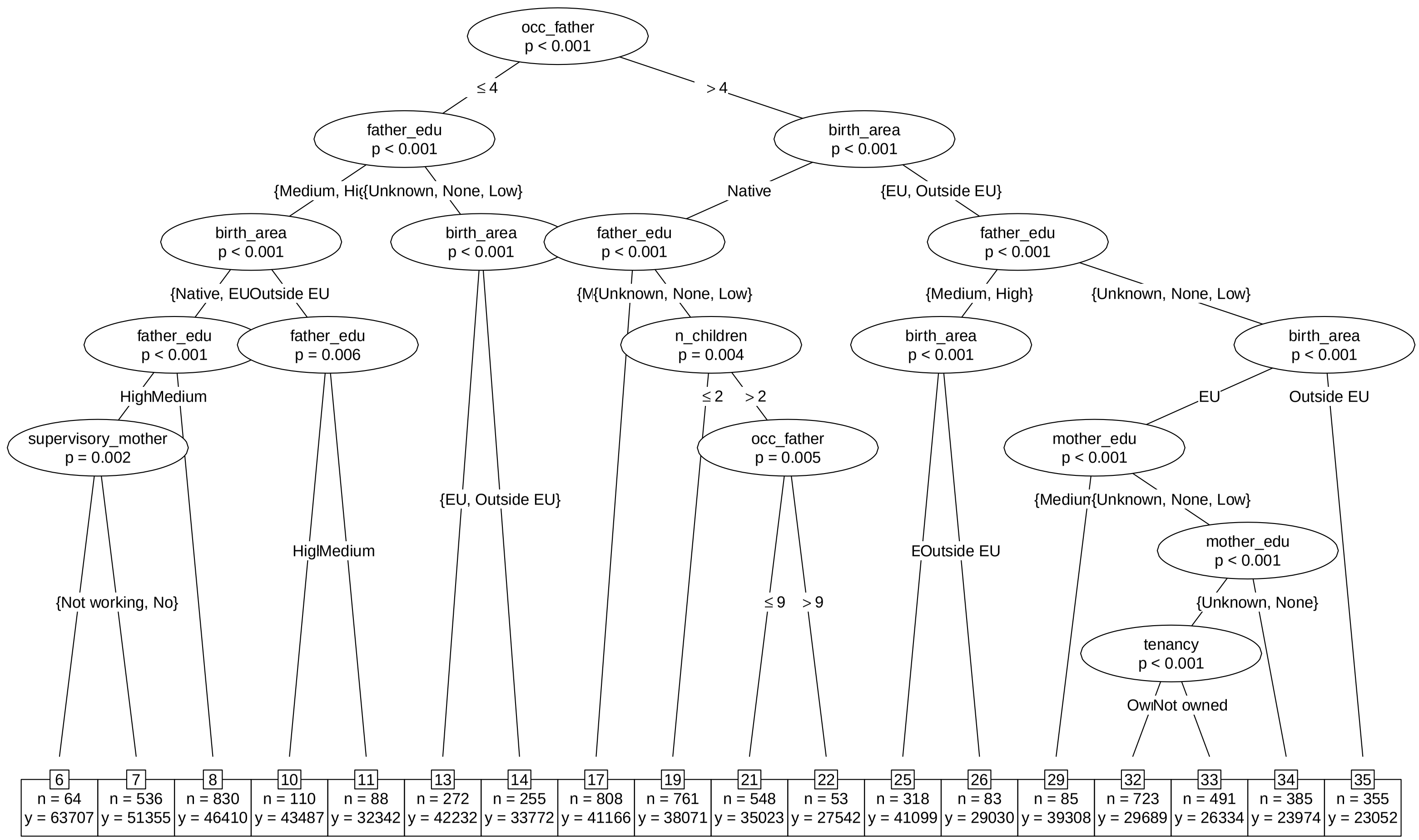


Figure 4A.19: Opportunity Tree: Lithuania

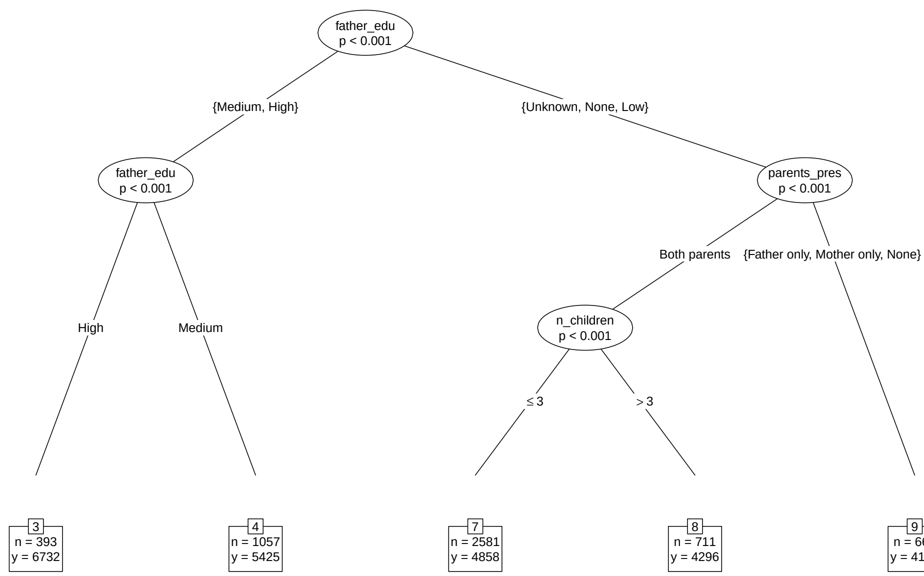


Figure 4A.20: Opportunity Tree: Latvia

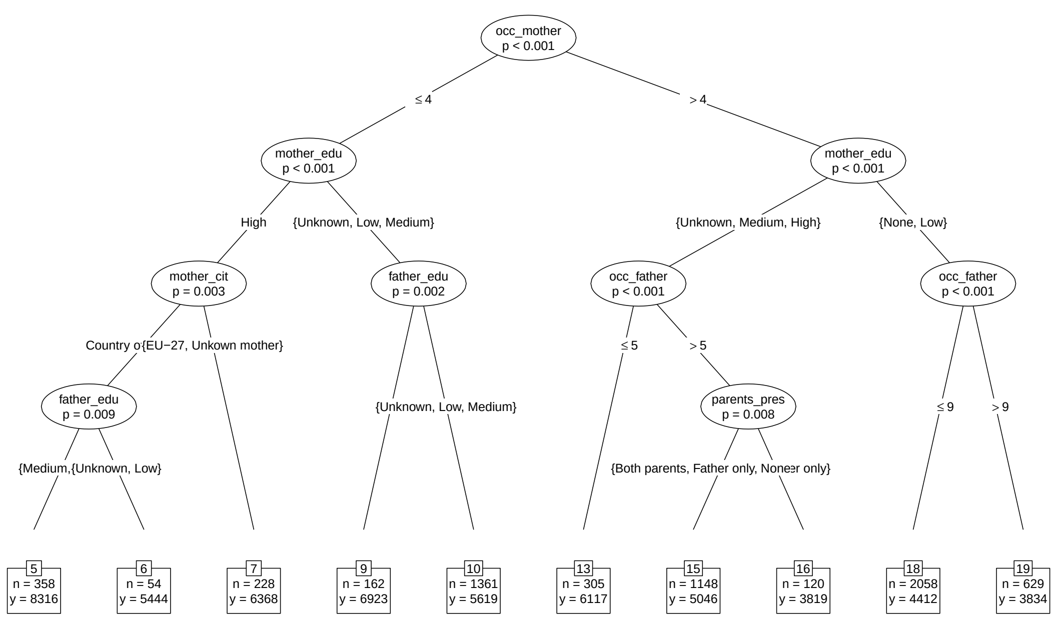


Figure 4A.21: Opportunity Tree: Montenegro

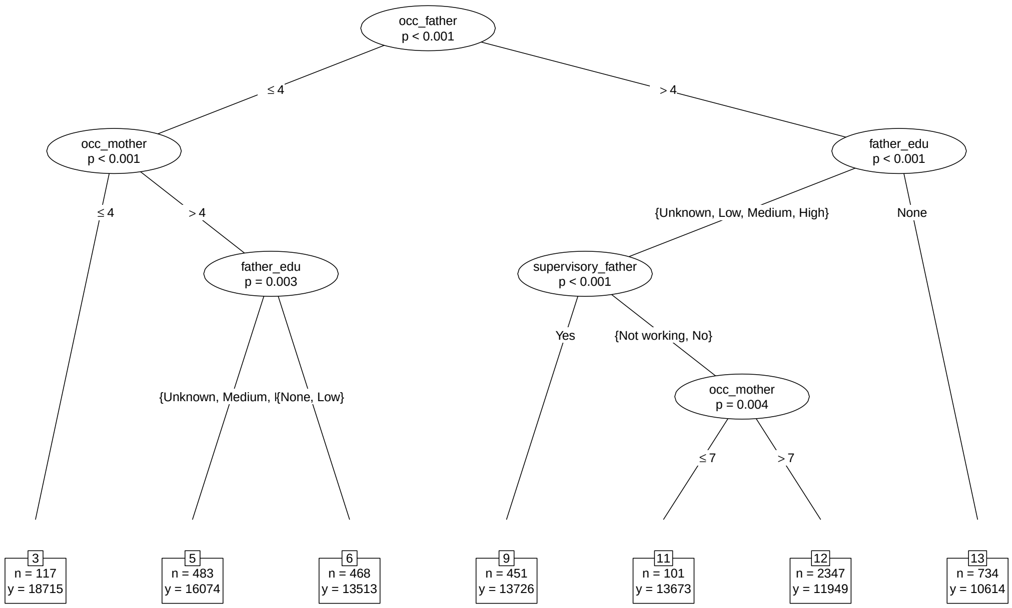


Figure 4A.22: Opportunity Tree: Netherlands

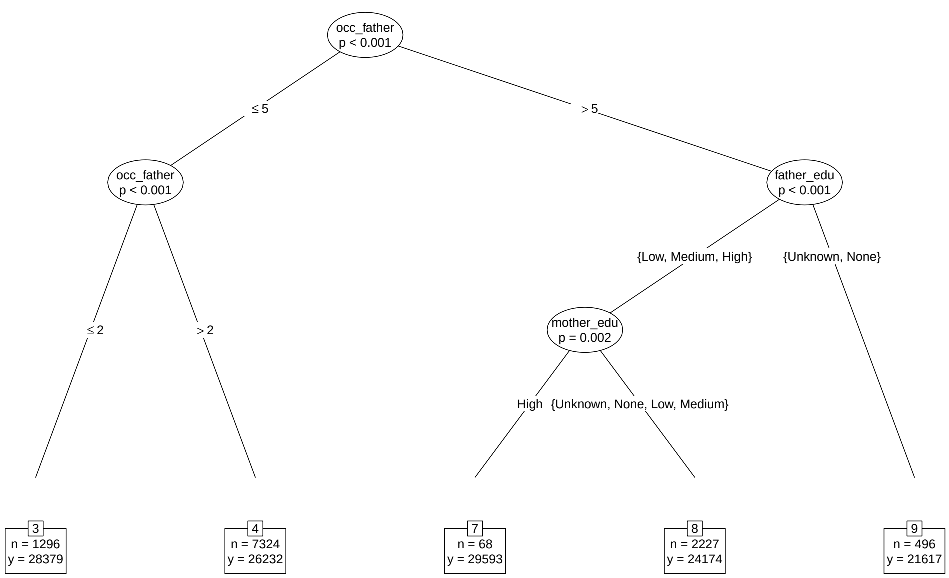


Figure 4A.23: Opportunity Tree: Norway

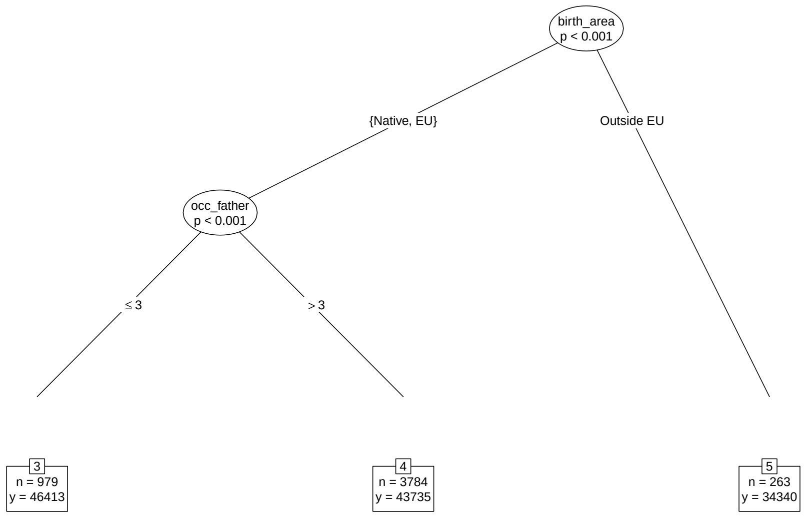


Figure 4A.24: Opportunity Tree: Slovenia

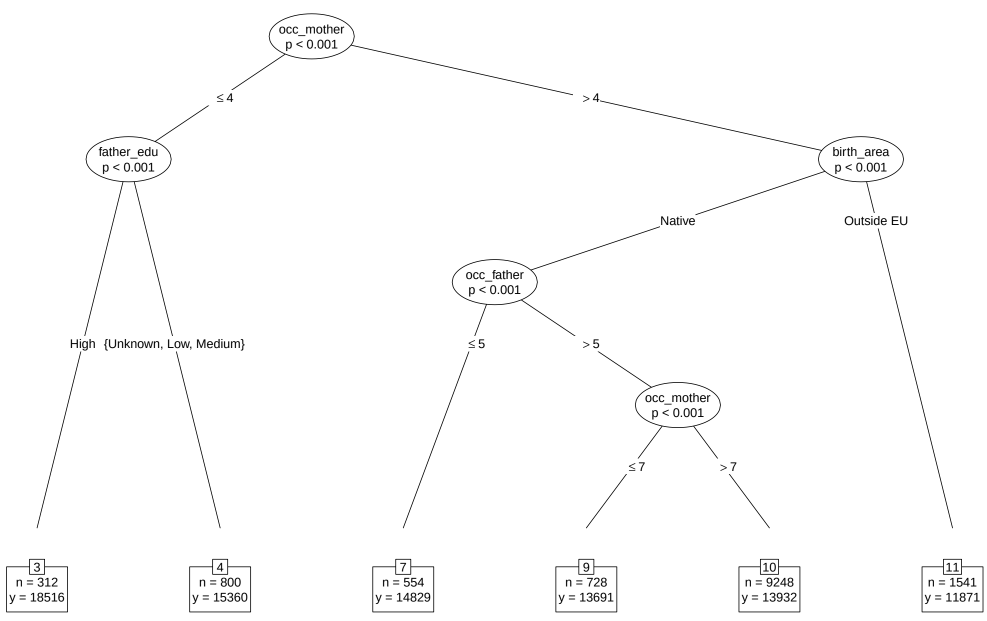


Figure 4A.25: Opportunity Tree: Poland

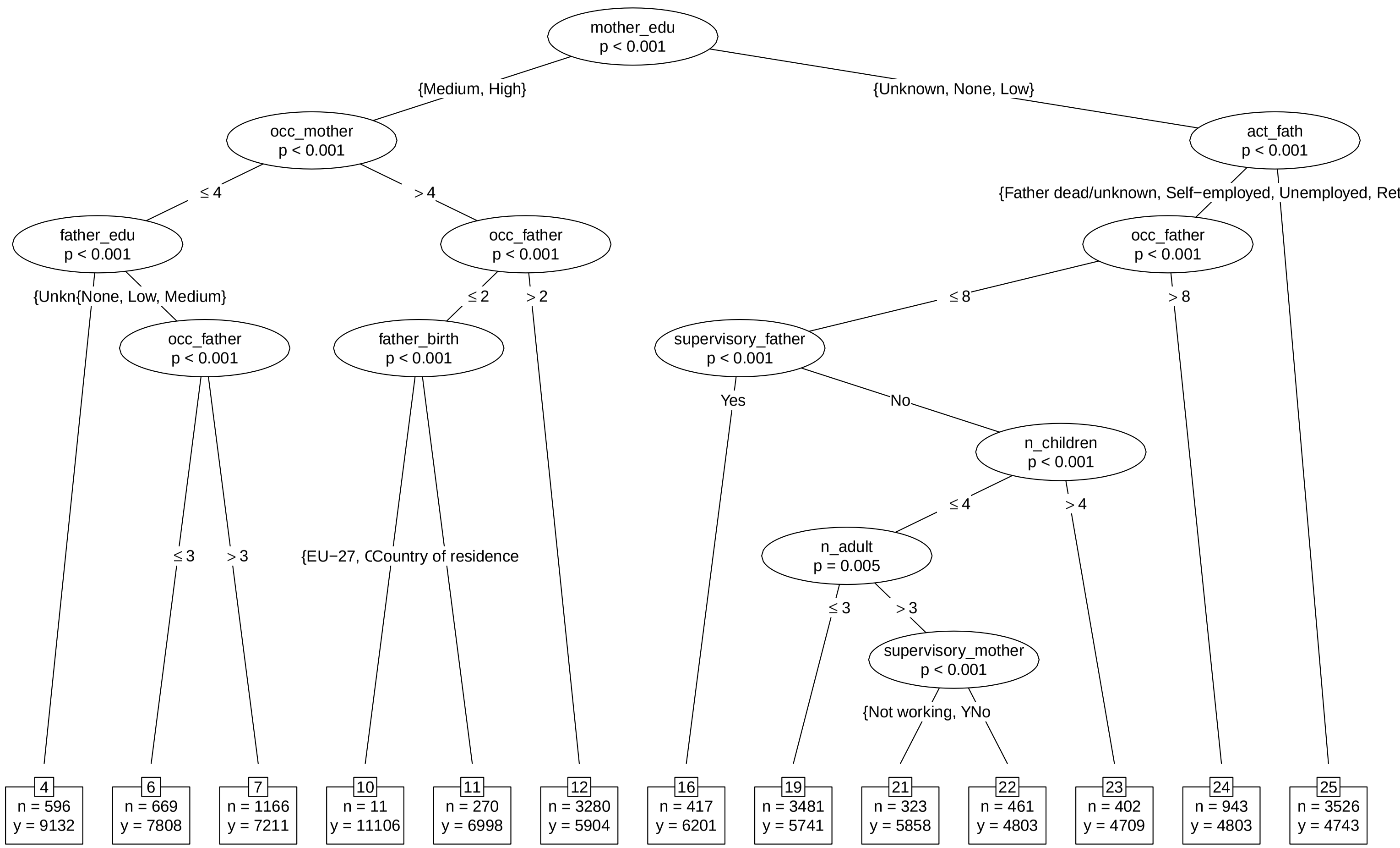


Figure 4A.26: Opportunity Tree: Portgual

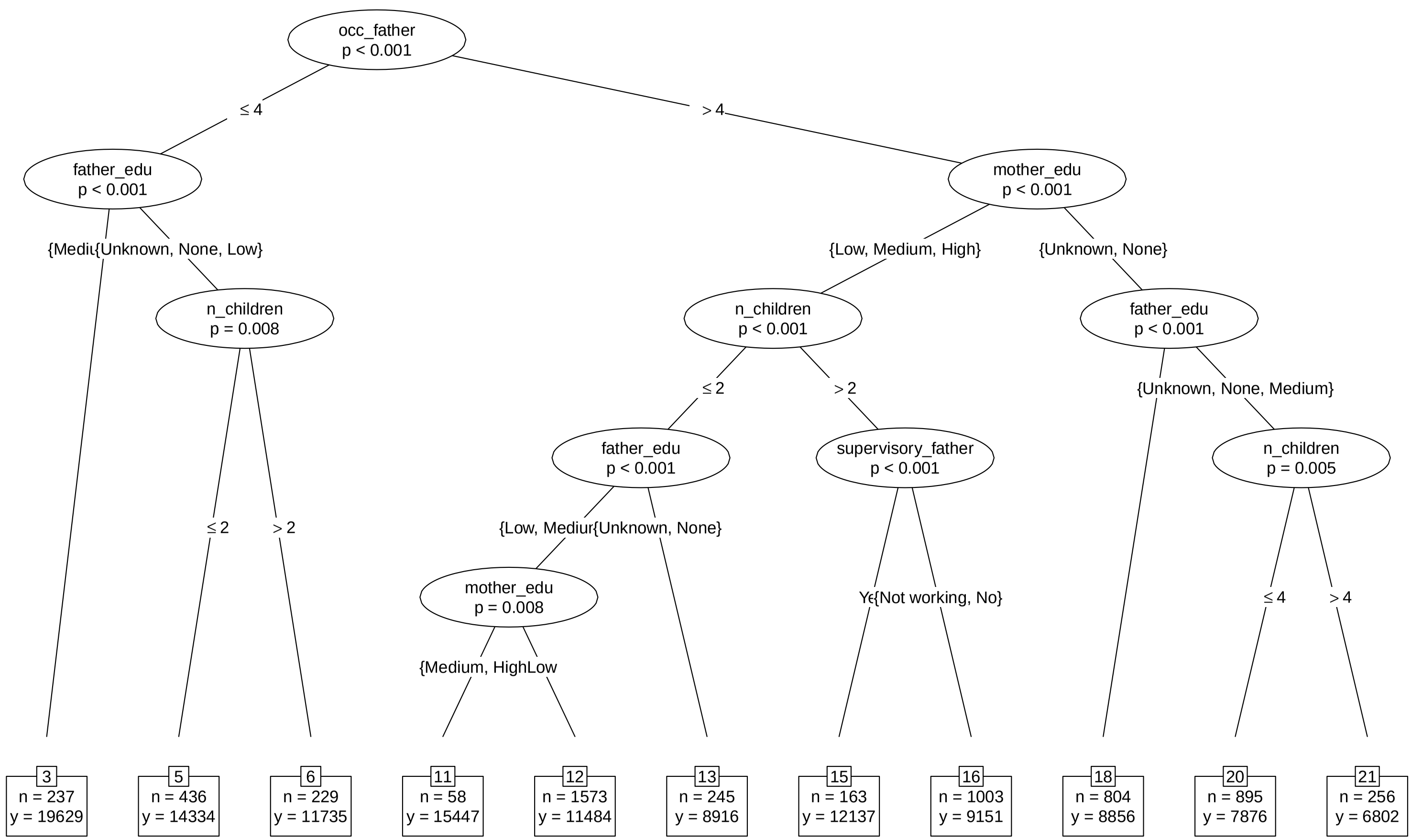


Figure 4A.27: Opportunity Tree: Romania

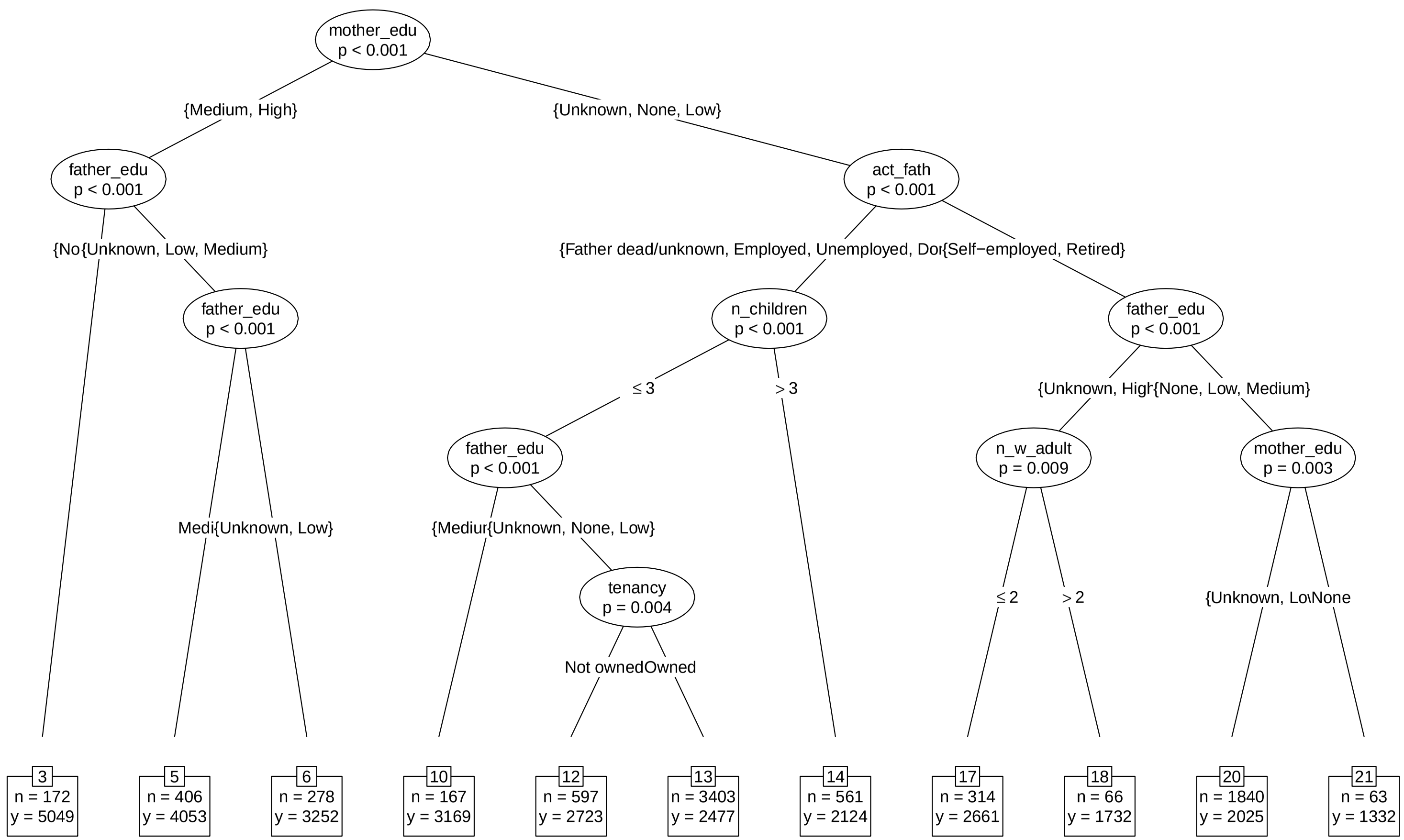


Figure 4A.28: Opportunity Tree: Slovakia

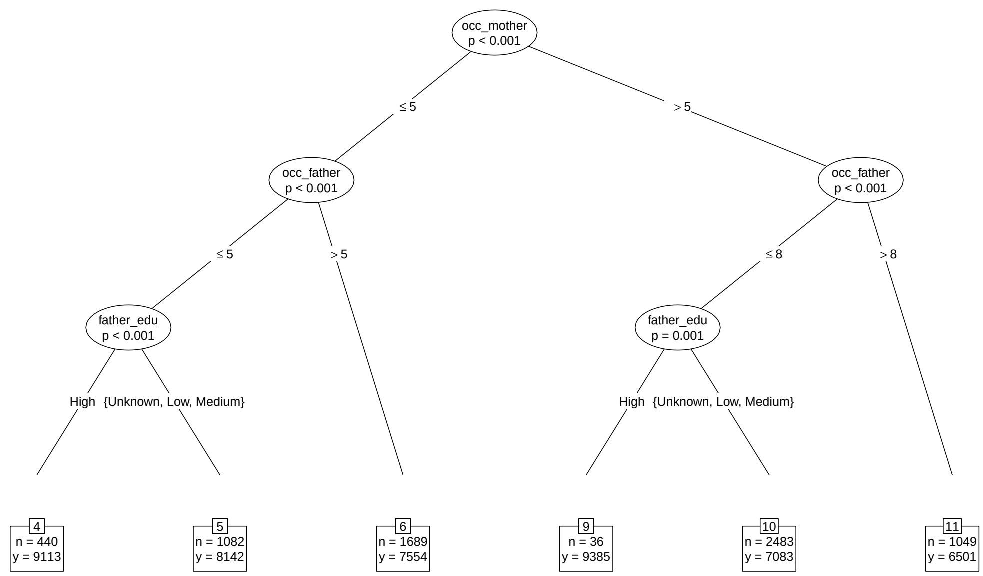
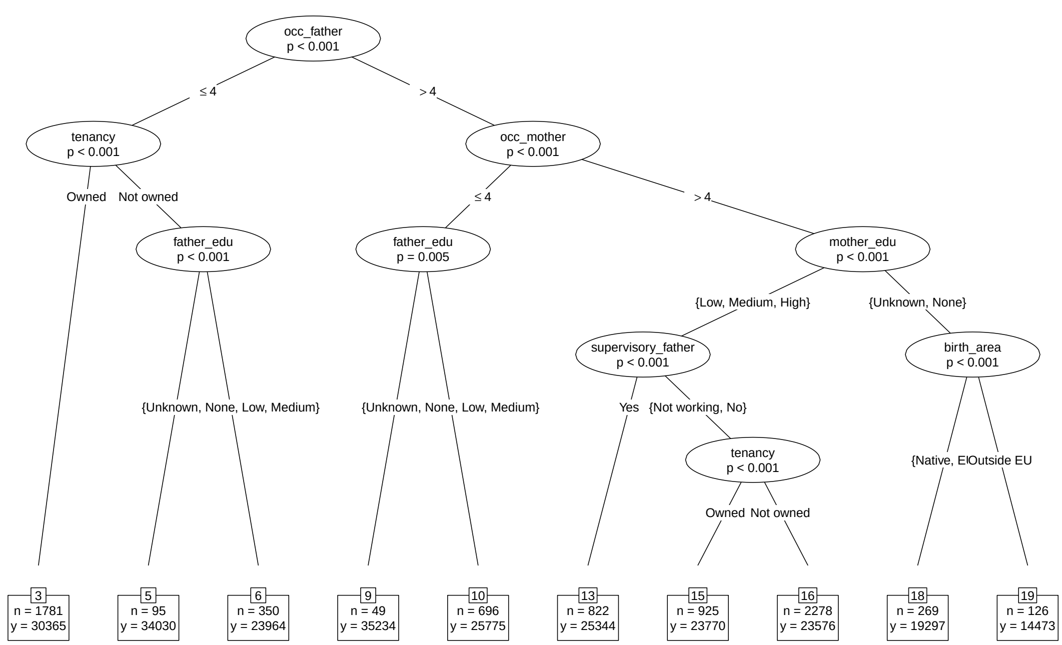
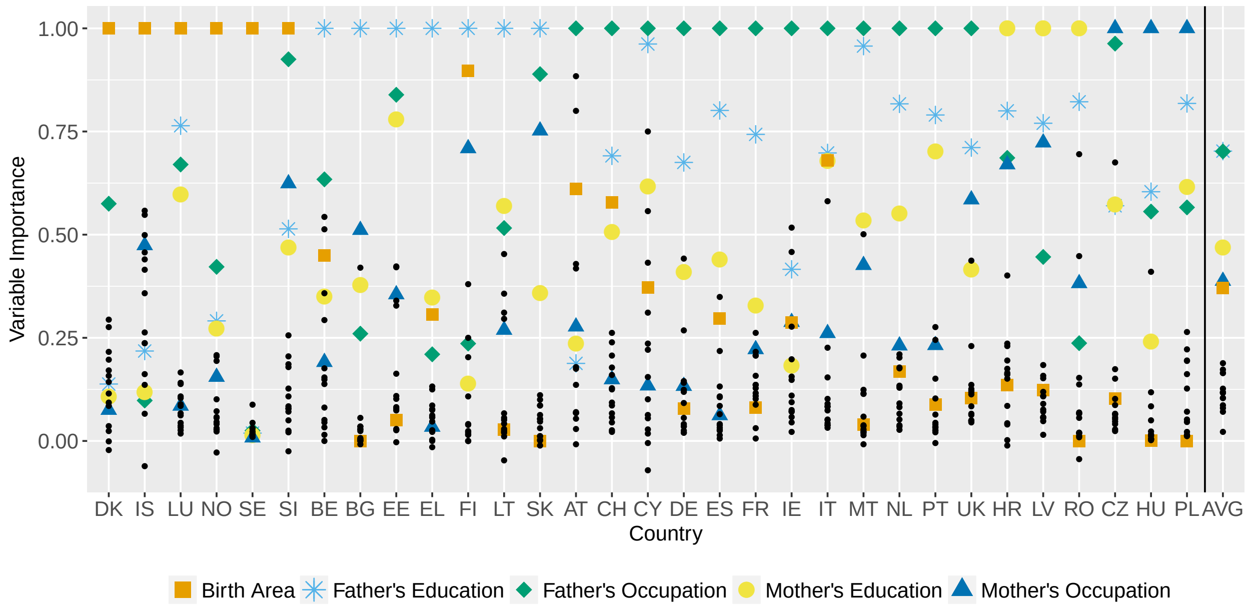


Figure 4A.29: Opportunity Tree: United Kingdom



Forests

Figure 4A.30: Variable Importance Plot

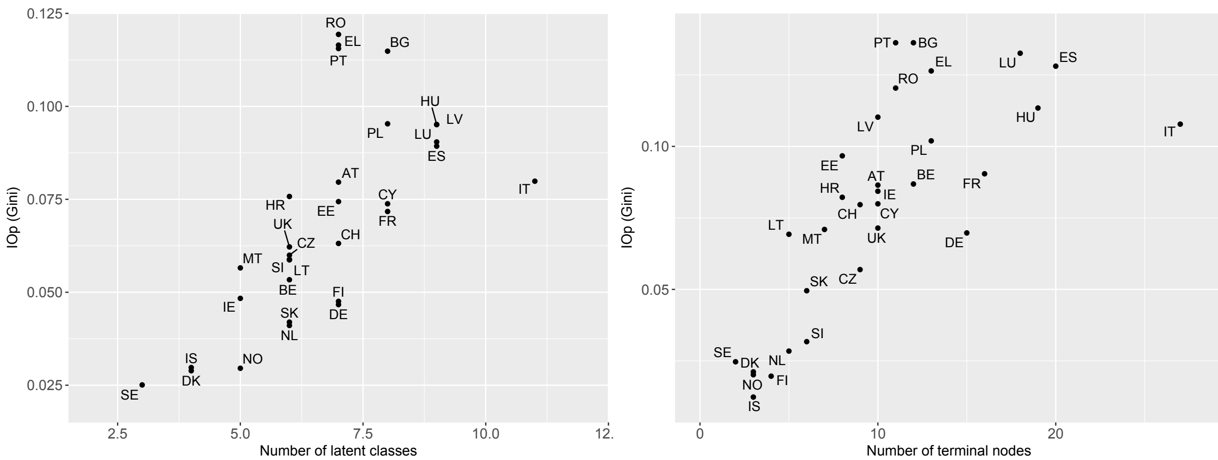


Note: Each dot shows the importance of a particular circumstance for the predictions from our random forest. The importance of a circumstance is measured by permuting the circumstance, calculating a new MSE^{OOB}, and computing the difference in the MSE^{OOB} between the original model and the model with the permuted circumstance. The importance measure is standardized such that the circumstance with the greatest importance in each country equals one.

4.A.5 Tree Complexity

Figure 4A.31: Correlation of Complexity and Magnitude of Estimate

(a) Latent Class Analysis (b) Conditional Inference Tree



Note: Complexity of the opportunity structure is proxied by the number of latent classes and the number of terminal nodes, respectively.

Chapter 5

Free to Choose or Free to Lose? Understanding Individual Attitudes towards Government Paternalism

Free to Choose or Free to Lose?
Understanding Individual Attitudes towards Government Paternalism[1]

David Dreyer Lassen
University of Copenhagen

Daniel Gerszon Mahler
University of Copenhagen

Abstract

In the past decades, behavioral economics has credibly identified numerous decision-making biases preventing people from acting in their own self-interest. This has given rise to a new reason for government interventions: internalities. In contrast to traditional reasons for government intervention, such as redistribution and externalities, overcoming internalities often involves the use of paternalistic policies. We investigate, theoretically and empirically, the formation of attitudes towards such paternalistic policies. Theoretically, we focus on the role of self-interest and distinguish self-interest as construed for the rational decision-maker from self-interest related to non-material benefits such as autonomy, and from self-interest arising from self-control problems.

Empirically, we employ two novel data sets – a Danish survey on political opinion combined with administrative data on actual behavior and a large-scale cross-country survey – to analyze attitudes towards coercive paternalistic policies in the health and financial domains. We show that targets of paternalism are more opposed to paternalism than non-targets both in Denmark and across nine Western democracies, and rely on our theoretical priors to explore mechanisms that can explain these attitudes. We find evidence for non-targets favoring paternalism for altruistic reasons and targets being relatively more in favor if they have self-control problems.

Keywords: Paternalism, internalities, public opinion, self-control problems
JEL codes: D61, H11, I18, I38

[1]Financial support from the Danish Council for Independent Research | Sapere Aude and the European Research Council through ERC Starting Grant no. 313673 is gratefully acknowledged. We thank Felix Sebastian Døssing and participants at University of Copenhagen, APSA, and Dublin for comments and suggestions.

"[W]e can't always have what we want. Sometimes we have to have what's good for us."
Margaret Thatcher, cited in Weiss (1991)

5.1 Introduction

Advances in psychology and behavioral economics have credibly identified numerous decision-making biases preventing people from acting in their own, long-run interest. This has led to calls for coercive government involvement in correcting these biased choices and revived the case for government paternalism in areas such as obesity and financial decision-making. Already, governments across the world are experimenting with such regulation. Both the Obama administration in the U.S. and the Cameron administration in the U.K. explicitly tried to harness behavioral economics in nudging or lightly coercing individuals to choose differently than they would have done on their own, by changing framings, default options or broader choice architectures. More direct measures have also been proposed and implemented. The former mayor of New York City, Michael Bloomberg, famously advocated a limit on the size of soda cups[2], the Danish government recently implemented and later, citing administrative difficulties, rescinded a so-called fat tax (Bødker *et al.*, 2015), and the UK recently introduced auto-enrollment pension schemes (Clark and Knox-Hayes, 2009). However, little is known about the formation and sources of attitudes towards policy where the government interferes in individual decisions on the premise of decision-making failures.

In this paper, we investigate theoretically and empirically the formation of attitudes towards a broad class of policies aimed at correcting individual decision-making failures by influencing or overruling voluntary individual decisions. Collectively, such policies are known as the exercise of (coercive) paternalism.[3] Classical liberalism has forcefully argued against this kind of interference with individual decision-making, insisting that individuals themselves are in the best position to make choices concerning their own welfare. Consequently, in cases not involving the possibility of harming others, it is argued that the state should refrain from overriding individual choice (Mill, 1869). This argument finds support in neoclassical economics, where rational decision-makers would trade off costs and benefits in ways that are individually optimal, and the state (or other outside actors) would be unable to improve upon these choices. However, behavioral economics, based on empirically founded

[2] Michael M. Grynbaum: "New York Plans to Ban Sale of Big Sizes of Sugary Drinks." *New York Times* May 30, 2012.

[3] We follow Le Grand and New (2015) and denote a government intervention towards an individual as paternalistic if 1) it is intended to address a failure of judgment by that individual and 2) it is intended to further the individual's own good. This does not exclude policies that also include considerations on costs borne by taxpayers, so-called fiscal externalities (discussed below), as long as the policy satisfies 1) and 2). We discuss the relation to nudging below.

research in economics and psychology, demonstrates that individuals as decision-makers may be boundedly rational (Simon, 1982), time inconsistent (Ainslie, 1975; Laibson, 1997) or suffer from various decision-making biases (Tversky and Kahneman, 1974; Kahneman, 2011). In these cases, individuals can make decisions that may harm themselves, often construed, theoretically, as a conflict between current and future selves (Thaler and Shefrin, 1981).

In our analysis, we consider attitudes towards legislation intended to affect smoking, consumption of unhealthy food, pension savings, and the take-up of risky mortgages. These choice domains are all characterized by potentially conflicting short-term and long-term goals *within* individuals: outcomes that are attractive in the short term – e.g., spending rather than saving for pension (Thaler and Benartzi, 2004) or the consumption of comforting, but unhealthy food (O'Donoghue and Rabin, 2006) – can influence long-term welfare negatively. For this reason, decisions that do not, due to the existence of choice biases, adequately account for long-term consequences are said to have *internalities* (Rabin, 2013; Herrnstein *et al.*, 1992)[4], and it is these internalities that are the subject of interest for policy-makers and regulators.[5]

Paternalistic policies can influence decisions and outcomes in two ways: Directly, by affecting individuals' choices and freedom, and indirectly, through public budgets or broader costs to society. The direct effects include the main effect on decisions by choice-biased individuals, but also effects on non-intended targets, typically individuals who are not subject to choice-biases, and threats to individual autonomy and freedom. Both the research and more policy-oriented literature are well aware of the challenge that paternalism present to individuals without choice-biases and, more generally, to individual liberty and agency. Consequently, various forms of light paternalism and nudging, designed to affect only those with biased behavior while leaving others unaffected, have been suggested (Camerer *et al.*, 2003; Thaler and Sunstein, 2003, 2008; Schmidt, 2017). In practice, many proposed and implemented policies aimed at changing behavior – such as pension savings mandates and smoking restrictions – are more coercive in nature. These policies inevitably also affect individuals without behavioral biases who are consistently making rational decisions, a key point of contention for classical liberalism. Rational decision-makers exist alongside people with choice biases, who in turn can be distinguished by their awareness of their biases. Fol-

[4]In economic theory, *externalities* are consequences on bystanders of actions not reflected in prices, typically due to missing markets, which create a wedge between individual and societal optimality. In the present case, choice biases create an internal wedge between short run and long run individual optimality; hence, the term *internalities*.

[5]Some of the suboptimal consumption, particularly in the case of tobacco, may be due to addiction. Our discussion applies also to people suffering from addiction problems if the following two conditions apply: 1) purchases caused by addiction are considered a failure of judgment by the individual and 2) policies are enacted with the objective to help the addicts.

lowing the behavioral economic literature, individuals aware of their own biases are denoted sophisticated, in contrast to those who are not cognizant of such biases, denoted naïve. This distinction between rational vs. sophisticated and naïve choice-biased individuals requires revisiting the theoretical notion of self-interest as used in the literature on public opinion and attitude formation. We develop this point below.

The potential costs and benefits of coercive paternalism are not limited to the direct effects on the individuals, choice-biased or not, whose behavior is influenced. Consequences of smoking and obesity, for example, are costly to treat, and such costs, known as fiscal externalities, are borne largely by taxpayers or co-insurees. Similarly, while a lack of pension savings is primarily an individual problem, in most modern societies a basic old age pension is provided by the state at the cost of taxpayers. Moreover, some paternalistic policy instruments may have consequences at odds with broader societal goals; for example, taxes on soda meant to address both internalities and fiscal externalities may be regressive and, as a consequence, counter efforts at redistribution (Galizzi and Loewenstein, 2016). Efforts to understand the formation of attitudes towards paternalism must address both direct and indirect effects.

Our investigation of attitudes towards paternalism is based on two large-scale data sets. The primary data set combines a survey carried out in Denmark in 2014 with third-party administrative register data. In the survey, we elicit attitudes towards paternalism across four policy domains, political attitudes, and self-reported behavior and outcomes, including smoking habits and body mass index. We combine this with administrative data on economic choices and outcomes, including pension payments, type of mortgage and borrowing behavior in the credit market. Methodologically, the presence of third-party reported data is important as social desirability bias and self-deception may introduce measurement error into self-reported behavior on sensitive issues (Hariri and Lassen, 2017). The second data set is a cross-country survey from 2017, where we elicit attitudes using the same survey instruments to examine the validity of the findings from the detailed study beyond Denmark. This is important, both because the relationship between individual views on the role of the state may differ across countries and since extant evidence has shown Denmark to be an outlier in attitudes towards the closely related concept of nudging (Reisch and Sunstein, 2016), which we consider separately below.

We show, across all policies in both the health and financial domains and both in Denmark and in the cross-country setting, that targets of paternalism – "paternalees" (Lusk *et al.*, 2014) – are more opposed to paternalistic policies than those whose behavior would not be affected by such policies. This holds true even for individuals with self-reported self-control problems. Paternalists – people who support implementing paternalist policies – are strongly motivated by altruism. Support for paternalism is strongly increasing in years of education and is more pronounced among left-wing voters and people who trust government. Concerns about the costs borne by tax payers, fiscal externalities, are important for support for paternalism in the health domain. Comparisons of attitudes towards hard paternalism and nudging suggests that paternalee resistance is largely coming from concerns over relinquishing the freedom to choose. While there is, as far as we know, no work on coercive paternalism across domains, a number of papers have considered public opinion towards policies aimed at addressing smoking with a self-interest perspective (Hersch, 2005; Green and Gerken, 1989), and, more recently, obesity (Gollust *et al.*, 2014; Lund *et al.*, 2011) separately. Oliver and Lee (2005), Pedersen *et al.* (2014) and Gyrd-Hansen and Kjær (2015) consider attitudes towards smoking and obesity jointly, but do not consider decision-making in the economic domain, nor broader motivations for attitudes towards paternalism. Recently, there has also been a number of studies considering attitudes towards nudges across domains (Felsen *et al.*, 2013; Arad and Rubinstein, 2017; Hagman *et al.*, 2015; Loewenstein *et al.*, 2015; Tannenbaum *et al.*, 2017; Jung *et al.*, 2016; Reisch and Sunstein, 2016; Sunstein, 2016). Our focus is on coercive paternalism.

