FACULTEIT ECONOMIE

EN BEDRIJFSWETENSCHAPPEN

KU LEUVEN

Essays on the Measurement of Social Inclusion and Social Exclusion in the European Union: a Multidimensional and Multilevel Analysis

Nummer 649

2019

Doctoral Committee

Advisors:

Prof. dr. Nicky Rogge	KU Leuven
Prof. dr. Tom Van Puyenbroeck	KU Leuven

Members:

Prof. dr. Laurens Cherchye	KU Leuven
Prof. dr. Koen Decancq	University of Antwerp
Prof. dr. Elena Bárcena-Martín	University of Malaga

Chair:

Prof. dr. Johan Eyckmans	KU Leuven

Acknowledgements

This PhD has taught me so much more than what is written in the next four chapters and I owe many thanks to the people who have supported me in different ways.

First and foremost I offer my heartfelt thanks to my supervisor, Nicky Rogge, and co-supervisor, Tom Van Puyenbroeck, for their constant encouragements and support. I truly feel blessed with their supervision. At any moment during the last five years, I could come to Nicky and Tom with any problem, big or small. Their doors were always open to me. Nicky gave me the freedom to work in my own way. If you would compare this dissertation with its initial research proposal, you would see that he trusted me to pursue my own research interests. I feel privileged for this, as there are not many people who have such liberty in their professional life. I am also very grateful for the fact that Nicky and Tom supported me in interrupting my PhD to work for the European Commission. At the same time, while he endorsed the pursuit of new research ideas, he would always stimulate me to remain pragmatic, efficient and realistic. But besides the excellent quality of their supervision, writing a PhD under the supervision of Nicky and Tom was most of all a pleasant experience. I really enjoyed the numerous of meetings the three of us had at Scott's, the bar next to the campus, where we would discuss the progress of my research or explore new ideas. It appears brainstorming and drinking Duvel is combination that goes particularly well together.

Besides my supervisors, I would like to thank the other members of my doctoral committee: Laurens Cherchye, Koen Decancq and Elena Bárcena-Martín. I found the pivate defense to be highly stimulating and I thank them for their careful reading of the chapters. Their valuable and constructive remarks and suggestions during the private defence have been most useful for improving the quality of my research. A special thanks goes out to Koen, who triggered my interest in well-being measurement when he was the supervisor of my master thesis at the University of Antwerp back in 2012. I further extend my gratitude to Johan Eyckmans for his willingness to chair the public defence.

Importantly, I am indebted to the co-authors that contributed greatly to this dissertation, Anne-Catherine Guio, who co-authored the first and second chapter, and Frank Vandenbroucke and Eric Marlier, who co-authored the second chapter. I have learned a lot about policy-orientated research, econometrics and working with the EU-SILC dataset during these cooperations. It was a true privilege to have Frank as the supervisor of my master thesis during the MASE doctum colloquium programme. I would like to thank Anne-Catherine and Eric for giving me the opportunity to attend and present the second chapter of this dissertation at the NET-SILC3 conference and workshop in Athens.

Writing a PhD can be a lonely endeavor, but I have been surrounded by great people who have made the process a nice experience. I was fortunate to be a part of a whole new generation of young PhD students who started working at the Brussels campus. It didn't take a long time for us to 'find each other', in offering advice and support and sharing frustrations. But most importantly, we had fun. We had countless of dinners and (happy hour) drinks at Scott's. I am sure our little PhD excursions to the Ardennes and Ostend will remain memorable for a long time. A special thanks for the good times go out to Rosine, Tobias, Prabesh, Lode, Jann, Sam, Dori, Liesbeth, Céline and Carolien.

Another special word goes out to two colleagues of mine that helped me a lot in the completion of this work: Machteld and Jens. We had a nice 'clique' in our little corner of the PhD office with some truly hilarious moments. It is nice to go to work and be able to be yourself (even if this includes being grumpy, sorry for that!). Machteld, when you entered the PhD office as a colleague, I immediately knew that I would like you. You have become a wonderful friend. Our discussion about econometrics and academic research in general helped me become a better researcher. Jens, we started our PhD at the same time and it has always been a pleasure to share an office with you. I am grateful for the advice I got from you in terms of optimization problems and I wish you good luck with finishing your PhD later this year.

I would like to thank my brothers and all my friends for their support in these last years. Without them, I would have never been able to complete this dissertation. It would be impossible to name everyone, but a warm thanks go to Britt, Goele, Karen, Marit, Lise, Sarah, Elise, Oli, Josip, Timo and Thomas.

Finally, the last words of appreciation go out to my father and mother who have always given me their unconditional love, help and encouragement in everything I do. I thank them with all my heart, though any words of expressing my gratitude for what they have done for me and my brothers would fall short.

Pim Verbunt

Brussels, May 2019

Contents

List of Tables

List of Figures

General Introduction

Social inclusion and social exclusion in the European Union

The overall aim of this dissertation is to improve the measurement and understanding of social inclusion and social exclusion in the European Union (EU). Before discussing the research objectives pursued in this dissertation in more detail, it is important to note that the concepts "social inclusion" and "social exclusion" will be used simultaneously and interchangeably with concepts such as "quality-of-life", "well-being", "poverty" and "deprivation" throughout this dissertation. These terms are used in a wide variety of different ways and fields and there seems to be a lack of consensus on an exact definition. Therefore, in line with the objective of contributing to a better measurement and understanding of social inclusion and exclusion in the European Union, I follow Atkinson, Cantillon, Marlier and Nolan (2002, p. 3) and "simply accept the use of the terms "social inclusion" and "social exclusion" as shorthand for a range of concerns considered to be important for the European social agenda". The authors further note that "there is broad agreement about the list of such concerns, which encompass poverty, deprivation, low educational qualifications, labour market disadvantages, joblessness, poor health, poor housing or homelessness, (digital) illiteracy and innumeracy, precariousness, and incapacity to participate in society" (Atkinson et al., 2002, p. 3). In point of fact, the European Union has consistently sought to broaden the concept of social exclusion, stressing its multidimensional nature and drawing a contrast with the sole reliance on a monetary poverty indicator. In its 1992 submission on 'intensifying the fight against social exclusion', the Commission stated of social exclusion that, "more clearly than the concept of poverty, understood far often as referring exclusively to income, it also states out the multidimensional nature of the mechanism whereby individuals and groups are excluded from taking part in the social exchanges, from the component practices and rights of social integration" (European Commission, 1992, p. 8; Atkinson et al., 2002, p. 33).

Given that the focus of this dissertation lies on the measurement of social inclusion and social exclusion in the European Union, it is important to first discuss the particularities of EU social policy setting. Marlier, Atkinson, Cantillon and Nolan (2007, p. 17) note that "even though the founding fathers of the EU had expected social progress to evolve naturally from the economic progress generated by the Common Market, for many years the Single European Market and the European Monetary Union largely eclipsed the social dimension of the EU". The subsidiarity principle is key to understand the division of competences within the EU social policy setting. According to this principle, the European Union can take action only if, and in so far as, the objectives of the proposed action cannot be sufficiently achieved by the Member States. This means that the policies to achieve social inclusion are first and foremost the responsibility of the Member States (Atkinson et al., 2002). The actions that are undertaken at the EU level to combat social exclusion, such as the Regional Development Fund and the Social Fund, are limited in both scale and scope.

Social policy has truly become a specific focus of attention for EU cooperation with the adaptation of the Lisbon Strategy in 2000 (Marlier et al., 2007, p.17). The incorporation of the promotion of social inclusion received a prominent position within the Lisbon Strategy. It was agreed upon to advance social policy on the basis of the open method of coordination (OMC) at the European Union level. In essence, the OMC is a mutual feedback process of planning, monitoring, examination, comparison and adjustment of national (and sub-national) policies, all of this on the basis of common objectives agreed for the EU as a whole. The so-called Laeken indicators were given an explicit role in the representation of the EU's common goals on social inclusion. The Laeken indicators are a set of social indicators that represent, in effect a toolbox of instruments that allows Member States to use a common language for assessment in the area of poverty and social exclusion (Atkinson et al., 2002, p. viii). Member States, while agreeing on the indicators by which performance is to be judged, are left to choose the methods by which these objective are realised (under the subsidiarity principle) (Marlier et al., 2007, p. 40). Some Member States perform better in one area, and others in a different area, so that there is a great deal of scope for mutual learning. The key supranational goal, however, is fundamentally the same for everyone: progress towards more social inclusion and social quality (Atkinson et al., 2002, p. viii).

This dissertation is a collection of four chapters. The research objectives pursued in this dissertation are dual. The overarching research objective of Chapter 1 and Chapter 2 is to obtain a better conceptual understanding of key European Union indicators of social exclusion. The overarching research objective of Chapter 3 and Chapter 4 is to present new methodologies that embrace the multidimensional nature of social inclusion in its measurement.

Chapter 1: Towards a better understanding of the Europe 2020 poverty and social exclusion indicator

In Chapter 1, the focus lies on the EU 2020 social inclusion target group. The EU 2020 social inclusion target quantifies the social inclusion goal of the ten-year EU 2020 strategy that aims to reduce the number of European citizens living in poverty or social exclusion by 20 million by 2020. It is the first quantitative social target agreed for the whole of the EU. The EU 2020 social exclusion target group is identified by meeting any of the following criteria: being at-risk-of-poverty or severely materially deprived or living in a jobless household. The at-risk-of poverty approach is the most common and traditional way to measure poverty in European countries. The indicator has been around for many decades. It represents an indirect and relative approach to poverty measurement: households are considered poor if their equivalised disposable income falls below the income poverty threshold, which is expressed as a share (60% per cent) of the national median income level. Material deprivation, the second criterion of the social exclusion target group, is a relative newcomer in the social policy constellation at the national and EU level (Copeland & Daly, 2012, p. 281). The material deprivation indicator was added to the EU portfolio of commonly agreed indicators in 2009, to complement income poverty figures and to better reflect differences in living standards in the EU after the enlargements (Guio, 2009). It takes a direct and absolute approach to poverty measurement: households are asked directly whether they can afford certain items that are deemed necessary for a comfortable material life. The deprivation threshold is the same for the whole of the EU. The third indicator used in the Europe 2020 social exclusion target group, viz., quasi-joblessness regroups people living in households that are jobless or very work poor. The inclusion of this indicator in the target group has not been evaluated positively by researchers. It is not always clear why jobless households that are not income poor, nor materially deprived should be included in the target population (Nolan & Whelan, 2011, p. 18). Living in a jobless household is rather seen as a risk factor that has a direct impact on the other two social exclusion indicators (De Graaf-Zijl & Nolan, 2011). A more precise definition of the three indicators is given in Section 1.2.1.

The three indicators reflect “the multiple facets of poverty and social exclusion across Europe”, but also “the diversity of situations and priorities across Member States” (European Commission, 2011, p. 9). In fact, under the principle of subsidiarity, Member States can set their national targets on the basis of one of the three indicators, a combination of all three indicators or suggest their own indicator(s). The majority of the Member States have used the same definition as the EU headline target, while others have set their target on the basis of national indicators in order to better reflect their national circumstances (e.g. Germany uses the number of long-term unemployed, while the United Kingdom uses the previously existing child poverty target) (European Commission, 2011).

Atkinson and co-authors (2010, p. 129) warn that “progress in terms of combating poverty and social exclusion will depend very much on the extent to which the chosen policies are directed at households where the criteria overlap”. Some targets can even be in conflict with each other, as policies targeting one indicator can worsen the situation according to the other indicators. Individuals that are both income poor and severely materially deprived are therefore often considered as a priority group (Nolan & Whelan, 2011). Households that are income poor are, however, not necessarily also severely materially deprived, and vice versa. In 2017, 101.1 million EU citizens (20.1 per cent of the population) were either at-risk-of poverty or severely materially deprived (89.4 per cent of the total social inclusion target group). Within this group, 85.3 million EU citizens (or 16.9 per cent of the population) were at-risk-of poverty, 33.1 million EU citizens were severely materially deprived (or 6.6 per cent of the population) and 17.3 million people were both income poor and severely materially deprived (3.4 per cent of the population). In sum, around half (52.3 per cent) of the severely materially deprived are also income poor and around a fifth (20.3 per cent) of the income poor are also severely materially deprived.

Chapter 1 assesses to what extent the risk of income poverty and severe material deprivation are shaped by the same drivers. In this regard, it is important to distinguish the drivers situated at the individual or household-level and drivers situated at the country-level. The income-deprivation relationship at the *individual-level* has been analysed along different perspectives. Some studies look at the 'causal' role of individual income as one of the determinants of material deprivation (Whelan et al., 2001; Whelan & Maître, 2007; Berthoud & Bryan, 2011). For instance, Whelan and co-authors (2002) point out that persistent income poverty (over time) is a key predictor of persistent deprivation, alongside a variety of resource related variables such as education, labour market experience and social class, and needs related variables such as marital status and household structure. Another approach consists in identifying the socio-economic characteristics of the consistently poor (Bradshaw & Finch, 2003; Lollivier & Verger, 1998; Guio et al., 2010). For instance, Fusco, Guio and Marlier (2011) compared the socio-economic characteristics of the income poor and/or materially deprived through the application of multinomial logistic regressions for 25 European countries (separately). Other studies focus on the *drivers situated at the country-level* that can explain the existence of income poverty or material deprivation. Broadly speaking, there are two types of approaches. A first line of research setups use panel studies with cross-national and longitudinal variation to determine income poverty rates (Moller et al., 2003; Brady, 2005; Brady & Kall, 2008; Diris et al., 2017). Panel studies on material deprivation rates seem, to best of my knowledge, inexistent. A second approach is the so-called "multilevel setup". These type of models regress both individual or household-level ('micro') and country-level ('macro') variables to account for cross-national variations in social exclusion.

This chapter follows the multilevel approach and jointly assesses the drivers that are situated at the household-level and country-level for both the risk of income poverty and severe material deprivation. The aim is to provide a comprehensive understanding of the policy levers that should be mobilised to fight social exclusion in the European Union. A multilevel multinomial regression setup, which allows to explain differences between those who suffer from income poverty 'only', those who suffer from severe material deprivation 'only', those who suffer from both problems and those who suffer from none, is employed. Chapter 1 is the first multilevel study that addresses the drivers of income poverty and material deprivation in a single model.[1] The chapter further contributes to the literature by testing and comparing the relationship of

[1] Saltkjel and Malmberg-Heimonen (2016) also consider the determinants of the risk of income poverty and severe material deprivation in a multilevel setting, but employ two separate binary logistic regression models.

different social spending concepts with deprivation and income poverty that have not been addressed in previous multilevel studies (such as the in-kind versus in-cash dichotomy within social spending and the size versus pro-poorness of social spending). The analysis is conducted on the EU Statistics on Income and Living Conditions (EU-SILC) 2012 cross-sectional dataset.

Chapter 1 also presents new innovative methodological strategies. First, for the first time in the literature, the single level and multilevel regression approaches are combined to model the risk of social exclusion. It shows that both types of models are needed to get a comprehensive and precise picture of the drivers of social exclusion. Single level models make it possible to identify specific national risk factors and offer a better understanding of within-country variations in the explanatory power of household determinants on social exclusion (as coefficients are by definition allowed to vary in each country, national specificities with regard to micro-drivers are better captured). The advantage of multilevel models is that they allow for a better understanding of the cross-national variations in social exclusion. Second, the dual use of single level and multilevel models is further exploited by calculating within and between-country explained variance measures, which allows to compare the explanatory of the employed model across countries and specifications. Third, the usual econometric approach of identifying significant relationships is complemented by decomposing these explained variance measures. This way, the issue of non-comparability of estimated coefficients in logistic models is circumvented.[2] In addition, quantified knowledge on how (in)effective the different household-level and country-level variables are in explaining the risk of material deprivation and income poverty is obtained.

Chapter 2: Towards a better understanding of the EU child deprivation indicator

In Chapter 2, the focus lies on a new official EU child deprivation indicator. Fighting child poverty and investing in children's well-being has featured on the agenda of the European Union for many years. In February 2013, a new step forward was taken when the European Commission published a Recommendation on "Investing in children: breaking the cycle of disadvantage" (European Commission, 2013a), subsequently adopted by the EU Council of Ministers. The Recommendation advocates that "preventing the transmission of disadvantage across generations is a crucial investment in Europe's future" and that "early intervention and prevention are essential for developing more effective and efficient policies, as public expenditure addressing the consequences of child poverty and social exclusion tends to be greater than that needed for intervening at an early age". Indeed, the literature on the relationship between family circumstances and future outcomes shows that the family that one is born into

[2] Mood (2010) points out that because of unobserved heterogeneity, the estimated coefficients in logistic regression models cannot be compared across model specifications, groups, samples or time points.

matters greatly for success in life.[3] High (income) poverty among children is especially alarming, as there are strong links between family resources early in a child's life and later school and labour market performance.[4] Diris and Vandenbroucke (2016) show that those children growing up in material deprivation have substantially lower chances of obtaining favourable later-life outcomes, even though the causal relationship weakens when other aspects of family background are controlled for.

Another key element of the EU Recommendation is that it calls on Member States to "(reinforce) statistical capacity where needed and feasible, particularly concerning child deprivation". A first significant step taken in this direction was the EU Luxembourg Presidency in the first half of 2005. It was argued that simple age group breakdowns of EU social indicators were insufficient to adequately capture the child-specific material and social living conditions, which may differ from those of their family. The 2009 wave of EU-SILC included an ad hoc module aimed at collecting such information. Guio and co-authors (2012) carried out a first in-depth analysis of these data. The authors identified an optimal set of children's deprivation items and proposed a child deprivation index. These items were then included again in the 2014 EU-SILC ad hoc module on deprivation, allowing additional analysis by Guio and co-authors (2018). In March 2018, a child-specific deprivation indicator was adopted at the EU level, following the work of Guio and co-authors (2018). The indicator consists of 17 items, covering both material and social aspects of deprivation, which are aggregated in a child-specific deprivation scale. The child-specific deprivation variables will be collected in the EU-SILC ad hoc module every three years, with the aim of measuring and monitoring child deprivation in a robust and comparative way for the whole EU.

Chapter 2 analyses the drivers of the new EU child deprivation indicator using the EU-SILC 2014 cross-sectional dataset. A similar methodological strategy as the one followed in Chapter 1 is employed in that both single level and multilevel regression models (with accompanying decompositions of their within- and between-country explained variance measures) are used to assess the drivers of (*in casu*, child-specific) social exclusion. Given the count nature of the child deprivation indicator, the methodological difference with Chapter 1 lies in the fact that a negative binomial regression model is employed. The choice for the negative binomial regression is motivated by the fact that the accumulation of deprivations suffered by children can be investigated. However, a key difference with the first chapter is that special attention is given to the role of income, measured both at the household-level and country-level, as a driver of child deprivation, but also as a control variable for other social spending policy indicators. In Chapter 1, (low relative) income is part of the dependent variable. It will be shown that the inclusion (or

[3] See Corak (2013) for an overview of the intergenerational transmission of income.