The paper proceeds as follows. The next section presents a theoretical framework for thinking about attitude formation towards paternalistic policies, with a particular focus on a broad concept of self-interest, including the possibilities of biased choice and the value of decision autonomy. Section 5.3 presents the Danish data, including measures of attitudes and variable construction. Section 5.4 presents the main results and Section 5.5 replicates these results with cross-country empirical evidence. Section 5.6 contains two key extensions: measures of personality traits and an analysis of nudging, while Section 5.7 concludes.

5.2 Theoretical Framework

Paternalistic policies seek to override individual choice with the specific aim of protecting people from the consequences of their own choices, that is, to save people from themselves. This means that the concept of self-interest is more complex than when individual choice is synonymous with self-interest, as in the neoclassical idea of revealed preference, where choices are made precisely because they are optimal for the individual and, as a consequence, reveal individual preferences.

What do behavioral shortcomings mean for how we conceptualize and understand self-interest in attitude formation? We distinguish between three types of self-interest: The "classical" self-interest, procedural or expressive self-interest and behavioral self-interest.[6]

[6] An alternative, or additional, way of considering different conceptions of self-interest is the distinction

Under classical self-interest, everyone who sees their choices constrained by paternalistic policies will disapprove, as they were already choosing optimally. Classical self-interest can also be invoked if individual behavior has external effects. If tax payers or co-insurees pay for health costs associated with other people's unhealthy life-styles, or if the lack of pension savings on part of the population leads government to pay out pensions instead, at the cost of tax payers (Le Grand and New, 2015). Such effects, called fiscal externalities, may cause people without self-control problems (and those who are not cognizant of having such problems) to support paternalism.

Outside of the classical framework, procedural or expressive self-interest can be a factor that influences attitudes towards policy (Saint-Paul, 2011). If people value making their own decisions separately from the outcome of these decisions, as suggested by recent experimental evidence (Bartling *et al.*, 2014; Owens *et al.*, 2014; Lusk *et al.*, 2014; Bobadilla-Suarez *et al.*, 2017), then individuals with behavioral biases may have to trade off the gain arising from paternalistic policies yielding better outcomes against the intrinsic value of autonomy and influence on one's own life.[7] Even paternalees with known self-control problems may in this case be against paternalism. Autonomy could also matter instrumentally if individuals believe that they through trial and error learn how to make better decisions, which paternalistic policies may impede upon (Wright and Ginsburg, 2012). Additionally, some people may value influencing others as a separate goal, getting benefits from being moralizing (Petersen, 2013) and being "meddlesome" (Blau, 1975). This could lead non-targets of paternalism to favor paternalism.

Finally, behavioral self-interest allows for the possibility that an individual's choices may not be the best available decision or consistent with long-term welfare. For example, Thaler and Shefrin (1981) contrast short-term goals of individuals as "myopic doers" with long-term goals of individuals as "farsighted planners." In our case, myopic doers may consume unhealthy food for short-term benefit without trading such consumption off against long-term health consequences. For people with self-control problems, it may be in their self-interest to save for pensions, even if they do not save at present. Respondents with self-control problems could favor paternalism, as it could be consistent with their own objectives. This crucially

between subjective and objective self-interest. Subjective self-interest reflects the preferences and perceptions of the individual in question while objective self-interest reflects policy-makers' (or experts') ideas or inferences about what individuals ought to prefer based on their economic and health circumstances and the known consequences of their choices (Chong *et al.*, 2001). In this literature, people are argued to be more likely to recognize their own self-interest and act upon it when their stakes in the policy are clear or they have been primed to personal costs and benefits. However, this argument still presumes the absence of choice-biases, where a key issue is that short-term and long-term benefits and costs do not coincide.

[7] A related concept is the psychological notion of reactance, which occurs when individuals go against the desired aim of a policy because they do not want to be controlled (Brehm, 1966). Hedlin and Sunstein (2016) and (Arad and Rubinstein, 2017) have shown that reactance can be a feature of attitudes towards nudging.

depends on whether individuals are aware of their self-control problems or not – whether they are "sophisticated" or "naïve." Hersch (2005) and Gyrd-Hansen and Kjær (2015) find in the health domain that respondents with self-reported self-control problems are indeed more supportive of paternalism than respondents who do not report self-control problems. The attitudes of individuals not subject to behavioral biases will depend on actual policy design. For example, if paternalistic policies aimed at helping individuals with self-control problems also affect the consumption of rational consumers, ceteris paribus, they will be more opposed to such policies. If policies can be designed in ways that do not affect rational consumers, they will be indifferent, unless they have other motives or factors that affect their preferences over paternalistic policies.

In addition to self-interest, we consider four factors that can affect attitudes towards paternalism: altruism, political values, trust and attributions of responsibility for individual behavior and outcomes.[8] Extensive evidence has shown that individuals intrinsically care about other people's well-being (see for example Fehr and Schmidt (2006)). Some non-targets of paternalism may be altruistic and believe that the paternalees are engaging in irrational behavior, which lowers their overall welfare. In this case, non-targets may favor paternalism on altruistic grounds. This would imply, for example, that non-smokers could favor paternalistic policies to combat smoking, as they believe these policies would be in the interest of the smokers themselves. Some evidence has pointed to the relevance of this channel. Jacobsson *et al.* (2007) find that most individuals who wish to improve the health of smokers with diabetes, do so in a paternalistic manner.

Political values are important as some individuals may not have well-established preferences over the appropriateness of paternalistic policies. In order for them to answer survey questions on paternalism, it is possible that they use representative heuristics (Kahneman and Tversky, 1972). That is, they may rely on attitudes towards related policy questions to form an attitude. Their opinion towards paternalism could then be dependent on their general attitude towards government interference with the market, such as their views on redistribution. This would be consistent with recent findings for the U.S. suggesting that self-identified Republican voters in general are more opposed to nudges than Democratic voters (Sunstein, 2016). We thus expect respondents with more left-wing attitudes with respect to distributional policies to be more in favor of paternalism.[9] Individuals' general level

[8] Arguably, these factors may reflect individuals' self-interest if, for example, individuals express altruistic attitudes because it makes them feel good (Andreoni, 1990). There is no clear divide between what engages individuals' self-interest and what does not. Here, we group the concepts according to whether we think they mostly relate to self-interest or other reasons for attitude-formation. See also Oliver and Lee (2005), who consider determinants of public opinion over health policy and obesity.

[9] Interestingly, Tannenbaum *et al.* (2017) find that attitudes towards nudges are not determined by political opinion. Rather, they find that attitudes towards nudges are positive whenever the objective of the nudge is

of trust in politicians can matter for attitudes towards paternalism in many ways. Firstly, individuals may not trust that politicians are capable of knowing what is in the best interest of the paternalees (Mill, 1869). Indeed, knowing what is good for individuals becomes increasingly hard in a behavioral world, in particular, since they may not know, or realize it, themselves. If the government's knowledge turns out to be incorrect, then paternalism is a futile exercise. Secondly, individuals may consider paternalistic policies a slippery slope for granting more authority to politicians. Even if they favor specific paternalistic policies in isolation, they may express negative attitudes towards them because they do not trust politicians would use this expanded authority optimally (Camerer *et al.*, 2003; Wright and Ginsburg, 2012). We expect individuals who have a low trust towards political parties to be less supportive of paternalism. Finally, Oliver and Lee (2005) argue that individuals who hold others responsible for their choices in a particular domain are less likely to favor paternalistic policies in this domain. For example, if individuals perceive obesity as a result of lack of self-control, they are less likely to support paternalistic policies than if they think obesity is genetic. This is related to the complexity of the choice that has to be made, as individuals may be less likely to hold individuals responsible for mistakes in complicated settings. Conly (2012) argues that support for coercive paternalism is particularly strong when individuals are unlikely to choose sensibly due to the complexity of the choice or due to cognitive incompetence. Similarly, Bhargava and Loewenstein (2015) argue that paternalistic policies are more appropriate when the aim is to limit the complexity of products. For these reasons, we hypothesize that individuals who hold others responsible for their choices are more likely to be opposed to paternalism.[10]

5.3 Data and Empirical Specification

We employ a telephone survey on savings, consumption behavior and political attitudes carried out in January 2014. 6,009 people selected randomly from the Danish population above the age of 25 were interviewed. The response rates were about 50%. The data were collected by a professional survey firm and linked to administrative data at Statistics Denmark after approval with the Danish Data Protection Agency. De-identified data are analyzed at secure servers. We remove respondents with missing values in our key variables leaving us

in accordance with the respondent's own political views and vice versa. They term this effect the *partisan nudge bias*.

[10]Attributions of responsibility can also be related to the argument that the problem with paternalism is its expressive content; the fact that it suggests that individuals are incapable of making good judgments for themselves (Cornell, 2015). If some people believe that individuals ought not to be held responsible for their choices in a certain domain, the expressive critique becomes less strong, and arguably so could the opposition to paternalism.

with a total sample of 5,411 respondents.

5.3.1 Dependent Variables

We asked four questions on regulation of individual choices:

In recent years, the extent to which the state should decide over people's actions has been discussed. Do you agree or disagree that the state should legislate in order to affect, among other things:

- *How much people smoke?*
- *Whether people eat unhealthy food?*
- *Whether people save for pensions?*
- *Whether people can take out interest-only mortgages?*[11]

The answer categories were: 1. Strongly disagree, 2. Partly disagree, 3. Neither agree nor disagree, 4. Partly agree, 5. Strongly agree (and 6. Don't know, which was not read aloud). The order of the four questions was randomized.

The four questions can all represent cases of government paternalism according to a long list of paternalistic policies surveyed in Le Grand and New (2015). In line with the chapter division in Thaler and Sunstein (2008), they cover the two major domains of paternalistic policies, health and financial decisions. To be paternalistic, one of the motivations for such policies should be to address a failure of judgment by the individual for the individual's own good. Naturally, these policies may have other motivations as well, such as combating negative externalities and equity concerns. It is hard to think of a policy which only has a paternalistic motivation behind it (Dworkin, 1972). Hausman (forthcoming) suggests that a policy is paternalistic if it would have been carried out even if it had no non-paternalistic justifications.

The answers to the questions depend on the specific framing that we used. We deliberately tried to assess opinions towards coercive paternalism by using the word "decide," rather than, say, "influence" in the phrase on whether "the state should decide over people's actions." We also deliberately asked the questions at a very high and abstract level rather than referring to specific policy suggestions, such as banning smoking at bars or levying fat

[11] We also asked a question having to do with regulation of businesses, "Whether there should be a limit on borrowing interest rates." We consistently got different results with this question than the other four questions, and accordingly decided to omit it from the main analysis. We hypothesize that respondents fail to understand that the burden of limiting borrowing interest rates will also fall partly on themselves, thus considering this policy a clear improvement without any paternalistic content.

taxes. We did so since such specific policies have been heavily discussed in Denmark and we expect our respondents have quite specific attitudes towards these policies that may not map unto their attitudes towards paternalism in general. In particular, attitudes towards policies that already have been discussed or implemented may be strongly correlated with attitudes towards the political parties that advocated for these policies (Tannenbaum *et al.*, 2017).

5.3.2 Paternalee Indicators

We match each question on attitudes to paternalism with one or more variables indicating whether the respondent would be a likely target of the policies in the particular domain. For the question on smoking, we simply elicit smoking behavior with the question: *We would like to know if you (1) are a smoker; (2) are a smoker, but have tried to quit; (3) have been a smoker, but have quit; (4) never have been a smoker.* For the question on unhealthy food, we calculate respondents' BMI based on answers to questions on height and weight.