[4] See Brooks-Gunn and Duncan (1997) for an overview of studies that specifically focus on income poverty in relation to child outcomes. The authors conclude that poverty early in life (preschool and early school years) is most strongly related to important future outcomes.

exclusion) of household income in the model has important consequences on the relationship of the other independent variables in the model. This strategy allows the confrontation of diverging results reported in the material deprivation literature and suggests reasons why some macro-variables (do not) have an impact on between-country differences in child deprivation. A second contribution is that a robust analysis is presented in that many relationships found in the material deprivation and child deprivation literature are replicated and presented jointly. In total, more than ten country-level variables are considered.

Chapter 3: Aggregating social inclusion performances in the European Union

The at-risk-of poverty or social exclusion indicator and the child deprivation indicator, the indicators analysed in Chapter 1 and Chapter 2, are mere examples of new social indicators at the EU level. In recent years, a substantive amount of EU indicators, covering a diverse set of topics and objectives in the areas of social inclusion and protection, have been developed. This has been one of the main achievements of the social Open Method of Coordination. The list of indicators is continuously being improved as statistics, data collection and policy needs evolve.

Since the sovereign debt crisis, social indicators are increasingly used by the European Commission as a tool to monitor social performances of Member States. For instance, since 2014, the Commission uses a scoreboard of key indicators in its draft Joint Employment Report to follow employment and social developments. The scoreboard serves as an analytical tool allowing better and earlier identification of major employment and social problems, especially any that risk generating effects beyond national borders (European Commission, 2013b). In addition, a set of auxiliary social and employment indicators has been included in the Alert Mechanism Report and In-Depth Reviews of the Macroeconomic Imbalance Procedure, "to better integrate the social implications of imbalances in the current framework of macroeconomic surveillance" (European Commission, 2013b, p.5). Next, a new monitoring and accountability scheme, the "Social Protection Performance Monitor" (SPPM) is put forward to identify annual "social trends to watch" and "positive recent social trends" in the EU Member States and in the EU-region as a whole. Finally, in 2017, the European Commission launched the European Pillar of Social Rights, with the aim of strengthening the social dimension within the EU by delivering new and more effective rights for EU citizens. The European Pillar of Social Rights is accompanied by a 'Social Scoreboard' which monitors trends and performances across EU countries in three areas related to the principles under the Pillar (i.e. equal opportunities and access to the labour market, dynamic labour markets and fair working conditions, public support / social protection and inclusion). The Scoreboard feeds into the European Semester of economic policy coordination. It uses 14 headline indicators which are since 2018 used to compare Member States' performances in the annual Joint Employment Report.

The richness of these scoreboards or dashboards of social indicators is often simultaneously considered as a strength and a drawback. In particular, whereas they offer policy makers and stakeholders a lot of policy information, they are often prone to the criticism that they deliver too much information to be efficient communication tools, even when their main messages are summed up in a limited set of headline indicators (Fleurbaey & Blanchet, 2013, p.33). Composite indicators are a popular alternative to dashboard. They provide a more parsimonious description of policy situations by aggregating performances on multiple indicators. Fleubaey and Blanchet (2013, p. 34) point out that single measures have a natural power of attraction against which excessively rich sets of detailed indicators cannot compete. Composite indicators are renowned for their ability to draw public attention by giving "the bigger picture" and by ranking performances (the "hit-parade phenomenon"). They are frequently used by important international organizations such as the United Nations, the OECD, the World Economic Forum, and the European Commission in wide ranging fields such as the economy (e.g. the Internal Market Index, the Competitiveness Index), the knowledge economy (e.g. the Knowledge Economy Indicator), human development (e.g. the Human Development Index), technological development (e.g. the Technology Achievement Index), health (e.g. the Health System Performance Index), and the environment (the Environmental Performance Index and the Environmental Vulnerability Index)" (Rogge, 2012, p. 143).

Composite indicators remain, however, controversial. Its opponents argue that they "summarise too much and communicate less than the separate sub-indicators" (Micklewright, 2002, p. 47). For example, a sharp change in a sub-indicator can remained concealed in the index value, as long as it is offset by an opposing change in a different sub-indicator. In addition, given that composite indicators require a large amount of modelling option and choices, they are often prone to the criticism that they are "exercises in measurement without theory". Recurring and interrelated concerns pertaining to the construction of composite indicators are (i) the selection of specific sub-indicators, (ii) the fact that these sub-indicators may have different measurement units in their original form, (iii) the importance one should attach to a sub-indicator in terms of its contribution to the final composite construct, (iv) the functional form of the aggregator function, (v) robustness of the obtained results to some of these modelling options, etc. (Van Puyenbroeck & Rogge, 2017; for a more elaborate discussion, see Nardo et al., 2005). Each of these steps involve value judgement that are necessarily controversial (Stiglitz et al., 2009, p. 207).

In Chapter 3, a new extension of the "Benefit-of-the-Doubt" method for constructing composite indicators is proposed. Before explaining the methodological contribution of the extension in more detail, and, more in particular, why I believe the model presented in this chapter is appropriate to aggregate social performances of EU Member States, I first discuss some key properties of the original “Benefit-of-the-Doubt” method and some of its extensions presented in the recent literature. The “Benefit-of-the-Doubt” methodology has become an accepted method in academics and international organizations for the construction of composite indicators (see Melyn and Moesen (1991) for the first (and most) basic Benefit-of-the-Doubt model and see Cherchye and co-authors (2007) for an intuitive introduction to the method). The flexibility and the optimistic stance in the determination of the weights is often praised as the most important advantage of the Benefit-of-the-Doubt method. In a setting in which objective knowledge on the true policy weights is usually lacking or incomplete, the Benefit-of-the-Doubt model derives for each country the set of optimal weights from the observed sub-indicator values themselves. More in particular, the Benefit-of-the-Doubt model defines importance weights for each country such that the impact of sub-indicators of relative strength is maximized and the impact of sub-indicators of relative weakness is minimized in the composite value. This quality explains much of the appeal of the Benefit-of-the-Doubt model: in what is usually a sensitive evaluation environment, disappointed countries can no longer blame a low composite indicator score on damaging or unfair weights. Any other weighting scheme than the one specified by the Benefit-of-the-Doubt model would worsen the score. Another key property of the original Benefit-of-the-Doubt model is that it is unit invariant, i.e. composite indicator scores are independent of the units in which the factors are measured.

The original “Benefit-of-the-Doubt” composite indicator model uses a linear aggregation function and optimistic weights (as described above). Several extensions of the method have been proposed in recent years. Two of these extensions are crucial to understand the methodological contribution of the new “Benefit-of-the-Doubt” model presented in Chapter 3. These are: the “Benefit-of-the-Doubt” model that uses pessimistic weights and the “Benefit-of-the-Doubt” model that uses a multiplicative aggregation function. The conceptual starting point of the pessimistic “Benefit-of-the-Doubt” counterpart is opposite to the one of the traditional Benefit-of-the-Doubt model in that the model computes weights that give a high (low) weights to sub-indicators on which the evaluated country performs relatively weakly (strongly), as compared to the other countries in the sample (Zhou et al., 2007). Specifically, the pessimistic Benefit-of-the-Doubt model assesses the policy performance of countries under a ‘worst-case’ evaluation scenario, in which weights for the sub-indicators are defined such that the composite value of each country is minimized relative to the other countries.

The other extension that is relevant to understand the methodological contribution presented in Chapter 3 is the “Benefit-of-the-Doubt” model that uses a multiplicative aggregation function. An interesting property of the multiplicative aggregation function is that it, contrary to their linear equivalents, penalizes inequality among sub-indicators and that poor performances on one sub-indicator cannot be fully compensated by sufficiently high values on other sub-indicators.[5] There are currently three "Benefit-of-the-Doubt" composite indicator models that employ a multiplicative aggregation function, which can, in turn, be clustered into two types of models. A first class of models use a direct approach, in the sense that they are based on programming problems looking for optimal weights such that the multiplicative composite indicator is maximized (Zhou et al., 2010; Tofallis, 2014). Both models essentially transform the multiplicative optimization problem to a linear equivalent by applying logarithmic transformations on the sub-indicators. However, both models have some unappealing drawbacks. The model of Zhou and co-authors (2010) is not unit invariant, whereas the model of Tofallis (2014) violates the linear homogeneity property (i.e. linear homogeneity imposes that a one per cent increase in all sub-indicators increases the composite value with one per cent). These drawbacks were a key motivation of Van Puyenbroeck and Rogge (2017) to present a third multiplicative Benefit-of-the-Doubt model, which differs from the other two in that uses a two-step approach. Specifically, in a first step, weights of the different sub-indicators are estimated using the original Benefit-of-the-Doubt model and, in a second step, these Benefit-of-the-Doubt-derived importance weights are used in a geometric mean quantity index. In other words, given that 'true' importance weights of the sub-indicator within a (geometric mean quantity) composite index are unknown, the optimal weights derived from the linear Benefit-of-the-Doubt serve as shadow price information.

The methodological contribution of Chapter 3 lies in the fact that a holistic framework that integrates optimistic Benefit-of-the-Doubt-based weighting jointly with pessimistic Benefit-of-the-Doubt-based weighting within a geometric mean quantity index, is presented. Although the proposed measurement framework is generic in the sense that it can be used to summarise performances in *any* policy setting, I believe the joint use of country-specific optimistic and pessimistic weights is particularly relevant for the EU social policy setting. As noted by Atkinson and co-authors (2002, p. 72), "in the context of the EU there are evident difficulties in reaching agreement on weights [within a composite indicator], given that each Member State has its own national specificity". That is, as nicely put by Rogge and Van Nijverseel (2018, p. 3): "there is a broadly shared EU-wide concern to improve the social inclusion in the EU and, at the same time, the different traditions and instruments to achieve this goal being, under the subsidiarity principle, still largely situated at the national level. The idea of using a weighting approach that imposes equality of weights across domains and/or countries (e.g., equal weighting, fixed weighting or other statistical approaches) is undesirably and unnecessarily restrictive".

[5] The 2010 change in the construction of UNDP's Human Development Index is a well-known illustration of recognizing and dealing with this issue (Van Puyenbroeck & Rogge, 2017, p. 1006). See, however, Ravallion (2010) for a critical review of the HDI's change in aggregation function.

For Cherchye and co-authors (2004) the presence of the subsidiarity principle in the field of European social policy making provides a prima facie reason for using country specific optimistic Benefit-of-the-Doubt weights.[6] The Benefit-of-the-Doubt model allows countries to retain some degree of flexibility in the design of policy instruments and measures and, thus, avoids penalizing countries for pursuing particular policy objectives, at the acknowledged expense of another conflicting objective. Such a data-oriented weighting method is justifiable in the typical composite indicator context of uncertainty about, and lack of consensus on, an appropriate weighting scheme. The favourable weighting approach inherent in the Benefit-of-the-Doubt model offers an appealing flexible alternative (Rogge & Nijverseel, 2018). That being said, given that countries' weakest social performances are typically ignored when optimistic weights are employed, solely relying on optimistic weights may, for the specific case of social inclusion, be somehow difficult to defend from a normative point of view. Instead, if one is concerned with poor social outcomes, in a Rawlsian spirit, one could argue to give a higher weight to the sub-indicators on which the country performs more poorly. This objective is perused in the Chapter 3. Specifically, a pessimistic version of the "Benefit-of-the-Doubt" model is employed to compute shadow price weights that reflect countries' weakest social performances.

Regarding this weighting debate (i.e., should optimistic weighting intrinsically be preferred over pessimistic weighting in the construction of a composite indicator for social inclusion, or vice versa), it is argued that both optimistic and pessimistic weighting schemes are justifiable, following fundamentally different intuitions. While both models interpret comparative performances as a revealed evidence of policy priorities, the pessimistic model attempts to incentivize policy makers to prioritize on the dimensions in which the country lags behind, whereas the optimistic model stipulates that the weights should explicitly reflect a country's policy priorities (i.e., the dimensions in which the country is doing well relative to the other countries). Optimistic weights can thus be defended in light of the subsidiarity principle, whereas pessimistic weights may be more appealing from a normative point of view.

Regardless of the position one takes in the optimistic versus pessimistic weighting debate, it is shown in Chapter 3 that synthesized, yet detailed information that may be useful for monitoring specific policy performance developments can be obtained, if both weighting schemes are integrated in the geometric mean quantity index. This way the loss of information that is inherent in the construction of composite indicators, a key conceptual criticism on the approach, can be, at least partially, mitigated. First, the inter-temporal, multi-factor decomposition of Van Puyenbroeck and Rogge's (2017) geometric composite indicator framework is adjusted such that a comprehensive and nuanced view on changes in countries' policy performances is provided.

[6] For other applications of the Benefit-of-the-Doubt methodology on the EU social indicators, see Giambona & Vassallo, 2014; Rogge & Konttinen, 2018; Rogge, 2017; Rogge & Self, 2018; Rogge & Van Nijverseel, 2018.

Second, a measure that assesses the degree of unbalance in a country's policy portfolio mix is proposed. This measure gives an idea about the inequality in performances across social policy dimensions on which the country holds a comparative social advantage and a comparative social disadvantage. A decomposition which allows to account for (changes in) the measure is further presented.

The proposed method is illustrated on commonly agreed EU indicators (period 2008–2013) from the overarching portfolio of social protection and social inclusion. The overarching EU objectives of social protection and social inclusion are to promote (i) social cohesion, equality between men and women and equal opportunities for all through adequate, accessible, financially sustainable, adaptable and efficient social protection systems and social inclusion policies, (ii) effective and mutual interaction between the Europe 2020 objectives of smart, sustainable and inclusive growth, taking full account of the relevant social provisions of the Lisbon Treaty, (iii) good governance, transparency and the involvement of stakeholders in the design, implementation and monitoring of policy (European Commission, 2015, p. 8). The nine overarching commonly agreed EU social inclusion indicators are: (i) at risk of poverty or social exclusion rate, (ii) relative median poverty risk gap, (iii) income quintile ratio (S80/S20), (iv) early school leavers, (v) aggregate replacement ratio, (vi) at-risk-of-poverty rate anchored at a fixed moment in time (2008), (vii) employment rate of older workers, (viii) in work at-risk-of poverty rate, and (ix) activity rate.

Chapter 4: The capability approach

In Chapter 4, a new measurement framework is presented that gauges social inclusion outcomes in terms of the capability approach. The capability approach is an economic theory developed by Amartya Sen as an alternative to welfare economics. The traditional economic notions of commodity and utility are replaced, respectively, with functioning and capability. A functioning vector is a description of the state of "being and doings" on various well-being dimensions. A capability set is the set of all the potential feasible functioning vectors from which an individual can choose. The capability approach conceives a person's well-being in terms of his/her freedom to choose various functionings (Sen, 1985, 1988, 1993; Nussbaum & Sen, 1993). The objective is the enlargement of the capabilities or opportunity freedoms of individuals. Limited choice among functionings can lead to capability deprivation and, in extremis, social exclusion (Papadopoulos & Tsakloglou, 2008).

Chapter 4 tackles two key issues raised to the empirical operationalization of the capability approach. These issues are: the unobservability of capability sets and the critique that the capability approach does not pay sufficient attention to groups. Before discussing the contribution of the chapter to the literature, I first discuss these issues in more detail. First, a longstanding difficulty in the empirical operationalization of the Capability Approach is that countries' or individuals' capabilities (capability sets) cannot be measured or observed. As a consequence, and despite of the fact that capabilities are often considered as a conceptually superior metric to measure well-being as compared to functionings, none of the empirical applications of the capability approach actually employ capability sets to make well-being assessments. Instead, most empirical capability approach studies pragmatically use functionings to make such evaluations. Social indicators that measure the average (achieved) functioning performance (e.g. average life expectancy, employment rate, etc.) are often seen as a (scalar-based) capability measure.

Second, the capability approach has been criticized of being an exclusively individual approach and not paying enough attention to groups (Burchardt & Vizard, 2005; Stewart, 2005, Steward, 2008; Alkire, 2008). These authors call attention to the importance that groups may have in individual well-being, in shaping individual preferences, and in generating social mobilization and collective action (Roche, 2009). In response to this criticism, Robeyns (2018) argues that groups and social structures can easily be accounted for in the capability approach, though she acknowledges that the role of groups should receive more consideration. She cites several studies that analyse the average functionings of one group compared to another, e.g. women and men (Kynch & Sen, 1983; Nussbaum, 2000; Robeyns, 2003, 2006) or the disabled versus those without disabilities (Kuklys, 2005; Zaidi & Burchardt, 2005). She also refers to the fact that the UNDP (1995, 2004) has produced Human Development Reports based on the capability approach with a focus on both gender and culture.

To take the first issue into account, a method is proposed to non-parametrically estimate capability sets at a group-level from achieved functioning bundles of group members. Similar to the social indicator approach, the proposed method builds on the idea that the observed functioning achievements of group members entail information on the capabilities of a group. However, the main difference lies in the fact that the average functioning achievements of group members is not equated with a (scalar) measure of group-level capabilities. Instead, the proposed method estimates the capability sets of groups by enveloping the observed functioning bundles of group members. This way, the original conceptualization of the capability approach that well-being should be gauged by capability *sets* is maintained. The second issue is taken into account by presenting a framework to evaluate the estimated group capability sets. The objective of this evaluation exercise is to quantify the extent to which groups differ in their ability of generating capabilities.

Similar to the methodology presented in Chapter 3, the methodology presented in Chapter 4 is generic in the sense that it can be employed to make well-being assessments in any setting. Given that the focus of this dissertation lies on the measurement of social inclusion and exclusion in the European Union, the approach is illustrated on European social inclusion data in the EU-SILC 2013 cross-sectional dataset. In particular, the capability sets of 32 European countries are estimated and evaluated, considering four functioning dimensions (income, health, housing quality and material living conditions). The evaluation exercise consists in the comparison of the extent of the estimated country-specific capability sets.

Chapter 1

Explaining differences within and between countries in the risk of income poverty and severe material deprivation: comparing single and multilevel analyses

(This chapter is published in *Social Indicators Research*; co-author: Anne-Catherine Guio)

1.1. Introduction

In the context of the Europe 2020 strategy, the European Union agreed – for the first time in its history – on a EU-wide poverty reduction target. The Europe 2020 strategy is a ten-year strategy that was proposed by the European Commission in 2010 to obtain smart, sustainable and inclusive growth. To measure the progress in reaching the Europe 2020 goals, five headline targets on employment, R&D, climate change, education and social exclusion have been agreed upon for the whole of the EU. The social exclusion target aims at lifting at least 20 million people out of the risk of poverty and social exclusion by 2020. The target group is made up of those meeting any of the following criteria: being at-risk-of-poverty, being severely materially deprived or living in (quasi-)jobless households.

The main goal of this study is to investigate the differences between the risk factors of income poverty and severe material deprivation by assessing to what extent both indicators are subject to the same determinants. Given that these two indicators encompass the majority of the Europe 2020 social exclusion target group (i.e. 89 % of the targeted population at the EU level), it is crucial for policy makers to better understand which determinants are effective in explaining the differences within and between countries in the risk of income poverty and severe material deprivation.

We will attempt to address several gaps in the current literature. Firstly, this is the *first multinomial logistic multilevel study to simultaneously address the risk factors of income poverty and material deprivation*. Fusco, Guio and Marlier (2011) compared the socio-economic characteristics of the income poor and/or materially deprived through the application of multinomial logistic regressions for 25 European countries separately. In this chapter, we extend their analysis by pooling all countries together and adding a multilevel structure. Specifically, we employ single level and multilevel multinomial logistic regression models, differentiating, with regard to the dependent variable, between three groups, namely the 'income poor only', the 'materially deprived only' and the 'consistently poor' (i.e. those being both income poor and materially deprived). We apply our method to the cross-sectional European Union Statistics on Income and Living Conditions (EU-SILC) for the year 2012.

Secondly, we *combine for the first time both a single level and multilevel regression approach* to predict the risk of income poverty and material deprivation. In the existing literature there is a distinction between the so-called "micro-level" and "macro-level" approaches. In the micro-level approach, the focus is on the role of socio-economic characteristics at individual or household-level to explain the prevalence of income poverty and/or material deprivation within countries.[7] For the macro-level approach, several authors focus on contextual differences to determine income poverty rates.[8] Multilevel studies stress the role of both individual and country characteristics in explaining the risk of either income poverty or material deprivation.[9] A main advantage of single level models is that they allow identifying specific national risk factors and offer a better understanding of within-country variations of the impact of household determinants. A main advantage of multilevel modelling is that it allows the investigation of cross-national variations in the risk income poverty and/or severe material deprivation within the EU pooled dataset. However, its drawbacks come from the fact that it does not allow differentiating the impact of micro-drivers at the national level, but only takes into account the national differences in the composition of individual or household-level risk factors. The comparative use of single level and multilevel models allows for the first time to confront the respective within- and between-country explanatory power of both types of models. This chapter will show that both types of models should be combined to offer a comprehensive understanding of the policy drivers needed to reach the EU social exclusion target in each EU country.

Thirdly, there is a lack of a *quantified knowledge on how (in)effective the different household-level and country-level variables are* in explaining the risk of material deprivation and income poverty. Usually, econometric models are used to identify significant relations and their sign. In our case, we go one step further and employ Shapley decompositions to establish the relative contribution of each explanatory variable to within- and between-country explained variance measures for the risk of income poverty and/or material deprivation (Shapley, 1953; McKelvey & Zainova, 1975; Snijders & Bosker, 2012).

[7] See Whelan et al., 2003; Whelan et al., 2004; Guio et al., 2010; Fusco et al., 2011.
[8] See Moller et al., 2003; Brady, 2005; Brady & Kall, 2008; Diris et al., 2017.
[9] For the risk of income poverty see Wiepking & Maas, 2005; Brady et al., 2009; Lohmann, 2009; Reinstadler & Ray, 2010; Bäckman & Ferrarini 2010; Chzhen & Bradshaw, 2012; Chzen, 2014; Saltkjel & Malmber-Heimonen, 2016. For the risk of material deprivation, see Kim et al., 2010; Nelson, 2012; Whelan & Maître, 2012; Whelan & Maître; 2013; Bárcena-Martín et al., 2014; Visser et al., 2014; Saltkjel & Malmber-Heimonen, 2016; Bárcena-Martín et al., 2017b.

Finally, we also *test and compare the impact of different social spending concepts* on the deprivation and income poverty risks, which have not been addressed in previous multilevel studies. Indeed, while several multilevel studies have indicated that income poverty and material deprivation are shaped by common institutional determinants related to the welfare state, there is a lack of knowledge regarding the impact of different types of social spending on both risks. Firstly, we pick up on the in-kind versus in-cash social spending dichotomy, as research has shown that these social spending components may have a different distributional impact (OECD, 2008; OECD, 2011). Secondly, given that the question around the optimal degree of universalism and targeting is still open to debate, we investigate whether the size or the pro-poorness of social spending is the most effective in levels in explaining country-to-country differences in the risk of income poverty and severe material deprivation.[10] Thirdly, we investigate some diverging results from the literature regarding the relative importance of institutional and macroeconomic variables in explaining country differences in the risk of severe material deprivation. That is, while Kenworthy and co-authors (2011) showed that material deprivation rates are correlated to a country's social policy generosity, but not to its level of affluence, other researchers have found a negative relationship between aggregate income levels and material deprivation (Whelan & Maître, 2012; Bárcena-Martín et al., 2014; Visser et al., 2014).

This chapter is structured as follows. Section 1.2 defines the social exclusion target group. Section 1.3 explains the conceptual differences between income poverty and material deprivation by discussing the within- and between-country determinants. Section 1.4 discusses the methodological approach. The results of our empirical analysis are presented in Section 1.5. Section 1.6 concludes.

1.2. Social exclusion target group

1.2.1. Definitions

In 1975 the EU Council of Ministers defined poverty as "individuals or families whose resources are so small as to exclude them from the minimum acceptable way of life of the Member State in which they live" (Council of the European Communities, 1975). Since 2001, the EU has a growing portfolio of commonly agreed social indicator to measure Member States' progress in the fight against poverty and social inclusion. In 2010, the European Union agreed on an EU-wide poverty reduction target in the context of the Europe 2020 strategy. This target is the union of three indicators which were part of the EU portfolio of social indicators:

[10] See Bárcena-Martín and co-authors (2018) for a first multilevel study investigating the relationship between benefit targeting and child poverty.

The *at-risk-of poverty indicator* ("IP") is the first criterion of the social exclusion target group. Specifically, the at-risk-of poverty rate refers to the situation of people whose disposable income is below 60 per cent of the median national income. It is a purely relative measure as the poverty line varies from country to country ("… of the Member State in which they live"). It is connected to inequality, although the main focus lies on the lower end of the income distribution. The indicator is an indirect approach to measuring poverty, since the focus lies on inputs or the means available to the household ("… due to a lack of resources").

(Severe) Material deprivation, the second criterion of the social exclusion target group, is a relative newcomer in the social policy constellation at a national and EU level (Copeland & Daly, 2012, p. 281). A material deprivation indicator was included in the EU portfolio of commonly agreed indicators in 2009, to better reflect differences in living standards in the EU after the enlargements (see Guio, 2009). The material deprivation ("MD") indicator is based on the enforced lack of the following items: (1) to pay their rent, mortgage or utility bills, (2) to keep their home adequately warm, (3) to face unexpected expenses, (4) to eat meat or proteins regularly, (5) to go on holiday, (6) to have a television set, (7) a washing machine, (8) a car and (9) a telephone. The enforced lack condition means that people would like to possess the items, but cannot afford them; choices or preferences should play no role. The composition of the indicator has recently been revised (see Guio et al., 2012; Guio et al., 2017). The *standard* material deprivation indicator adopted in 2009 uses a threshold of at least three deprivations to identify people considered as deprived, although the *severe* concept adopted in the context of the Europe 2020 strategy set the threshold at four deprivations. The material deprivaiton indicators differ in two important ways from the income poverty approach. Firstly, material deprivation takes a direct approach to measuring poverty, as the focus lies on outcomes. Secondly, it is an absolute measure, since the same threshold and items are used for the whole EU.

The third indicator used in the Europe 2020 social exclusion target (*quasi-joblessness*) regroups people living in households whose work intensity (WI, hereafter) is lower than 20%. The household WI is the ratio of the total number of months that all working-age (18-59) household members have worked and the total number of months the same household members theoretically could have worked. Household members over the age of 60 and students between the age of 18 and 24 do not influence a household's work intensity. The inclusion of "living in a jobless household" as the third criterion of the social exclusion target group has not always been positively evaluated by researchers. It is sometimes unclear why jobless households that are neither income poor nor materially deprived should be included in the target population (Nolan & Whelan, 2011). Instead, living in a jobless household is rather seen as a risk factor that has a direct impact on the other two target components (De Graaf-Zijl & Nolan, 2011). This indicator will thus be used as a household-level determinant of income poverty and material deprivation, not as a dependent variable. Even if it could be defended to include the third component of the

Europe 2020 social exclusion target as a dependent variable, to analyse possible reverse causality effects, this would fall outside the scope of this paper.

1.2.2. EU-SILC dataset

The headline poverty target of reducing the number of people living in poverty and social exclusion by 20 million in 2020 has been defined on the basis of the EU-SILC dataset. This source provides detailed information on income, labour, health and education at the individual level, while deprivation variables are collected at the household level. In line with the EU indicators approach, it is assumed that household members share the same standard of living and pool household resources together. As a consequence, the individual is used as the unit of analysis and the household is used as the unit of measurement.

The analysis in this chapter uses EU-SILC 2012 data. The full dataset contains 613,151 individuals over a total of 31 countries (all 28 EU Member States plus Norway, Switzerland and Iceland). The main regression analyses are carried out on a dataset of 492,122 individuals (*N*). The individuals that were excluded from the analysis mainly live in households in which all members are over the age of 59 (i.e. households for which the WI could not be calculated).[11] The percentage population excluded for the WI variable is 18.6 per cent with a standard deviation of 0.04. The highest percentage population excluded is found in Germany (27 %) and the Romania (23.3%), while Slovenia (11%) and Slovakia (12.1%) have the lowest rate. These numbers reflect differences in the countries' demographic structure and composition of the household. The focus of this study thus lies on the working age population. The item non-response (i.e. the percentage of individuals who miss the variable) is negligible for the other variables in the model.

The income poverty rates, the severe material deprivation rates and the overlaps between both measures are shown in Table 1.1. In this table, one immediately notices the differences between severe material deprivation and income poverty rates in terms of range across countries. There are clear differences in the material deprivation rates between the old and new EU Member States and – to a lesser extent – between Southern and Northern Member States. The variation in income poverty rates is much more narrow, as income poverty is relatively defined. Specifically, the income poverty rates vary between 4 (i.e. Denmark) and 27.1 (i.e. Greece) per cent, while the severe material deprivation rates vary between 0.6 (i.e. Switzerland) and 41.8 (i.e. Bulgaria). The overlap between income poverty and severe material deprivation also differs greatly across countries (Table 1.1, column 5). In less affluent countries, a large share of the population suffer

[11] Note that *all* individuals in the dataset (i.e. including the individuals that are missing in the regression analysis) were used for the computation of the at-risk-of poverty indicator. This is in line with the standard EU approach.

from both income poverty and severe material deprivation (consistent poverty), i.e. more than 10% in Bulgaria, Greece, Hungary, Latvia and Romania.

Table 1.1 – Income poverty and material deprivation rates across Europe

	Income poverty (IP)	Severe material deprivation (MD)	IP only	MD only	Both MD and IP (Cons. Pov.)
Austria	13.6	4.3	11.4	2.1	2.2
Belgium	15.0	6.9	10.8	2.7	4.2
Bulgaria	18.8	41.8	3.8	26.8	15.0
Switzerland	11.9	0.6	11.5	0.3	0.3
Cyprus	13.4	15.2	8.6	10.4	4.8
Czech Republic	8.5	6.0	6.0	3.6	2.5
Germany	12.9	4.3	10.3	1.7	2.6
Denmark	4.0	1.2	3.6	0.8	0.4
Estonia	19.5	10.6	13.5	4.6	6.0
Greece	27.1	21.2	13.7	7.8	13.4
Spain	19.7	5.5	16.5	2.2	3.3
Finland	9.1	1.9	8.2	1.0	0.9
France	14.2	5.4	11.3	2.5	2.9
Croatia	21.0	16.7	12.6	8.3	8.4
Hungary	16.9	29.1	5.1	17.3	11.8
Ireland	16.2	11.2	12.2	7.2	4.0
Iceland	5.9	1.8	5.5	1.5	0.4
Italy	17.8	12.2	11.8	6.3	5.9
Lithuania	16.2	18.4	9.3	11.5	6.9
Luxembourg	16.2	1.6	15.2	0.5	1.0
Latvia	21.8	27.0	9.7	14.9	12.1
Malta	14.7	8.5	11.6	5.4	3.1
The Netherlands	4.8	1.3	4.3	0.9	0.4
Norway	6.1	1.3	5.6	0.7	0.5
Poland	19.9	14.3	13.3	7.6	6.6
Portugal	19.2	9.8	14.4	5.0	4.8
Romania	23.9	28.6	10.8	15.5	13.1
Sweden	11.1	1.3	10.1	0.4	0.9
Slovenia	9.4	5.5	7.3	3.5	2.0
Slovakia	11.1	9.6	7.1	5.7	3.9
United Kingdom	15.6	9.5	11.5	5.4	4.1
Average	14.7	10.7	9.9	5.9	4.8

Source: EU-SILC (2012) cross-sectional data, authors' computation.

1.3. Determinants of income poverty and severe material deprivation

1.3.1. Household-level determinants (within-country analysis)

As material deprivation gained importance as an indicator of social exclusion, its association with income (poverty) has been well researched in recent years (Whelan et al., 2001; Perry, 2002; Whelan & Maître, 2006; Whelan & Maître, 2007; Berthoud & Bryan, 2011; Fusco et al., 2011). It is now well established that income poverty and material deprivation tap into different phenomena. Income poverty is solely based on a household's current income, whereas material deprivation is determined by a household's "command over resources" (or its "permanent income") and the costs a household faces (Perry, 2002). Fusco and co-authors (2011, p. 139) note that "two individuals with the same current income can have very different standards of living if their income does not adequately measure all the resources that are available to each of them and/or if they face different costs". We briefly review the household-level determinants that can explain the mismatch between income poverty and material deprivation within countries. All summary statistics of the household-level variables can be found in Table 1.A1 in Appendix 1.

A. A household's command over *resources* forms a first important factor that influences the probability of material deprivation. The resources of a household include its current income, but also previous incomes, accumulated saving or debts, previous investment in durables or housing or any other source of economic or social support. We do not include current income as an explanatory variable at the household-level, as the income poverty approach is defined on this variable. Instead, we are interested in the underlying mechanisms that influence households' ability to generate resources on the labour market.

   We introduce two variables that – albeit partially and indirectly – capture the resources available to households', i.e. households' work intensity and educational attainments. The *work intensity* variable (see Section 1.2.1 for a definition) capture households' short term ability to generate resources on the labour market. Two education dummies measure a household's long-term ability to generate resources on the labour market. A higher education dummy measures whether someone in the household has a tertiary education degree (*Higher education (hh)*). The second education dummy takes the value of one if at least one household member has a primary degree or less, provided that no one has a tertiary degree (*Lower education (hh)*). Averagely educated households (i.e. no household member with a tertiary degree, nor with a primary degree) form the reference category. The educational attainments of household members aged over 60 do not influence the dummies. An individual's education is expected to permanently influence his/her ability to generate resources on the labour market and has an impact on his wage level. We therefore expect that education has a stronger association with a household's living standard factors (and hence, the risk of material deprivation) than the work intensity variable (see also Fusco et al., 2011).

B. *Costs* increase the level of resources necessary for a household to maintain its standard of living. Diverging household needs are a key element to explain differences in costs. Needs are influenced by household structure, marital status, number and age of children, health problems or tenure status (Whelan et al., 2004, p. 294). The income poverty approach only takes account of differences in household composition by equalizing the household disposable income. The OECD-modified scale used in the EU approach gives a weight of one to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 14. Households with the same income level, adjusted to this scale, but who face different costs, can have very different living standards, i.e. in terms of material deprivation suffered.

   Household costs and needs are operationalized by a health and a tenure status variable. The health variable is a dummy that takes the value of one if at least one person in the household reports having bad or very bad health (*Bad health (hh)*). A tenure dummy takes the value of one if the household rents its house on the private market or with a social (free or reduced) tariff, as compared to owning its own house (*rent*).

C. We include households' *socio-demographic diversity* as a third cluster of household-level variables. These variables influence both the resources available to household and its costs/needs. Households' socio-demographic diversity is captured by five household structure dummies and three demographic variables. The different household structures include households with two adults and no children (as the reference group), families with two adults and one or two children (*Couple, one or two children*), families with two adults and more than two children (*Couple, more children*), single adult households (*Single, no children*), single parent families (*Single parent*) and a rest category (*Other hh*).[12] The demographic variables capture the age and gender of the household head and a household's migration background. The impact of age on social exclusion often follows a non-linear pattern and varies greatly from country to country (Boarini & Mira d'Ercole, 2006). We therefore only test whether household heads (HRP, hereafter) younger than 29 have a higher risk of social exclusion (*Young age, <29 (HRP)*).[13] Similarly, we include a dummy that indicates if the HRP is a female (*Female (HRP)*). Lastly, a migrant dummy

[12] All individuals younger than 18 are considered children.
[13] The household reference person or household head is defined as the individual responsible for the accommodation. If more than one person bears this responsibility, the oldest person is chosen.

takes the value of one if someone in the household was born outside the EU (*Migrant (hh)*).

1.3.2. Country-level determinants (between-country analysis)

Besides the differences in the composition of the population in terms of household-level risk factors, national performances may also be influenced by institutional and macroeconomic country-level variables. Multilevel studies have shown that both concepts are shaped by common institutional determinants that relate to the welfare state.[14] Given that we do not control for current income at the household-level, the interpretation of the country-level effects requires some additional elaboration. Specifically, it is expected that the country-level variables capture most of the 'direct' material deprivation-reducing effects that are normally captured by current income at the household-level. These effects include the cushioning effect of social transfers (both in terms of its size and pro-poorness) and market income, which are all components of current income measured at the household-level. As we do not include current income at the household-level, this 'direct' effect is only imperfectly proxied by the country-level variables that we introduce in our model. The aggregated country-level variables may also include additional, 'indirect' aggregated effects, as some multilevel studies found a negative relationship between aggregated variables (in terms of levels of affluence or social policy context) and deprivation, even after controlling for income at the household-level (see Whelan & Maître, 2012; Bárcena-Martin et al., 2014). However, given that we cannot include current income as an independent variable in our model (i.e. the income poverty approach is based on this variable), it is impossible to distinguish between the 'direct' effects situated at the household-level and potential additional 'indirect' effects situated at the country-level. We did, however, test and compare several relationships that have not been addressed in previous multilevel studies. All summary statistics of the country-level variables can be found in Table 1.A2 in Appendix 1.