We measure pension savings behavior by the total pension contributions relative to disposable income, including both mandatory payments, as part of collective labor market agreements, and private pension savings declared as such in order to enjoy privileged tax treatment, both from administrative data. We use data from the 13 years we have information on their pension savings to smooth out potential annual fluctuations. This will serve as a proxy for the degree to which individuals forego current consumption in order to increase pensions. Mortgage type for homeowners is measured in the administrative data. We look at whether respondents have variable or fixed interest rate mortgages and whether they have interest-only or interest and repayment mortgages. In addition, we compute the marginal interest rate faced by the respondent (Kreiner *et al.*, 2012).[12]

We also use the administrative data to get sociodemographic details on the respondents; their gender, age, family status, income, education, whether they are immigrants and whether they have a mortgage.

[12]To calculate the marginal interest rate, firstly, the interest rate for every account held by a household is derived. If the household has a loan account, the marginal interest rate for an individual is the highest interest rate faced by the household. If the household only has deposit accounts, the marginal interest rate is the lowest interest rate among these accounts. Concretely, we use the highest marginal interest rate observed in the past three years. We do so in order to have more variation and to better capture whether the respondents in recent years have faced bad borrowing prospects. If we use the most recent marginal interest rate or the mean of the last three years' marginal interest rate we get qualitatively similar results but with slightly lower precision.

5.3.3 Variables for Hypothesis Testing

The survey contains specific questions that allow us to test which of the hypothesized theories that have empirical support. Most of our survey variables are gauged by standard survey questions adapted from the GSS and European Social Survey.

We measure whether attitudes are driven by classical self-interest by eliciting respondents *time and risk preferences*.[13] Time preferences are self-reported on a scale from 0-10 (10 being very patient) and risk preferences on a similar scale (10 being very risk averse).

We measure *political preferences* by a question on attitudes towards distribution on a scale from 1-5. Trust in politicians is elicited by combining answers to a question on whether the respondents trust the government and whether they trust the main opposition party. We do so to assure that the trust variable does not capture political preferences for the current government.

In the smoking domain, we define people with *self-control problems* as individuals who have tried to quit smoking but still smoke. One interpretation is that these smokers are sophisticated behavioral, i.e. cognizant of their self-control problems, but feeling unable to quit smoking. Obviously, some of these individuals may no longer wish to quit. Conversely, others may have self-control problems as well, but may not recognize it or may try to justify their actions in order to reduce cognitive dissonance. In the financial domain, we identify individuals as having self-control problems if they answer positively to a question on whether it is difficult for them to control their expenses. Again, others may have self-control problems as well but may misreport their answers or fail to acknowledge their shortcomings.

We measure *attributions of responsibility* by a question on whether individuals believe that success requires more luck than hard work. Presumably, individuals that think success requires mostly hard work generally hold individuals more responsible for their choices, also when it comes to health and financial decisions. Everything else equal, we expect that this would make them less inclined to support paternalistic interventions.

We elicit concerns about *fiscal externalities* by asking people whether they think that smokers should pay for smoking-related hospital bills. If individuals answer yes to this question, they are presumably concerned about costs to society. Since we will be controlling

[13] As noted above, according to the neoclassical economic model, self-interest is revealed in what people do. If people smoke, it is because smoking is optimal for them. Economists view choices as being largely driven by underlying preference parameters, including differences in attitudes towards risk and impatience (Frederick *et al.*, 2002). Previous evidence has shown that time preferences are a predictor of individuals' credit card borrowing, saving behavior, smoking behavior, alcohol consumption, and BMI (Chabris *et al.*, 2008; Meier and Sprenger, 2010; Sutter *et al.*, 2013). Risk preferences are less predictive but are correlated with BMI and smoking behavior among adolescents (Sutter *et al.*, 2013). In the classical economic setting, smokers and frequent eaters of unhealthy food simply have less patient time preferences or are more willing to engage in risky behavior.

for preferences for redistribution, the coefficient on the hospital bills question should not be driven by political preferences. In order for the channel from fiscal externalities to have support, individuals who think smokers should pay for smoking-related bills should be more in favor of paternalism.

We attempt to elicit whether *altruism* can drive the results by distinguishing people by their views on a very heavy-handed policy intervention: banning smoking. While radical, smoking bans have recently been implemented in Denmark outside schools, in public workplaces and, somewhat controversially, on railway stations. We ask respondents: “Do you think that a general smoking ban will benefit smokers?” This question indirectly gets at the cause of supporting paternalism. If individuals in favor of paternalism answer yes to this question, then altruism might be a motivating factor. Further, we classify individuals' as being *moralizing* (Petersen, 2013) or meddlesome (Blau, 1975; Sadrieh and Schröder, 2012) based on their answers to a question asked as part of the Big-Five inventory discussed below. The particular question we use asks respondents whether they have a tendency to focus on other people's flaws.

Finally, we address whether *freedom* and *personal autonomy* matter for attitudes by comparing attitudes to hard paternalism with attitudes towards nudging. We normalize all of the hypothesis variables to lie in the interval between zero and one to foster comparisons in later regressions. Summary statistics of all the variables are shown in Table 5A.1 in the Appendix.

5.3.4 Empirical Specification

We model attitudes towards paternalistic policies for individual i in domain d as

$$attitude_i^d = f(sociodemographics_i, behavior_i^d, hypothesis_i) \tag{5.1}$$

$hypothesis_i$ denotes the aforementioned survey questions, which allow us to test which of the theories that are capable of explaining attitudes. The dependent variables are measured using a Likert scale and, as such, could be analyzed using an ordered discrete choice model. However, for ease of interpretation, we use OLS with robust standard errors. Throughout, results are qualitatively similar in terms of sign and degree of statistical significance when applying ordered logit models (see Table 5A.2 in the Appendix). A concern, both with ordered logit models and OLS, is that results could be driven by, say, differences between those who disagree strongly and those who disagree (where most of the mass of answer distributions is located). For robustness, we constructed a binary variable equal to one if a respondent agrees or strongly agrees, and zero in case of disagreement or no strong opinion).

This, too, yields results that are qualitatively similar to the OLS results (see Table 5A.3 in the Appendix).

5.4 Main Results

In this section, we first describe the distributions of attitudes towards paternalistic policies in different domains and how individuals' attitudes covary across domains. Subsequently, we test our key hypotheses in a standard, multivariate regression framework and explore how different explanations for approval of or resistance to paternalism are supported in the data.

5.4.1 Descriptive Results

Figure 5.1 shows the distribution of answers to the key questions on paternalism. In general, respondents are skeptical of government paternalism in the form of legislation and regulation. About 60 percent disagree with the statement that the government should legislate to regulate smoking, and 65 percent disagree with legislation to influence consumption of unhealthy food. Respondents are more favorable towards government legislation in the economic domain, with about $35 - 40$ percent agreeing that the state should affect whether people save for pension or take out risky, interest-only mortgages.

Table 5.1 shows the Spearman's rank correlation matrix for the answers. Attitudes to all categories of paternalistic policies are significantly positively correlated. Attitudes to legislation towards smoking and unhealthy food are considerably stronger correlated.

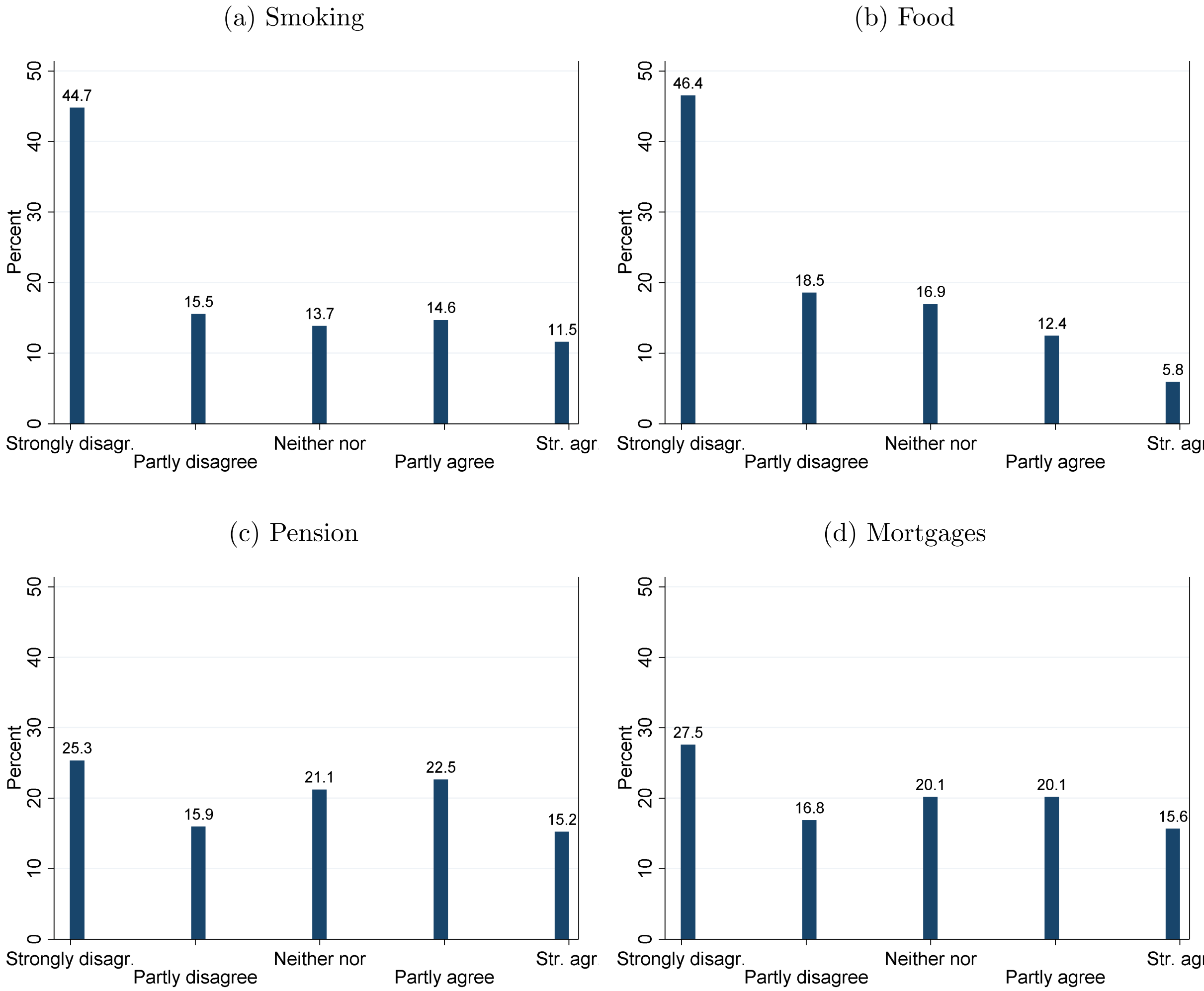
Table 5.1: Spearman's Rank Correlation of Attitudes towards Paternalism

	Smoking	Food	Pension	Mortgages
Smoking	1			
Food	0.64***	1		
Pension	0.28***	0.29***	1	
Mortgages	0.21***	0.23***	0.26***	1

Notes: * p<0.10, ** p<0.05, *** p<0.01.

The theories outlined in the previous section may help explain why we find a greater support for paternalistic policies in the economic domain than in the health domain. In recent years, Danish politicians have discussed and implemented various laws with regards to smoking and unhealthy products. Some of these policies, such as sugar and fat taxes, were so unpopular that the government had to repeal them (Bødker *et al.*, 2015). Due to individuals' likely exposure to these debates, they may be more aware of the loss of autonomy these

Figure 5.1: Attitudes Towards Paternalistic Policies



Notes: Histograms over dependent variables. $n = 5411$.

policies entail, and consequently display a stronger reactance to furthering paternalism in this domain. Paternalism in the economic realm, on the other hand, has been less discussed. This could lead individuals to be unaware of their choice-diminishing content and hence to support these policies to a greater extent.

Another possibility is that individuals possess more knowledge about food consumption decisions than they do about financial decisions. As a result, they may have more established preferences in this area and see less of a need for the government to intervene. In contrast, in the economic domain the complexity of financial products may lead people to acknowledge that they do not fully understand the financial world and accept government interference. Hence, it is possible that individuals hold others responsible for poor food and cigarette choices but not for poor financial choices.

The divergence in attitudes may also be explained by behavioral self-interest if there is a larger share of individuals with cognizant self-control problems in the economic realm than in the food and smoking realm. In the financial domain, due to large degrees of financial illiteracy and the complexity mentioned before, some individuals may acknowledge that they make suboptimal decisions and hence approve of government assistance.

5.4.2 Paternalees vs. paternalists

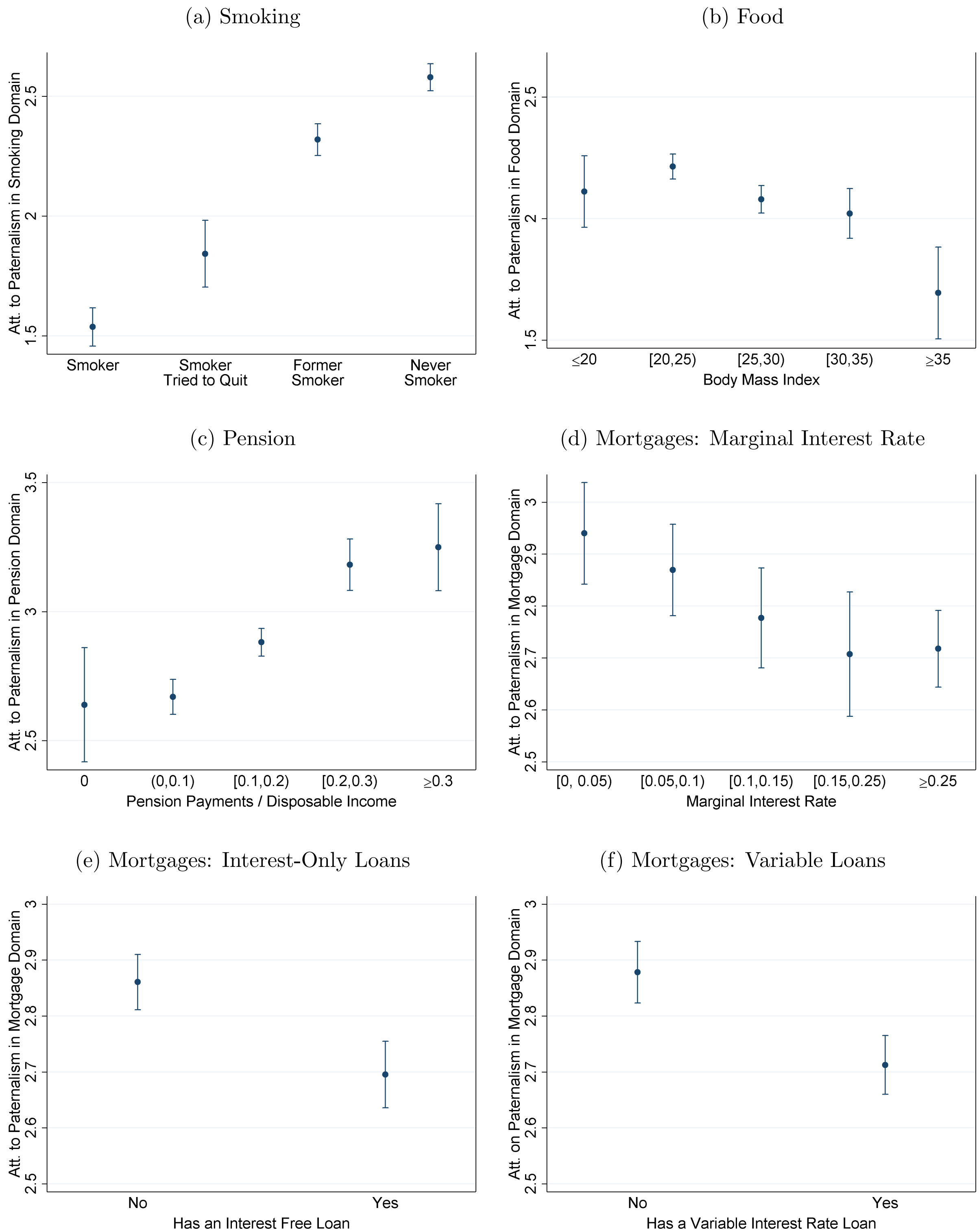
Next, we assess the bivariate relationships between paternalee indicators in each of the four policy domains and attitudes towards paternalism in these domains. Figure 5.2 shows the mean attitude to paternalism categorized by whether the respondent is a likely paternalee or not.

The figure reveals a consistent pattern. In all cases, the likely targets – and intended beneficiaries – of paternalism are significantly more opposed to paternalism than the non-targets. Smokers are much less supportive of paternalistic policies aimed at reducing smoking. Respondents with a high BMI are less supportive of paternalistic policies targeted unhealthy eating. Respondents with smaller pension payments relative to disposable income, i.e. those who sacrifice the least consumption for pension savings for a given income, are strongly against paternalism with regards to pension savings despite likely being the ones who under-save. Individuals who have the most risky and shortsighted mortgage loans in the form of variable interest rates and interest-only loans, or who face high marginal interest rates, are the least supportive of regulating the availability of interest-free loans. In all cases, Wilcoxon-Mann-Whitney tests or Kruskal-Wallis tests reject identical distributions at the 1 pct. level. This is consistent with revealed preference on behalf of potential paternalees: If they believe they are already choosing optimally, they see no need for government intervention. However, it is also consistent with choice-biased individuals being naïve, i.e., not aware of their own biases. We return to this below.

Table 5.2 shows that the links between behavior in one domain and attitudes towards paternalism in the same domain are clearly present in a multivariate OLS regression where we control for socio-demographic characteristics generally thought to be important for political attitudes.

When we look at the variables capturing whether the respondents would be likely targets of the particular paternalistic policies, the picture is the same as before. Across the board, paternalees are less likely to support paternalistic policies in their domain. Regarding the sociodemographic variables, we observe patterns that are interesting in themselves. Ceteris paribus, women are more skeptical of paternalistic legislation in the health domain. This is in contrast to widespread perceptions that women generally are more left wing and more in favor

Figure 5.2: Paternalees' Attitudes towards Paternalism



Notes: Attitudes towards paternalistic policies categorized by individuals who are likely targets of these policies and individuals who are unlikely targets of these policies. The y-axis reports the mean answer on a scale from 1-5 where 5 indicators support for paternalism while 1 indicates opposition. Bars indicate 95 pct. confidence interval. In all cases, Wilcoxon-Mann-Whitney tests or Kruskal-Wallis tests reject identical distributions at the 1 pct. level. $n = 5411$.

Table 5.2: Attitudes towards Paternalistic Policies

Independent variable	Dependent variable			
	Smoking	Food	Pension	Mortgages
Paternalee variables				
Smoker, tried to quit	0.295***			
	(0.080)			
Former smoker	0.726***			
	(0.054)			
Never smoker	0.930***			
	(0.052)			
Body Mass Index		-0.016***		
		(0.004)		
Pension payments / disposable income			1.476***	
			(0.237)	
Has an interest free loan				-0.126***
				(0.049)
Has a variable interest rate loan				-0.131**
				(0.052)
Marginal interest rate				-0.006**
				(0.002)
Sociodemographic characteristics				
Female	-0.196***	-0.151***	-0.065	-0.033
	(0.043)	(0.040)	(0.043)	(0.044)
Age	-0.002	-0.001	0.011***	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Single	-0.037	0.026	-0.028	-0.022
	(0.050)	(0.047)	(0.051)	(0.053)
Annual gross income (100,000 DKKs)	-0.000	-0.003	-0.015**	-0.024***
	(0.008)	(0.006)	(0.007)	(0.006)
Short education	0.075	0.130***	0.164***	-0.048
	(0.053)	(0.048)	(0.056)	(0.057)
Medium education	0.307***	0.291***	0.146**	0.093
	(0.068)	(0.061)	(0.068)	(0.070)
Long education	0.449***	0.423***	0.329***	0.322***
	(0.087)	(0.077)	(0.080)	(0.083)
Immigrant	0.116	0.157	0.049	-0.009
	(0.127)	(0.120)	(0.122)	(0.127)
Has a mortgage	0.016	0.021	-0.009	0.122**
	(0.047)	(0.042)	(0.047)	(0.060)
Occupation dummies	Yes	Yes	Yes	Yes
Observations	5,411	5,411	5,411	5,411
r^2	0.093	0.046	0.032	0.030

Notes: * p<0.10, ** p<0.05, *** p<0.01. OLS regressions with robust standard errors. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. The reference level of education is basic education while reference level for the smoking variable is smoker.

of redistribution (Iversen and Rosenbluth, 2006), something which is also true in the present body of data. Age does not have large effects except in the pension question, where older individuals are more in favor of paternalism towards retirement savings, presumably due to a higher salience of pension savings for older generations. Support for paternalistic policies increases strongly in education across domains. Our data shows that the highly educated are less likely to be paternalees in all dimensions but the mortgage domain, where education plays a smaller role. Neither being single nor an immigrant is predictive of attitudes.

Overall, there is little support for the proposition that those exercising potentially harmful or impatient behavior desire to be subjected to paternalistic policies. These results give rise to two questions: Why are paternalees, some of whom presumably would benefit from paternalism, against these policies? And, why are non-targets, whose welfare to a smaller extent depends on the paternalistic policies, relatively supportive of them? With regards to the former question, the results are consistent with the notion that smokers are less patient than non-smokers and with the proposition that individuals care about autonomy. With regards to the latter finding, it is consistent with non-targets being driven by concerns about fiscal externalities, moral prescriptive behavior, or altruistic paternalism.

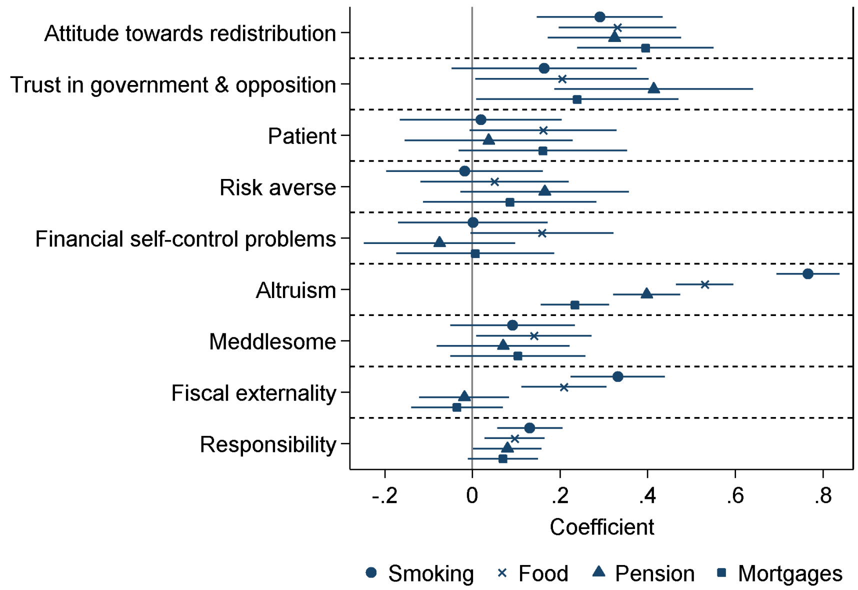
5.4.3 Beyond Narrow Self-Interest?

In this section, we examine which of the competing theories that are capable of explaining our empirical findings. We will focus on the issue of smoking due to the additional smoking-related questions that the respondents answered. We do so by adding our theory variables to the regression from before. All theory variables are standardized to be between 0 and 1 to foster comparisons of the coefficients. Regression results are plotted in Figure 5.3 and reported fully in Table 5A.4 in the Appendix.

Attitudes towards paternalistic policies are strongly associated with political preferences. People who are in favor of redistribution are also more likely to be in favor of paternalistic interventions across domains. Trust in politicians is less important, but it does seem to matter in the economic domain, where individuals with greater levels of trust are more supportive of paternalism. This could imply that some individuals find paternalistic policies welfare improving but have no trust in the government executing these policies appropriately or fear that it might lead to a slippery slope of interventionist policies. In contrast, we find no support for individual time and risk preferences explaining attitudes towards paternalism in any domain. Hence, differences in preferences do not seem able to explain why paternalees are less supportive of paternalism.

Altruistic reasons for engaging in paternalism has the most explanatory power; respondents who think banning smoking would make smokers better off are more supportive of

Figure 5.3: Hypothesis Testing



Notes: Coefficients from OLS regressions with robust standard errors. Bars indicate 95 pct. confidence interval. The regressions control for paternalee indicators, sociodemographic characteristics and occupation dummies. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. All variables for hypothesis testing are scaled to be in the interval between 0 and 1 to foster comparisons. *Attitude towards redistribution*=1 indicates support for redistribution. *Trust in government & opposition*=1 indicates high trust. *Patient/Risk averse*=1 indicates being very patient/risk averse. *Financial self-control problems*=1 for respondents who find it difficult to control their expenses. *Altruism*=1 for respondents who think smoking bans will benefit smokers. *Meddlesome*=1 for respondents who tend to find faults with others. *Fiscal externality*=1 for respondents who think smokers should pay their own hospital bills. *Responsibility*=1 for respondents who believe success requires luck.

paternalism. The effect is largest in the smoking domain, but significant and positive in the three other domains as well. At the same time, we observe small, but significant, support for attribution of responsibility: respondents who think success is a result of luck rather than hard work are more supportive of paternalism. There is, in contrast, only little support for individuals' being supportive of paternalism since they like to moralize or be meddlesome.

Respondents who think smokers should pay their own hospital bills are more in favor of paternalism. We take this to be support for the notion that fiscal externalities matter; individuals who answer yes to this question are likely guided by the societal costs of smoking. The fact that they are more supportive of paternalism suggests that they find it unfair that they should cover the expenses of people with behavioral biases. This holds also in the food domain, but not for the economic domains. One likely reason for this is that the survey instrument is clearly focused on the health domain and that external effects in the economic domain are less clear to respondents.

Individuals with self-reported financial self-control problems are not much more in favor of paternalism. This may be because only very few acknowledge that they have self-control problems. In contrast, smokers who tried to quit (not in the figure), are significantly more favorable towards paternalistic policies for smoking than other smokers. Nevertheless, respondents that tried to quit are less supportive of paternalism than non-targets. Given that the gap remains after we have controlled for the various hypotheses, this may imply that the difference cannot fully be accounted for by the theories we have tested so far. The only theory which we have not been able to test directly is that individuals value their autonomy and freedom. Since freedom is only at stake for the paternalees, it is possible that this can explain why paternalees who are cognizant of their self-control problems still are more against paternalism than non-targets. We return to this when considering nudges below.