A. The *distinction between in-kind and in-cash social benefits* has received more attention in recent years, as there is a clear trend in Western welfare states to spend more on in-kind benefits relative to cash transfers. Firstly, social benefits paid in-kind have a direct relationship with the risk of material deprivation, but have an indirect relationship with the risk of income poverty. In-kind benefits ("public services" e.g. housing, health care, child care etc.) do not directly affect household income, but rather reduce certain calls on household expenditure, allowing poor households to spend their limited resources on other necessities (Kenworthy, 2011). They also indirectly boost the earnings of those at

[14] For the risk of income poverty, see Moller et al., 2003; Brady, 2005; Brady & Kall, 2008; Brady et al., 2009; Lohmann, 2009; Kim et al., 2010; Diris et al., 2017; Saltkjel & Malmberg-Heimonen, 2016. For the risk of material deprivation, see Dewilde, 2008; Nelson, 2012; Bárcena-Martín et al., 2014; Visser et al., 2014; Saltkjel & Malmberg-Heimone, 2016.

the low end of the income distribution by enhancing human capital, assisting with job search and placement, and facilitating work-family balance" (Kenworthy, 2011, p.64). Secondly, in-kind and in-cash social benefits have a different distributional impact. OECD (2008, 2011) studies revealed that publicly provided services reduce inequality by one fifth, whereas net cash transfers reduce overall inequality by one third. Moreover, the distributive pattern of cash transfers (excluding pensions) is more oriented towards lower incomes than that of services: 26 per cent of all cash benefits go to the bottom quintile, 15 per cent in the case of services (Verbist & Matsaganis, 2014).

The social spending levels are expressed as a percentage of the Gross Domestic Product (GDP) and are obtained from Eurostat (ESSPROSS database). The in-cash social benefits include sickness/health care, disability, family/children and unemployment benefits, housing allowances and social exclusion benefits (*In-cash social benefit levels*).[15] Pensions and survivor benefits are excluded. The reasons are twofold. Firstly, as a significant proportion of the elderly are not included in the analysis, integrating pensions into the social protection concept may lead to potentially biased results and policy conclusions. Secondly, by excluding pension spending from the transfer concept, we avoid that transfer levels might be strongly driven by demographical factors (e.g. countries with more elderly spend more on pensions). Instead, pensions are assumed to be a deferred wage. The social benefits in-kind include in-kind healthcare services, disability, family/children, unemployment and housing benefits (*In-kind social benefit levels*).[16] The in-cash benefit levels vary moderately in the EU (between 3 (i.e. for Romania) and 8.5 (i.e. for Belgium) per cent of GDP), whereas the variation in in-kind spending rates is slightly wider (between 3.5% (i.e. for Latvia) and 10.7% (i.e. for Denmark, Iceland and France)). The Western and Scandinavian countries spend a much larger percentage of their GDP on social benefits, both in-cash and in-kind, compared to the Central and Eastern European countries. In fact, these two social spending variables are strongly positively correlated (i.e. Pearson correlation coefficient of 0.62). To avoid potential multicollinearity issues and to establish the general impact of social protection levels, we also include an alternative variable that sums average in-kind and in-cash benefit levels, expressed as a percentage of GDP (*Total social benefit levels*).

[15] In more detail, the cash benefits include cash benefits for sickness/health care (paid sick leave), disability (pension, early retirement, care allowance and economic integration of the handicapped), family/children (income maintenance benefit in the event of childbirth, birth grant, parental leave benefit, family or child allowance) and unemployment (full and partial unemployment benefit, early retirement, vocational training allowance, redundancy compensation).

[16] In more detail, spending on in-kind benefits includes healthcare services (direct provision or reimbursement), disability (accommodation, assistance in carrying out daily tasks, rehabilitation), family/children (child care, accommodation, home help), unemployment (mobility and resettlement, vocational training, placement services and job-search assistance) and housing benefits (rent benefit). It is also important to note that we do not include public expenditure on education in our analysis. Although it is a very important in-kind benefit, aggregate data on education was unavailable for some countries in our dataset.

B. Between-country differences in the risk of income poverty and material deprivation may not only be explained by levels of spending on social protection, but also by the *pro-poorness of social benefits*. If social benefits are more targeted towards the poor, one intuitively expects that poverty is reduced, both in absolute and relative terms. However, Korpi and Palme (1998) claim in their influential work that "the more we target benefits at the poor and the more concerned we are with creating equality via equal public transfers to all, the less likely we are to reduce poverty and inequality." The authors argue that increased pro-poorness of social benefits comes at the expense of spending size. In essence, the authors explain this "redistribution paradox" by the reasoning that in countries where social benefits are strongly geared towards the poor, fewer resources tend to be available for redistribution because there is less widespread and less robust political support for redistribution. When social benefits are more universally distributed, there is stronger support among the middle class (i.e. median voters) for redistribution, including to the most needy. The relationship between pro-poorness of social benefits and inequality has been re-examined in a multitude of studies, with some authors conforming support for Korpi and Palme's thesis, but others finding a non-existent or positive association between the two measures (for an extensive literature review, see Marx et al., 2013). There is, however, a lack of studies investigating the relationship between benefit targeting and material deprivation.

   We define pro-poorness of in-cash social spending as the share of transfers (excluding pensions) that is distributed to the lowest five deciles in the pre-transfer household income distribution (including pensions), following Marx and co-authors (2013) (*pro-poorness bottom 50*). In the pre-transfer household income (including pensions) income distribution, households are ranked according to the position they would be in a hypothetical situation with no social redistribution (excluding pensions). By including pensions in the income concept for ranking households, we avoid households in which a significant part of the income comes from pension benefits being considered as extremely poor and, consequently, populating the bottom end of the distribution. The pro-poorness of social benefits is strongly positively correlated with total spending on social benefits (i.e. Pearson correlation coefficient of 0.70). Among the countries that are included in the dataset, the Scandinavian countries, but also the United Kingdom and The Netherlands, are the countries in which social benefits are very strongly targeted towards the poor. Eastern European countries such as Romania, Bulgaria, Latvia and Estonia do the least to target the poorest.

C. Next to institutional variables related to the welfare state, some authors stress the importance of *aggregated income levels* as a crucial country-level determinant to predict material deprivation. Bárcena-Martín and co-authors (2014), and Visser and co-authors

(2014), find a small negative association between GDP per capita and material deprivation, even after controlling for current income at the individual-level. Other researchers argue that this relationship is less clear-cut. Whelan and Maître (2012) showed that national median income levels are only indirectly associated with the risk of material deprivation after controlling for individuals' current income, through an extensive set of cross-level interactions with social class variables, educational qualification, number of children and marital status. Kenworthy and co-authors (2011) showed that material deprivation rates are correlated to a country's social policy generosity, but not to its level of affluence.

The macroeconomic variables include the OECD equivalised median income levels and the average rate of unemployment. The OECD equivalised median income levels are expressed in purchasing power standard (PPS) per 1,000 and are directly derived from the EU-SILC micro dataset (*Median income*). Median income levels vary extensively across Europe, between 4,252 (Romania) and 27,659 PPS (Luxembourg).

D. Even though we control for work intensity on the individual-level, we further introduce the unemployment rate to account for the possible effect of the business cycle on the size and pro-poorness of social benefits. The unemployment rate is defined as the number of people unemployed as a percentage of the active population (*Unemployment rate*). The active population includes the employed (fully and part-time and both employees and self-employed) and unemployed people, but not the economically inactive. The variable is directly derived from the EU-SILC micro dataset.

1.4. Methods

1.4.1. Single level and multilevel multinomial logistic regression

We use a multinomial logistic regression model to assess to what extent income poverty and material deprivation are subject to the same determinants. The dependent variable of the multinomial model has four categories: being 'materially deprived only', being 'income poor only', being 'consistently poor' and being none of these, as the reference category (*r*). We run single level models to investigate the impact and relative importance of the household-level variables in explaining the extent of income poverty and material deprivation within different EU countries. The single level multinomial model (1a) is given by the following formula:

$$\ln\left(\frac{\pi_{ki}}{\pi_{ri}}\right) = \beta_{0k} + \sum_{h=1}^{H} \beta_{hk} x_{hi} \qquad (1a, single\ level)$$

with:

π_{ki} as the probability for individual i ($i=1,..., N$) to belong to the k^{th} ($k = 1,...,K$) category of the dependent variable
π_{ri} as the probability for individual i ($i=1,..., N$) to belong to the reference category r of the dependent variable (i.e., not income poor, not severely materially deprived)
β_{0k} as the overall intercept of the k^{th} ($k = 1,...,K$) category of the dependent variable
x_{hi} as the value of the h^{th} ($h=1,...,H$) independent variable defined at the household-level for individual i ($i=1,..., N$)
β_{hk} as the coefficient of the h^{th} ($h=1,...,H$) independent variable defined at the household-level for the k^{th} ($k = 1,...,K$) category of the dependent variable

We add a multilevel structure to the model to evaluate the impact of country-level variables that can explain the between-country differences in the risk of income poverty and material deprivation. Multilevel models are particularly appropriate to study nested data designs, where respondents are organized within more than one level. In our study, individuals (i) are nested within countries (j). Contextual differences in the prevalence of income poverty/material deprivation across countries are measured by (observable) country-level variables (z_{cj}) and (unobservable) random intercepts (U_{kj}). The multilevel multinomial model (1b) is given by the following formula:

$$\ln\left(\frac{\pi_{kij}}{\pi_{rij}}\right) = \beta_{0k} + \sum_{h=1}^{H} \beta_{hk} x_{hij} + \sum_{c=1}^{C} \beta_{ck} z_{cj} + U_{jk} \qquad (1b, multilevel)$$

with:

π_{kij} as the probability for individual i ($i=1,..., N$) living in country j ($j=1,...,31$) to belong to the k^{th} ($k = 1,...,K$) category of the dependent variable
π_{rij} as the probability for individual i ($i=1,..., N$) living in country j ($j=1,...,J$) to belong to the reference category r of the dependent variable (i.e., not income poor, not severely materially deprived)
β_{0k} as the overall intercept of the k^{th} ($k = 1,...,K$) category of the dependent variable
x_{hij} as the value of the h^{th} ($h=1,...,H$) independent variable defined at the household-level for individual i ($i=1,..., N$) living in country j ($j=1,...,J$)
β_{hk} as the coefficient of the h^{th} ($h=1,...,H$) independent variable defined at the household-level for the k^{th} ($k = 1,...,K$) category of the dependent variable
z_{cj} as the value of the c^{th} ($c = 1,...,C$) independent variable defined at the country-level for country j ($j=1,...,J$)
β_{ck} as the coefficient of the c^{th} ($c = 1,...,C$) independent variable defined at the country-level for the k^{th} ($k = 1,...,K$) category of the dependent variable
U_{jk} as the error term for country j ($j=1,...,J$) for the the k^{th} ($k = 1,...,K$) category of the dependent variable, $\sim N(0, \mu_k)$

The left-hand side of the equation (1a) and (1b) refer to the log odds of being in category k, rather than in the reference category r (i.e. not income poor not severely deprived). Similarly, β_{0k} refers to the overall intercept of category k. In the multilevel model, it is interpreted as the log odds when the country residual ("random intercept") is not taken into account (U_{kj}=0). An important notion that we use to interpret the results are exponentiated coefficients (i.e. exp(β_{hk}), exp(β_{ck})). Exponentiated coefficients give the amount by which the relative risk (i.e. the ratio between the probability of belonging to category k of the dependent variable (π_{kij}) and the probability of belonging to reference category r of the dependent variable (π_{rij})) is multiplied when an independent variable (x_{hij}, z_{cj}) increases by one unit (keeping all other variables constant).

Many authors are critical about the use of multilevel models, claiming that country-level effects are often imprecisely measured. For Meuleman and Billiet (2009) the limited number of groups are an important cause of estimation errors, suggesting that a logistic nested model requires at least 30 countries. Stegmueller (2013) and Bryan and Jenkins (2015) stress the fact that the maximum likelihood estimation method often leads to biased estimates of group effects. Both authors suggest that a Bayesian estimation approach can drastically improve estimates of the country effects. All our models are therefore run with Bayesian Markov Chain Monte Carlo simulations (MCMC) with hierarchical centring and with Iterative Generalised Least Squares (IGLS) estimates as starting values for each parameter.

1.4.2. Pseudo R^2-measures

We use *McKelvey and Zavoina's pseudo R^2-measure* to obtain the degree of the variance that is explained by the models. McKelvey and Zavoina's (1975) pseudo R^2 (R^2_{MZ}) is the only pseudo R^2-value that captures the notion of explained variance and is thus conceptually similar to the traditional R^2-value in OLS regressions (Veall & Zimmerman, 1996). The R^2_{MZ} concept is based on the variance of an underlying latent (unobserved) variable. According to the latent variable approach, each nominal outcome category of the dependent variable is conceived as the result of an underlying non-observed continuous latent variable $\check{y}_{kij}$. The latent variable version of models (1a) and (1b) are given in formula (2a) and (2b). Both versions depend on an observable non-random part (($\beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{sij}$) for (2a) and ($\beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{sij} + \sum_{c=1}^{C} \beta_{ck} z_{cj}$) for (2b)) and on a non-observable random part ((ε_{ik}) for (2a)) (($\varepsilon_{ik} + U_{jk}$) for (2b)):

$$\breve{y}_{kij} = \beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{sij} + \varepsilon_{ik} \quad (2a, single\ level)$$

$$\breve{y}_{kij} = \beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{sij} + \sum_{c=1}^{C} \beta_{ck} z_{cj} + U_{jk} + \varepsilon_{ik} \quad (2b, multilevel)$$

The category of the dependent variable (k=*1*,*2*,*3*) with the highest latent value ($\breve{y}_{kij}$) is the actual realized (Y_k) outcome of the dependent variable, as given in (3a) and (3b):

$$Y_{ki} = \begin{cases} 1\ if\ \breve{y}_{1i} > \breve{y}_{2i}, \breve{y}_{3i}, \breve{y}_{ri} \\ 2\ if\ \breve{y}_{2i} > \breve{y}_{1i}, \breve{y}_{3i}, \breve{y}_{ri} \\ 3\ if\ \breve{y}_{3i} > \breve{y}_{1i}, \breve{y}_{2i}, \breve{y}_{ri} \\ r\ if\ \breve{y}_{ri} > \breve{y}_{1i}, \breve{y}_{2i}, \breve{y}_{3i} \end{cases} \quad (3a, single\ level)$$

$$Y_{kij} = \begin{cases} 1\ if\ \breve{y}_{1ij} > \breve{y}_{2ij}, \breve{y}_{3ij}, \breve{y}_{rij} \\ 2\ if\ \breve{y}_{2ij} > \breve{y}_{1ij}, \breve{y}_{3ij}, \breve{y}_{rij} \\ 3\ if\ \breve{y}_{3ij} > \breve{y}_{1ij}, \breve{y}_{2ij}, \breve{y}_{rij} \\ r\ if\ \breve{y}_{rij} > \breve{y}_{1ij}, \breve{y}_{2ij}, \breve{y}_{3ij} \end{cases} \quad (3b, multilevel)$$

We adjust McKelvey and Zavoina's pseudo R^2_{MZ} measure to the multinomial (multilevel) version of the logistic regression model by calculating a separate R^2-measure for each category of the dependent variable.[17] McKelvey and Zavoina's R^2-measures, similarly to the traditional OLS R^2-measure, the proportion in the dependent variable that is explained by the independent variables. The total variance, given in (4a) and (4b), is based on the variance of the underlying latent variable in formula (2a) and (2b):

$$var(\breve{y}_k) = var(\beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{\mathrm{h}ij} + \varepsilon_{ik}) \qquad (4a, single\ level)$$

$$var(\breve{y}_k) = var(\beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{\mathrm{h}ij} + \sum_{c=1}^{C} \beta_{ck} z_{\mathrm{cj}} + U_{jk} + \varepsilon_{ik}) \qquad (4b, multilevel)$$

Because the error terms on the individual-level ε_{ik} are assumed to be independent across the independent variables, the total variance can be decomposed into an explained part (σ_k) and an

[17] See Snijders and Bosker (2012) for the multilevel equivalent of McKelvey and Zavoina's pseudo R^2-measure for the binary logistic multilevel model.

unexplained part at the individual level (ϖ), as given in (5a).[18] In the multilevel context, the total variance can be decomposed into an explained part (σ_k) an unexplained part at the country-level (μ_k) and an unexplained part at the individual level (ϖ), as given in (5b):

$$var(\check{y}_k) = \sigma_k + \varpi \qquad (5a, single\ level)$$

$$var(\check{y}_k) = \sigma_k + \mu_k + \varpi \qquad (5b, multilevel)$$

For the calculation of the explained part of the variance σ_k, given in (6a) and (6b), it is important to note that we use the coefficients obtained from the model in (1a) and (1b), which is not conditional on the realized value of Y_{kij}:

$$\sigma_k = var(\beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{hij}) \qquad (6a, single\ level)$$

$$\sigma_k = var(\beta_0 + \sum_{h=1}^{H} \beta_{hk} x_{hij} + \sum_{c=1}^{C} \beta_{ck} z_{cj}) \qquad (6b, multilevel)$$

The unobserved variance at the individual-level ϖ, given in (7), is assumed to have a logistic distribution and has a fixed variance equal to $\pi^2/3 \approx 3.29$. More formally, the measures is defined as:

$$\varpi = var(\varepsilon_{ik}) = \frac{\pi^2}{3} \qquad (7)$$

The unobserved variance at the country-level ϖ, given in (8), is an estimate from the multilevel model in (1b):

$$\mu_k = var\left(U_{jk}\right) \qquad (8)$$

Finally, the McKelvey and Zavoina's pseudo R^2_{MZ} measure is defined as ratio of the variance that is explained by the model (σ_k) and the total (i.e. explained and unexplained) variance ($var(\check{y}_k)$), as given in (9a) and (9b):

18 The independence assumption implies that the covariance between the error terms and the independent variables is equal to zero.

$$R^2_{k,MZ} = \frac{\sigma_k}{\sigma_k + \varpi} \qquad (9a, single\ level)$$

$$R^2_{k,MZ} = \frac{\sigma_k}{\sigma_k + \mu_k + \varpi} \qquad (9b, multilevel)$$

We are also interested in the importance of the unobserved errors at country-level. We therefore calculate the *Variance Partition Coefficient* (VPC) and an explained country-level variance measure. The VPC, given in (10), is defined as the proportion of total unexplained variance ($\mu_k + \varpi$) that is attributable to the country-level (μ_k):

$$VPC_k = \frac{\mu_k}{\mu_k + \varpi} \quad (10)$$

The *explained country-level variance*, given in (11), is defined as the difference between the unexplained country variance of the 'micro' model ($\mu_{k,micro}$, i.e. including only household-level variables) and the unexplained country variance of the 'full' model ($\mu_{k,full}$, i.e. including both household-level and country-level variables) divided by the unexplained country variance of the 'micro' model ($\mu_{k,micro}$). By using the unexplained country variance of the 'micro' model as the benchmark, we ensure that the measure only reflects the variance that is explained by the variables at country-level (i.e. after controlling for national differences in the composition of household-level risk factors). More formally, the explained variance measures are defined as:

$$Explained\ country\ variance_k = \frac{\mu_{k,micro} - \mu_{k,full}}{\mu_{k,micro}} \quad (11)$$

Finally, we calculate the *Likelihood Ratio R²* (R^2_{LR}) and the *Deviance Information Criterion* (DIC) to compare the fit of the models we estimate. The Likelihood Ratio R^2 is a popular Pseudo R^2-measure. It measures the proportional reduction in the Deviance (*D*) value:

$$R^2_{LR} = \frac{D_{null} - D_{fitted}}{D_{null}} \quad (12)$$

with D_{null} as the Null Deviance (i.e. the difference in -2xlog(likelihood) values between a model with only the intercept (and no random intercept) and a saturated model (i.e. a model with a theoretically perfect fit)) and D_{fitted} as the Fitted Deviance (i.e. the difference in -2xlog(likelihood) values for the fitted model with at least one predictor (or random intercept) and a saturated model). Smaller Deviance values indicate a better fit as the fitted model deviates less from the saturated model. Higher values in R^2_{LR} thus represent a better fit of the model to the data. The DIC is a compromise between the model's fit with the data and overall complexity, where lower DIC values indicate a better model (Spiegelhalter et al., 2002).[19] More formally, the DIC value is defined as:

$$DIC = \overline{D} + p_D \quad (13)$$

with $\overline{D}$ as the average Deviance from 5000 iterations ('fit') and p_D the effective number of parameters ('complexity').[20]

1.4.3. Decomposing the explained variance measures

We use a Shapley decomposition to establish the relative importance of the different variables on the household and country-level in explaining differences within and between countries in the risk of income poverty and/or material deprivation. The *Shapley value* is a way to fairly distribute the total gains of a game to players in cooperative game theory (Shapley, 1953). As Charpentier and Mussard (2011, p. 538) put it, "the Shapley value is appealing for economists since it brings out the contribution of each variable without recourse to econometric models for which the quality of fit and all tests have to be checked". It has been used to decompose the goodness-of-fit measure in both linear and logistic regression models.[21] In this decomposition the exact contribution of each independent variable to the total R^2-value is calculated. Specifically, the contribution of each variable *v (v = 1,…,n)*, C_v, to the total R^2-value (14) is defined as:

$$C_v = \sum_{\substack{S \in \mathbb{N} \\ |S| \leq n \\ v \in S}} \frac{(|S|-1)!\,(n-|S|)!}{n!} * [R^2(S) - R^2(S \backslash \{v\})] \quad (14)$$

where *S* is a sub-model of the full (i.e. containing all independent variables) regression model, |*S*| is the number of variables in the sub-model *S* and *n* is the number of variables *v* of the full model.

The Shapley decomposition calculates the average marginal contribution of each variable included in the regression, by considering both its unique contribution ($|S|$=1) and the

[19] See Browne (2015, p. 28 and p. 177) for a formal definition and discussion of the Deviance and DIC criteria in a multinomial logistic regression context.
[20] Note that the effective number of parameters (p_D) is an estimate and might differ from the actual (nominal) number of parameters (which is always an integer number) due to dependencies between the random effects and the actual number of parameters (Browne, 2015).
[21] See Deutsch and Silber (2006) for a Shapley decomposition of the likelihood ratio in a binary logistic model estimating the odds on poverty.

contribution when combined with (all possible combinations of) other variables ($n \geq |S| \geq 2$). The marginal contribution $[R^2(S) - R^2(S\backslash\{v\})]$ represents the difference in the R²-value for sub-model S containing variable v and the model which is identical to S except not containing variable v (read: "model S without variable v"). The factor $\frac{(|S|-1)!(n-|S|)!}{n!}$ weights the marginal contributions by the number of permutations represented by the sub-model S.[22][23]

We apply Shapley decompositions on several measures. We decompose McKelvey and Zavoina's explained variance measure R^2_{MZ} (for each category of the dependent variable) and the overall goodness-of-fit measure R^2_{LR} to assess the relative importance of the independent household-level variables in the single level models. We cluster certain variables together and calculate the relative contribution of each *group of variables*, instead of each independent variable of the model. The reasons are twofold. Firstly, we are mainly interested in the influence of an entire group of indicators that can explain the prevalence of income poverty and/or material deprivation. Secondly, clustering the independent variables into groups of interest also significantly reduces the total number of sub-models that need to be estimated to calculate the contribution of each variable C_v. Finally, we decompose the explained country variance measure to compare the explanatory power of the country-level variables for the different categories of the dependent variable.

1.5. Results

1.5.1. Household-level determinants

In this section we discuss the relationship between the household-level variables and the risk of income poverty and severe material deprivation. Firstly, we establish *for which category of the dependent variable the household-level variables are the most effective in explaining within-country differences*. Secondly, we describe *in which countries the household-level variables*

[22] For example, in the case of three variables (A,B,C) there are seven possible submodels (ABC, AB, AC, BC, A, B, C). Submodels S with 3 variables (|S|=3; ABC) obtain a weight of 2/6, submodels S with two variables (|S|=2; AB, AC, BC) obtain a weight of 1/6 and submodels S with one variable (|S|=1; A, B, C) obtain a weight of 2/6. The average contribution of variable A to the R²-value (C_A) equals $\frac{1}{3} * (R^2_{ABC} - R^2_{BC}) + \frac{1}{6} * (R^2_{AB} - R^2_{B}) + \frac{1}{6} * (R^2_{AC} - R^2_{C}) + \frac{1}{3} * (R^2_{A} - 0)$. The average contribution of variable B to the R²-value (C_B) then equals $\frac{1}{3} * (R^2_{ABC} - R^2_{AC}) + \frac{1}{6} * (R^2_{AB} - R^2_{A}) + \frac{1}{6} * (R^2_{BC} - R^2_{C}) + \frac{2}{6} * (R^2_{B} - 0)$. The average contribution of variable C to the R²-value (C_C) equals $\frac{1}{3} * (R^2_{ABC} - R^2_{AB}) + \frac{1}{6} * (R^2_{AC} - R^2_{A}) + \frac{1}{6} * (R^2_{BC} - R^2_{B}) + \frac{2}{6} * (R^2_{C} - 0)$. The R²-value then neatly equals the sum of the individual contributions of the variables ($R^2 \ = C_A + C_B + C_C$).

[23] Young (1985) showed that the Shapley value is the only solution that simultaneously satisfies three attractive properties: efficiency, symmetry and monotonicity. Firstly, the efficiency axiom stipulates that the R² of the full model is decomposed among the independent variables). Secondly, the symmetry axiom ensures that substitutes with the same explanatory power obtain equal valuation (equal treatment). Finally, the monotonicity axiom ensures that an increase in the R²-value does not reduce the explanatory value attributed to any variable.

have the strongest explanatory power. Finally, we assess *to what extent income poverty and material deprivation are subject to the same underlying household-level variables*. We use the Shapley decompositions in Table 1.2 to compare the relative contribution of the different clusters of household-level covariates to the within-country explained variance measure from the single level analysis. The detailed results of the multilevel ‘micro’ model, including exponentiated coefficients, are provided in Table 1.A3 in Appendix 1. Specifically, the household-level variables capture differences in households’ short term ability to generate resources on the labour market (i.e. *work intensity*), long term ability to generate resources on the labour market (i.e. *educational variables*), costs (i.e. *bad health, rent*) and socio-demographic characteristics (i.e. *household structure and other demographic variables*).

1.5.1.1. Within-country explanatory power across categories

The results from the single level models indicate that the *explanatory power of the model varies considerably across the different categories of the dependent variables* (Table 1.2). For all countries except Switzerland, Cyprus and Norway, the explained variance for the ‘consistent poverty’ category is significantly larger than for the ‘income poverty only’ and ‘material deprivation only’ categories. The household-level variables explain, on average across countries, 59 per cent of the within-country differences for the risk of ‘consistent poverty’, 36 per cent of the within-country differences for the risk of ‘income poverty only’, and 32 per cent of the within-country differences for the risk of ‘material deprivation only’. This result indicates that the predictive power of the socio-economic characteristics at the household-level are higher for those cumulating both income poverty and severe material deprivation, than for those suffering from ‘only’ one of these problems.

Table 1.2 – Shapley decompositions of the household-level variables on the explained within-country differences

Country		R^2	Decomposition of pseudo R^2-measures (relative contribution) WI	Education	Costs	Socio-Demo-graphics
Average of single level models	R^2_{MZ} IP only	0.36	49%	20%	11%	19%
	R^2_{MZ} MD only	0.32	24%	34%	22%	20%
	R^2_{MZ} Cons. Pov.	0.59	41%	25%	18%	16%
	R^2_{LR} Likelihood Ratio	0.21	45%	18%	17%	19%
Austria	R^2_{MZ} IP only	0.27	50%	8%	21%	22%
	R^2_{MZ} MD only	0.42	34%	20%	31%	16%
	R^2_{MZ} Cons. Pov.	0.66	31%	15%	33%	21%
	R^2_{LR} Likelihood Ratio	0.21	44%	7%	25%	24%
Belgium	R^2_{MZ} IP only	0.49	46%	11%	19%	23%
	R^2_{MZ} MD only	0.47	25%	15%	37%	23%
	R^2_{MZ} Cons. Pov.	0.69	44%	10%	28%	18%
	R^2_{LR} Likelihood Ratio	0.32	43%	9%	26%	22%
Bulgaria	R^2_{MZ} IP only	0.43	61%	28%	3%	8%
	R^2_{MZ} MD only	0.22	35%	47%	6%	12%
	R^2_{MZ} Cons. Pov.	0.66	51%	40%	4%	5%
	R^2_{LR} Likelihood Ratio	0.22	54%	36%	4%	7%
Switzerland	R^2_{MZ} IP only	0.31	38%	22%	9%	31%
	R^2_{MZ} MD only	0.63	7%	42%	23%	28%
	R^2_{MZ} Cons. Pov.	0.49	28%	15%	31%	26%
	R^2_{LR} Likelihood Ratio	0.19	35%	19%	11%	35%
Cyprus	R^2_{MZ} IP only	0.44	26%	24%	18%	32%
	R^2_{MZ} MD only	0.17	32%	30%	22%	16%
	R^2_{MZ} Cons. Pov.	0.42	44%	26%	19%	11%
	R^2_{LR} Likelihood Ratio	0.18	31%	21%	19%	29%
Czechia	R^2_{MZ} IP only	0.33	58%	15%	7%	20%
	R^2_{MZ} MD only	0.28	18%	40%	21%	21%
	R^2_{MZ} Cons. Pov.	0.58	35%	40%	14%	10%
	R^2_{LR} Likelihood Ratio	0.19	52%	13%	16%	19%
Germany	R^2_{MZ} IP only	0.38	48%	12%	23%	16%
	R^2_{MZ} MD only	0.38	20%	17%	41%	23%
	R^2_{MZ} Cons. Pov.	0.63	37%	18%	31%	14%
	R^2_{LR} Likelihood Ratio	0.26	47%	11%	24%	18%
Denmark	R^2_{MZ} IP only	0.29	22%	12%	7%	59%
	R^2_{MZ} MD only	0.34	14%	39%	25%	21%
	R^2_{MZ} Cons. Pov.	0.62	20%	14%	28%	39%
	R^2_{LR} Likelihood Ratio	0.20	24%	12%	15%	49%
Estonia	R^2_{MZ} IP only	0.30	74%	17%	4%	6%
	R^2_{MZ} MD only	0.15	27%	44%	11%	18%
	R^2_{MZ} Cons. Pov.	0.55	51%	32%	6%	11%
	R^2_{LR} Likelihood Ratio	0.17	65%	18%	7%	10%
Greece	R^2_{MZ} IP only	0.29	49%	30%	2%	18%
	R^2_{MZ} MD only	0.22	17%	49%	13%	20%
	R^2_{MZ} Cons. Pov.	0.45	45%	39%	2%	14%
	R^2_{LR} Likelihood Ratio	0.14	45%	30%	7%	19%

Spain	R^2_{MZ} IP only	0.38	55%	18%	11%	16%
	R^2_{MZ} MD only	0.27	23%	33%	18%	26%
	R^2_{MZ} Cons. Pov.	0.65	48%	28%	12%	12%
	R^2_{LR} Likelihood Ratio	0.22	49%	21%	13%	17%
Finland	R^2_{MZ} IP only	0.35	47%	20%	14%	19%
	R^2_{MZ} MD only	0.39	18%	29%	29%	24%
	R^2_{MZ} Cons. Pov.	0.56	43%	13%	29%	15%
	R^2_{LR} Likelihood Ratio	0.24	44%	14%	21%	21%
France	R^2_{MZ} IP only	0.39	37%	22%	23%	19%
	R^2_{MZ} MD only	0.37	26%	12%	39%	23%
	R^2_{MZ} Cons. Pov.	0.61	40%	13%	29%	18%
	R^2_{LR} Likelihood Ratio	0.25	41%	15%	25%	19%
Croatia	R^2_{MZ} IP only	0.42	81%	9%	1%	9%
	R^2_{MZ} MD only	0.17	15%	47%	20%	19%
	R^2_{MZ} Cons. Pov.	0.63	65%	29%	2%	5%
	R^2_{LR} Likelihood Ratio	0.20	69%	15%	6%	10%
Hungary	R^2_{MZ} IP only	0.44	63%	27%	3%	7%
	R^2_{MZ} MD only	0.20	27%	41%	17%	15%
	R^2_{MZ} Cons. Pov.	0.63	51%	34%	6%	9%
	R^2_{LR} Likelihood Ratio	0.21	55%	27%	8%	10%
Ireland	R^2_{MZ} IP only	0.43	78%	7%	6%	9%
	R^2_{MZ} MD only	0.31	49%	12%	20%	20%
	R^2_{MZ} Cons. Pov.	0.62	65%	9%	12%	14%
	R^2_{LR} Likelihood Ratio	0.21	65%	8%	12%	15%
Iceland	R^2_{MZ} IP only	0.19	33%	5%	27%	35%
	R^2_{MZ} MD only	0.38	22%	17%	14%	47%
	R^2_{MZ} Cons. Pov.	0.94	5%	12%	23%	59%
	R^2_{LR} Likelihood Ratio	0.17	26%	6%	32%	36%
Italy	R^2_{MZ} IP only	0.36	61%	16%	8%	15%
	R^2_{MZ} MD only	0.21	33%	30%	26%	11%
	R^2_{MZ} Cons. Pov.	0.52	48%	25%	17%	10%
	R^2_{LR} Likelihood Ratio	0.17	53%	18%	16%	13%
Lithuania	R^2_{MZ} IP only	0.38	50%	31%	5%	13%
	R^2_{MZ} MD only	0.25	33%	34%	19%	14%
	R^2_{MZ} Cons. Pov.	0.59	49%	39%	7%	6%
	R^2_{LR} Likelihood Ratio	0.19	52%	27%	11%	10%
Luxembourg	R^2_{MZ} IP only	0.42	33%	24%	21%	23%
	R^2_{MZ} MD only	0.52	4%	47%	6%	43%
	R^2_{MZ} Cons. Pov.	0.62	29%	11%	42%	18%
	R^2_{LR} Likelihood Ratio	0.26	31%	20%	26%	23%
Latvia	R^2_{MZ} IP only	0.43	60%	24%	3%	13%
	R^2_{MZ} MD only	0.23	26%	35%	14%	25%
	R^2_{MZ} Cons. Pov.	0.62	47%	31%	9%	14%
	R^2_{LR} Likelihood Ratio	0.21	52%	23%	9%	16%
Malta	R^2_{MZ} IP only	0.54	50%	39%	1%	10%
	R^2_{MZ} MD only	0.19	44%	36%	15%	5%
	R^2_{MZ} Cons. Pov.	0.67	57%	27%	6%	11%
	R^2_{LR} Likelihood Ratio	0.22	62%	20%	6%	12%
The Netherlands	R^2_{MZ} IP only	0.28	34%	15%	20%	31%
	R^2_{MZ} MD only	0.41	17%	28%	38%	17%
	R^2_{MZ} Cons. Pov.	0.58	30%	5%	35%	30%
	R^2_{LR} Likelihood Ratio	0.22	34%	9%	31%	26%

Norway	R^2_{MZ} IP only	0.31	35%	6%	23%	36%
	R^2_{MZ} MD only	0.63	18%	48%	14%	20%
	R^2_{MZ} Cons. Pov.	0.55	29%	18%	16%	37%
	R^2_{LR} Likelihood Ratio	0.30	29%	10%	30%	31%
Poland	R^2_{MZ} IP only	0.23	37%	48%	1%	14%
	R^2_{MZ} MD only	0.24	20%	48%	21%	11%
	R^2_{MZ} Cons. Pov.	0.47	39%	42%	12%	6%
	R^2_{LR} Likelihood Ratio	0.14	38%	36%	15%	11%
Portugal	R^2_{MZ} IP only	0.37	61%	25%	4%	10%
	R^2_{MZ} MD only	0.37	11%	42%	34%	13%
	R^2_{MZ} Cons. Pov.	0.62	50%	27%	16%	7%
	R^2_{LR} Likelihood Ratio	0.21	50%	19%	20%	11%
Romania	R^2_{MZ} IP only	0.24	26%	55%	3%	16%
	R^2_{MZ} MD only	0.12	9%	62%	16%	13%
	R^2_{MZ} Cons. Pov.	0.43	12%	76%	1%	11%
	R^2_{LR} Likelihood Ratio	0.10	17%	58%	7%	18%
Sweden	R^2_{MZ} IP only	0.36	35%	9%	24%	32%
	R^2_{MZ} MD only	0.47	26%	23%	14%	37%
	R^2_{MZ} Cons. Pov.	0.50	47%	3%	22%	28%
	R^2_{LR} Likelihood Ratio	0.27	39%	6%	25%	30%
Slovenia	R^2_{MZ} IP only	0.40	54%	25%	12%	10%
	R^2_{MZ} MD only	0.25	17%	50%	20%	14%
	R^2_{MZ} Cons. Pov.	0.60	44%	36%	11%	8%
	R^2_{LR} Likelihood Ratio	0.24	52%	17%	13%	18%
Slovakia	R^2_{MZ} IP only	0.34	66%	13%	6%	15%
	R^2_{MZ} MD only	0.15	42%	9%	26%	23%
	R^2_{MZ} Cons. Pov.	0.62	50%	35%	8%	8%
	R^2_{LR} Likelihood Ratio	0.21	64%	13%	10%	13%
United Kingdom	R^2_{MZ} IP only	0.24	62%	16%	8%	13%
	R^2_{MZ} MD only	0.46	25%	17%	45%	14%
	R^2_{MZ} Cons. Pov.	0.50	42%	13%	36%	10%
	R^2_{LR} Likelihood Ratio	0.20	45%	11%	30%	14%

Source: EU-SILC (2012) cross-sectional data, authors' computation. *Note*: 'IP only' , 'MD only' and 'Cons. Pov.' denote, respectively, 'income poverty only', 'severe material deprivation only' and 'consistent poverty'.