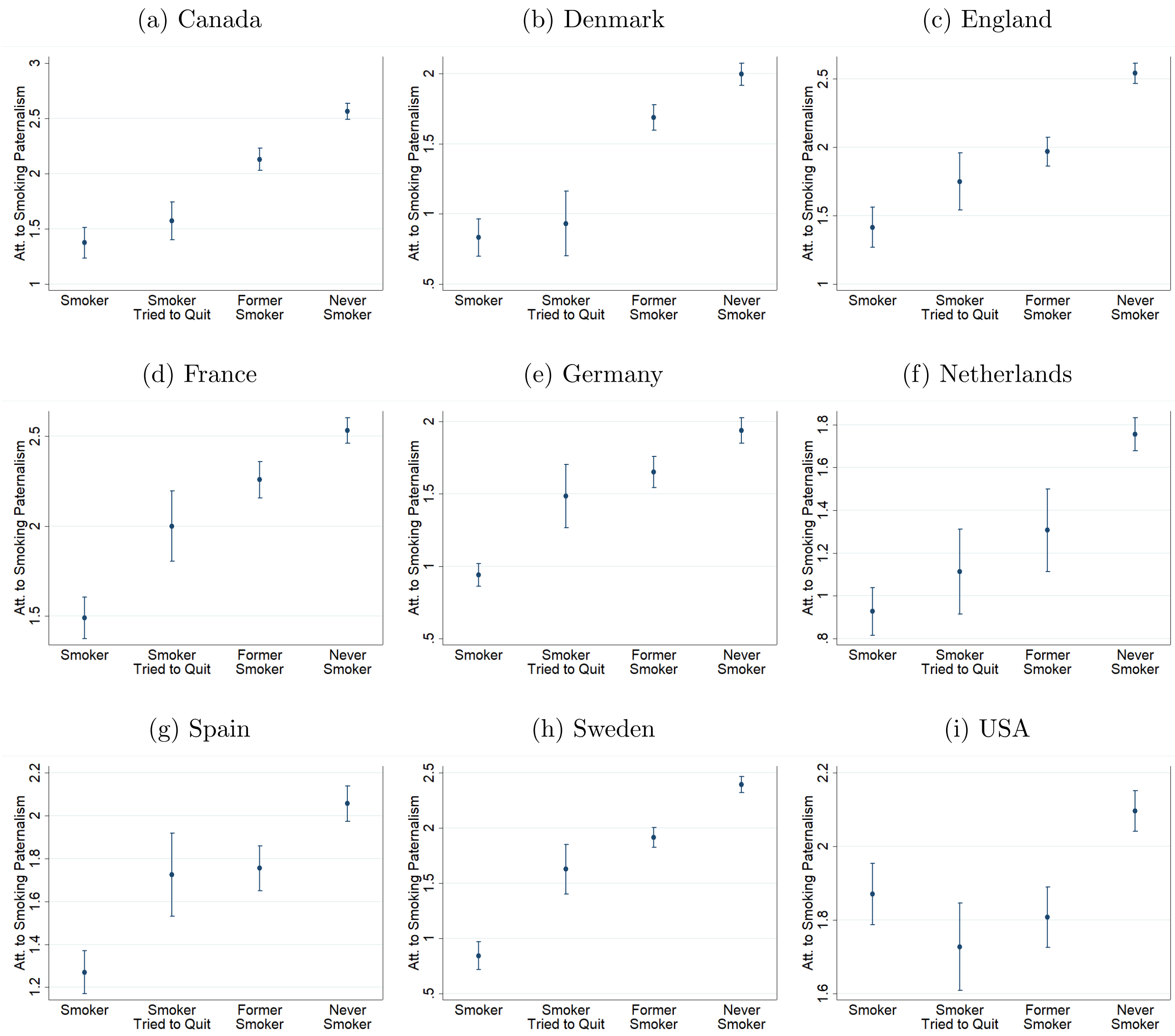
5.5 Cross-Country Results

In existing cross-country evidence, Danes have shown to be more hostile towards nudging than like-minded European countries (Reisch and Sunstein, 2016). Although it is uncertain that this also applies to coercive paternalism, in order to provide some assurance that our results are not particular to the case of Denmark, we utilize another survey we conducted in 2017 in Denmark and eight other advanced capitalist democracies: the U.S., Canada, England, Spain, France, Germany, Netherlands and Sweden. Each survey is representative of the adult population of the country in question and contains 2000 respondents (with the exception of the U.S. sample, which contains 5000 respondents). We ask the respondents our main questions on attitudes towards paternalism in the smoking domain and in the unhealthy food domain, as well as about their smoking behavior, height and weight. This allows us to recreate the figures relating attitudes towards paternalism to indicators on whether the respondent is a likely paternalee. We omit the figures of the two economic dimensions since we do not have access to administrative economic data in the eight other countries, but show in Table 5A.5 in the Appendix that survey questions capturing savings and borrowing behavior yield results qualitatively similar to those reported for the economic domains above.

Figure 5.4 reports the attitudes towards smoking paternalism broken down by the respondents' smoking status. In all nine countries but the U.S., point estimates suggest that smokers are the most opposed to paternalism in the smoking domain. They are followed by smokers that have tried to quit, former smokers, and never smokers.

A similar, but slightly less pronounced, picture emerges if we compare attitudes towards food paternalism broken down by BMI (Figure 5.5). Whenever statistical differences appear between BMI groups, heavier BMI's are associated with greater resistance to paternalism.

Figure 5.4: Cross-Country Attitudes towards Paternalism in Smoking Domain

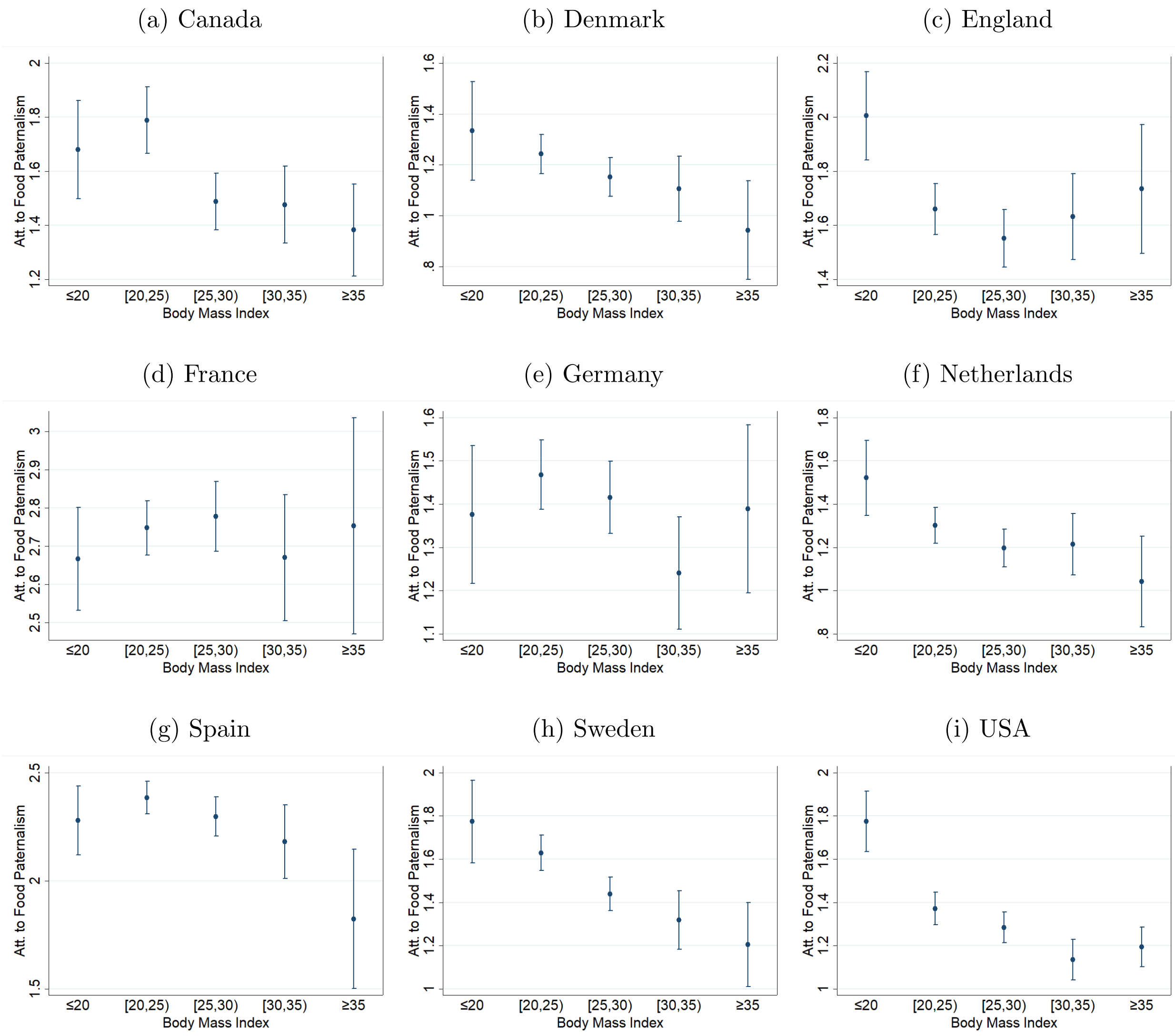


Notes: Attitudes towards paternalism in the smoking domain for nine different countries. The y-axis reports the mean answer on a range from 1-5, where 5 indicates in favor of paternalism in the smoking domain while 1 indicates being against paternalism. Bars indicate 95 pct. confidence interval.

Noticeable exceptions include France, where attitudes towards food paternalism seem unrelated to BMI.

The broad picture is the same as in our main Danish sample and remains the same if we control for socio-demographic variables (see Table 5A.5 in the Appendix). As noted above, this remains true also in the economic domain, when we proxy administrative economic data with survey measures: current pension savings are proxied by current savings measured as the respondents' assessment of the number of weeks they can get by if hit by an economic shock; administrative data on borrowing are proxied by self-reported marginal interest rates.

Figure 5.5: Cross-Country Attitudes towards Paternalism over Unhealthy Food



Notes: Attitudes towards paternalism about unhealthy food for nine different countries. The y-axis reports the mean answer on a range from 1-5, where 5 indicates in favor of paternalism in the food domain while 1 indicates being against paternalism. Bars indicate 95 pct. confidence interval.

Those with more savings support paternalism more than those with less savings; those with high interest rates support paternalism less than those facing low interest rates.

Regarding the socio-demographic characteristics, the cross-country regressions show the same pattern across all four domains as we observed in the Danish case: women are more critical of paternalism, while respondents with tertiary education are more supportive of paternalistic policies. This makes us confident that our results also replicate to most other Western countries.

5.6 Further perspectives

In this section, we examine two additional explanations for individual differences in attitudes towards paternalism: Psychological traits and the value of autonomy.

5.6.1 Psychologial traits

It is well-established that individual personality is an important predictor of political behavior and attitudes (Mondak, 2010). We extend our analysis with measures of the Big 5 personality traits. This is of interest for two reasons. Firstly, to see if this affects our core finding that being a paternalee predicts opposition to paternalism. This could be the case if Big Five measures correlate with self-control problems. Secondly, the relationship between attitudes towards paternalism and personality traits is of interest in its own right.

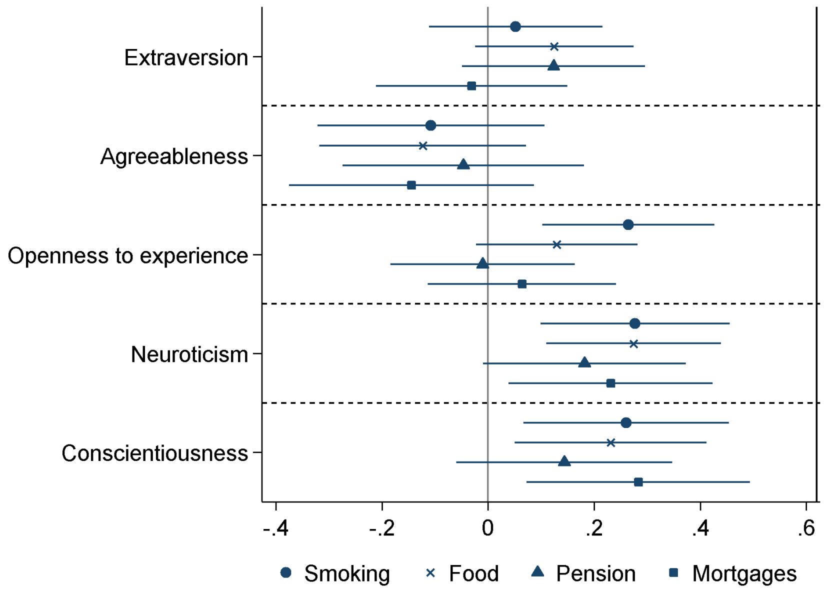
We elicit personality facets by a modified ten-item personality inventory proposed by (Gosling *et al.*, 2003) as adapted in the World and European Value Surveys. This is a brief measure of the Big Five personality domains (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) used extensively in personality research, including in research on personality and political behavior (Mondak, 2010). The Gosling *et al.* (2003) inventory is shown to satisfy standard psychometric criteria, while obviously being less precise than, say, the full 240-item Big Five measures. The questions were translated into Danish and asked in random order. No validated Danish language version exists at present.[14]

Personality traits are not driving the observed differences in attitudes between paternalees and paternalists (regression results are shown in Table 5A.6 in the Appendix). At the same time, however, personality traits have effects on attitudes beyond what is already captured by sociodemographic characteristics, paternalee indicators and our hypothesis variables. This is displayed in Figure 5.6.

We find consistent results with regard to neuroticism, with neurotic people being more in favor of paternalism. Neuroticism is sometimes characterized as the negative of emotional stability; these are individuals that characterize themselves as (reverse of) "relaxed, handles stress well" and "gets nervous easily." In our case, individuals that get nervous easily could be defined by ill-established preferences and hence more willing to accept government interventions. Interestingly, neuroticism is typically not found to be strongly associated with

[14]Respondents answer the following: I see myself as someone who (1) is reserved (E-R); (2) is generally trusting (A); (3) tends to be somewhat lazy (C-R); (4) is relaxed, handles stress well (N-R); (5) has few artistic interests (O-R); (6) is outgoing, sociable (E); (7) tends to find fault with others (A-R); (8) does a thorough job (C); (9) gets nervous easily (N); (10) has an active imagination (O) on a 1-5 scale from strongly disagree to strongly agree. The letters OCEAN refer to each of the traits, and each is gauged by a positive and a reverse, marked R, statement.

Figure 5.6: Attitude to Paternalism and Personality Traits



Notes: Relationship between attitudes towards paternalistic policies and personality traits. Based on OLS regressions with robust standard errors. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. Personality variables run from 0-1, where 1 indicates a high level of the particular trait. The regressions control for sociodemographic characteristics, paternalee indicators and hypothesis variables. Bars indicate 95 pct. confidence interval.

political behavior, including voting, and attitudes along a left-right dimension (Mondak, 2010). There is also some evidence for paternalistic support being high among individuals with high levels of conscientiousness (individuals that do a thorough job and do not consider themselves lazy). This could reflect that people who score high on conscientiousness are less likely to be subject to paternalistic policy themselves and, at the same time, more concerned about fiscal externalities.

5.6.2 The value of freedom and autonomy: Nudging vs. hard paternalism

One indirect way to assess whether the gap between the attitudes of paternalees and non-targets is driven by the value of freedom and autonomy is to look at nudges. By construction, nudges do not limit individuals' freedom and autonomy to the same degree as more coercive forms of paternalism. If the difference in attitudes between paternalees and paternalists partly is due to the intrinsic value of freedom, we should find much smaller differences when looking at nudges. The respondents were asked two questions about nudges:

In recent years, it has been discussed whether the state should try to affect people's behavior *by using psychologists' knowledge about how we make decisions.*

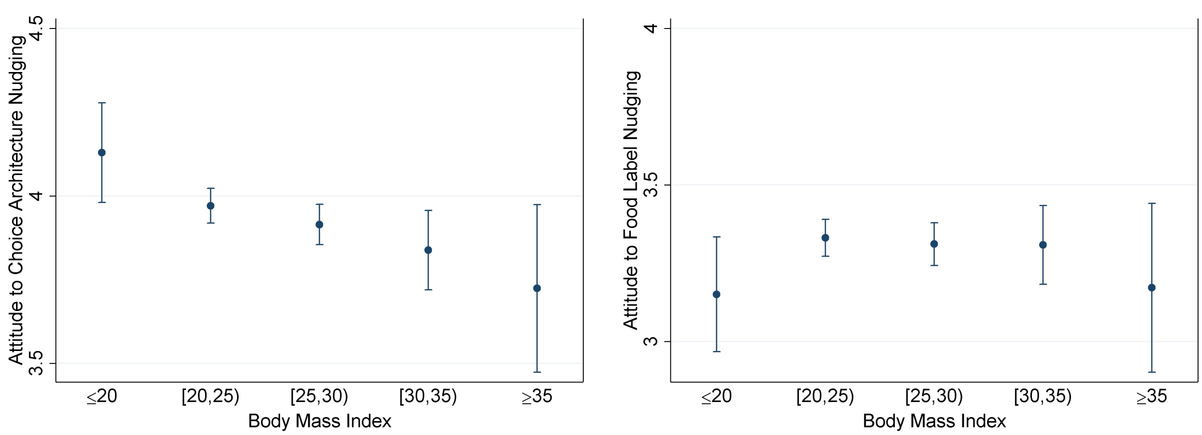
- *Do you agree or disagree that fruit and vegetables should have a more prominent position in the supermarket and that unhealthy products at the same time should be hidden a bit?*

- *Do you agree or disagree that we should try to solve the problem about unhealthy lifestyles by requiring that fat and sugar content is stated with big letters on the packaging?*

The answer categories were, like our other main dependent variables, on a scale from 1-5 with 5 indicating strongly agree. In Figure 5.7 we look at whether likely paternalees (people with a high BMI) are more negative towards these policies.

Figure 5.7: Attitudes towards Nudges

(a) Food Placement (b) Food Labels



Notes: Attitudes towards nudges categorized by individuals who are likely to be targets of these policies and individuals who are unlikely to be targets by these policies. The y-axis reports the mean answer on range from 1-5 where 5 indicates in favor of nudge while 1 indicates against paternalism. Bars indicate 95 pct. confidence interval. Wilcoxon-Mann-Whitney tests or Kruskal-Wallis tests reject identical distributions at the 2% level in the first case. In the second case, there is no evidence for dissimilar distributions (p=0.35).

The differences across groups are smaller than in the case of hard paternalism above. For the food placement question, a Kruskal-Wallis test rejects identical distributions at the 2% level, but not at any conventional level if we exclude the people with a very low BMI (5.2% of the sample). With the label question, there is no evidence for dissimilar distributions ($p = 0.35$). We test this more formally by running regressions similar to our baseline regression but with these two new dependent variables. Table 5A.7 in the Appendix displays the results. The coefficient on BMI is insignificant in the two nudge questions but significant at the 1 pct. level in our baseline regulation question. This is consistent with the idea that the value of freedom and choice is an important factor in attitudes towards coercive paternalism.

In general, we also see a different pattern in the determinants of nudges relative to paternalism. Other things equal, women, older people and respondents with little or no education are more in favor of nudges relative to coercive paternalism. These patterns also replicate in a cross-country setting. Table 5A.8 in the Appendix shows that BMI is insignificant in explaining attitudes towards the nudging questions. The cross-country results also confirm the finding from the Danish data that women are more supportive of nudges than men. By looking at the country fixed effects estimates, the finding from Reisch and Sunstein (2016), that Danes are generally more skeptical of nudges than other European countries, also carries through. In sum, while the level of support for nudging is lower in Denmark, the associations between both domain behaviors and socio-demographics are similar across countries.

5.7 Conclusion

Paternalistic policies are increasingly being used by Western governments, yet very little is known about the demand side for paternalism. In this paper, we have sought to get a better understanding of what determines whether paternalistic policies have popular support, which individuals are supportive of paternalism, and whether this can be explained by self-interest.

In particular, we studied attitudes towards coercive paternalism across four domains; smoking, unhealthy food, pension savings and interest free loans. We found that targets of paternalistic policies are less favorable towards these policies than non-targets. That is, the people these policies are constructed to help, for the most part, do not want them in place. This does not seem to be because these people have different time preferences, as traditional economic discourse would hypothesize. Rather, their opposition may reflect the value they place on freedom. The fact that non-targets generally support paternalism is primary driven by altruistic concerns; they are supportive because they believe it will benefit the lives of the people engaging in myopic behavior. The findings suggest that policies trying to overrule individuals' choices are "two-sided paternalistic" for the individuals they are constructed to help. Not only do they go against their revealed preferences in the form of their choices made, they are also highly in conflict with their attitudes towards the design of policies.

Bibliography

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, **82** (4), 463. (page 154)

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, **100** (401), 464–477. (page 158)

Arad, A. and Rubinstein, A. (2017). The people's perspective on libertarian-paternalistic policies. *Working paper*. (page 156, 157)

Bartling, B., Fehr, E., Herz, H. *et al.* (2014). The intrinsic value of decision rights. *Econometrica*, **82**, 2005–2039. (page 157)

Bhargava, S. and Loewenstein, G. (2015). Behavioral economics and public policy 102: Beyond nudging. *American Economic Review*, **105** (5), 396–401. (page 159)

Blau, J. H. (1975). Liberal values and independence. *The Review of Economic Studies*, **42** (3), 395–401. (page 157, 163)

Bobadilla-Suarez, S., Sunstein, C. R. and Sharot, T. (2017). The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *Journal of Risk and Uncertainty*, **54** (3), 187–202. (page 157)

Bødker, M., Pisinger, C., Toft, U. and Jørgensen, T. (2015). The rise and fall of the world's first fat tax. *Health policy*, **119** (6), 737–742. (page 153, 164)

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press. (page 157)

Camerer, C., Issacharoff, S., Loewenstein, G., O'donoghue, T. and Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism". *University of Pennsylvania Law Review*, **151** (3), 1211–1254. (page 154, 159)

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P. and Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty*, **37** (2-3), 237–269. (page 162)

Chong, D., Citrin, J. and Conley, P. (2001). When self-interest matters. *Political Psychology*, **22** (3), 541–570. (page 157)

Clark, G. L. and Knox-Hayes, J. (2009). The 'new' paternalism, consultation and consent: Expectations of UK participants in defined contribution and self-directed retirement savings schemes. *Pensions: An International Journal*, **14** (1), 58–74. (page 153)

Conly, S. (2012). *Against autonomy*. New York: Cambridge University Press. (page 159)

Cornell, N. (2015). A third theory of paternalism. *Michigan Law Review*, **113**. (page 159)

Dworkin, G. (1972). Paternalism. *The Monist*, pp. 64–84. (page 160)

Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism: Experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, **1**, 615–691. (page 158)

Felsen, G., Castelo, N. and Reiner, P. B. (2013). Decisional enhancement and autonomy: Public attitudes towards overt and covert nudges. *Judgment and Decision Making*, **8** (3). (page 156)

Frederick, S., Loewenstein, G. and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, **40** (2), 351–401. (page 162)

Galizzi, M. M. and Loewenstein, G. (2016). The soda tax as a measure for sustained change in consumption. *VOX*. (page 155)

Gollust, S. E., Barry, C. L. and Niederdeppe, J. (2014). Americans' opinions about policies to reduce consumption of sugar-sweetened beverages. *Preventive Medicine*, **63**, 52–57. (page 156)

Gosling, S. D., Rentfrow, P. J. and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, **37** (6), 504–528. (page 174)

Green, D. P. and Gerken, A. E. (1989). Self-interest and public opinion toward smoking restrictions and cigarette taxes. *Public opinion quarterly*, **53** (1), 1–16. (page 156)

Gyrd-Hansen, D. and Kjær, T. (2015). Government interventions to aid choice: Help to self-help or paternalism? *Health Policy*, **119** (7), 874–881. (page 156, 158)

Hagman, W., Andersson, D., Västfjäll, D. and Tinghög, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, **6** (3), 439–453. (page 156)

Hariri, J. G. and Lassen, D. D. (2017). Income and outcomessocial desirability bias distorts measurements of the relationship between income and political behavior. *Public Opinion Quarterly*, **81** (2), 564–576. (page 155)

Hausman, D. (forthcoming). Behavioral economics and paternalism. *Economics and Philosophy*. (page 160)

Hedlin, S. and Sunstein, C. R. (2016). Does active choosing promote green energy use? Experimental evidence. *Ecology Law Quarterly*, **43** (1), 107–142. (page 157)

Herrnstein, R., Prelec, D., Loewenstein, G. and Elster, J. (1992). Melioration. In G. Loewenstein and J. Elster (eds.), *Choice over time*, The Russell Sage Foundation, pp. 235–263. (page 154)

Hersch, J. (2005). Smoking restrictions as a self-control mechanism. *Journal of Risk and Uncertainty*, **31** (1), 5–21. (page 156, 158)

Iversen, T. and Rosenbluth, F. (2006). The political economy of gender: Explaining cross-national variation in the gender division of labor and the gender voting gap. *American Journal of Political Science*, **50** (1), 1–19. (page 169)

Jacobsson, F., Johannesson, M. and Borgquist, L. (2007). Is altruism paternalistic? *The Economic Journal*, **117** (520), 761–781. (page 158)

Jung, J. Y., Data, B. A. M. *et al.* (2016). American attitudes toward nudges. *Judgment and Decision Making*, **11** (1), 62. (page 156)

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan. (page 154)

— and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, **3** (3), 430–454. (page 158)

Kreiner, C. T., Lassen, D. D. and Leth-Petersen, S. (2012). Consumer responses to fiscal stimulus policy and households' cost of liquidity. *Working paper*. (page 161)

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, pp. 443–477. (page 154)

Le Grand, J. and New, B. (2015). *Government paternalism: Nanny state or helpful friend?* Princeton University Press. (page 153, 157, 160)

Loewenstein, G., Bryce, C., Hagmann, D. and Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science and Policy*, p. 35. (page 156)

Lund, T. B., Sandøe, P. and Lassen, J. (2011). Attitudes to publicly funded obesity treatment and prevention. *Obesity*, **19** (8), 1580–1585. (page 156)

Lusk, J. L., Marette, S. and Norwood, F. B. (2014). The paternalist meets his match. *Applied Economic Perspectives and Policy*, **36** (1), 61–108. (page 155, 157)

MEIER, S. and SPRENGER, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, **2** (1), 193–210. (page 162)

MILL, J. S. (1869). *On liberty*. Longmans, Green, Reader, and Dyer. (page 153, 159)

MONDAK, J. J. (2010). *Personality and the foundations of political behavior*. Cambridge University Press. (page 174, 175)

O'DONOGHUE, T. and RABIN, M. (2006). Optimal sin taxes. *Journal of Public Economics*, **90** (10), 1825–1849. (page 154)

OLIVER, J. E. and LEE, T. (2005). Public opinion and the politics of obesity in America. *Journal of Health Politics, Policy and Law*, **30** (5), 923–954. (page 156, 158, 159)

OWENS, D., GROSSMAN, Z. and FACKLER, R. (2014). The control premium: A preference for payoff autonomy. *American Economic Journal: Microeconomics*, **6** (4), 138–161. (page 157)

PEDERSEN, S. K., KOCH, A. K. and NAFZIGER, J. (2014). Who wants paternalism? *Bulletin of Economic Research*, **66** (S1). (page 156)

PETERSEN, M. B. (2013). Moralization as protection against exploitation: Do individuals without allies moralize more? *Evolution and Human Behavior*, **34** (2), 78–85. (page 157, 163)

RABIN, M. (2013). An approach to incorporating psychology into economics. *American Economic Review*, **103** (3), 617–22. (page 154)

REISCH, L. A. and SUNSTEIN, C. R. (2016). Do Europeans like nudges? *Judgment and Decision Making*, **11** (4), 310–325. (page 155, 156, 171, 177)

SADRIEH, A. and SCHRÖDER, M. (2012). The desire to influence others. *Working paper*. (page 163)

SAINT-PAUL, G. (2011). *The tyranny of utility: Behavioral social science and the rise of paternalism*. Princeton University Press. (page 157)

SCHMIDT, A. T. (2017). The power to nudge. *American Political Science Review*, **111** (2), 404–417. (page 154)

SIMON, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason*. MIT press. (page 154)

SUNSTEIN, C. R. (2016). Do people like nudges? *Administrative Law Review*. (page 156, 158)

SUTTER, M., KOCHER, M. G., GLÄTZLE-RÜTZLER, D. and TRAUTMANN, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, **103** (1), 510–531. (page 162)

Tannenbaum, D., Fox, C. R. and Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour*, **1** (7), 0130. (page 156, 158, 161)

Thaler, R. H. and Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of political Economy*, **112** (S1), S164–S187. (page 154)

— and Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy*, pp. 392–406. (page 154, 157)

— and Sunstein, C. R. (2003). Libertarian paternalism. *The American Economic Review*, **93** (2), 175–179. (page 154)

— and — (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press. (page 154, 160)

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, **185** (4157), 1124–1131. (page 154)

Weiss, D. M. (1991). Paternalistic pension policy: Psychological evidence and economic theory. *The University of Chicago Law Review*, **58** (4), 1275–1319. (page 153)

Wright, J. D. and Ginsburg, D. H. (2012). Behavioral law and economics: Its origins, fatal flaws, and implications for liberty. *Northwestern University Law Review*, **106** (3), 12–63. (page 157, 159)

5.A Appendix

Table 5A.1: Summary Statistics

	mean	sd	min	max
Dependent variables				
Smoking	2.33	1.45	1	5
Food	2.13	1.28	1	5
Pension	2.86	1.41	1	5
Mortgages	2.80	1.43	1	5
Nudge: Food placement	3.94	1.35	1	5
Nudge: Food label	3.31	1.52	1	5
Sociodemographic variables				
Female	0.48	0.50	0	1
Age	52	11.3	20	72
Single	0.20	0.40	0	1
Annual gross income (100,000 DKK's)	4.20	2.93	-2.81	62.5
Basic education	0.17	0.38	0	1
Short education	0.45	0.50	0	1
Medium education	0.25	0.44	0	1
Long education	0.12	0.33	0	1
Immigrant	0.03	0.16	0	1
Has mortgage	0.75	0.44	0	1
Paternalee variables				
Smoker	0.12	0.32	0	1
Smoker, tried to quit	0.06	0.23	0	1
Former smoker	0.34	0.48	0	1
Never smoker	0.48	0.50	0	1
BMI	25.5	4.18	15	64
Pension payments / disposable income	0.13	0.09	0	0.5
Marginal interest rate	14.9	6.03	0.06	25.2
Has interest-only loan	0.40	0.49	0	1
Has variable interest loan	0.50	0.50	0	1
Variables for hypothesis testing				
Risk aversity (1 = most risk averse)	0.52	0.21	0	1
Patience (1 = most patient)	0.67	0.21	0	1
Attitude towards redistribution (1 = in favor of redistribution)	0.56	0.26	0	1
Trust in government and opposition (1 = highest trust)	0.41	0.18	0	1
Smoking ban makes smokers better off (*altruism*)	0.56	0.50	0	1
Smokers should pay own hospital bills (*fiscal externalities*)	0.17	0.37	0	1
Sucess requires mostly luck (*responsbility*)	0.60	0.49	0	1
Find it difficult to control own expenses (*financial self-control*)	0.05	0.21	0	1
Tends to find faults with others (*meddlesome*)	0.29	0.26	0	1

Notes: Summary statistics of all variables used in the main analysis. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. $n = 5,411$. Based on privacy requirements on the use of Danish register data, the *min* and *max* columns reflext the average value of the four observations with respectively highest and lowest values.