1.5.1.2. Within-country explanatory power across countries

The *variation in the within-country explained variance measures across European countries* is very large for the risk 'material deprivation only', and more limited for the income poverty categories ('only' and 'consistent poverty'). The first one varies between 12 (Romania) and 63 per cent (Norway). The model is much more effective in explaining the risk of material deprivation among the non-income poor in Western and Northern European countries than in the Southern, Central and Eastern European countries. In the latter countries, differences in socio-economic characteristics of households play a much smaller role in explaining the risk of material deprivation amongst the non-income poor. Presumably, this is because the general standard of living is lower than in the former countries and, as a consequence, everyone has a greater likelihood of being severely materially deprived. It is more difficult to establish some cross-country patterns regarding the effectiveness of the model for the income poverty categories (‘only’ and ‘consistent poverty’). The explanatory power of the model for the ‘income poverty only’ category varies between 19 (Iceland) and 54 (Malta) per cent. The model is most effective for Belgium, Bulgaria, Cyprus, Croatia, Hungary, Ireland, Luxembourg, Latvia and Malta and least effective for Austria, Denmark, Greece, Estonia, Iceland, Poland, Romania, the Netherlands and the United Kingdom. The explained variance measure of the ‘consistent poverty’ category varies between 42 (i.e. Cyprus) and 94 (i.e. Iceland) per cent. The model is most effective for Austria, Belgium, Bulgaria, Spain, Iceland and Malta, while it is least effective for Switzerland, Cyprus, Greece, Italy, Poland, Romania, Sweden and the United Kingdom.

1.5.1.3. Households' short-term ability to generate resources on the labour market (work intensity)

The work intensity variable is, unsurprisingly, extremely important in predicting the outcome of all categories of the dependent variable. It is, however, much more important for the income poverty categories (‘only’ and ‘consistent poverty’) than for the ‘material deprivation only’ category. Living in a household in which all members are fully employed, as compared to a household where no one works, multiplies, respectively, the relative risk to be ‘income poor only’, ‘consistently poor’ and ‘materially deprived only’ (with not being income poor, nor materially deprived as the reference) by a factor of 0.02, 0.07 and 0.3 (see Model M2 in Table 1.A3 in Appendix 1). The Shapley decomposition in Table 1.2 further reveals that the work intensity variable is extremely effective in explaining within-country differences in the risk of income poverty (‘only’ and ‘consistent poverty’), but much less effective for the risk of ‘material deprivation only’. Specifically, the work intensity variable makes, on average across countries, a relative contribution of 49 per cent to the explained variance in the risk of ‘income poverty only’, a relative contribution of 41 per cent to the explained variance in the risk of ‘consistent poverty’, and a relative contribution of only 24 per cent to the explained variance in the risk of ‘material deprivation only’. The Shapley decomposition also reveals some diverging patterns across countries in the relative contributions of a household’s work intensity to the R^2-measures. Work intensity is very effective in explaining within-country differences in the risk of income poverty/material deprivation in some Central and Eastern European countries (i.e. Bulgaria, Czech Republic, Estonia, Croatia, Hungary, Latvia, Lithuania, Slovenia and Slovakia), but also in Italy, Ireland and Malta. It is least effective in countries like Switzerland, Denmark, Iceland, Luxembourg and Norway. This result seems to imply that work-poor households are better protected from income poverty/material deprivation in countries with a strong welfare state and/or in the more affluent countries.

1.5.1.4. Households' long term ability to generate resources on the labour market (educational dummies)

The results indicate that even though level of education is strongly associated with the prevalence of income poverty, it is a much stronger determinant in predicting material deprivation. The high education variable has exponentiated coefficients of 0.42 for the 'income poor only' category, of 0.39 for the 'materially deprived only' category and of 0.23 for the 'consistently poor' category. The ratio between the probability to be 'income poor only', 'materially deprived only' and 'consistently poor' and the probability of not being income poor, nor materially deprived is, respectively, multiplied by a factor of 1.61, 1.6 and 2.91 if the household is lowly educated, compared to if the household is moderately educated (see Model M2 in Table 1.A3 in Appendix 1). In fact, of all variables included in the model, the educational dummies are the most effective in explaining within-country differences in the 'risk of material deprivation only', while they are the second most effective for the other two income poverty categories ('only' and 'consistent poverty'). Specifically, the education variables make a relative contribution of 34 per cent to the explained variation in the risk of 'material deprivation only', a relative contribution of 20 per cent to the explained variation in the risk of 'income poverty only', and a relative contribution of 25 per cent to the explained variation in the risk of 'consistent poverty'. Furthermore, it is interesting to note that there is a clear difference in the relative impact of the educational variables on the explained within-country. The educational dummies play a much smaller role in Western and Northern European countries (e.g. Austria, Belgium, Czech Republic, Germany Denmark, Finland, Ireland, Iceland, the Netherlands, Norway, Sweden, Slovakia and the United Kingdom), while they are very effective in explaining the risk of income poverty/material deprivation in Southern, Central and Eastern European countries (e.g. Bulgaria, Greece, Hungary, Lithuania, Poland, Romania and Slovenia). A reasonable explanation for these results is that higher education is scarcer in less affluent countries and thus more valuable on the labour market.

1.5.1.5. Household costs (tenant, bad health)

Differences in households' *costs* form another important theoretical explanation for the limited overlap between income poverty and severe material deprivation within countries. The bad health variable has, respectively, exponentiated coefficients of 0.85, 1.86 and 1.32 for the risk of 'income poverty only', 'material deprivation only' and 'consistent poverty' (see Model M2 in Table 1.A3 in Appendix 1). The exponentiated coefficient smaller than one for the 'income poverty only' category may seem odd at first sight, yet can be rationally explained. The income loss due to not working because of having poor health is controlled for through the work intensity variable. At the same time, losses in a household's income can be partially compensated for by social transfers. Another interesting result is that the relative risk ratios of the deprivation categories are increased by a stronger factor for tenants than for house owners (with exponentiated coefficients of 2.57 for the ‘material deprivation only’ category, 3.48 for the ‘consistent poverty’ category and 1.8 for the ‘income poverty only’ category). Clearly, the additional health care cost of having a household member with (very) bad health or being a tenant puts pressure on the household budget, which explains the strong association with the prevalence of material deprivation. In fact, these two cost variables make, respectively, on average across countries, a relative contribution of 22 and 18 per cent to the explained within-country differences in the risk of ‘material deprivation only’ and ‘consistent poverty’. The two dummies are least effective for the ‘income poverty only’ category, with a relative contribution of 11 per cent to the explained within-country variance. Variables that capture costs play a much larger role in the more affluent countries (e.g. Austria, Belgium, Germany, Finland, France, Iceland, the Netherlands, Italy, Luxembourg, the Netherlands and the United Kingdom), as compared to the less affluent countries (e.g. Bulgaria, Estonia, Greece, Croatia, Hungary, Lithuania, Latvia, Malta, Romania). A plausible interpretation lies in the fact the households living in more affluent countries also face higher costs of living.

1.5.1.6. Households’ socio-demographic diversity (household structure and other demographic variables)

The results indicate that a household’s composition and structure have a statistically significant relationship with the risk of income poverty/material deprivation. Couples with more than two children, singles and single parents are much more likely to be socially excluded, as compared to couples without children. Two-adult families with one or two children have a higher risk of being income poor (with, respectively, exponentiated coefficients of 1.72 and 2.09 for the ‘income poverty only’ and the ‘consistent poverty’ categories), but the increased risk of being ‘materially deprived only’ is almost negligible (with an exponentiated coefficient of 1.08), as compared to couples without children. The other socio-demographic variables include the age and gender of the household’s reference person and the household’s ‘migration background’. The presence of a female household head slightly increases the risk of being ‘only materially deprived’ and ‘consistently poor’ (with exponentiated coefficients of 1.46 and 1.31, respectively), while the association with the ‘income poverty only’ category is negligible. These results might be lower than one would initially expect, but it is important to note that we already controlled extensively for other socio-economic characteristics, such as single parenthood. Living in a family with a young household head (younger than 29) is the factor most strongly related to the ‘income poverty only’ category (with an exponentiated coefficient of 1.56). Indeed, young people, at the start of their career, often have to rely on a lower income than people from older age groups. Having a young household head is less strongly associated with the risk of ‘material deprivation only’ and ‘consistent poverty’ (with exponentiated coefficients of 1.21 and 1.3, respectively). Socio-demographic variables make a relative contribution to the explained within-country differences of 19 per cent for the risk ‘income poverty only’ category, of 20 per cent for the risk of ‘material deprivation only’, and of 16 per cent for the risk of ‘consistent poverty’. Variables that capture households’ socio-demographic diversity are much more important in the Scandinavian countries, Switzerland, and the Netherlands, than in most Central and Eastern European countries.

1.5.2. Country-level determinants (between-country analysis)

We use the multilevel structure to account for between-country differences in the risks of income poverty and severe material deprivation. Firstly, we assess the r*elative importance of the country-level for the different categories of the dependent variable by comparing the estimated variance in random intercepts and VPC measures*. Secondly, we *compare the effectiveness of the household-level and country-level variables in explaining between-country differences*. Thirdly, we discuss the *in-cash versus in-kind dichotomy within social spending* by regressing both variables and comparing their relative effectiveness in explaining between-country differences. Fourthly, we *compare the explanatory power of the pro-poorness and the size of social spending levels* on country-to-country differences. Finally, we *contrast the relative importance of macroeconomic and institutional* variables on between-country differences in the risk of material deprivation.

The results of the random intercept multilevel model are shown in Table 1.3 Model 1 (M1) is an empty model, while Model 2 (M2) includes the household-level variables (see also Table 1.A3 in Appendix 1). Models 3 to 6 are ‘full’ models that include the household-level variables, two macroeconomic variables (i.e. *median income*, *unemployment rate*) and an institutional variable (i.e. *total social benefit levels* (M3), *in-cash social benefit levels* (M4), *in-kind social benefit levels* (M5), *pro-poorness bottom 50* (M6)). Model 7 is a ‘full’ model that include the household-level variables, two macroeconomic variables (i.e. *median income, unemployment rate*), and two institutional variables (i.e. *total social benefit levels* and *pro-poorness bottom 50* (M7)).

We set the chain length to 5000 iterations, with a burn-in length of 500 iterations that are not used to describe the final parameter distribution (i.e. these iterations are only used to initialise the Markov chains). The iteration traces show that the sampler is mixing well, as no large white patches are visible on the traces for all estimated parameters and Deviance criteria. We also ran more than sufficient chains to satisfy the Rafter-Lewis criterion in all multilevel models (Raftery & Lewis, 1992). The Raftery-Lewis diagnostic reflects the lengths of the Markov chain required to estimate a given quantile of the posterior distribution to a given accuracy. The diagnostic was satisfied for all country-level variables and random intercepts with quantiles and accuracy set at 2.5% and 97.5% and 0.005, respectively. This means that the actual Monte Carlo coverage of the nominal 95% interval estimate for the parameters should differ by no more than 1% point with a Monte Carlo probability of 95%). In addition, all country-level variables and random intercepts show high Effective Sample Sizes.

1.5.2.1. Importance of the country-level for income poverty and severe material deprivation

We analyse the importance of the unobserved country-level residuals for the different categories of the dependent variables by comparing the variance in random intercepts and VPC measures. Firstly, the variance in random intercepts of all models are statistically significant for all categories of the dependent variables, indicating that considerable unobserved differences in the risk of income poverty and severe material deprivation exist between countries, even after controlling for the household-level and country-level variables (Table 1.3). The variance in random intercept of the 'empty' model are much larger for the risk of 'material deprivation only' (i.e. 1.78) and the risk of 'consistent poverty' (i.e. 1.69), as compared to the risk of 'income poverty only' (i.e. 0.22) (M1, Table 1.3). In addition, the VPC measure indicates that a large share of the unexplained variance in the risk of material deprivation ('only' and 'consistent poverty') can be attributed to the country-level. For the 'risk of income poverty only', the share of the unexplained country-level variance in the total unexplained variance is much more limited. Specifically, the VPC measures of the 'empty' model equals 35 per cent for the risk of 'material deprivation only', 34 per cent for the risk of 'consistent poverty', and only 6 per cent for the risk of 'income poverty only'.

1.5.2.2. Comparing the effectiveness of the household-level and country-level variables in explaining between-country differences

After introducing the household-level variables (M2) to the 'empty' model (M1), the unexplained between-country differences for the risk of income poverty 'only' decreases, whereas the unexplained between-country differences for the risk of material deprivation ('only' and 'consistent poverty') increases. Specifically, the variance in random intercepts decreases by 20 per cent for the risk of income poverty 'only' (i.e. from 0.22 in M1 to 0.18 in M2) and increase by 11 and 38 per cent for the risk of material deprivation 'only' and 'consistent poverty', respectively (i.e. from 1.78 in M1 to 1.98 in M2 and from 1.69 in M1 to 2.34 in M2). These figures indicate that the association between the household-level risk factors and material deprivation differs quite strongly across countries. This result does not come as a surprise, as we found a strong cross-country variation in the explanatory power of the household-level variables for the material deprivation categories ('only' and 'consistent poverty') in the single level models.

The DIC criteria of the household-level variable model (M2) and the full models (M3-M7) are highly similar, indicating that these models show a similar balance between the fit and complexity of the model to the data. However, models including country-level variables (M3-M7) form an important contribution to the 'micro' model (M2) (i.e. containing only household-level variables), as they are much more effective in explaining unobserved country-to-country differences than the household-level variables. Specifically, the variance in random intercepts decreases by 52 per cent for the risk of income poverty 'only' (i.e. from 0.18 in M2 to 0.09 in M7), by 73 per cent for the risk of material deprivation 'only' (i.e. from 1.98 in M2 to 0.53 in M7), and by 84 per cent for the risk of 'consistent poverty' (i.e. from 2.34 in M2 to 0.37 in M7). It is clear from this result that the contextual variables are very effective in explaining the unobserved between-country differences in the risk of material deprivation ('only' and 'consistent poverty'), whereas their impact on the between-country differences in the risk of 'income poverty only' is more limited.

Table 1.3 – Multinomial logistic multilevel model, country-level variables

	Model 1 (empty)			Model 2 (micro)			Model 3 ('full')			Model 4 ('full')		
	IP only	MD only	Cons. Pov.	IP only	MD only	Cons. Pov.	IP only	MD only	Cons. Pov.	IP only	MD only	Cons. Pov.
Country variables												
Median income							-0	-0.15***	-0.15***	0	-0.18***	-0.16***
Unemployment rate							0.01	0.02	0.03	0.02*	0.02	-0.03
Total social benefit levels							-0.07***	-0.07	-0.11**			
In-cash social benefits levels										-0.16***	-0.01	-0.19
In-kind social benefit levels												
Pro-poorness bottom 50												
Variance measures												
Random intercept	0.22***	1.78***	1.69***	0.18***	1.98***	2.34***	0.1***	0.53***	0.5***	0.1***	0.57***	0.54***
Explained country variance							0.44	0.73	0.79	0.47	0.71	0.77
VPC	0.06	0.35	0.34	0.05	0.38	0.42	0.03	0.14	0.13	0.03	0.15	0.14
Model diagnostics												
R^2_{LR} Likelihood Ratio		0.07			0.24			0.24			0.24	
DIC		662,223.63			545200.95			545,200.92			545,202.35	
Effective number of parameters (pD)		92.18			131.08			130.84			131.51	
N		492,122			492,122			492,122			492,122	
Number of countries		31			31			31			31	

Source: EU-SILC (2012) cross-sectional data, authors' computation. *Note*: The reference category of the dependent variable is 'not being poor, nor severely materially deprived'. * Significant at 10% level, **Significant at 5% level, ***Significant at 1% level. The household-level variables are regressed in M2-M8, but not shown (all coefficients are similar to the coefficients of the 'micro' model (M2) in Table 1.A3). 'IP only', 'MD only' and 'Cons. Pov.' denote, respectively, 'income poverty only', 'severe material deprivation only' and 'consistent poverty'

Table 1.3 – Multinomial logistic multilevel model, country-level variables (continued)

	Model 5 ('full')			Model 6 ('full')			Model 7 ('full')		
	IP only	MD only	Cons. Pov.	IP only	MD only	Cons. Pov.	IP only	MD only	Cons. Pov.
Country variables									
Median income	-0.02	-0.15***	-0.17***	-0.01	-0.16***	-0.16***	0	-0.15***	-0.15***
Unemployment rate	0.08	0.02	-0.02	0.01	0.02	0.01	0.01	0.02	0.02
Total social benefit levels							-0.04	-0.04	-0.02
In-cash social benefits levels									
In-kind social benefit levels	-0.06*	-0.12*	-0.13*						
Pro-poorness bottom 50				-2.04***	-2.17	-4.69***	-1.55**	-1.64	-4.39***
Variance measures									
Random intercept	0.11***	0.5***	0.52***	0.09	0.52	0.36	0.09	0.53	0.37
Explained country variance	0.37	0.75	0.78	0.51	0.74	0.85	0.52	0.73	0.84
VPC	0.03	0.13	0.14	0.03	0.14	0.10	0.03	0.14	0.10
Model diagnostics									
R^2_{LR} Likelihood Ratio	0.24			0.24			0.24		
DIC	545201.74			545202.08			545201.51		
Effective number of parameters (pD)	131.22			131.25			130.90		
N	492122			492122			492122		
Number of countries	31			31			31		

Source: EU-SILC (2012) cross-sectional data, authors' computation. *Note*: The reference category of the dependent variable is 'not being poor, nor severely materially deprived'. * Significant at 10% level, **Significant at 5% level, ***Significant at 1% level. The household-level variables are regressed in M2-M8, but not shown (all coefficients are similar to the coefficients of the 'micro' model (M2) in Table 1.3). 'IP only', 'MD only' and 'Cons. Pov.' denote, respectively, 'income poverty only', 'severe material deprivation only' and 'consistent poverty'.