Table 5A.2: Main Regressions - Ordered Logit

Independent variable	Dependent variable			
	Smoking	Food	Pension	Mortgages
Paternalee variables				
Smoker, tried to quit	0.569***			
	(0.140)			
Former smoker	1.160***			
	(0.099)			
Never smoker	1.425***			
	(0.097)			
Body Mass Index		-0.025***		
		(0.006)		
Pension payments / disposable income			1.921***	
			(0.314)	
Has an interest free loan				-0.161***
				(0.061)
Has a variable interest rate loan				-0.163**
				(0.065)
Marginal interest rate				-0.007**
				(0.003)
Sociodemographic characteristics				
Female	-0.249***	-0.234***	-0.084	-0.045
	(0.059)	(0.060)	(0.056)	(0.056)
Age	-0.003	-0.003	0.014***	0.004
	(0.003)	(0.003)	(0.003)	(0.002)
Single	-0.068	0.000	-0.032	-0.041
	(0.071)	(0.071)	(0.067)	(0.068)
Annual gross income (100,000 DKKs)	0.001	-0.004	-0.023*	-0.036***
	(0.011)	(0.010)	(0.012)	(0.011)
Short education	0.158**	0.256***	0.211***	-0.055
	(0.079)	(0.077)	(0.074)	(0.074)
Medium education	0.471***	0.474***	0.190**	0.126
	(0.095)	(0.093)	(0.089)	(0.090)
Long education	0.612***	0.635***	0.424***	0.416***
	(0.116)	(0.112)	(0.101)	(0.106)
Immigrant	0.131	0.192	0.070	-0.007
	(0.175)	(0.182)	(0.166)	(0.165)
Has a mortgage	0.012	0.030	-0.007	0.158**
	(0.064)	(0.063)	(0.062)	(0.078)
Occupation dummies	Yes	Yes	Yes	Yes
Observations	5,411	5,411	5,411	5,411

Notes: * p<0.10, ** p<0.05, *** p<0.01. Ordered logit regressions with robust standard errors. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. The reference level of education is basic education while reference level for the smoking variable is smoker.

Table 5A.3: Main Regressions - Binary Logit

Independent variable	Dependent variable			
	Smoking	Food	Pension	Mortgages
Paternalee variables				
Smoker, tried to quit	0.552**			
	(0.222)			
Former smoker	1.286***			
	(0.154)			
Never smoker	1.568***			
	(0.152)			
Body Mass Index		-0.026***		
		(0.009)		
Pension payments / disposable income			1.611***	
			(0.348)	
Has an interest free loan				-0.221***
				(0.074)
Has a variable interest rate loan				-0.134*
				(0.077)
Marginal interest rate				-0.008**
				(0.004)
Sociodemographic characteristics				
Female	-0.364***	-0.265***	-0.167***	-0.130**
	(0.072)	(0.084)	(0.064)	(0.065)
Age	-0.003	0.001	0.018***	0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
Single	0.030	0.178*	-0.037	-0.072
	(0.087)	(0.093)	(0.076)	(0.077)
Annual gross income (100,000 DKKs)	-0.002	-0.008	-0.016	-0.032**
	(0.011)	(0.015)	(0.012)	(0.014)
Short education	0.054	0.161	0.128	-0.113
	(0.100)	(0.116)	(0.084)	(0.084)
Medium education	0.374***	0.463***	0.153	0.084
	(0.116)	(0.134)	(0.103)	(0.102)
Long education	0.521***	0.640***	0.442***	0.364***
	(0.136)	(0.155)	(0.122)	(0.123)
Immigrant	0.274	0.452**	0.107	0.018
	(0.196)	(0.202)	(0.174)	(0.177)
Has a mortgage	0.009	0.027	-0.056	0.136
	(0.081)	(0.089)	(0.071)	(0.086)
Occupation dummies	Yes	Yes	Yes	Yes
Observations	5,410	5,397	5,410	5,406

Notes: * p<0.10, ** p<0.05, *** p<0.01. Logit regressions with robust standard errors. The dependent variables equal 1 if the respondent agrees or strongly agrees with the policy, and 0 otherwise. The reference level of education is basic education while reference level for the smoking variable is smoker.

Table 5A.4: Beyond Self-Interest?

Independent variable	Dependent variable			
	Smoking	Food	Pension	Mortgages
Variables for hypothesis testing				
Attitude towards redistribution	0.291***	0.332***	0.324***	0.395***
	(0.073)	(0.068)	(0.078)	(0.080)
Trust in government & opposition	0.164	0.205**	0.414***	0.240**
	(0.108)	(0.101)	(0.116)	(0.118)
Patient	0.019	0.162*	0.037	0.161
	(0.094)	(0.085)	(0.098)	(0.098)
Risk averse	-0.018	0.051	0.165*	0.086
	(0.091)	(0.086)	(0.098)	(0.101)
Financial self-control problems	0.001	0.159*	-0.075	0.007
	(0.087)	(0.083)	(0.088)	(0.092)
Altruism	0.766***	0.530***	0.398***	0.234***
	(0.037)	(0.034)	(0.039)	(0.040)
Meddlesome	0.092	0.141**	0.070	0.104
	(0.073)	(0.067)	(0.077)	(0.079)
Fiscal externality	0.332***	0.209***	-0.019	-0.035
	(0.055)	(0.050)	(0.052)	(0.054)
Responsibility	0.131***	0.097***	0.080**	0.070*
	(0.038)	(0.035)	(0.040)	(0.041)
Paternalee variables				
Smoker, tried to quit	0.186**			
	(0.078)			
Former smoker	0.522***			
	(0.053)			
Never smoker	0.662***			
	(0.053)			
Body Mass Index		-0.013***		
		(0.004)		
Pension payments / disposable income			1.369***	
			(0.234)	
Marginal interest rate				-0.005**
				(0.002)
Has an interest free loan				-0.114**
				(0.049)
Has a variable interest rate loan				-0.126**
				(0.052)
Observations	5,403	5,403	5,403	5,403
r^2	0.175	0.104	0.059	0.044

Notes: * p<0.10, ** p<0.05, *** p<0.01. OLS regressions with robust standard errors. All regressions contain sociodemographic controls and occupation dummies. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. All variables for hypothesis testing are scaled to be between 0 and 1 to foster comparisons. *Attitude towards redistribution*=1 indicates in favor for redistribution. *Financial self-control problems*=1 for respondents who find it difficult to control their expenses. *Altruism*=1 for respondents who think smoking bans will benefit smokers. *Meddlesome*=1 for respondents who tend to find faults with others. *Fiscal externality*=1 for respondents who think smokers should pay their own hospital bills. *Responsibility*=1 for respondents who believe success requires mostly luck.

Table 5A.5: Cross-country Attitudes towards Paternalism

Independent variable	Dependent variable			
	Smoking	Food	Pension	Limit
Paternalee variables				
Smoker, tried to quit	0.210***			
	(0.038)			
Former smoker	0.545***			
	(0.027)			
Never smoker	0.849***			
	(0.024)			
Body Mass Index		-0.011***		
		(0.002)		
Weeks to financial trouble			0.002***	
			(0.001)	
Interest rate				-0.004*
				(0.002)
Sociodemographic characteristics				
Female	-0.197***	-0.132***	-0.170***	-0.067***
	(0.019)	(0.018)	(0.019)	(0.020)
Age	-0.006***	-0.005***	-0.004***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Single	0.071***	0.049**	0.074***	0.034
	(0.020)	(0.019)	(0.020)	(0.021)
Log net income (USD)	-0.025***	-0.013	-0.016**	-0.003
	(0.007)	(0.008)	(0.007)	(0.007)
High school	0.037	-0.002	0.068	0.063
	(0.052)	(0.051)	(0.056)	(0.064)
Vocational	0.076	0.019	0.128**	0.104*
	(0.051)	(0.049)	(0.055)	(0.063)
College	0.238***	0.110**	0.245***	0.120*
	(0.052)	(0.050)	(0.055)	(0.063)
Post-graduate	0.435***	0.230***	0.409***	0.217***
	(0.055)	(0.054)	(0.058)	(0.066)
Mortgage	0.030	0.022	0.062	0.008
	(0.054)	(0.058)	(0.054)	(0.058)
Country dummies				
Canada	0.212***	0.242***	0.265***	0.227***
	(0.035)	(0.039)	(0.034)	(0.035)
England	0.152***	0.333***	0.370***	0.544***
	(0.037)	(0.038)	(0.035)	(0.037)
Spain	-0.177***	0.926***	0.403***	0.020
	(0.035)	(0.036)	(0.036)	(0.037)
France	0.269***	1.390***	0.513***	0.448***
	(0.034)	(0.034)	(0.033)	(0.038)
Denmark	-0.148***	-0.086**	0.285***	-0.041
	(0.038)	(0.034)	(0.039)	(0.042)
Germany	-0.269***	0.074**	0.102***	0.172***
	(0.035)	(0.034)	(0.036)	(0.040)
Netherlands	-0.434***	-0.091**	0.188***	0.550***
	(0.040)	(0.037)	(0.041)	(0.042)
Sweden	0.077**	0.179***	-0.152***	0.401***
	(0.035)	(0.034)	(0.034)	(0.037)
Observations	17.706	15.682	15.311	13.919
r^2	0.126	0.192	0.060	0.046

Notes: * p<0.10, ** p<0.05, *** p<0.01. OLS regressions with robust standard errors. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. The reference level of education is basic education, the reference level for the smoking variable is smoker, and the reference country is USA. *Limit* elicits whether respondents agree or disagree that there should be a limit on borrowing interest rates.

Table 5A.6: Accounting for Personality Traits

Independent variable	Dependent variable			
	Smoking	Food	Pension	Mortgages
Psychological traits				
Extraversion	0.052	0.125	0.124	-0.031
	(0.083)	(0.076)	(0.088)	(0.092)
Agreeableness	-0.108	-0.123	-0.047	-0.145
	(0.109)	(0.100)	(0.116)	(0.118)
Openness to experience	0.264***	0.130*	-0.010	0.064
	(0.083)	(0.078)	(0.089)	(0.090)
Neuroticism	0.277***	0.275***	0.182*	0.231**
	(0.091)	(0.084)	(0.098)	(0.098)
Conscientiousness	0.261***	0.231**	0.144	0.283***
	(0.099)	(0.092)	(0.104)	(0.107)
Paternalee variables				
Smoker, tried to quit	0.166**			
	(0.078)			
Former smoker	0.516***			
	(0.054)			
Never smoker	0.654***			
	(0.054)			
Body Mass Index		-0.012***		
		(0.004)		
Pension payments / disposable income			1.374***	
			(0.234)	
Marginal interest rate				-0.005**
				(0.002)
Has an interest free loan				-0.110**
				(0.049)
Has a variable interest rate loan				-0.125**
				(0.052)
Occupation dummies	Yes	Yes	Yes	Yes
Sociodemographic controls	Yes	Yes	Yes	Yes
Hypothesis variables	Yes	Yes	Yes	Yes
Observations	5,381	5,381	5,381	5,381
r^2	0.179	0.107	0.060	0.046

Notes: * p<0.10, ** p<0.05, *** p<0.01. OLS regressions with robust standard errors. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. The reference level for the smoking variable is smoker. Personality variables run from 0-1, where 1 indicates a high level of the particular trait.

Table 5A.7: Attitudes towards Nudges/Paternalism on Unhealthy Food

Independent variable	Dependent variable		
	Nudge: Placement	Nudge: Labels	Regulation
Paternalee variable			
Body Mass Index	-0.007	-0.001	-0.013***
	(0.005)	(0.005)	(0.004)
Sociodemographic characteristics			
Female	0.385***	0.206***	-0.120***
	(0.042)	(0.047)	(0.039)
Age	0.003*	0.008***	-0.003
	(0.002)	(0.002)	(0.002)
Single	0.021	-0.043	0.023
	(0.047)	(0.055)	(0.045)
Annual gross income (100,000 DKKs)	0.003	-0.013*	-0.000
	(0.007)	(0.008)	(0.006)
Short education	0.110**	-0.004	0.102**
	(0.054)	(0.060)	(0.046)
Medium education	0.039	-0.011	0.260***
	(0.064)	(0.071)	(0.059)
Long education	-0.111	-0.223***	0.396***
	(0.078)	(0.086)	(0.075)
Immigrant	-0.081	-0.005	0.189*
	(0.122)	(0.130)	(0.113)
Has a mortgage	0.031	0.079	0.013
	(0.044)	(0.050)	(0.041)
Variables for hypothesis testing			
Attitude towards redistribution	0.432***	0.568***	0.329***
	(0.072)	(0.083)	(0.068)
Trust in government & opposition	-0.224**	-0.185	0.203**
	(0.106)	(0.122)	(0.101)
Patient	0.492***	0.137	0.142*
	(0.093)	(0.103)	(0.084)
Risk averse	-0.019	0.087	0.050
	(0.091)	(0.103)	(0.086)
Financial self-control problems	-0.014	0.112	0.163*
	(0.096)	(0.099)	(0.083)
Altruism	0.400***	0.449***	0.534***
	(0.037)	(0.042)	(0.034)
Fiscal externality	0.270***	0.376***	0.210***
	(0.047)	(0.054)	(0.049)
Responsibility	-0.042	-0.011	-0.094***
	(0.038)	(0.043)	(0.035)
Observations	5,411	5,411	5,411
r^2	0.083	0.070	0.102

Notes: * p<0.10, ** p<0.05, *** p<0.01. OLS regressions with robust standard errors. All regressions contain occupation dummies. The dependent variables range from 1 to 5 with 5 indicating great support for paternalism. *Nudge: Placement* indicates attitudes towards placing healthy products at a prominent place in supermarkets, and vice versa for unhealthy food. *Nudge: Labels* indicates attitudes towards requiring fat and sugar content stated on food products. *Regulation* is the baseline question on attitudes towards regulation of unhealthy eating habits.

Table 5A.8: Cross-country Attitudes towards Nudges

Independent variable	Dependent variable		
	Nudge: Placement	Nudge: Labels	Regulation
Paternalee variable			
Body Mass Index	-0.002	-0.003*	-0.011***
	(0.001)	(0.001)	(0.002)
Sociodemographic characteristics			
Female	0.154***	0.118***	-0.132***
	(0.018)	(0.016)	(0.018)
Age	-0.001*	0.004***	-0.005***
	(0.001)	(0.001)	(0.001)
Single	0.058***	0.031*	0.049**
	(0.019)	(0.017)	(0.019)
Log net income (USD)	-0.011	0.002	-0.013
	(0.008)	(0.007)	(0.008)
High school	0.036	-0.072	-0.002
	(0.051)	(0.046)	(0.051)
Vocational	0.079	-0.012	0.019
	(0.049)	(0.045)	(0.049)
College	0.093*	-0.015	0.110**
	(0.050)	(0.046)	(0.050)
Post-graduate	0.167***	0.014	0.230***
	(0.054)	(0.049)	(0.054)
Mortgage	0.035	0.025	0.022
	(0.055)	(0.051)	(0.058)
Country dummies			
Canada	0.266***	0.228***	0.242***
	(0.038)	(0.034)	(0.039)
England	0.327***	0.378***	0.333***
	(0.036)	(0.032)	(0.038)
Spain	0.480***	0.623***	0.926***
	(0.032)	(0.028)	(0.036)
France	0.487***	0.513***	1.390***
	(0.033)	(0.029)	(0.034)
Denmark	0.068*	-0.305***	-0.086**
	(0.036)	(0.035)	(0.034)
Germany	0.120***	0.288***	0.074**
	(0.035)	(0.032)	(0.034)
Netherlands	0.113***	0.209***	-0.091**
	(0.036)	(0.034)	(0.037)
Sweden	0.178***	-0.017	0.179***
	(0.033)	(0.032)	(0.034)
Observations	15.869	15.880	15.682
r^2	0.035	0.075	0.192

Notes: * p<0.10, ** p<0.05, *** p<0.01. OLS regressions with robust std. errors. The dependent variables range from 1 to 5 with 5 indicating support for paternalism. The ref. level of education is basic education, the ref. level for the smoking variable is smoker, and the ref. country is USA. *Nudge: Placement* indicates attitudes to placing unhealthy food at a non-prominent place in supermarkets. *Nudge: Labels* indicates attitudes to requiring fat and sugar content stated on food products. *Regulation* is the baseline question on attitudes to regulating unhealthy food.

Chapter 6

Do Altruistic Preferences Matter for Voting Outcomes?

Do Altruistic Preferences Matter for Voting Outcomes?[1]

Daniel Gerszon Mahler
University of Copenhagen

Abstract

Extensive evidence has shown that some people vote for altruistic reasons while others vote for selfish reasons. This paper analyzes how, if at all, altruistic preferences matter for voting outcomes. To this end, a Danish survey is conducted ($n = 2000$) where respondents are asked to identify (1) the party they would vote for if elections were held tomorrow, (2) the party they would vote for if they only were to consider what is best for themselves, and (3) the party they would vote for if they were to consider what is best for society as a whole. Differences in where individuals cast their altruistic, selfish, and actual votes are analyzed by locating the Danish political parties in a political compass. Altruistic preferences are found to drive votes to the left and away from extreme candidates. A smaller U.S. survey on the 2016 presidential candidates ($n = 400$) yields similar results. The results suggest that political candidates may be able to increase their vote share by capitalizing on the duality of voting behavior and influencing whether voters vote selfishly or altruistically.

Keywords: Altruism, elections, preferences, voting
JEL: A13, D64, D71, D72

[1] I am thankful for comments received from participants at Harvard's Political Economy Workshop, Harvard's Working Group in Political Psychology and Behavior, University of Copenhagen's Public Economics and Political Economy Seminar, the Social Choice and Welfare Conference in Lund, the 74th Annual Midwest Political Science Association Conference in Chicago, and the 9th Annual Experimental Political Science Conference at New York University. In addition, I am grateful for comments and advice from Ryan Enos, Mikkel Høst Gandil, Chase Harrison, and David Dreyer Lassen. The usual disclaimer applies. This project has received support from the European Research Council (Starting Grant).

"Once upon a time two boys found a cake. One of them said, *'Splendid! I will eat the cake.'* The other one said, *'No, that is not fair! We found the cake together, and we should share and share alike, half for you and half for me.'* The first boy said *'No, I should have the whole cake!'* Along came an adult who said, *'Gentlemen, you shouldn't fight about this: you should compromise. Give him three quarters of the cake.'*"

(Smullyan, 1980, p. 56)

6.1 Introduction

Some of the most important economists and political philosophers over the past century have argued that individuals are able to make judgments based on what they think is best for themselves, and based on what they think is best for society as a whole (Arrow, 1951; Harsanyi, 1955; Sen, 1977; Dworkin, 1978; Elster, 2006). This dichotomy of judgments has been the subject of intense study in the context of voting. Do individuals vote based on what they think is best for themselves or based on what they think is best for society as a whole? Downs' canonical model of economic voting assumes that when confronted with different political platforms, *"a rational man always takes the one which yields him the highest utility ceteris paribus; i.e. he acts to his own greatest benefit"* (Downs, 1957, p. 36). However, recent literature has suggested that some individuals are motivated by altruistic or other ethical concerns when voting (Feddersen and Sandroni, 2006; Feddersen *et al.*, 2009; Messer *et al.*, 2010). Indeed, one of the possible explanations for why we see such high turnouts despite the negligible chance of a single vote to be decisive, is that people vote for altruistic reasons (Fowler, 2006; Edlin *et al.*, 2007; Jankowski, 2007).

Prior studies have primarily analyzed *whether* individuals vote altruistically or selfishly. This paper takes as a premise that some people vote altruistically and others selfishly, and investigates if this duality of voting behavior matters for voting outcomes. If there are no systematic differences between individuals' selfish and altruistic preferences, then altruistic voting need not matter for election outcomes. To analyze whether altruistic preferences matter, a Danish survey is designed where a representative sample of more than 2000 Danish voters are asked to identify:

- the party they would vote for if elections were held tomorrow
- the party they would vote for if they only were to consider what is best for themselves
- the party they would vote for if they were to consider what is best for society as a whole

By comparing the answers to these three questions, it is possible to analyze how, if at all, voting outcomes are impacted by altruistic preferences (here proxied by answers to the last question). In order to analyze this in detail, the Danish political parties are located in a political compass such that differences in *where* the respondents place their actual, altruistic, and selfish votes can be scrutinized. This compass is constructed by scraping answers from two extensive online candidate tests that more than 90% of all Danish MPs answered.

It is not possible to infer whether individuals vote altruistically or selfishly based on the three questions. This, however, is not necessary to determine if altruistic preferences matter for the outcome. To illustrate this, consider a voter that casts his ballot for the party he thinks is best for society but not for the party he thinks is best for himself. Although such behavior is consistent with altruistic voting, since his motivations are not inferred, it cannot be concluded that he votes altruistically. He might vote for a third reason, which happens to result in a vote for the candidate he also thinks is best for society as a whole. It can be inferred, however, that he does not vote selfishly. This divergence between his actual and selfish vote choice can be used to analyze what *would have happened* if he had voted selfishly. At the aggregate level, similar observations can be used to analyze what the election outcome *would have looked like* if more people had voted selfishly or altruistically. Although such an outcome is a hypothetical construct, it may be possible for candidates and parties to influence the election in direction of this hypothetical outcome.

Of the 2000 respondents, 64% vote for the party they think is best for society *and* best for themselves. It is unclear whether these respondents vote altruistically, selfishly, or for a third reason. Since these respondents provide no variation, they will not drive any of the results. This does not matter for the purpose of this study, as altruistic voting behavior of these individuals by construction will not influence the outcome. Of the 2000 respondents, 15% do not vote for the party they think is best for society, while 29% do not vote the party they think is best for themselves. Hence, there is according to the survey at most 85% altruistic voters and at most 71% selfish voters.

When facing the three questions, individuals may want to appear consistent by aligning their vote choices more than what their true preferences reflect. If this is the case, then any differences between the three vote distributions should be considered a lower bound. Individuals may also be subject to a social desirability bias, implying that they would like to be seen as altruistic rather than selfish. This could lead them to (1) deliberately select a party in the selfish question that is far away from their actual vote and (2) falsely claim that their actual vote is the same as what they think is best for society as a whole. This occurrence would be strongest when the respondents answer the third question, as the purpose of the study may have become clearer at this point. Two points argue against the presence of such a social desirability bias. Firstly, the question order was partly randomized and there were no significant order effects. Secondly, in 12 cognitive interviews carried out to assess the quality of the questions, this behavior was not detected. Indeed, the purpose of the survey was not apparent to the respondents.

The paper finds that if more people had voted selfishly, the outcome would have been more right-winged and more votes would have fallen on ideologically extreme candidates. If more people had voted altruistically, the outcome would have been more left-winged and more centered, but the differences here are quite small. This suggests that most people already vote altruistically. A smaller survey on the 2016 U.S. presidential candidates ($n = 400$) generates very similar results.

The findings suggest that if policy makers design policies based on the preferences of the electorate, they need to decide on whether to use individuals' altruistic or selfish preferences as inputs. If individuals' altruistic preferences are used, then optimal policies will have a more left-wing agenda. The results also suggest that political candidates may be able to increase their vote share by capitalizing on the duality of voting behavior and attempting to change whether voters vote selfishly or altruistically.

6.2 Definitions & Related Literature

Different streams of literature are related to this study. Before reviewing the literature, a concise definition of altruistic voting is needed. As a starting point, consider Jencks (1990, p. 53) who defines individuals as altruistic "*when they feel and act as if the long-term welfare of others is important independent of its effects on their own welfare.*" Individuals could thus be said to *vote* altruistically if they vote based on the welfare of others independent of what is best for themselves. Importantly, in the context of this study a stronger version of altruism is needed, which one could label societal altruism. When voting, individuals may consider the welfare of certain others of society such as their region, ethnicity, or socio-economic class etc. Altruistic voting behavior in this study requires individuals to incorporate the entire society as others. Hence, *individuals will be said to vote altruistically if they vote based on the welfare of society as a whole independent of what is best for themselves.* It will not be claimed that any individuals *are* altruistic, only that they *vote* altruistically. Whereas altruistic behavior often contains some element of self-sacrifice, this dimension is hardly present in altruistic voting.

Notice that this definition is much stronger than the definitions of sociotropic voting or social preferences. Sociotropic voting is defined as *"taking some account [...] of the collective's interest"* when voting (Meehl, 1977, p. 14). Social preferences loosely imply that individuals' put weight on other people's payoff when making decisions (Charness and Rabin, 2002). It cannot be inferred whether individuals with social preferences or sociotropic behavior act for the sake of their own benefit or whether they have a genuine concern for other people's welfare. Hence, neither need be altruistic under the definition used here.

A first relevant stream of literature is precisely the literature on sociotropic voting. Following Kinder and Kiewiet (1979), a substantial literature has investigated whether individuals' vote choices best can be explained by evaluations of personal finances or by evaluations of the national economy. This literature generally finds that individuals primarily are guided by evaluations of the national economy and less so by their own pocketbook or other selfish concerns (see e.g. Sears *et al.*, 1979, 1980; Markus, 1988; Sears and Funk, 1990; Lewin, 1991). This is consistent with some individuals possessing altruistic preferences and invoking these in the voting booth. However, as argued in the previous paragraph, people may be concerned with the national economy and still be selfish if they think that the best way to advance their own interest is to create a healthier national economy. Therefore, although this kind of voting is consistent with altruistic voting behavior, it need not reflect altruism (Kiewiet and Lewis-Beck, 2011; Kinder and Kiewiet, 1981).

Other studies impute individuals' likely material self-interest and check if this aligns with their reported attitudes. These studies generally find that individuals' attitudes do not align much with their material interests (Fisher, 1985; Sudit, 1988; Funk, 2000; Chong *et al.*, 2001). Other studies try to detect if some individuals vote altruistically by eliciting different kinds of preferences. This is close to the approach taken in the present paper. Hudson and Jones (1994, 2002) ask their respondents (i) how they would vote with respect to a specific government policy, (ii) how they would vote if the goal was their self-interest, and (iii) how they would vote if considering the public interest. They find that public concerns are a stronger determinant of voting behavior than selfish concerns, suggesting a substantial presence of altruistic voting.

A final relevant stream of literature illustrates that altruistic behavior matters when evaluating outcomes. One example is cost-benefit analyses and particularly contingent valuation studies. When individuals are asked to state their willingness to pay for a public good, some people may take on their "public hat" while others may take on their "private hat" (Nyborg, 2000; McConnell, 1997). That is, some people might invoke their altruistic preferences and answer what the public good contributes to the average person or to the entire community. As the story in the epigraph illustrates, if one aggregates such altruistic preferences with selfish preferences, the final evaluation mixes two different concepts and has little relevance. Altruism also matters for policy evaluations based on revealed preferences. Consider two individuals that have the same selfish preferences over the consumption of a set of goods that includes a public good. Suppose further that the individuals also are capable of ranking their own consumption bundles according to what they think is best for society as a whole. One of the individuals may invoke such altruistic preferences when doing purchases (Lusk *et al.*, 2007). This individual would then consume more of the public good than is in his own interest. Often this choice behavior is used to recover preferences for each individual, which are used to derive the welfare of each individual. In this example, one would falsely infer that the individual motivated by altruism gets a high welfare boost from purchasing the public good even though his selfless behavior in fact could harm his level of well-being.[2]

In the context of elections, only a few papers have tried to determine if altruistic preferences matter for evaluating outcomes. Feddersen *et al.* (2009) show that if some people have ethical expressive preferences (which is similar to the notion of altruism adopted here), then the voting outcome will be morally biased in large elections. Morally biased means that morally superior alternatives will get more votes. Feddersen *et al.* (2009) do not attempt to look at the political direction of this bias. The studies cited earlier by Hudson and Jones (1994, 2002) do look at the direction of the bias in terms of government spending. They find that preferences based on the public interest on average are more supportive of increasing public spending than preferences based on self-interest. This paper will differ from these studies by looking at how altruistic preferences matter for voting outcomes at large considering all relevant political dimensions.

6.3 Hypotheses

The equilibrium effects of altruistic voting behavior is a sparsely studied subject. I will distinguish between three ways in which altruistic preferences may matter and based on previous literature generate hypotheses to be tested by the data.

The first matter of concern is whether the actual outcome is more in line with the hypothetical outcome where all vote altruistically or the hypothetical outcome where all vote selfishly (henceforth the altruistic outcome and the selfish outcome). Notice that the actual outcome does not have to be related to the selfish or altruistic outcome at all. The story in the epigraph about the two boys and the cake reflects this possibility. If both boys had been selfish, they would demand the cake for themselves and a compromise would be to split it evenly. If both boys had been altruists, supposedly, they would both want to split the cake evenly and a settlement would be to do so. Once one is selfish and the other altruistic, the

[2] Some individuals may be purchasing altruistically because of the warm glow effect of doing so (Andreoni, 1990), in which case it indeed could be beneficial to their own well-being. As argued, this paper looks at a more genuine notion of altruism which requires that acts are carried out independently of considerations about personal welfare.

outcome reflects neither the aggregation of selfish preferences nor the aggregation of altruistic preferences. For this reason, Elster (1997, p. 14) argues that it can be "a dangerous thing" if some but not all vote according to what they believe constitutes the common good. If some vote selfishly and others not, then the preferences of the selfish voters will be counted more than once, which may generate a socially inferior outcome.

The first aim of the empirical analysis will be to investigate whether such a pattern exists. The literature summarized in the previous section suggests that more people vote altruistically than selfishly. Therefore, I expect the actual outcome to lie closer to the altruistic outcome than the selfish outcome. This would imply that if all people were to vote altruistically, the effect would be relatively small as most people already vote altruistically. If, on the other hand, all people were to vote selfishly, the effect could be quite substantial. Based on this I generate the following hypothesis:

Hypothesis 1: The actual outcome will be a combination of the altruistic outcome and the selfish outcome but more aligned with the altruistic outcome.

Secondly, one can look at whether the outcome is more left-winged or right-winged if more people vote altruistically or selfishly, respectively. In order to derive a hypothesis in this regard one could set up a simple model. A natural starting point would be modeling one political axis where voters derive utility from their income alone, and parties or candidates propose different income distributions. Individuals could be said to vote selfishly if they maximize their own utility, and altruistically if they maximize the utility of the entire society. There would be several shortcomings to this approach. First of all, there is more than one dimension in politics. Individuals with a low income may vote for the right wing for selfish reasons because cultural, religious, and social values outweigh economic concerns (Frank, 2007). Secondly, the set-up could change substantially if the model was dynamic rather than static, as individuals may be willing to tolerate inequality if they believe this is necessary to foster growth and future well-being. Thirdly, there is no reason to believe that the altruistic judgments would be utilitarian. Individuals may vote on what they think is best for society based on other concepts of justice or fairness. Individuals may also vote altruistically based on, for example, environmental concerns or other non-person specific outcomes. Finally, individuals may vote altruistically by voting for candidates based on valence judgments, such as their character traits and ability to govern (Stokes, 1963). In sum, it is very hard to make a credible hypothesis about the altruistic outcome through a simple model.

I will instead rely on previous literature. Zettler *et al.* (2011) show that altruistic people are more inclined to support left-wing agendas due to a concern for social equality. Hudson and Jones (1994, 2002) find that preferences for the public interest on average are more supportive of increasing public spending than votes based on self-interest. Norton and Ariely (2011) and Kiatpongsan and Norton (2014) report that people find the ideal distribution of respectively wealth and wages more equal than the actual distribution. Based on these findings I generate a second hypothesis:

Hypothesis 2: The altruistic outcome will be more left-winged that the actual outcome, and the selfish outcome will be more right-winged.

Finally, after locating the votes in a political dimension, it may be interesting to see if the variance of the vote distributions are impacted by altruistic voting. A lower variance can be interpreted as a greater degree of political agreement and – holding the mean location of the votes fixed – a smaller support for extreme candidates. Little previous literature is available for guidance in this perspective and it is therefore not possible to create a credible hypothesis.

6.4 Context & Data

6.4.1 Context

In order to analyze the impact of altruistic preferences on voting outcomes, a representative sample of 2000 Danish voters are surveyed. The survey was conducted four months following the 2015 Danish general election. Denmark has a multi-party parliamentary democracy with nine political parties represented in the parliament. The Danish political system has a relatively small degree of polarization. In the past decades, shifting center-right and center-left governments have taken office. The 2015 election saw the center-left government headed by the first female Prime Minister, Helle Thorning Schmidt, from the Social Democrats be replaced by a right-wing government. The new government was led by Lars Løkke Rasmussen from the Liberals with support from the Danish People's Party, the Conservatives, and Liberal Alliance. The election outcome was notable in that it resulted in a minority government containing only the Liberals even though another right-wing party, the Danish People's Party, obtained a higher vote share. The nine political parties and their vote shares in the election are shown in Table 6.1.

Danish politics is focused on economic policies (employment, taxation etc.) and value policies (immigration, environment etc.). The nine political parties offer combinations of views in both of these dimensions. The relatively low degree of polarization in Danish politics means that the voter can choose between parties that are not too different from each

Table 6.1: Vote Shares in the 2015 Danish General Elections

Left-wing parties	Right-wing parties
A: Social Democrats (26.3%)	O: Danish People's Party (21.1%)
Ø: Red-Green Alliance (7.8%)	V: Liberals (19.5%)
Å: The Alternative (4.8%)	I: Liberal Alliance (7.5%)
B: Social Liberal Party (4.6%)	C: Conservative People's Party (3.4%)
F: Socialist People's Party (4.2%)	

other. This, in turn, makes it less likely that voters who have different selfish and altruistic judgments still believe that the same party is most capable of advancing both of these concerns. This makes the Danish political system a good unit of analysis. To check whether the Danish findings can generalize to other Western democracies, a smaller survey on the 2016 U.S. presidential candidates is also conducted. The U.S. democracy in many ways represents a most different case to the Danish one.

6.4.2 Survey Design

The survey was conducted by the survey agency, Epinion, which delivers the opinion polls to the Danish Broadcasting Corporation, DR. The respondents were a representative sample of the Danish electorate based on gender, region, and age. Survey weights have been used throughout the analysis to assure full representativeness over these variables. The respondents were asked the following three central questions:

- Who would you vote for if general elections were held tomorrow?
- Who would you vote for if you only were to consider what is best for yourself?
- Who would you vote for if you were to consider what is best for society as a whole?

I refer to the answers to these questions as the "actual votes," the "selfish votes," and the "altruistic votes." Naturally, the answers only reflect hypothetical voting intentions and not votes that were carried out. As such, it may be appropriate to add "intended" or "hypothetical" prior to the three labels. For simplicity, I will just use "actual," "selfish," and "altruistic."

The framing "society as a whole" was chosen since it contains less political bias than using phrases such as "the country" or "Denmark." Besides these three questions, the questionnaire contained about 25 questions on voting behavior, political preferences, political knowledge, altruism, and demographics. The questions were to the extent possible adapted from the

2015 Danish Election Survey.[3] The full questionnaire is available in Appendix 6.A.1. The questionnaire was tested using best practices in cognitive interviewing (Willis, 2004). In total 12 cognitive interviews were conducted. The cognitive interviewing resulted in minor rephrasing of questions but revealed no noteworthy problems in terms of understanding the three essential questions. Only individuals that do not regularly vote in elections were unable to answer the questions on selfish and altruistic voting. When individuals had to answer what is best for themselves, they mostly thought about their own economic or occupational situation. When answering what was best for society, they referred to ideology and ethical values. None spoke of strategic concerns or misplacing their views so as to appear either selfish or altruistic. Indeed, it was not clear to the interviewees that the point of the questionnaire was to analyze altruistic voting behavior.

Of the 2000 respondents, 1600 successfully answered all three central questions. 96% of the respondents who knew who to vote for if elections were held tomorrow were also able to answer either the altruistic or the selfish question, while 85% were able to answer both. This suggests that the respondents were able to comprehend and answer the questions.

When answering the three central questions, it is possible that the respondents aligned their answers in order to appear consistent. This would diminish the differences between the various vote distributions and imply that if any differences are found, these should be considered a lower bound. It is also possible that the respondents falsely placed their selfish choice far from their actual choice to appear non-selfish, or that they falsely aligned their altruistic choice with their actual choice to appear altruistic. This would overestimate the difference between the selfish votes and the actual votes, and underestimate the difference between the altruistic votes and the actual votes. Although none of the cognitive interviews pointed to this being a threat, to make this kind of misrepresentation less likely, the respondents were randomly allocated into two groups. One group was asked the "best for yourself" question before the "best for society as a whole" question, while the other had the order reversed. It was not possible for the respondents to go back and change their previous answer. The respondents' incentives to misplace their true stances become more salient when they already have answered one or two of the three questions. Initially, the intention was only to use the answers to the first question asked. However, a chi-squared test for whether the selfish (altruistic) party choices are independent of the question order results in a p-value of 0.17 (0.27). Relatedly, a chi-squared test for whether the likelihood of voting non-selfishly (non-altruistically) is independent of the question order results in a p-value of 0.85 (0.29). Consequently, the question ordering, and thereby the deliberate misplacement of selfish votes,

[3] One question on whether voting is a duty or a choice was adapted from the 2010 American National Election Study. Another question on distributive preferences was taken from the TV2 candidate test (http://politik.tv2.dk/valg2015/kandidattest).

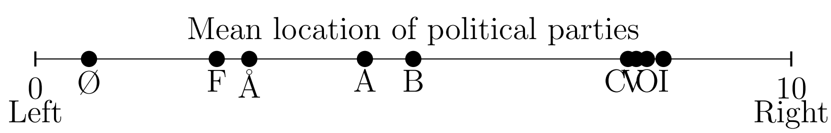
does not seem to matter and the whole sample will be used when making inferences.

6.4.3 Placing the Political Parties in Political Axes

The answers to the three central questions make it possible to detect differences in vote distributions by political parties and political blocs. In order to gain a deeper understanding of where the votes move, the political parties will also be located in political axes. This will be done in two ways:

1. One Dimension: The respondents are asked to locate each political party on a scale from 0-10, where 0 means extreme left and 10 means extreme right. Using the average of these locations, the different political parties can be placed in a one-dimensional space. In order to ensure that the assessments are somewhat reliable, only the respondents who display sufficient political knowledge (by correctly reporting the number of members of parliament and the parties partaking in the government) are used. The resulting placements are given in Figure 6.1.

Figure 6.1: Placement of Political Parties



Note: Average location of the various political parties on a scale from 0 (left) to 10 (right) using the mean assessment of respondents who displayed a certain level of political knowledge. Ø: Red-Green Alliance, F: Socialist People's Party, Å: The Alternative, A: Social Democrats, B: Social Liberal Party, C: Conservative People's Party, V: Liberals, O: Danish People's Party, I: Liberal Alliance.

The four right-wing parties, C, V, O, and I cluster around the same point despite being quite different politically. For example, the Danish People's Party (O) is normally considered left of center on economic issues but far right-wing when it comes to attitudes towards immigrants. Liberal Alliance (I), on the other hand, is the most right-wing party economically, but more moderate when it comes to immigration policies. The method fails to capture this due to the reduction of dimensionality. When individuals are forced to locate parties on a one-dimensional scale they might not share the same interpretation of the scale. The second method deals with this problem.

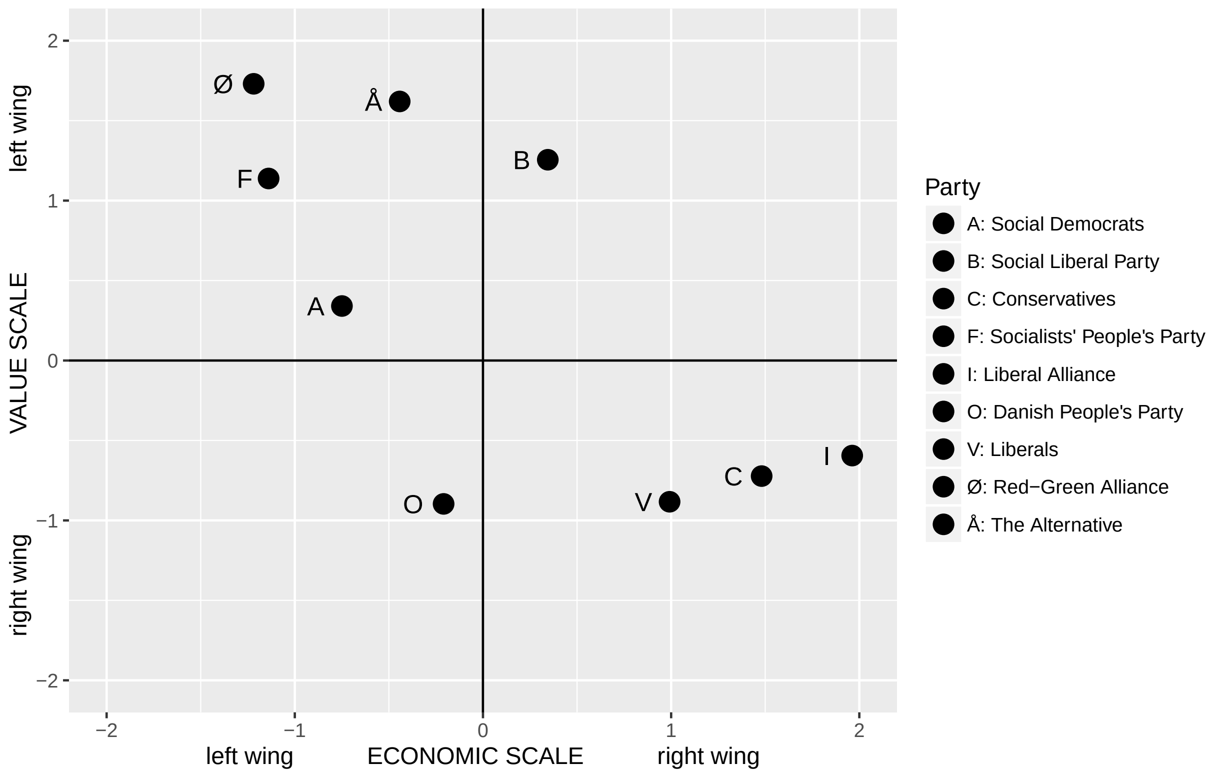
2. Multiple Dimensions: To capture the multiple dimensions present in Danish politics, I utilize online candidate tests conducted by the two main Danish television stations,

DR and TV2, prior to the 2015 elections. These tests asked all politicians running for parliament to display if they agree or disagree with respectively 15 and 42 statements on various political issues such as "more jobs should be created in the public sector." The statements had five possible answer categories ranging from completely disagree to completely agree. Remarkably, 161 of the 175 elected candidates to parliament answered all of these questions. Scraping their answers makes it possible to conduct factor analysis to determine the number of relevant dimensions in Danish politics. In general, there appears to be three dimensions; the economic dimension, the value dimension (containing issues such as immigration, the environment, and crime), and EU politics. Political discourse in Denmark has it that the EU is not an important determinant of voting behavior, so only the first two dimensions will be dealt with in this analysis. Every politician is given a factor score in each of the two dimensions and party averages are computed such that the parties can be located in a political compass. The resulting compass is shown in Figure 6.2. A detailed derivation of the compass is given in Appendix 6.A.2. This approach easily captures the difference between the Danish People's Party (O) and Liberal Alliance (I). It should be noted, though, that the questions may have been chosen to deliberately highlight political differences that may not be very salient or important for the voters. The final compass deals with this issue by weighting the loadings with the importance of the statements as assessed in a separate poll.

After the political positions have been calculated using these two methods, the statistical analysis amounts to comparing i) the share of votes by party, ii) the mean location of the votes, and iii) the variances of the votes. With regards to comparing the share of votes by party, McNemar's test will be used. When comparing the mean location of the votes, paired t-tests will be used. For comparisons of variances, Levene's test will be used to account for the high degree of non-normality.

Even though the median voter likely will determine the outcome, I prefer to look at the mean and variance of the votes. The party the median voter favors will at times not change even though there is a great shift of votes. Conversely, it may change even though there is only a slight change of votes (Höchtl *et al.*, 2012). With only one sample to analyze, this binary measure will not be very informative. The size of the change in the mean location, on the other hand, will be suggestive of how often one can expect the median vote to change.

Figure 6.2: Political Compass



Note: Political compass of the nine political parties in the Danish parliament in two political dimensions, the economic scale and the value scale. The figure is constructed from an exploratory factor analysis based on Danish politicians' answers to DR and TV2's candidate test (see Appendix 6.A.2 for details). The scales have been standardized such that a value of 1 means one standard deviation from the mean position of the Danish MP's.

6.5 Results

6.5.1 Descriptive Statistics

Before analyzing whether altruistic voting matters, it is useful to establish how big an overlap there is between the altruistic votes, selfish votes, and actual votes. Table 6.2 breaks down the degree of overlap between the answers to the three central questions.

Table 6.2: Overlap Between Altruistic, Selfish, and Actual Votes

		Actual vote = selfish vote		
		No	Yes	Sum
Actual vote = altruistic vote	No	8.0%	6.9%	14.8%
	Yes	21.5%	63.7%	85.2%
	Sum	29.4%	70.6%	100%

Note: Breakdown of overlap between the answers given to the actual voting question, the selfish question, and the altruistic question. $n = 1600$.

63.7% of the respondents selected the same party as their actual choice, altruistic choice, and selfish choice. This is hardly surprising as many individuals may convince themselves that what is best for themselves is also best for society at large (Edlin *et al.*, 2007). 29.4% voted for a party they did not believe was best for themselves and 14.8% voted for a party they did not believe was best for society.

It is not possible to infer if individuals voted selfishly or altruistically for two reasons. Firstly, it cannot be inferred whether the 63.7% voted for altruistic reasons, selfish reasons, or for a third reason. Secondly, suppose a voter votes for party A, selfishly prefers party B, and altruistically prefers party A. It is not possible to infer that this person voted altruistically, since the person may have voted for party A for reasons not related to altruism. However, it can be inferred that the person did not vote selfishly. Hence, it is possible to obtain upper bounds on the share that voted selfishly and altruistically. At most 70.6% voted selfishly and at most 85.2% voted altruistically. 8.0% voted neither selfishly nor altruistically. Of these 8.0%, 39% voted for a party that they spatially placed in between their altruistic choice and selfish choice on a left-right wing scale.

We can next try to characterize the voters who are likely to have voted altruistically (here proxied by the 21.5% that vote for what they think is best for society and not for what they think is best for themselves) rather than selfishly (the corresponding 6.9%). The first columns of Table 6.3 use a binary logit regression to characterize these individuals. Respondents that are interested in politics are predicted to be more likely to vote altruistically. This may be because people interested in politics are more inclined to vote for ideological reasons which need not overlap with their personal interests. Two proxy variables for altruistic behavior (whether the person has donated blood or donated to charity) and a variable about preferences for redistribution come out insignificant. The reason for the many insignificant variables may be that strictly speaking only non-selfish versus non-altruistic voting is detected, which blurs the picture.

Table 6.3: Who Are the Altruistic and Consistent Voters?

	Altruistic Voters		Consistent Voters	
	Coef.	Std. err.	Coef.	Std. err.
Female	-0.39	(0.35)	0.15	(0.16)
Age	0.00	(0.01)	0.02***	(0.01)
Education level (8-point scale)	0.05	(0.11)	-0.09**	(0.04)
Personal income (8-point scale)	0.20	(0.13)	-0.10	(0.05)
Unemployed (yes=1/no=0)	0.76	(0.77)	0.24	(0.42)
Employee in the public sector (yes/no)	-0.51	(0.37)	0.04	(0.19)
Interested in politics (4-point scale)	0.50**	(0.24)	-0.39***	(0.11)
Supporter of a political party (yes/no)	-0.17	(0.32)	0.63***	(0.16)
Political standing (0=left to 10=right)	0.45	(0.25)	-0.43***	(0.11)
Political standing^2	-0.04	(0.02)	0.05***	(0.01)
In doubt of who to vote for (yes/no)	1.27	(1.15)	-0.24	(0.40)
Political knowledge (yes/no)	-0.08	(0.44)	0.27	(0.18)
Donated blood (yes/no)	-0.12	(0.33)	0.23	(0.15)
Donated money to charity (yes/no)	-0.26	(0.35)	0.06	(0.17)
Pref. for redistribution	-0.09	(0.18)	0.04	(0.08)
Voting is a duty (yes/no)	0.22	(0.33)	-0.23	(0.16)
Would vote strategically (yes/no)	-0.46	(0.32)	-0.33**	(0.16)
Observations	287		963	

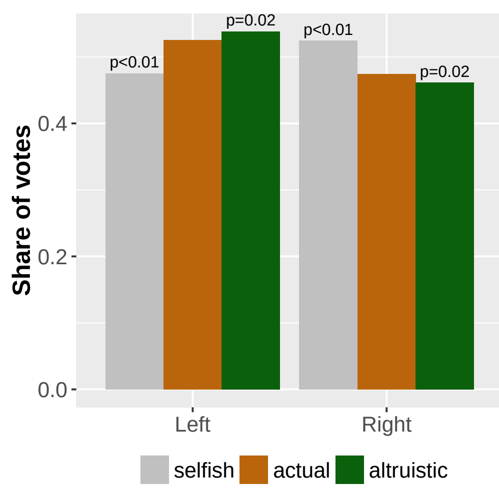
Note: ** $p<0.05$, *** $p<0.01$. Characterization of altruistic and consistent voters based on binary logit regressions. *Altruistic Voters* is a binary variable equaling 1 if the respondents vote for a party they think is best for society but not best for themselves and 0 if they vote for a party they think is best for themselves but not best for society. *Consistent Voters* is a binary variable equaling 1 if the respondent chooses the same party for the actual, altruistic, and selfish question and 0 otherwise. The indepenent variable *Political knowledge* equals 1 if the respondent knows the number of seats in parliament and the parties in government and 0 otherwise.

One can also see if there are differences between the ones whose votes overlap (the 63.7% that vote for the same in all three questions) and the ones whose votes do not overlap (the remaining 36.3%). This is shown in the right part of Table 6.3. Older people are more likely to report the same answer to all three questions. The same applies to supporters of a political party. The latter could be because individuals become interested in politics for ideological reasons, but once they belong to a political party, the success of this party becomes their self-interest. Individuals with political knowledge are less likely to give overlapping answers. This may be because they are more able to understand the nuances between the parties and hence select different ones for the different questions. Respondents that place themselves in the middle of a left-right scale are the least likely to give overlapping answers. People that place themselves in the extreme ends are more likely to be consistent. Perhaps this is because individuals who are at the extreme ends of the political spectrum can gain the most from politics and are most likely to convince themselves that their own gains are also to the benefit of society as a whole. Individuals that would consider voting strategically give less overlapping vote choices. One could fear that this will drive the differences between the altruistic, selfish, and actual vote choices. A robustness check will show that this is not the case.

6.5.2 Main Results

The distribution of votes by political bloc is shown in Figure 6.3. The right wing receives 52.5% of the selfish votes, 47.4% of the actual votes, and 46.1% of the altruistic votes. This suggests that altruistic voting generates a more left-winged outcome in line with hypothesis 2.

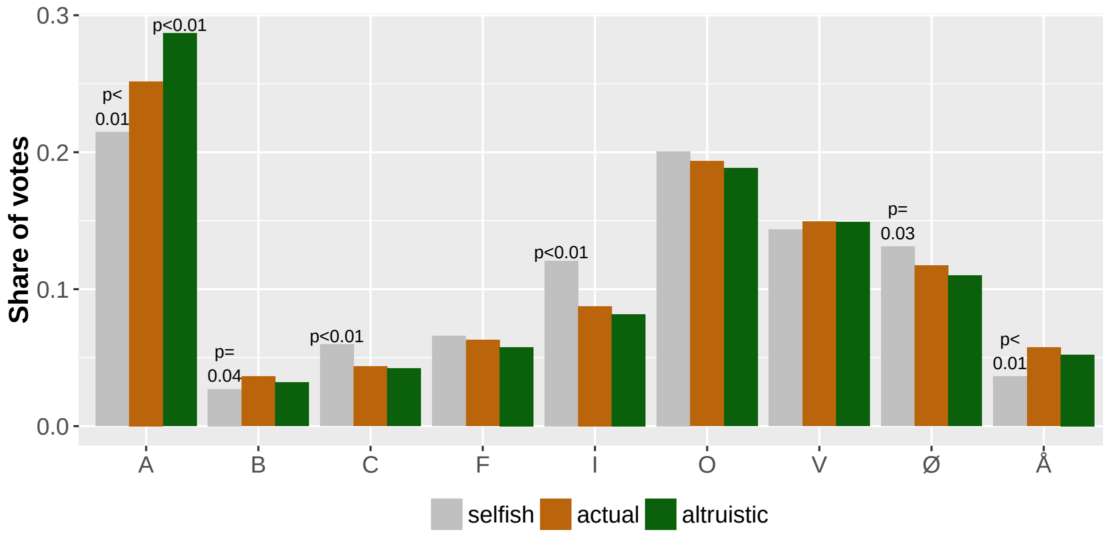
Figure 6.3: Distribution of Votes by Bloc



Note: Distribution of selfish, actual, and altruistic votes by political bloc. The p-values indicate difference from the share of actual votes using McNemar's test. $n = 1600$.

This can be broken down by political party as shown in Figure 6.4. The altruistic votes appear to be much more aligned with the actual votes than the selfish votes. This is in line with hypothesis 1. Only for the Social Democrats (A) is there a significant difference between the actual votes and the altruistic votes. This does not imply that there are no movements between the altruistic and actual votes for the other parties, but only that many of these movements balance out. The Social Democrats (A) is the major party in the left-wing bloc, which has governed Denmark in the majority of the past century. Conservatives (C), Liberal Alliance (I), and the Red-Green Alliance (Ø) receive more selfish votes than actual votes. The first two of these have the most right-wing economic policies while the latter has the

Figure 6.4: Distribution of Votes by Party



Note: Distribution of selfish, actual, and altruistic votes by political party. P-values indicate difference from the share of actual votes using McNemar's test. $n = 1600$. A: Social Democrats, B: Social Liberal Party, C: Conservative People's Party, F: Socialist People's Party, I: Liberal Alliance, O: Danish People's Party, V: Liberals, Ø: Red-Green Alliance, Å: The Alternative.

most left-wing economic policies.

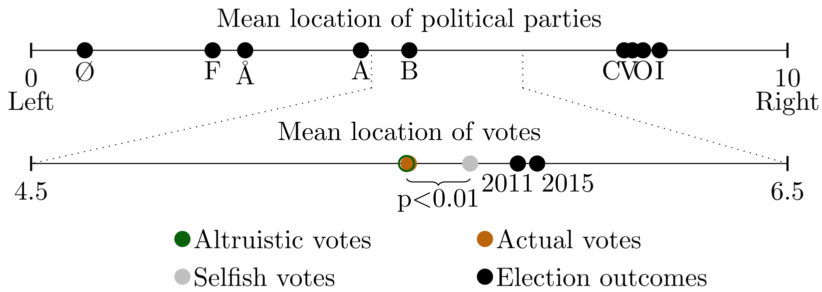
One Dimension

In order to break down the results in more detail, the political parties are placed on a scale from 0 (left) to 10 (right) using the mean placement by respondents who displayed a minimal political knowledge. The top part of Figure 6.5 shows these average placements. The bottom part shows the average location of the actual, altruistic, and selfish votes using the placements of the political parties.