Table 1.4 – Shapley decompositions of the country-level variables on the explained between-country differences

		Explained country variance	Decomposition of explained part of the between country variance					
			Median income	Unemployment rate	Total social benefit levels	Cash social benefit levels	Inkind social benefit levels	Pro-poorness bottom 50
M3 ('full')	IP only	0.44	0.13 (29)%	0.04 (10)%	0.27 (61)%			
	MD only	0.73	0.43 (59)%	0.1 (13)%	0.2 (28)%			
	Cons. Pov.	0.79	0.44 (56)%	0.09 (12)%	0.25 (32)%			
M4 ('full')	IP only	0.47	0.13 (28)%	0.06 (13)%		0.28 (59)%		
	MD only	0.71	0.46 (65)%	0.1 (15)%		0.15 (21)%		
	Cons. Pov.	0.77	0.44 (57)%	0.1 (13)%		0.23 (30)%		
M5 ('full')	IP only	0.37	0.16 (43)%	0.04 (10)%			0.17 (47)%	
	MD only	0.75	0.47 (64)%	0.1 (13)%			0.17 (23)%	
	Cons. Pov.	0.78	0.51 (65)%	0.09 (12)%			0.18 (23)%	
M6 ('full')	IP only	0.51	0.14 (28)%	0.03 (6)%				0.34 (66)%
	MD only	0.74	0.49 (67)%	0.09 (13)%				0.15 (20)%
	Cons. Pov.	0.85	0.48 (57)%	0.09 (10)%				0.25 (30)%
M7 ('full')	IP only	0.52	0.09 (18)%	0.03 (6)%	0.17 (32)%			0.23 (44)%
	MD only	0.73	0.39 (53)%	0.09 (12)%	0.15 (21)%			0.1 (14)%
	Cons. Pov.	0.84	0.38 (45)%	0.08 (9)%	0.18 (21)%			0.21 (25)%

Source: Calculations from EU-SILC (2012) cross-sectional data, authors' computation. *Note*: 'IP only' , 'MD only' and 'Cons. Pov.' denote, respectively, 'income poverty only', 'severe material deprivation only' and 'consistent poverty'.

1.5.2.3. In-cash versus in-kind dichotomy within social spending

We now test the impact of different country-level variables on the deprivation and income poverty risks (M3-M7). The DIC criteria indicate that all full models have a very similar balance between the fit and complexity of the models to the data. This result can be explained by the fact the model specifications are highly similar (i.e. the models are exactly the same with the exception of one or two variables, which are, in turn, correlated with each other). To compare the effect of the country-level variables, we will use the usual econometric approach of identifying significant relations and their signs and complement it with decompositions of the explained between-country variance measures.

Average spending on social benefits in proportion of GDP (i.e. including both in-kind and in-cash benefits) is negatively associated with the risk of income poverty ('only' and 'consistent poverty'), but not with the risk of 'material deprivation only', after controlling for median income and the unemployment rate (Table 1.3, M3). In other words, living in a country with higher social spending levels reduces the risk of severe material deprivation only for those individuals that also have a low income ('consistent poverty'). Models 4 and 5 give a more nuanced picture to this result. These models show that in-cash benefits are negatively associated with the risk of 'income poverty only', while in-kind benefit levels are marginally negatively associated with all categories of the dependent variable. These results thus show that it is important to distinguish between cash and in-kind social spending, as their relationship differs across the various categories of the social exclusion target group. In fact, it nuances the use of a global variable that combines cash and in-kind transfers, as the significancy of the relationship with social spending may change if a single spending component (cash or in-kind) is regressed (i.e. the relationship of total social spending with 'consistent poverty' (M3) loses its significancy if only the cash benefit component is regressed (M4); the insignificant relationship of total spending levels with 'material deprivation only' (M3) becomes significant if only the in-kind component is regressed (M5)). The insignificant relationship of in-cash social spending levels with the risk of material deprivation ('only' and 'consistent poverty') can be explained by the fact that material deprivation is influenced by the general standard of living in the country (median income) and the provision of in-kind services which decreases costs, and not specifically by the composition of median income (market income versus cash transfers)[24]. In-kind transfer levels are also negatively associated with the risk of income poverty ('only' and 'consistent poverty'). The Shapley decompositions reveal that in-cash spending is more effective in explaining between-country differences than in-kind benefits (M4-M5, Table 1.4) for the risk of 'income poverty only' as well as for the risk of 'consistent

[24] When median income levels and the unemployment rate are not co-regressed, in-cash social benefit levels have a strong statistically significant negative relationship with all the categories of the dependent variable.

poverty'. Specifically, cash and in-kind spending levels explain, respectively, 28 and 17 per cent of between-country differences in the risk of 'income poverty only', 23 and 18 per cent of the between-country differences in the risk of 'consistent poverty', and 15 and 17 per cent of the between-country differences in the risk of 'material deprivation only'.

A plausible interpretation of all these results is that cash transfers directly protect the income of households from setbacks, whereas in-kind services indirectly boost the household budget by decreasing costs. This can explain why cash transfer levels are effective in explaining country differences in the risk of income poverty, and why in-kind spending explains a larger amount of the between-country differences in the 'risk of material deprivation only', as services such as health care and social housing, support the permanent income of the non-income poor households. The fact that in-kind benefits decrease the risk of income poverty can be explained by the fact that they include unemployment benefits (mobility and resettlement, vocational training, placement services and job-search assistance) which can have a direct impact on household income, by increasing the chance of finding a job for those unemployed.

1.5.2.4. Size versus pro-poorness of social spending

The results indicate that pro-poor social transfers have a statistically significant and strong negative association with the risk of income poverty ('only' and 'consistent poverty'), but not with the risk of 'material deprivation only' (Table 1.3, M6). In other words, pro-poorness has an impact on 'material deprivation', but only for those individuals that also have a low income ('consistent poverty'). A plausible interpretation of this finding is that the more pro-poor the transfer system becomes, the less likely individuals in the category 'material deprivation only' are to receive cash transfers (by definition, as they do not have a low income). The relationship between pro-poorness and income poverty ('only' and 'consistent poverty') remains statistically significant after controlling for the total social spending variables (Table 1.3, M7). However, total social benefit spending loses its significancy with income poverty ('only' and 'consistent poverty') when pro-poorness is co-regressed, which is likely to be a multicollinearity issue (Table 1.3, M3 and M7). To compare the relative importance of both variables, we analyse the Shapley decomposition of the explained between-country variance measure of the model that includes Pro-poorness bottom50 and Total social benefit levels (Table 1.4, M7). The results indicate that, once differences in median income are controlled for, the pro-poorness of social spending explain a bit more between-country differences than the size of total social spending for the risk of income poverty ('only' and 'consistent poverty'), although both variables are important. Specifically, the pro-poorness of cash transfers and the level of total spending on social benefits explain, respectively, 23 and 17 per cent of the between-country differences in the risk of 'income poverty only', and 21 and 18 per cent of the between-country differences in the risk of 'consistent poverty' (Table 1.4, M7). This indicates that the pro-poorness of social benefits is not only an effective tool to combat income poverty, but also that it should not be understood independently from the size of social spending. This in line with the recommendation in the recent literature of "targeted universalism" (i.e. universalism with a large degree of targeting of the poorest).[25]

1.5.2.5. Relative importance of macroeconomic and institutional variables on between-country differences

Finally, we establish whether *median income levels or institutional variables* are the most effective factors to explain between-country differences in the risk of material deprivation. Kenworthy and co-authors (2011) showed that material deprivation rates are correlated to a country's social policy generosity, but not to its level of affluence. Other researchers have found a negative relationship between aggregate income levels and material deprivation (Whelan & Maître, 2012; Bárcena-Martín et al., 2014; Visser et al., 2014). We regressed both types of variables simultaneously in models 3 to 7 (M3-M7) in Table 1.4. The results indicate that both median income levels and institutional variables have a statistically significant and negative association with the risk of material deprivation ('only' and 'consistent poverty'). These variables capture differences in household income between countries, as well as other potential other missing variables in the model that could impact the deprivation risk (quality of services, level of wealth, etc.). To compare the relative importance of the macroeconomic and institutional variables, we again turn to the Shapley decomposition of the explained between-country variance (Table 1.4, M3-M7). The results reveal that median income levels explain the largest share of the between-country differences in the risk of material deprivation for all models. However, the results from the final model show that the institutional variables (i.e. including total spending on social benefits and the pro-poorness of cash transfers) are equally effective as median income levels in explaining between-country differences for the risk of 'consistent poverty', while median income levels explain much more between-country differences than the institutional variables for the risk of 'material deprivation only' (Table 1.4, M7). Specifically, the median income levels and the institutional variables explain, respectively, 38 and 39 (18 plus 21) per cent of the between-country differences in the risk of 'consistent poverty' and 39 and 25 (15 plus 10) per cent of the between-country differences in the risk of 'material deprivation only'. This result somehow contradicts the hypothesis of

[25] We also used an alternative specification of the pro-poorness measure to verify the robustness of our results (*Pro-poorness bottom20*). The alternative pro-poorness variable is defined as the share of transfers that is distributed to the lowest two deciles in the pre-transfer household income distribution, instead of the share that is distributed to the broader bottom five deciles. The results we obtained using *Pro-poorness bottom20* were similar to the results obtained using *Pro-poorness bottom50*.

Kenworthy and co-authors (2011). The difference in results is most likely because a larger group of countries – and, specifically, less affluent countries – were selected in our analysis. Finally, the unemployment rate explains approximately 9 per cent of the between-country differences in the risk of material deprivation ('only' and 'consistent poverty').

1.6. Conclusion

In this chapter we investigated the differences between the risk factors of income poverty and severe material deprivation by assessing to what extent both indicators are subject to the same determinants. Given that these two indicators encompass the majority of the Europe 2020 social exclusion target group, it is crucial for policy makers to better know which determinants are effective in explaining the differences within and between countries in the risk of both types of problems.

This paper is innovative on several methodological aspects. Firstly, the estimation strategy allowed us to provide an extensive overview of the household-level and country-level risk factors of three distinct groups of people in the Europe 2020 social exclusion target group, i.e. those that are 'income poor only', 'severely materially deprived only' and 'consistently poor'. Secondly, we combined for the first time both a single level and multilevel regression approach to predict the risk of income poverty and material deprivation. We argued that both methods have their strengths and weaknesses and should – ideally – be combined to offer a comprehensive understanding of the policy drivers needed to reach the EU social exclusion target in each EU country. In the single level models we analysed the effectiveness of the household-level variables in explaining the within-country differences for different European countries. In the multilevel model the focus lied on the effectiveness of the contextual variables in explaining unobserved between-country differences. Thirdly, we employed the Shapley decomposition method to establish the relative contribution of each explanatory variable to within and between country explained variance measures. This approach allowed us to obtain quantified knowledge on the (in)effectiveness of the regressed independent variables, in addition to the usual econometric approach of identifying signs and significances of relevant variables.

The results from the single level models showed that there are considerable differences in the explanatory power of the household-level variables across the categories of the dependent variable and across countries. The household-level variables were the most effective for the 'consistent poverty' category (i.e. explaining on average 59 of the within-country differences across European countries versus 36 per cent of for the risk of 'income poverty only' and 32 per cent for 'material deprivation only'). This indicates that the predictive power of the socio-economic characteristics at the household-level are higher for those cumulating both income poverty and severe material deprivation, than for those suffering from 'only' one of these problems. In terms of cross-country variation of the within-country explanatory power of the household-level variables, we found that the differences in socio-economic characteristics of households are much more effective in explaining material deprivation among the non-income poor in Western and Northern European countries, than in Southern, Central and Eastern European countries.

The single level analysis further showed that the (relative) effectiveness of certain socio-economic household-level risk factors in explaining within-country differences differs strongly across European countries. On the one hand, we found that a household's work intensity, education have a large relative contribution to the explained within-country variance measures in Central and Eastern European countries. Their explanatory power is much more limited in Western and Northern European countries. On the other hand, a household's socio-demographic characteristics and costs have a much stronger association with income poverty/material deprivation in some Western and Northern European countries than in most Central and Eastern European countries. We argued that the association of the household-level risk factors is likely to be mediated by variables at country-level. We did not analyse these relationships, as they fall outside the scope of the paper. One could, however, address this issue by adding random slopes and cross-level interactions and applying Shapley decompositions to determine which cross-level interaction explains the largest part of the cross-country variation in the coefficients of the household-level variables. We leave this as an interesting avenue for future research.

We used the multilevel model to account for between-country differences in the risks of income poverty and severe material deprivation. We found that more than one third of the total unexplained variance in the risk of material deprivation ('only' and 'consistent poverty') is situated at the country-level. The country-level is less important for the risk of income poverty 'only' (only 6 per cent of the total unexplained variance is situated at the country-level). The macroeconomic and institutional variables explained about 80 per cent of the between-country differences in the risk of material deprivation ('only' and 'consistent poverty'), and about half of the between-country differences in the risk income poverty 'only', after controlling for the household-level variables.

We also tested and compared the effectiveness of different social spending concepts and macroeconomic variables in explaining between-country differences. Firstly, we showed that it is, in contrast to the usual approach in the multilevel poverty and deprivation literature, important to distinguish between cash and in-kind social spending. In-cash benefits are negatively associated with the risk of 'income poverty only', while in-kind benefit levels are negatively associated with all categories of the dependent variable (although in-kind social spending explain more between-country differences for risk of material deprivation (‘only’ and ‘consistent poverty’) than for the risk of ‘income poverty only’). Secondly, we found that living in a country in which social benefits are strongly targeted towards the poor significantly reduces the risk of income poverty (‘only’ and ‘consistent poverty’). The results further indicated that the pro-poorness of social spending explain more between-country differences than the size of social spending for the risk of income poverty (‘only’ and ‘consistent poverty’), although both variables are important and should, hence, not be understood independently of each other. This in line with the recommendation in the recent literature of “targeted universalism” (i.e. universalism with a large degree of targeting of the poorest). Thirdly, in all models we estimated, we found that median income levels are the most important country-level determinant for the risk of severe material deprivation. The role of institutional variables should, however, not be underestimated. In fact, total social spending levels and the pro-poorness of cash transfers together explain as much of the between-country differences for the risk of ‘consistent poverty’ as median income levels and a considerable (but smaller) share of the between-country differences for the risk of ‘material deprivation only’. Finally, it is important to note that we conducted a sensitivity analysis, in which we excluded the two most outlying countries, Romania and Bulgaria (one at a time and simultaneously). The results were found to be insensitive to the inclusion of both countries.

Some of the results obtained in this chapter should be read with the necessary caution. Firstly, an important aspect that should be further investigated is related to the way the multivariate structure of the indicators could be better analysed jointly with the multilevel structure of the data. While the multinomial multilevel model presented in this chapter is a first attempt in the literature to take this issue into account, other models (e.g. models taking into account the intensity of overlap between income poverty and severe material deprivation) come to mind. Secondly, this chapter mainly focussed on children and working age population, as households in which all members over the age of 59 were excluded from the analysis (the work intensity variable is not defined for people aged 60 years or more). While this approach allowed us to identify households’ labour market activity as a crucial driver of the risk of income poverty and severe material deprivation, it is clear that the conclusions of this chapter do not hold for the complete old age population. Thirdly, the country-level variables are effective in explaining cross-sectional between-country differences in the risk of income poverty/material deprivation. One should, however, be reluctant to use these results to explain the impact of changes in the independent variables on changes in the risk of income poverty/material deprivation. The determination of the risk of income poverty and material deprivation could be improved by moving towards a dynamic model. As Whelan and co-authors (2001, p. 364) suggest, longitudinal information can “allow us to examine the relationship between current and permanent income and the manner in which resources are accumulated and eroded”. In a multilevel context, it could prove interesting to analyse how changes in both country-level (e.g. economic crisis, changes in the pro-poorness of social spending) and household-level (e.g. moving into unemployment, divorce) variables over time affect the probability of moving in and out of material hardship and/or income poverty (for a first dynamic logistic multilevel study on income poverty see, Bosco & Poggi, 2016; for a first dynamic probit multilevel study on child poverty, see Bárcena-Martín et al., 2017a).

Appendix 1.

Table 1.A1 – Summary statistics of the independent household-level variables

Country	WI	Higher education (hh)	Lower education (hh	Bad Health (hh)	Rent	Couple, one or two children	Couple, more children
Austria	71.1	30.7	1.4	11.6	39.6	29.3	8.0
Belgium	66.8	52.5	13.5	12.3	27.6	28.7	11.4
Bulgaria	67.6	33.4	7.6	22.3	11.0	15.2	1.3
Switzerland	65.1	53.5	1.6	3.7	48.2	30.9	10.1
Cyprus	70.7	53.2	15.5	11.1	22.4	22.3	8.9
Czechia	76.2	26.7	0.2	12.1	17.7	28.5	4.4
Germany	70.5	50.8	2.8	10.0	41.7	32.5	6.3
Denmark	81.9	54.7	0.3	4.9	20.8	35.4	11.8
Estonia	66.2	42.7	4.8	17.6	13.9	23.3	6.9
Greece	57.5	35.9	24.6	13.2	19.5	29.0	6.4
Spain	61.3	46.5	18.8	12.7	18.1	29.8	4.2
Finland	72.8	56.2	0.0	3.4	18.7	30.3	13.8
France	74.3	46.3	11.6	10.7	35.8	32.4	11.7
Croatia	55.5	23.6	4.3	24.0	9.4	14.4	5.5
Hungary	65.3	29.5	4.0	23.6	9.5	22.6	6.9
Ireland	51.9	57.6	10.3	4.9	33.3	28.0	19.1
Iceland	79.3	48.5	1.2	3.0	18.0	26.7	14.6
Italy	65.3	25.7	8.9	14.3	25.1	29.6	4.5
Lithuania	70.4	45.4	4.5	22.8	6.3	18.9	3.1
Luxembourg	67.9	35.5	32.7	12.6	26.2	32.0	11.5
Latvia	66.1	39.4	2.6	22.6	18.7	20.7	4.6
Malta	62.5	27.3	13.3	5.2	15.6	26.3	5.1
The Netherlands	74.9	52.9	5.8	2.9	17.6	35.3	13.7
Norway	80.7	58.1	0.7	4.7	10.1	34.3	14.5
Poland	65.4	30.8	15.6	23.7	16.8	22.6	5.5
Portugal	70.9	23.4	53.9	25.1	25.4	28.2	3.4
Romania	69.6	21.4	5.1	17.7	3.1	18.6	2.3
Sweden	80.0	51.9	2.0	3.0	25.2	36.3	11.4
Slovenia	77.4	40.9	3.2	9.0	18.7	20.3	4.4
Slovakia	76.4	41.2	0.4	23.5	9.4	15.9	3.5
United Kingdom	67.6	48.3	0.0	11.6	36.0	31.0	10.2
Average	69.3	41.4	8.7	12.9	21.3	26.8	8.0

Source: EU-SILC (2012) cross-sectional data, authors' computation.