The figure shows that the selfish votes were significantly more right-winged than the actual votes. The altruistic votes overlap with the actual votes. This suggests that if more people had voted altruistically, the mean outcome of the votes would hardly have changed. The figure also plots the election outcomes of the two most recent elections using the party placements weighted with the share of seats each party obtained in the given election. As Denmark shifted from a center-left to a right-wing government from 2011 to 2015, and the difference between the selfish and actual votes is of greater magnitude, this difference is of meaningful size.[4]

[4] There is a relatively big difference between the 2015 outcome and the actual votes for a few reasons. First

Figure 6.5: Mean Location of Votes



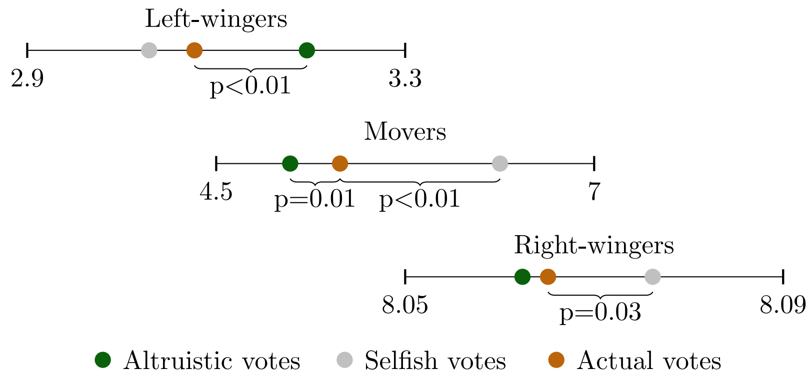
Note: The top part shows the average location of the various political parties on a scale from 0 (left) to 10 (right) using the mean assessment of respondents who displayed political knowledge. Given these placements, the bottom part shows the mean location of the respondents' actual, altruistic, and selfish votes. The altruistic votes and actual votes overlap almost entirely. The p-value is based on a paired t-test. The election outcome circles are constructed by using the placements of the political parties weighted with the share of seats each party obtained in the given election.

We can break down the results by different types of voters as shown in Figure 6.6. The top (bottom) part of the figure shows the average location of the altruistic, selfish, and actual votes for the subgroup of respondents who selected a left-wing (right-wing) party for all three choices. The middle part shows the location of the votes for respondents who selected both a right-wing party and a left-wing party in one or more of the three questions. A clear pattern emerges. Supporters of either side of the political spectrum believe that what is best for society is towards the center whereas what is best for themselves is further to the extreme in their own bloc.

Figure 6.6 indicates that the variances of the votes differ. This is tested in Figure 6.7, which shows the standard deviation of the actual, altruistic, and selfish votes. The selfish votes have a greater variance than the actual votes, and the altruistic votes have a smaller variance. Given that the mean location of the votes is near the center of the scale in all three cases, this suggests that the altruistic votes cluster more around centrist candidates, or in other words, that extreme candidates are less chosen.

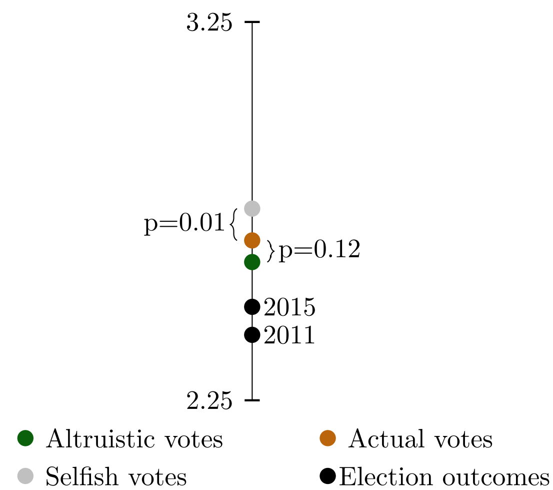
of all, the survey was conducted 4-5 months after the election, which allows some voters to have changed their mind. In the intervening months Denmark was met by a refugee crises, a change of leadership in the main opposition party, and a national budget proposal with drastic changes in funding for development aid, environmental issues, education and more. Secondly, although the sample is representative by age, gender, and region, this may not be sufficient to capture all types of voters. Finally, the 15% of the sample who do not know who to vote for in either the selfish or altruistic question may give further issues with regards to representativeness.

Figure 6.6: Mean Location of Votes by Type



Note: Mean location of the actual, altruistic, and selfish votes by voter type. *Left-wingers (Right-wingers)*: Individuals whose actual, selfish, and altruistic vote were for a left-wing (right-wing) party. *Movers*: Individuals whose actual, selfish, and altruistic votes contained both a left-wing party and a right-wing party. The scale is from 0 (left) to 10 (right). The p-values are from paired t-tests.

Figure 6.7: Standard Deviation of Votes

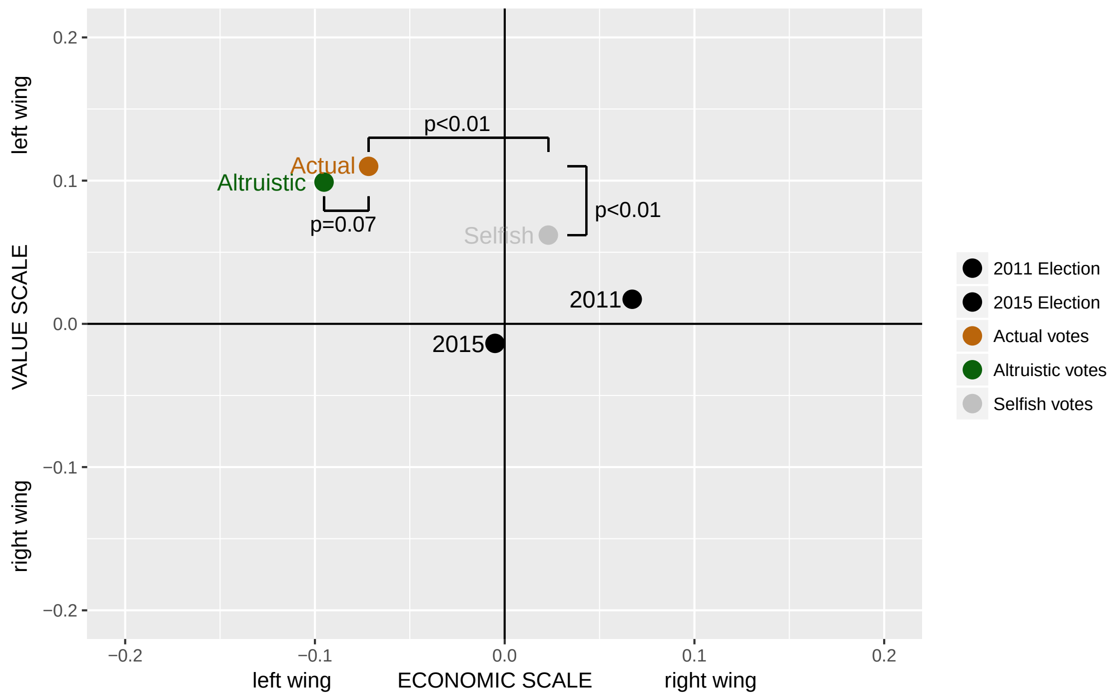


Note: Standard deviation of the actual, altruistic, and selfish votes using the placements of the political parties from Figure 6.1. The p-values are based on Levene's test.

Two Dimensions

A problem with the results presented thus far is that analyzing politics in one dimension may hide relevant information. To deal with this, the mean location of the selfish, actual, and altruistic votes can be compared using the political compass showed in Figure 6.2. This allows me to break down the party movements by the economic axis and the value axis. The results are shown in Figure 6.8.

Figure 6.8: Political Compass - Means

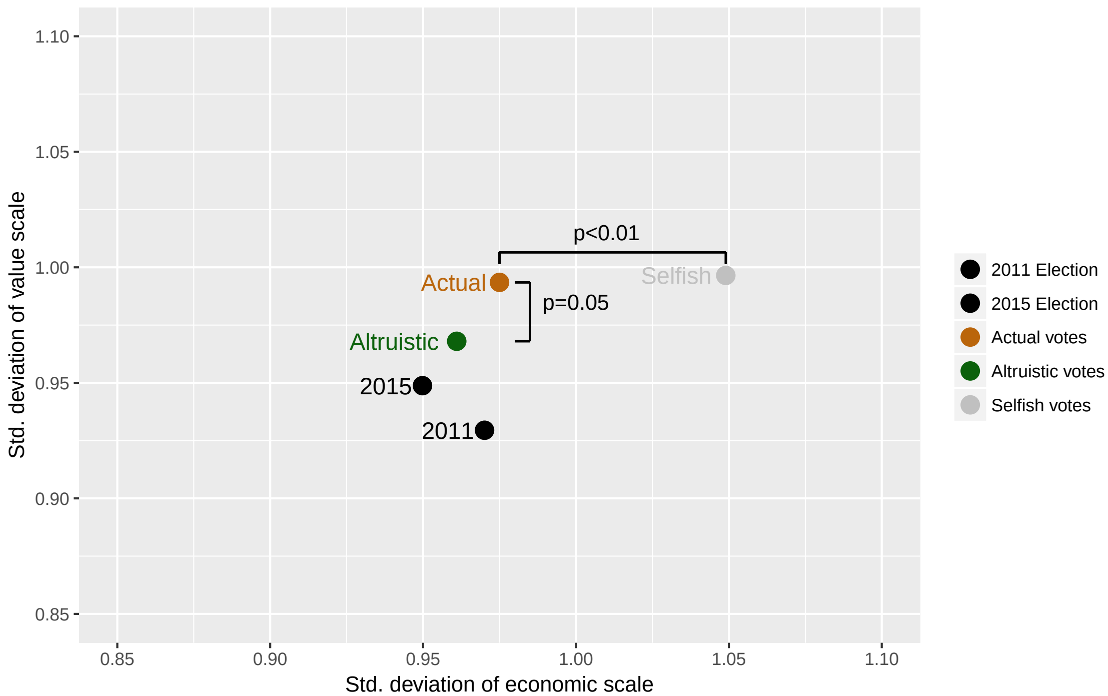


Note: Mean location of the selfish, actual, and altruistic votes in a two-dimensional political compass using data from DR and TV2's candidate test. 2011 and 2015 reflect the average location of votes in these two elections. The p-values are from paired t-tests.

By comparing the distance between the three points on the economic dimension with the distance between the three points on the value dimension, one quickly sees that economic concerns seem to drive most of the difference. The selfish votes are more right-winged in both dimensions, particularly in the economic scale. The altruistic votes are slightly more left-winged economically, but with no noteworthy difference in value politics. There might be less of a difference in value politics because values in general incorporate views about society as a whole and to a lesser extent private concerns. It may also be harder for individuals to evaluate what values are best for themselves in contrast to evaluating what economic policies are best for themselves. Again, when comparing the differences with the 2011 and 2015 election outcomes, they appear to be of important magnitude.

Figure 6.9 breaks down the results in the two dimensions by variance. The distance between the points is again greater in the economic scale than in the value scale. Selfish votes have a much greater variance in the economic dimension. Hence, when people vote selfishly they tend to choose parties in the extreme ends of the economic dimension. This is consistent with the story that if you have an above average income, your selfish vote from an economic perspective is the one that proposes the lowest tax rate. If you have a below average income, your selfish interest from an economic perspective is to vote for the party that favors the most redistribution. The altruistic votes are once again less extreme, this time particularly in the value dimension. This may suggest that when individuals think of which values are best for society as a whole, they choose what the average person believes.

Figure 6.9: Political Compass - Standard Deviation



Note: Standard deviation of the selfish, actual, and altruistic votes in two political dimensions. 2011 and 2015 reflect the average location of votes in these two elections. The p-values are from Levene's test.

To sum up, the analysis gives the following four insights: 1) The actual votes align closely with the altruistic votes, indicating that the respondents vote more altruistically than selfishly. This suggests that if more people had voted altruistically, it would hardly have impacted the election outcome. This is in line with hypothesis 1 and implies that the inferior pattern discussed in the epigraph has no empirical leverage. 2) The selfish votes are much more right-winged, and the altruistic votes slightly more left-winged than the actual votes. This is in line with hypothesis 2 and suggests that individuals on average become more egalitarian when thinking about what is best for society. 3) The selfish votes are placed on more extreme candidates, while there is more agreement within the altruistic votes. Hence, the more altruistically individuals vote, the more consensus there will be among the selected candidates and, supposedly, the better equipped they will be to reach compromises. 4) The latter two results apply primarily to the economic dimension of politics; when individuals think about what is best for themselves or best for society they mostly refer to tax and redistributive policies.

6.5.3 Checking for External Validity with U.S. Data

One concern is that the results only apply to a Danish context. To check whether this is the case, a similar survey adapted to U.S. circumstances was conducted in October 2015. The U.S. arguably makes a most different case to the Danish democracy based on the size of the democracy, the two-party system, and the presidential system. In addition, the American electorate is less homogeneous and contains more varied views. The survey was conducted through Harvard University's Digital Lab for the Social Sciences (DLABSS). In total 400 respondents took the survey. The pool of respondents is not representative of the U.S. electorate but rather comparable to online panels such as MTurk.

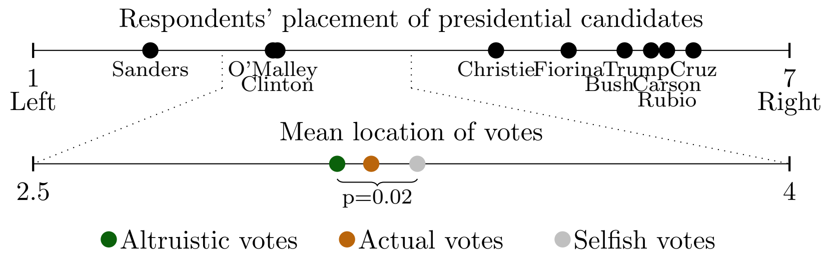
In a U.S. context, the two-party system may hide nuances between the selfish, altruistic, and actual choice. To deal with this issue, respondents were asked to select whom they would vote for among the presidential candidates at the time of the survey (fall 2015). This is naturally not how a ballot would look like, but it deals with the issue that partisanship likely generates only few changes across the two parties (Green *et al.*, 2004).

Each respondent was asked to place some of the candidates on a scale from 1 (extremely liberal) to 7 (extremely conservative). The top part of Figure 6.10 shows the mean location of a selection of the candidates. The bottom part compares the mean location of the respondents' votes using this scale. As was the case in Denmark, the selfish votes are more to the right and the altruistic votes are more to the left. Due to the smaller sample, the power of this finding is less strong.

Figure 6.11 compares the standard deviation of the votes. The selfish votes have the largest variance as was the case in Denmark.

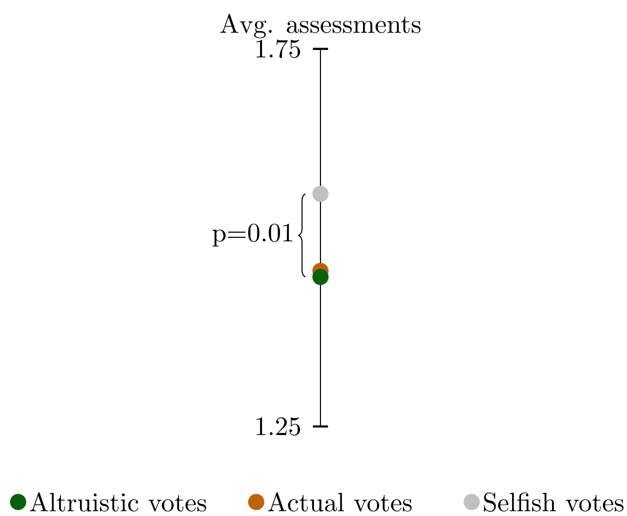
Naturally, these findings are suggestive. Nevertheless, as they are congruent with the Danish findings, there does not seem to be great concerns about external validity. The fact that both countries move to the middle when voting non-selfishly suggests that the altruistic outcome is defined in relative terms rather than in absolute terms. In both countries individuals may think that the median voter represents what is good for the country, even though the median voter is very different in the two countries.

Figure 6.10: U.S. Results - Mean Location of Votes



Note: The top part shows the location of the presidential candidates on a scale from 1 (extremely liberal) to 7 (extremely conservative) using the mean assessment by the respondents. Some candidates are omitted from the figure. The bottom part shows the mean location of the various votes based on the locations of the presidential candidates. The p-value is from a paired t-test. $n = 400$.

Figure 6.11: U.S. Results - Standard Deviation of Votes



Note: Standard deviation of the respondents' actual, altruistic, and selfish votes using the average placements of the presidential candidates as taken from Figure 6.10. The p-value is based on Levene's test. $n = 400$.

6.5.4 Robustness Checks

Another way to test how individuals' votes shift when voting selfishly and altruistically is by using their own assessment of where the political parties stand on the left-right wing scale. Using these locations one can detect if individuals choose a party more to the left or to the right *according to their own beliefs*. Figure 6A.1 in Appendix 6.A.3 does this for the Danish data. The results are very similar to the main results: the actual votes align closely to the altruistic votes, while the selfish votes are more left-winged and have a greater variance.

Bargsted and Kedar (2009) have shown that multiparty systems can be prone to strategic voting. If this drives the results, they are of less relevance for other elections. To indirectly check this, the respondents were asked if they would ever consider voting strategically (preceded by an explanation of what strategic voting is). Dividing the sample into two parts depending on the answer to this question, it is possible to check if the results for the ones that would not vote strategically are similar to the overall findings. Figure 6A.2 in Appendix 6.A.3 shows that the results for the sub-sample of respondents that would not vote strategically are in line with the overall results. This makes it unlikely that the results are driven by strategic voting. The findings for the sub-sample that would vote strategically are either insignificant or in line with the general results. The fact that similar findings were found in the U.S. and Denmark even though the voters face very different strategic incentives in these two democracies also suggests that the results are not driven by strategic voting.

A concern with the postulated results is that people may vote altruistically but care for *more* than the society at large. Individuals that care about the environment, refugees, and global poverty may not constrain themselves to the welfare of the citizens of the society they live in when voting. Suppose for example that some voters think party A is best for society, but vote for party B because they think it is best for the world at large. In that case the analysis until now would falsely have classified these voters as voting non-altruistically. The respondents were also asked which party they would vote for if they were to consider what is best for the entire world. Obviously, more people had a hard time answering this question. Nevertheless, vote distributions have been reanalyzed for the respondents that answered all four questions in Figures 6A.3-6A.8 in Appendix 6.A.3. In general, votes for the world are much more left-winged than the societal altruistic votes, in particular in the value dimension. The standard deviation is quite comparable to the actual votes. Hence, votes for the world pull some people to parties on the far left-wing of the value dimension.

6.6 Conclusion

This paper analyzed whether the fact that some people vote altruistically and others selfishly matters for voting outcomes. This was tested by conducting a survey in Denmark ($n = 2000$) and in the U.S. ($n = 400$) where respondents were asked to identify (1) the party they would vote for if elections were held tomorrow, (2) the party they would vote for, if they only were to consider what is best for themselves, and (3) the party they would vote for if they were to consider what is best for society as a whole. The results showed that if more people had voted selfishly, the election outcome would have been more right-winged and extreme candidates would have garnered more votes. If more people had voted altruistically, the outcome would become a bit more left-winged and a bit more concentrated around the centrist choices.

Overall, the paper finds that it does in fact matter whether individuals vote for selfish reasons or for altruistic reasons. For political parties, this means that vote shares can be significantly increased if more people are compelled to vote for a different reason. The emerging question that arises from this paper is whether political candidates can capitalize on the duality of voting motivations. A number of studies have shown that framing can invoke altruistic attitudes and behavior in the form of preferences over the management of forest areas (Russell *et al.*, 2003), willingness to pay for conservation areas (Ovaskainen and Kniivilä, 2005), and willingness to pay for public goods in general (Ajzen *et al.*, 1996). It is uncertain if these findings map unto voting behavior.

For policymakers the findings are important whenever policies are created on the bases of the electorate's preferences. Since individuals possess both altruistic and selfish preferences, and since these differ systematically, a policymaker has to determine which preferences to use. If altruistic preferences are used the suggested policies should be more left-winged than if selfish preferences are used.

Bibliography

AJZEN, I., BROWN, T. C. and ROSENTHAL, L. H. (1996). Information bias in contingent valuation: Effects of personal relevance, quality of information, and motivational orientation. *Journal of Environmental Economics and Management*, **30** (1), 43–57. (page 216)

ANDREONI, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, **100** (401), 464–477. (page 197)

ARROW, K. (1951). Individual values and social choice. *New York: Wiley*, **24**. (page 193)

BARGSTED, M. A. and KEDAR, O. (2009). Coalition-targeted duvergerian voting: How expectations affect voter choice under proportional representation. *American Journal of Political Science*, **53** (2), 307–323. (page 214)

CHARNESS, G. and RABIN, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, **117** (3), 817–869. (page 196)

CHONG, D., CITRIN, J. and CONLEY, P. (2001). When self-interest matters. *Political Psychology*, **22** (3), 541–570. (page 196)

DOWNS, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, pp. 135–150. (page 193)

DWORKIN, R. (1978). *Taking rights seriously*. Harvard University Press. (page 193)

EDLIN, A., GELMAN, A. and KAPLAN, N. (2007). Voting as a rational choice: Why and how people vote to improve the well-being of others. *Rationality and Society*, **19** (3), 293–314. (page 193, 205)

ELSTER, J. (1997). The market and the forum: Three varieties of political theory. *Deliberative democracy: Essays on reason and politics*, pp. 3–33. (page 198)

— (2006). Altruistic behavior and altruistic motivations. *Handbook of the economics of giving, altruism and reciprocity*, **1**, 183–206. (page 193)

Feddersen, T., Gailmard, S. and Sandroni, A. (2009). Moral bias in large elections: Theory and experimental evidence. *American Political Science Review*, **103** (02), 175–192. (page 193, 197)

— and Sandroni, A. (2006). A theory of participation in elections. *The American Economic Review*, **96** (4), 1271–1282. (page 193)

Fisher, R. C. (1985). Taxes and expenditures in the U.S.: Public opinion surveys and incidence analysis compared. *Economic Inquiry*, **23** (3), 525–550. (page 196)

Fowler, J. H. (2006). Altruism and turnout. *Journal of Politics*, **68** (3), 674–683. (page 193)

Frank, T. (2007). *What's the matter with Kansas? How conservatives won the heart of America.* Macmillan. (page 198)

Funk, C. L. (2000). The dual influence of self-interest and societal interest in public opinion. *Political Research Quarterly*, **53** (1), 37–62. (page 196)

Green, D. P., Palmquist, B. and Schickler, E. (2004). *Partisan hearts and minds: Political parties and the social identities of voters.* Yale University Press. (page 213)

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility. *Journal of Political Economy*, **63** (4), 309–321. (page 193)

Höchtl, W., Sausgruber, R. and Tyran, J.-R. (2012). Inequality aversion and voting on redistribution. *European Economic Review*, **56** (7), 1406–1421. (page 203)

Hudson, J. and Jones, P. (2002). In search of the good samaritan: Estimating the impact of altruism on voters' preferences. *Applied Economics*, **34** (3), 377–383. (page 196, 197, 198)

— and Jones, P. R. (1994). The importance of the ethical voter: An estimate of altruism. *European Journal of Political Economy*, **10** (3), 499–509. (page 196, 197, 198)

Jankowski, R. (2007). Altruism and the decision to vote explaining and testing high voter turnout. *Rationality and Society*, **19** (1), 5–34. (page 193)

Jencks, C. (1990). Varieties of altruism. *Beyond self-interest*, pp. 53–67. (page 195)

Kiatpongsan, S. and Norton, M. I. (2014). How much (more) should CEOs make? A universal desire for more equal pay. *Perspectives on Psychological Science*, **9** (6), 587–593. (page 199)

Kiewiet, D. R. and Lewis-Beck, M. S. (2011). No man is an island: Self-interest, the public interest, and sociotropic voting. *Critical Review*, **23** (3), 303–319. (page 196)

Kinder, D. R. and Kiewiet, D. R. (1979). Economic discontent and political behavior: The role of personal grievances and collective economic judgments in congressional voting. *American Journal of Political Science*, **23** (2), 495–527. (page 196)

— and — (1981). Sociotropic politics: The American case. *British Journal of Political Science*, **11** (02), 129–161. (page 196)

Lewin, L. (1991). *Self-interest and public interest in Western politics*. Oxford University Press. (page 196)

Lusk, J. L., Nilsson, T. and Foster, K. (2007). Public preferences and private choices: Effect of altruism and free riding on demand for environmentally certified pork. *Environmental and Resource Economics*, **36** (4), 499–521. (page 197)

Markus, G. B. (1988). The impact of personal and national economic conditions on the presidential vote: A pooled cross-sectional analysis. *American Journal of Political Science*, **32** (1), 137–154. (page 196)

McConnell, K. E. (1997). Does altruism undermine existence value? *Journal of Environmental Economics and Management*, **32** (1), 22–37. (page 196)

Meehl, P. E. (1977). The selfish voter paradox and the thrown-away vote argument. *American Political Science Review*, **71** (01), 11–30. (page 195)

Messer, K. D., Poe, G. L., Rondeau, D., Schulze, W. D. and Vossler, C. A. (2010). Social preferences and voting: An exploration using a novel preference revealing mechanism. *Journal of Public Economics*, **94** (3), 308–317. (page 193)

Norton, M. I. and Ariely, D. (2011). Building a better America - One wealth quintile at a time. *Perspectives on Psychological Science*, **6** (1), 9–12. (page 199)

Nyborg, K. (2000). Homo economicus and homo politicus: Interpretation and aggregation of environmental values. *Journal of Economic Behavior & Organization*, **42** (3), 305–322. (page 196)

Ovaskainen, V. and Kniivilä, M. (2005). Consumer versus citizen preferences in contingent valuation: Evidence on the role of question framing. *Australian Journal of Agricultural and Resource Economics*, **49** (4), 379–394. (page 216)

Russell, C. S., Bjørner, T. B. and Clark, C. D. (2003). Searching for evidence of alternative preferences, public as opposed to private. *Journal of Economic Behavior & Organization*, **51** (1), 1–27. (page 216)

Sears, D., Lau, R. and Tyler, T. (1980). Self-interest vs. symbolic politics in policy attitudes and presidential voting. *American Political Science Review*, **74** (3), 670–684. (page 196)

Sears, D. O. and Funk, C. L. (1990). The limited effect of economic self-interest on the political attitudes of the mass public. *Journal of Behavioral Economics*, **19** (3), 247–271. (page 196)

—, Hensler, C. P. and Speer, L. K. (1979). Whites' opposition to busing: Self-interest or symbolic politics? *American Political Science Review*, **73** (2), 369–384. (page 196)

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, **6** (4), 317–344. (page 193)

Smullyan, R. M. (1980). *What is the name of this book? The riddle of Dracula and other logical puzzles*. Dover Publications. (page 193)

Stokes, D. E. (1963). Spatial models of party competition. *American Political Science Review*, **57** (2), 368–377. (page 198)

Sudit, M. (1988). Ideology or self-interest? Medical students' attitudes toward national health insurance. *Journal of Health and Social Behavior*, **29** (4), 376–384. (page 196)

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications. (page 201)

Zettler, I., Hilbig, B. E. and Haubrich, J. (2011). Altruism at the ballots: Predicting political attitudes and behavior. *Journal of Research in Personality*, **45** (1), 130–133. (page 198)

6.A Appendix

6.A.1 Questionnaire

Q1: Would you say that you are very interested in politics, somewhat interested in politics, only slightly interested in politics, or not at all interested in politics?

- Very interested
- Somewhat interested
- Only slightly interested
- Not interested at all

[new page]

Q2: Who would you vote for if general elections were held tomorrow?

- A: Social Democrats
- B: Social Liberal Party
- C: Conservative People's Party
- F: Socialist People's Party
- I: Liberal Alliance
- O: Danish People's Party
- V: Liberals
- Ø: Red-Green Alliance
- Å: The Alternative
- An independent candidate
- Another party
- I would cast a null vote
- I would not vote
- I am not eligible to vote
- Don't know

[If Q2 is answered "I am not eligible to vote" → end of survey]
[If Q2 is answered "Don't know" → Q2b]
[Else → Q3]

Q2b: Even though you are in doubt we would like to ask you if there is a party you are more inclined to vote for?

- A: Social Democrats
- B: Social Liberal Party
- C: Conservative People's Party

- F: Socialist People's Party
- I: Liberal Alliance
- O: Danish People's Party
- V: Liberals
- Ø: Red-Green Alliance
- Å: The Alternative
- An independent candidate
- Another party
- Don't know/still in doubt

[new page]

[Randomize order of Q3 and Q4]

Q3: Who would you vote for if you only were to consider what is <u>best for yourself?</u>

- A: Social Democrats
- B: Social Liberal Party
- C: Conservative People's Party
- F: Socialist People's Party
- I: Liberal Alliance
- O: Danish People's Party
- V: Liberals
- Ø: Red-Green Alliance
- Å: The Alternative
- An independent candidate
- Another party
- Don't know

[new page]

Q4: Who would you vote for if you were to consider what is <u>best for society as a whole?</u>

- A: Social Democrats
- B: Social Liberal Party
- C: Conservative People's Party
- F: Socialist People's Party
- I: Liberal Alliance
- O: Danish People's Party
- V: Liberals
- Ø: Red-Green Alliance

- Å: The Alternative
- An independent candidate
- Another party
- Don't know

[new page]

Q5: Who would you vote for if you were to consider what is <u>best for the entire world?</u>

- A: Social Democrats
- B: Social Liberal Party
- C: Conservative People's Party
- F: Socialist People's Party
- I: Liberal Alliance
- O: Danish People's Party
- V: Liberals
- Ø: Red-Green Alliance
- Å: The Alternative
- An independent candidate
- Another party
- Don't know

[new page]

Q6: Do you consider yourself a supporter of a particular political party?

- Yes, I consider myself a supporter of a particular political party
- No, I do not consider myself a supporter of a particular political party
- Don't know

Q7: Different people feel differently about voting. For some, voting is a duty - they feel they should vote in every election no matter how they feel about the parties. For others voting is a choice. For you personally, is voting mainly a duty, mainly a choice, or neither a duty nor a choice?

- Mainly a duty
- Mainly a choice
- Neither a duty nor a choice
- Don't know

Q8: Some people vote "strategically." That is, they vote for someone else than their preferred party in an attempt for their vote to have a bigger impact on the result of the elections. Would *you vote strategically?*

- Yes, I would vote strategically
- No, I would not vote strategically
- Don't know
- I rarely vote

Q9: People vote for different reasons. Some people vote mainly to influence the outcome of elections while others vote mainly to express their opinion. Why do you vote?

- Mainly to influence the outcome of elections
- Mainly to express my opinion
- Both
- Neither
- Don't know
- I rarely vote

[new page]

Q10: Have you ever donated blood?

- Yes
- No
- Don't know

Q11: To what extent do you agree or disagree that the economic inequality in the Danish society should be reduced?

- Very much agree
- Agree
- Neither agree nor disagree
- Disagree
- Very much disagree
- Don't know

Q12: In Denmark you get tax deductions for certain charitable donations. Did you receive such a tax deduction last year?

- Yes
- No
- Don't know

[new page]

Q13: Who did you vote for at the general elections the 18th of June 2015?

- A: Social Democrats
- B: Social Liberal Party
- C: Conservative People's Party
- F: Socialist People's Party
- I: Liberal Alliance
- O: Danish People's Party
- V: Liberals
- Ø: Red-Green Alliance
- Å: The Alternative
- An independent candidate
- Another party
- I cast a null vote
- I did not vote
- I was not eligible to vote

[new page]

Q14: In politics people often talk about left and right.
Where would you place yourself on this scale?

	0. left	1	2	3	4	5	6	7	8	9	10. right	DK
Yourself												

Q15: Where would you place the political parties?

	0. left	1	2	3	4	5	6	7	8	9	10. right	DK
A: Social Democrats												
B: Social Liberal Party												
C: Conservative People's Party												
F: Socialist People's Party												
I: Liberal Alliance												
O: Danish People's Party												
V: Liberals												
Ø: Red-Green Alliance												
Å: The Alternative												

[new page]

Now there will be some questions about politics in general. There can be many complicated questions in politics. Remember that it is always possible to answer "don't know."

Q16: Which parties formed the government in the months leading up to the general elections of 2015?

- A: Social Democrats
- B: Social Liberal Party
- C: Conservative People's Party
- F: Socialist People's Party
- I: Liberal Alliance
- O: Danish People's Party
- V: Liberals
- Ø: Red-Green Alliance
- Å: The Alternative
- Don't know

[new page]

Q17: How many members of parliament are there if we ignore the four from Greenland and the Faeroe Islands?

- Note: ____
- Don't know

[new page]

Now there will be some questions about you.

Q18: Are you male or female?

- Male
- Female

Q19: What year were you born (i.e. 1982)

- Note: ____

Q20: What is your zip code?

- Note: ____
- I live abroad
- Don't know

Q21: What is your highest completed education level?

- Primary education (i.e. 9th or 10th grade)

- Secondary education (i.e. regular high school or vocational high school)
- Vocational training (i.e. carpenter, health care assistant, nursery assistant)
- Shorter higher education (i.e. laboratory technician, educator, building technician)
- Medium higher education (i.e. teacher, nurse)
- Bachelor' s degree (i.e. BSc, BA)
- Longer higher education (i.e. architect, MA, MD)
- Doctoral degree (i.e. PhD)
- Other
- Don't know

Q22: What description best matches your labor market status?

- Employee in the private sector
- Employee in the public sector
- Self-employed
- Student
- Retired/on early retirement benefits
- Unemployed
- Long-term sick
- On maternity or paternity leave
- Other

Q23: What is your yearly gross income?

- Less than 100.000 kr.
- 100.000-199.999 kr.
- 200.000-299.999 kr.
- 300.000-399.999 kr.
- 400.000-499.999 kr.
- 500.000-749.999 kr.
- 750.000-999.999 kr.
- 1.000.000 kr. or above
- Don't know / will not disclose

Q24: What is your household's yearly gross income?

- Less than 100.000 kr.
- 100.000-199.999 kr.
- 200.000-299.999 kr.
- 300.000-399.999 kr.

- 400.000-499.999 kr.
- 500.000-749.999 kr.
- 750.000-999.999 kr.
- 1.000.000 kr. or above
- Don't know / will not disclose

6.A.2 Derivation of Political Compass

DR and TV2 are the two major television networks in Denmark. Prior to the election of 2015 they asked all candidates running for parliament to answer the following questions (translated from Danish) on a 5-point scale from “completely disagree” to “completely agree” *(my labels added prior to the questions).*

TV2

- *Employment1*: More jobs should be created in the public sector
- *Employment2*: Politicians should do more to ensure that workers from other EU countries do not undercut Danish wages
- *Employment3*: The time required to be re-eligible for unemployment benefits should be lowered from 12 to 6 months
- *Health1*: It is economically necessary to introduce private fees in selected areas of the health sector.
- *Health2*: One of the main priorities for the hospitals should be to create a more coherent treatment of patients, whereby the same doctor and the same nurses follow the patient throughout the hospitalization
- *Health3*: Increased competition from the private sector makes the public health system perform better
- *Immigrants1*: The Danish policy towards immigrants is too strict
- *Immigrants2*: It should be easier than is the case today to expel immigrants/refugees, who have violated the criminal law
- *Immigrants3*: There should be more differentiation with regards to immigration than is the case today, such that it will be easier for certain nationalities to obtain residence permits than others
- *Social1*: Too many people are stuck with social transfers because the allowances are too high
- *Social2*: More economic support should be given to families with children than is the case today
- *Social3*: Economic inequality in the Danish society should be reduced
- *Social4*: The requirement of mutual dependencies for benefit recipients should be removed
- *Children1*: There is too large a focus on tests in the Danish primary school
- *Children2*: There should be a greater political focus on socially vulnerable families with children and less on well-functioning families than is the case today
- *Children3*: There should be a greater emphasis on discipline in day-care centers
- *Economy1*: In the long run it is economically necessary to introduce private fees in certain

selected areas of the elderly care

- *Economy2*: The Danish level of wages is so high that it hurts the Danish economy
- *Economy3*: Public investments should increase in order to strengthen the economy
- *Economy4*: If the conditions for private companies improve, our competitiveness and the Danish economy will benefit
- *Taxes1*: The top marginal tax bracket should be maintained
- *Taxes2*: The property tax should be raised and the revenue should be used to lower the tax on labor
- *Taxes3*: There should be a differentiated VAT, such that for example healthy food will have a low VAT while other goods will have a much higher VAT
- *Foreign1*: Denmark should participate less in international military operations
- *Foreign2*: Denmark should increase its defense spending
- *Foreign3*: Development assistance should be lowered
- *Crime1*: Sentences for crimes involving violence should be increased
- *Crime2*: The age of criminal liability, which is 15 years today, should be lowered
- *Crime3*: There should be less emphasis on punishment and greater emphasis on rehabilitation in Danish law
- *Environment1*: Corporates' green taxes should be increased
- *Environment2*: Politicians should create incentives for more farmers to transition from traditional to organic animal breeding
- *Environment3*: It is an important political task to get Danes to recycle more than they do today
- *Environment4*: It benefits both the environment and the economy if there is a greater political emphasis on the transition to renewable energy than is the case today
- *EU1*: Denmark has given up too much power to the EU
- *EU2*: Denmark should abolish its opt-out on justice matters in the next election cycle
- *EU3*: In the long run Turkey should join the EU
- *Education1*: Students in primary schools should receive grades earlier than is the case today
- *Education2*: There should be more discipline in the primary schools
- *Elderly1*: More resources should be directed towards elderly care even if this means that other welfare areas will have to receive fewer resources
- *Elderly2*: The elderly care should be income adjusted such that wealthy elderly pay for some of the services they receive
- *Animals1*: Fur farming should be banned
- *Animals2*: The practice of keeping hens in cages should be banned

DR

- *Education3*: After the school reform, students have too long days at school
- Taxes4: The tax on cigarettes should be increased
- *Health4*: A visit to the general practitioner should cost e.g. 100 kr.
- *Elderly3*: More of the elderly care should be outsourced to private companies
- *Employment4*: The time required to be re-eligible for unemployment benefits should be decreased
- *Employment5:* Companies should be held accountable for whether their foreign subcontractors in Denmark comply with Danish rules on wages, taxes and VAT or not
- *Taxes5*: Growth in the public sector is more important than tax cuts
- *Environment5*: Investment in public transportation should be given priority over investment to the benefit of private cars
- *Crime4*: The punishment for severe violence and rape must be increased
- *Economy5*: Unemployment benefits should be lowered such that the financial gains from working increase.
- *Immigrants4*: Public institutions in Denmark take religious minorities too much into consideration
- *EU4*: EU decides too much compared to Danish law
- *Foreign4*: The development aid should be lowered
- *Environment6*: Efforts to improve the environment should take priority over economic growth
- *Culture1*: The public subsidy for culture should be lowered

Political discourse in Denmark has it that Danish politics consists of two or three dimensions, these being an economic axis, a value axis (concerning issues such as immigration, environmentalism, and crime), and an EU policy axis. To let the data speak for itself, Exploratory Factor Analysis is used to get a qualified understanding of how many axes exist and which items load unto which axis. The most reasonable result has the three speculated factors. The resulting loadings using a promax rotation are displayed in Table 6A.1. Only loadings over 0.6 and only questions with at least one loading above 0.6 are displayed. With a fourth factor only one question, *Children3*, has a loading above 0.5. As this question has little relevance in political discourse, three factors were deemed appropriate.

The pattern from the factor loadings highly resembles the expectations. Questions on employment, the economy, and health care load unto the economy factor, questions on crime,

Table 6A.1: Factor Loadings from Exploratory Factor Analysis

Question	Factor 1 (economy)	Factor 2 (values)	Factor 3 (EU)
Employment1	0.80		
Employment2	1.00		
Employment3	0.77		
Employment4	0.80		
Employment5	0.82		
Social1	0.61		
Social3	0.73		
Social4	0.74		
Economy1	0.70		
Economy3	0.79		
Economy4		0.65	
Economy5	0.65		
Health1	0.92		
Health4	0.93		
Taxes1	0.73		
Taxes2		0.79	
Taxes4		0.61	
Taxes5	0.85		
Elderly3	0.82		
Children1		0.81	
Children3		0.63	
Crime1		0.87	
Crime2		0.68	
Crime3		0.86	
Crime4		0.82	
Environment1		0.78	
Environment5		0.72	
Environment6		0.73	
Foreign2		0.68	
Foreign3		0.73	
Foreign4		0.77	
Immigrants1		1.00	
Immigrants2		0.94	
Immigrants3		0.80	
Immigrants4		0.92	
Education1		0.63	
Education2		0.81	
Education3			0.78
EU1			0.84
EU2			0.87
EU3		0.74	
EU4			0.82

immigration, and the environment load unto the values factor, and questions on EU load unto the EU factor. A few tax questions load unto the values factor, but since these questions by and large concern value politics (such as whether cigarettes should be taxed), this is hardly surprising. Perhaps the only surprise is that an education question loads unto the EU factor. This question pertains to the assessment of a school reform, which the centrist parties favored and the more extreme parties opposed. This happens to be the divide of EU politics, and hardly anything more should be attributed to this finding.

A problem with the factor loadings is that they need not reflect the political importance and salience of the issues. To accommodate this, when calculating the factor scores for each of three factors, I weight the factor loadings with a measure of how important each question is. These weights are obtained from an opinion poll at the time of the election conducted by the same agency that was responsible for the TV2 candidate test. The weights reflect which topics the respondents found most important.

Table 6A.2: Topic Weights

Topic	Weight
Health	0.30
Immigrants	0.29
Employment	0.22
Economy	0.21
Social	0.14
Elderly	0.14
Taxes	0.13
Children	0.13
Environment	0.10
Education	0.07
Animals	0.05
EU	0.04
Foreign	0.03
Crime	0.03

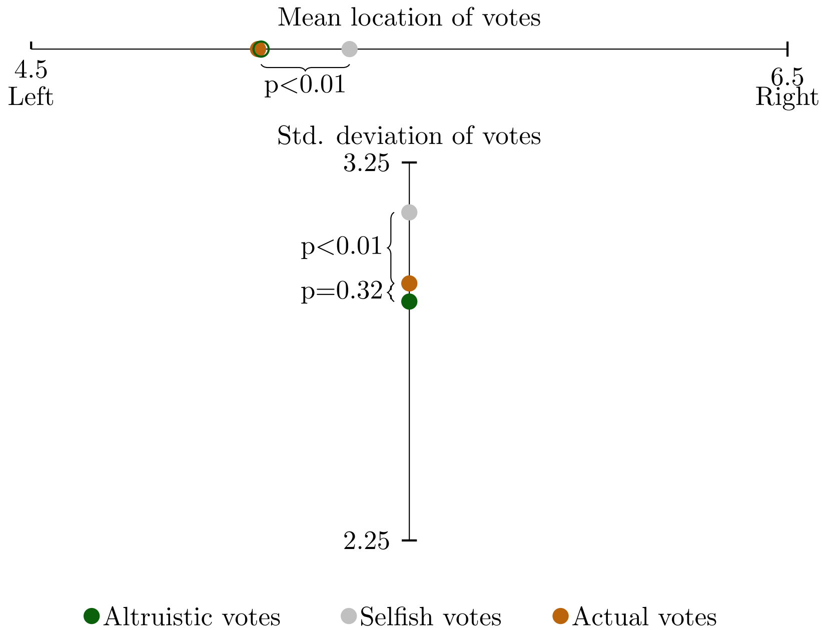
Note: Share of respondents in an opinion poll that thought specific topics were among the most important for the election (respondents could select more than one option). The numbers are obtained from http://politik.tv2.dk/valg2015/2015-05-27-danskernes-valg-sundhed-og-hospitaler-er-det-vigtigste-tema.

To obtain the final political compass showed in the main text, the resulting individual factor scores are standardized and party averages are taken from the two first factors.

6.A.3 Robustness Checks

Using Subjectively Assessed Positions

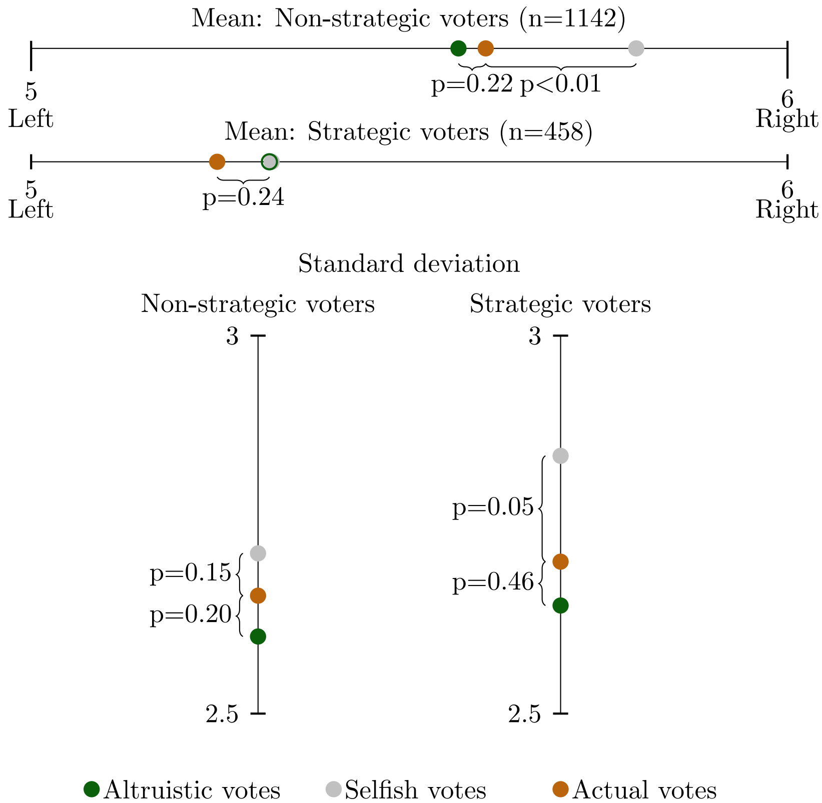
Figure 6A.1: Using Subjective Scales



Note: Mean and standard deviation of respondents' actual, altruistic, and selfish votes based on where the respondents themselves located the party they were voting for on a scale from 0 (left) to 10 (right). The p-value in the upper part is from a paired t-test. P-values in the lower part are from Levene's test.

Accounting for Strategic Voting

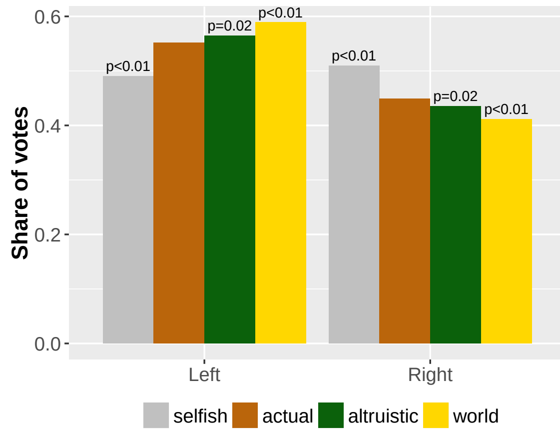
Figure 6A.2: Strategic Voting



Note: Mean and standard deviation of respondents' actual, altruistic, and selfish votes broken down by subgroups. *Non-strategic voters* answered negatively to a question of whether they would consider voting strategically. *Strategic voters* answered positively to this question. P-values in the upper part are from paired t-tests. P-values in the lower part are from Levene's test.

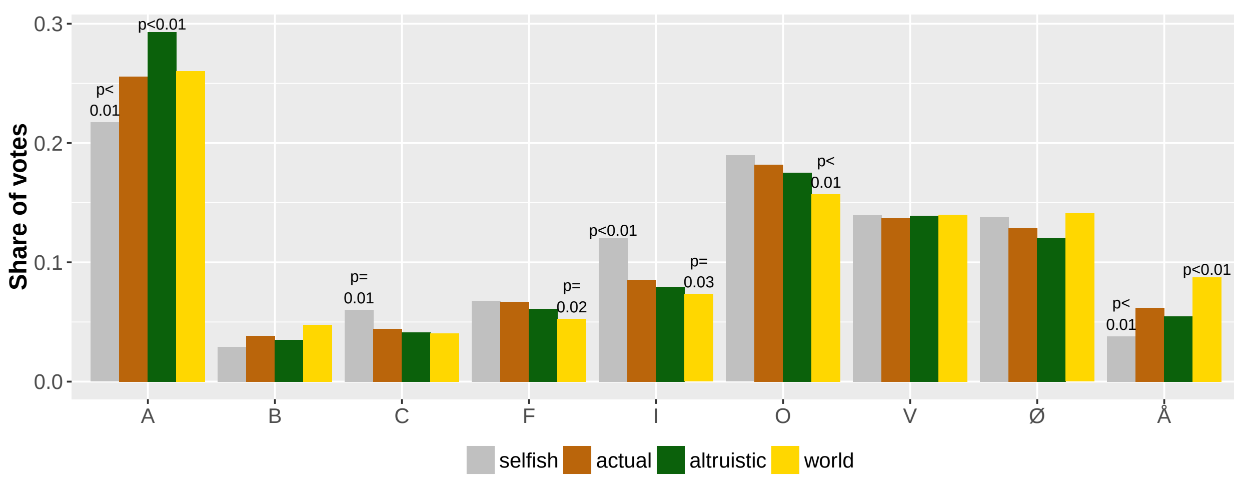
The World

Figure 6A.3: Distribution of Votes by Bloc



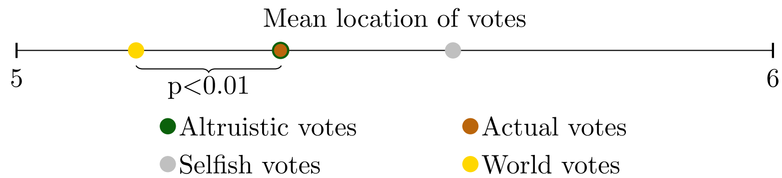
Note: Distribution of selfish, actual, altruistic, and world votes by political bloc. World votes are answers to the question 'Who would you vote for if you were to consider what is best for the entire world?' P-values indicate difference from the share of actual votes using McNemar's test. $n = 1377$.

Figure 6A.4: Distribution of Votes by Party



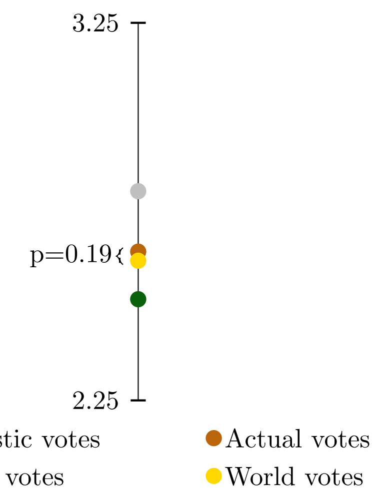
Note: Distribution of selfish, actual, altruistic, and world votes by political party. P-values indicate difference from the share of actual votes using McNemar's test. $n = 1377$. A: Social Democrats, B: Social Liberal Party, C: Conservative People's Party, F: Socialist People's Party, I: Liberal Alliance, O: Danish People's Party, V: Liberals, Ø: Red-Green Alliance, Å: The Alternative.

Figure 6A.5: Mean Location of Votes



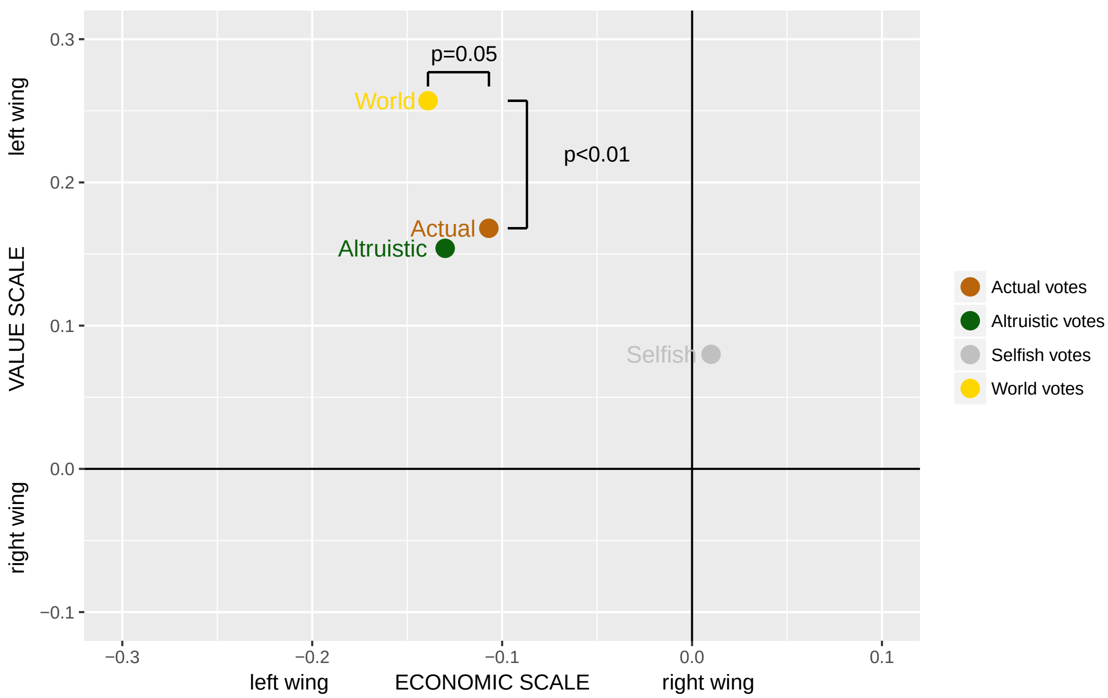
Note: Mean location of the respondents' actual, altruistic, selfish, and world votes. The p-values are from paired t-tests. $n = 1377$.

Figure 6A.6: Standard Deviation of Votes



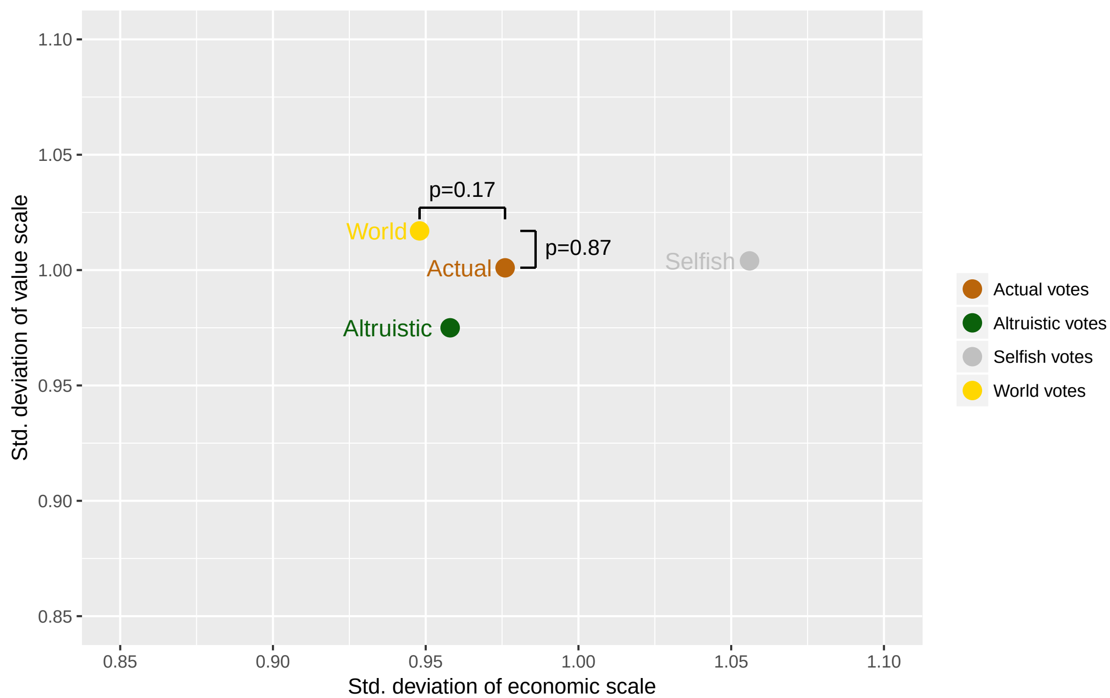
Note: Standard deviation of the respondents' actual, altruistic, selfish, and world votes. The p-value is from a Levene's test. $n = 1377$.

Figure 6A.7: Political Compass - Means



Note: Mean location of the selfish, actual, altruistic and world votes in a two-dimensional political compass. P-values are from paired t-tests. $n = 1377$.

Figure 6A.8: Political Compass - Variance



Note: Standard deviation of the selfish, actual, altruistic, and world votes in a two-dimensional political compass. P-values are from Levene's test. $n = 1377$.