Table 1.A1 – Summary statistics of the independent household-level variables (continued)

Country	Single parent	Single, no children	Other hh	Young age, <29 (HRP)	Female (HRP)	Migrant (hh)
Austria	10.1	5.6	26.7	6.2	36.5	14.1
Belgium	7.4	6.0	28.8	4.7	38.2	17.9
Bulgaria	3.0	1.5	63.2	6.1	58.0	0.8
Switzerland	7.4	3.9	26.7	4.7	36.4	16.2
Cyprus	3.0	2.0	51.4	3.6	20.4	21.7
Czechia	5.6	3.7	38.1	3.8	33.0	1.3
Germany	11.1	5.0	21.3	3.7	37.0	10.9
Denmark	4.5	3.3	24.4	4.4	38.4	7.4
Estonia	3.8	2.6	47.2	6.5	47.1	18.8
Greece	4.1	1.8	43.3	1.7	18.9	9.3
Spain	3.7	2.4	46.0	1.7	34.4	10.1
Finland	6.4	3.5	23.6	7.4	36.0	3.1
France	6.6	5.6	25.0	8.7	37.5	11.8
Croatia	2.9	0.8	61.3	1.3	24.0	18.1
Hungary	4.9	3.1	44.7	2.5	38.6	0.5
Ireland	4.9	9.3	26.8	5.3	45.0	9.0
Iceland	3.2	4.5	39.5	4.9	38.1	6.7
Italy	6.2	3.0	42.6	2.0	33.2	8.8
Lithuania	4.9	2.8	49.3	2.5	48.4	11.5
Luxembourg	5.5	4.5	32.7	3.0	34.4	16.6
Latvia	5.5	4.8	45.7	5.0	56.2	22.1
Malta	2.6	2.0	51.2	1.5	39.4	9.7
The Netherlands	6.4	4.5	20.9	3.5	32.8	7.6
Norway	5.5	3.6	25.2	6.6	35.3	8.5
Poland	2.5	2.1	53.1	3.1	35.6	0.7
Portugal	2.6	3.3	47.4	2.3	38.6	10.5
Romania	4.6	1.1	52.5	1.6	26.2	0.1
Sweden	4.9	4.2	25.4	8.4	47.0	15.2
Slovenia	1.9	1.2	62.1	2.1	34.4	16.9
Slovakia	2.9	0.8	67.3	1.1	34.5	0.3
United Kingdom	6.9	10.6	22.2	7.1	46.1	12.4
Average	5.0	3.6	39.9	4.1	37.4	10.3

Source: EU-SILC (2012) cross-sectional data, authors' computation.

Table 1.A2 –Summary statistics of the independent country-level variables

Country	median income, 1000 PPS	Unemploy-ment rate, %	in-cash social benefit levels, %	in-kind social benefit levels, %	total social benefit levels, %	pro-poorness bottom 50, %	pro-poorness bottom 20, %
Austria	19.7	7.8	6.2	8.1	14.3	51.6	12.6
Belgium	18.3	10.8	8.5	8.5	17.0	57.6	22.5
Bulgaria	6.0	19.3	3.2	4.9	8.1	39.7	6.3
Switzerland	25.6	3.1	4.8	6.9	11.7	61.4	18.4
Cyprus	18.4	14.1	6.3	4.3	10.6	50.2	13.8
Czechia	10.8	9.8	4.4	5.9	10.3	58.4	14.3
Germany	18.8	8.7	6.2	10.1	16.3	53.2	8.3
Denmark	19.7	9.5	7.7	10.7	18.4	85.3	46.3
Estonia	8.6	10.2	4.2	4.5	8.7	44.9	12.9
Greece	10.3	25.8	4.3	9.7	14.0	56.3	19.8
Spain	14.8	27.4	6.5	7.6	14.1	59.2	19.9
Finland	18.9	10.4	7.1	9.8	16.9	70.0	29.4
France	18.4	10.5	5.9	10.7	16.6	59.9	20.3
Croatia	8.0	28.6	6.4	6.0	12.4	58.2	25.8
Hungary	8.0	15.2	4.8	6.3	11.1	53.9	14.5
Ireland	17.3	20.8	8.2	7.6	15.8	65.2	30.7
Iceland	17.5	6.9	7.0	10.7	17.7	75.6	39.3
Italy	15.7	14.4	4.2	7.0	11.2	53.3	12.1
Lithuania	7.2	15.8	4.0	5.0	9.0	51.4	13.0
Luxembourg	27.7	7.4	7.2	6.6	13.8	53.7	17.0
Latvia	6.7	18.5	3.1	3.5	6.6	39.4	10.1
Malta	15.2	6.2	3.3	4.9	8.2	69.2	28.9
The Netherlands	18.7	4.1	7.3	9.3	16.6	75.0	41.0
Norway	26.2	3.5	8.1	8.1	16.2	76.1	37.2
Poland	8.6	12.6	3.2	4.4	7.6	56.9	17.3
Portugal	10.1	20.4	4.5	6.3	10.8	61.2	23.6
Romania	4.3	5.0	3.0	4.6	7.6	38.7	10.8
Sweden	19.0	6.5	5.3	10.6	15.9	67.0	22.5
Slovenia	14.5	17.3	5.2	7.7	12.9	65.7	25.6
Slovakia	10.1	14.0	4.6	5.3	9.9	63.6	25.6
United Kingdom	17.9	6.3	6.1	10.6	16.7	73.5	36.2
Average	14.9	12.6	5.5	7.3	12.8	59.5	21.8

Source: EU-SILC (2012) cross-sectional data, authors' computation.

Table 1.A3 – Multinomial logistic multilevel model, household-level variables

	Model 1 ('empty')			Model 2 ('micro')		
Household variables						
Constant	-2.15 (0.12)***	-3.13 (0.04)***	-3.28 (0.04)***	-1.05 (0.35)***	-2.85 (0.06)***	-2.64 (0.07)***
Work intensity				-0.03 (0.07)***	-0.01 (0.3)***	-0.04 (0.02)***
Higher education (hh)				-0.87 (0.42)***	-0.94 (0.39)***	-1.47 (0.23)***
Lower education (hh)				0.48 (1.61)***	0.47 (1.6)***	1.07 (2.91)***
Bad health (hh)				-0.16 (0.85)***	0.62 (1.86)***	0.28 (1.32)***
Rent				0.59 (1.8)***	0.94 (2.57)***	1.25 (3.48)***
Couple, one or two children				0.59 (1.81)***	0.08 (1.08)***	0.86 (2.37)***
Couple, more children				1.15 (3.17)***	0.41 (1.5)***	1.65 (5.21)***
Single, no children				1.01 (2.75)***	0.7 (2.01)***	1.51 (4.54)***
Single, children				1.39 (4)***	1.15 (3.15)***	1.96 (7.09)***
Other hh				0.3 (1.35)***	0.17 (1.19)***	0.53 (1.69)***
Young age <29 (HRP)				0.44 (1.56)***	0.19 (1.21)***	0.26 (1.3)***
Female (HRP)				0.05 (1.05)***	0.38 (1.46)***	0.27 (1.31)***
Migrant (hh)				0.62 (1.86)***	0.51 (1.66)***	0.87 (2.38)***
Variance measures						
Random intercept	0.22***	1.78***	1.69***	0.18***	1.98***	2.34***
Explained country variance						
VPC	0.06	0.35	0.34	0.05	0.38	0.42
Model diagnostics						
R^2 Likelihood Ratio		0.07			0.24	
DIC		662223.63			545200.95	
Effective number of parameters (pD)		92.18			131.08	
N		492122			492122	
Number of countries		31			31	

Source: EU-SILC (2012) cross-sectional data, authors' computation. *Note*: The reference category of the dependent variable is 'not being poor, nor severely materially deprived'. *Significant at 10% level, **Significant at 5% level, ***Significant at 1% level. Exponentiated coefficients are shown in brackets ($\exp(\beta_{hk})$). For the work intensity variable, a work potential of 100 per cent is used for the calculation of the exponentiated coefficients. 'IP only' , 'MD only' and 'Cons. Pov.' denote, respectively, 'income poverty only', 'severe material deprivation only' and 'consistent poverty'.

Chapter 2

Better understanding child deprivation: a 31-country comparative analysis

(co-authors: Anne-Catherine Guio (LISER), Eric Marlier (LISER), Frank Vandenbroucke (UvA))

2.1. Setting the scene

Fighting child poverty and investing in children's well-being has featured on the agenda of the European Union for many years. In February 2013, a new step forward was taken when the European Commission published a Recommendation on "Investing in children: breaking the cycle of disadvantage" (European Commission, 2013a), subsequently adopted by the EU Council of Ministers. An important element of the EU Recommendation is that it calls on Member States to "(reinforce) statistical capacity where needed and feasible, particularly concerning child deprivation".

The best way to provide accurate information on the actual living conditions of children in the EU, without making assumptions about the sharing of resources within the household, is to develop child-specific deprivation indicators - i.e. indicators based on information on the specific situation of children, which may differ from that of their parents. In 2018, the EU made a significant step in this direction by adopting the child-specific deprivation indicator proposed by Guio and co-authors (2018), using EU Statistics on Income and Living Conditions (EU-SILC) data.

This chapter analyses the determinants of child deprivation in 31 European countries (28 EU countries as well as Iceland, Serbia and Switzerland), using this indicator (see Section 2.2 for a definition). It combines analyses based on both single level and multilevel models, following the methodology presented in Chapter 1.[26] In doing so, it seeks to obtain a better and robust understanding of the joint relationship of micro-determinants (household's labour market attachment, household income, household composition, costs [due to needs related to housing, bad health…] etc.), macro-drivers and contextual determinants with child deprivation.

[26] Norway could not be included due to the large amount of missing data on child deprivation.

The main contribution of this chapter is that it both replicates and confronts a broad spectrum of (sometimes diverging) results reported in the literature and suggests reasons why variables, measured both at the micro- and macro-level, (do not) have a relationship with child deprivation. In most of the multilevel models described in the literature, the inclusion of *macro-level variables* (national social transfers in-cash, Gross Domestic Product (GDP) etc.) is justified by the fact that more generous welfare systems or more prosperous economies lead to lower levels of deprivation in the country. However, once micro-level (household-level) determinants that capture individual resources and social transfers received by the household are included in the model, the reason why such macro-level variables would still have a significant relationship with deprivation is not discussed. A priori, one would expect that solely macro-drivers that are not included at the micro level, such as the national amount of transfers in-kind, should explain between-country differences in deprivation in the multilevel model. However, many papers show the significant relationship of other aggregated variables, such as national social transfers in-cash or GDP per capita, after controlling for individual household income and other relevant household-level variables. The crucial question is therefore *why* a variable whose full impact is already taken into account at the household level, is expected to have an additional explanatory power at the country level. To disentangle the relationship of micro- and macro-drivers, we present a robust analysis, in which we replicate a number of analyses presented in other papers using a large variety of macro-variables. In addition, we explicitly argue why we expect certain *micro-level variables*, such as parents' education or migrant status or (quasi-)joblessness of the household to have a relationship with deprivation, next to the household's current income. Often, the expectation that such "social stratification" variables are related to deprivation is taken for granted without further argument.

This chapter is organised as follows. Section 2.2 defines the EU child deprivation indicators. Section 2.3 reviews the macro- and micro-determinants of (child) deprivation. Section 2.4 presents the models and estimation strategy. Section 2.5 presents the results of both the single level and multilevel models. Section 2.6 concludes.

2.2. A robust EU measure of child-specific deprivation

The optimal set of child deprivation items agreed at the EU level is both theory and data driven. From a theoretical point of view, it largely relies on Townsend's concept of relative deprivation:

"*Poverty can be defined objectively and applied consistently only in terms of the concept of relative deprivation. […] Individuals, families and groups in the population can be said to be in poverty when they lack the resources to obtain the type of diet, participate in the activities* *and have the living conditions and amenities which are customary, or at least widely encouraged or approved, in the societies to which they belong. Their resources are so seriously below those commanded by the average individual or family that they are, in effect, excluded from ordinary living patterns, customs or activities*." (Townsend, 1979, p. 31)

From a data analysis point of view, the retained items successfully passed statistical tests (suitability, validity, reliability, additivity).[27]

The final list consists of 12 "children" and 5 "household" items, which cover both material and social aspects of deprivation:[28]

<u>Children items:</u>

1. Some new (not second-hand) clothes
2. Two pairs of properly fitting shoes
3. Fresh fruit and vegetables daily
4. Meat, chicken, fish or vegetarian equivalent daily
5. Books at home suitable for the children's age
6. Outdoor leisure equipment
7. Indoor games
8. Regular leisure activities
9. Celebrations on special occasions
10. Invitation of friends to play and eat from time to time
11. Participation in school trips and school events
12. Holiday

<u>Household items:</u>

13. Replace worn-out furniture
14. Arrears
15. Access to Internet
16. Home adequately warm
17. Access to a car for private use

Only children lacking an item for affordability reasons (and not by choice or due to any other reasons) are considered as deprived of this item (enforced lack concept). Those lacking the item for "other reasons" are treated, together with those who have the item, as not deprived. There are, however, a number of questions raised by the notion of enforced lack (McKay, 2004; McKnight, 2013). The "other reasons" modality can encompass a large range of possible situations: people may not want/need an item, or they may be prevented from having an item

[27] See Gordon and co-authors (2000), who developed these tests in the UK.

[28] Besides the items relating directly to the deprivation situation of children, the above 17-item list includes some household items. As emphasised by Guio and co-authors (2012, 2018), not only items directly impacting children's immediate well-being should be considered in the children's index, but also items likely to have an indirect impact on their well-being. Indeed, qualitative studies have shown that children in households suffering from financial strain often do not ask their parents for the things they need in order to try to protect their parents from stress and feelings of guilt (Ridge, 2002 and 2011).

for many different reasons (e.g. lack of time of the parents due to caring responsibilities or due to work, no vehicle/ public transport, feeling unwelcome, etc.). Some of these “other reasons” may be correlated with their living standards, in the case of adaptive preferences, or shame to admit that children lack the item because it is unaffordable (Guio et al., 2012, p.34). That is the reason why Guio and co-authors (2018) investigated the characteristics of children living in households replying that they do not have the item for “other reasons”. They show that using the concept of enforced lack (rather than simple lack) makes it possible to control for individual preferences due to differences in cultures, age of children or parental practices. They also show that measures based on the enforced lack concept discriminate better between the worse-off and better-off children than those based on simple lack, and that the use of enforced lack ensures a higher reliability of the index.

In the analysis presented below, it is important to keep in mind some elements related to data collection and processing. First, in EU-SILC data relating to the living conditions of children are not collected from the children themselves, but from the adult answering the “household questionnaire” (household respondent). Secondly, according to the survey protocol to be followed by countries, if in a given household at least one child does not have an item, it is then assumed that all the children belonging to that household lack that item. It would of course be preferable to know the deprivation levels of each child in a household separately; it would then be possible to study differences in child deprivation within individual households, as well as between households (e.g. are girls more likely than boys to suffer from deprivation within a same household, or teenagers more likely than younger children?). However, collecting this type of information would be quite delicate and would also lengthen significantly the EU-SILC questionnaire. Thirdly, for most “children’s items”, the information relates to children aged between 1 and 15 (i.e. children’s items are collected in households with at least one child in this age bracket). Therefore, the child-specific deprivation indicator covers only children aged between 1 and 15. Yet, one item is collected in households with at least one child attending school (school trips).

The main child-specific indicator adopted at EU level is the proportion of children suffering from at least three items. In the rest of the chapter, we will analyse the full scale of deprivation (ranging from 0 to 17), i.e. the *child-specific deprivation intensity*.[29] The incidence of each individual deprivation item is presented in Table 2.1 and compared to the EU-28 average. This heat map highlights countries showing consistently high deprivation levels, such as Bulgaria and Romania, or on the contrary low levels (Nordic countries, Austria, Netherlands, Luxembourg). It also highlights countries where there is a mixed picture depending on the item,

[29] A second child-specific EU indicator has also been adopted at EU level: the average number of items lacked by deprived children. This measure is different from the child-specific deprivation *intensity* considered here, which looks at all children rather than only deprived children.

i.e. countries suffering from a relative disadvantage for some items, and a relative advantage for others.

Table 2.1 – "Heat map" providing for each item the proportion of children lacking the item in the country

	Fruit & vegetables	Books	Shoes	Indoor games	Proteins	Internet	Celebratic	Outdoor equipment	Clothes	School trips	Friends	Car	Home warm	Leisure	Arrear	Holidays	Furniture
Sweden	0,1	0,6	0,3	0,3	0,0	0,4	1,3	0,8	0,9	0,8	0,7	3,1	0,8	2,5	8,8	5,5	5,6
Finland	0,3	0,5	0,8	0,2	0,2	0,4	0,3	0,3	3,5	0,6	0,1	3,6	0,7	1,3	16,5	7,2	11,6
Iceland	0,4	0,3	1,9	0,2	0,8	0,5	0,3	0,6	0,9	0,6	0,1	2,7	2,2	4,3	24,1	3,6	20,4
Denmark	0,5	2,5	2,3	0,8	0,6	0,6	1,3	2,2	2,0	1,4	1,5	5,1	2,5	3,3	9,5	9,1	14,6
Switzerland	0,5	0,4	0,3	0,7	1,3	0,9	1,4	0,4	1,6	0,8	0,4	4,5	1,0	5,1	10,8	4,9	12,5
Austria	0,5	1,3	1,1	1,1	1,8	1,0	1,8	3,1	1,9	2,5	3,6	7,4	4,3	10,2	10,6	17,8	15,7
Netherlands	0,6	0,5	3,6	0,4	2,5	0,2	1,9	1,6	1,6	1,4	1,2	6,5	2,8	6,4	9,5	16,2	25,2
Luxembourg	0,8	0,8	1,0	1,5	1,1	1,4	1,9	2,7	2,9	3,6	2,3	2,1	1,0	2,7	6,3	9,4	20,9
Slovenia	1,0	1,1	1,2	1,3	1,4	1,3	2,5	2,0	5,9	2,3	3,4	3,3	4,0	10,7	28,0	7,2	15,8
Spain	1,7	2,3	3,0	3,5	2,9	13,5	11,4	5,8	7,7	10,6	12,8	6,6	12,0	13,1	17,8	34,5	46,4
Germany	1,8	0,7	2,2	0,6	3,6	0,9	1,5	1,3	2,1	0,6	1,7	4,4	5,3	6,2	9,7	17,4	17,8
Malta	1,9	2,0	5,9	2,1	6,9	4,4	4,9	4,1	6,1	2,7	4,9	4,5	21,6	6,0	22,0	34,9	29,7
Cyprus	2,1	5,4	1,3	3,6	2,4	8,7	10,8	7,7	5,4	2,5	12,3	1,4	25,4	21,2	41,7	40,2	60,9
Belgium	2,3	4,4	3,6	2,5	2,7	3,8	5,8	4,2	8,2	3,8	6,0	7,4	4,8	9,0	12,1	19,2	18,4
Italy	2,6	7,7	2,9	5,6	5,7	10,8	7,1	6,0	8,5	9,5	7,5	2,3	18,4	13,7	20,6	29,5	38,8
Ireland	2,6	1,0	6,5	1,4	3,1	4,8	3,0	3,2	12,3	3,3	3,2	6,6	9,4	7,3	25,6	53,1	28,6
France	2,7	1,2	5,2	1,0	2,3	1,8	5,2	1,7	8,9	4,8	2,4	2,8	5,1	6,2	15,0	11,6	28,0
Portugal	2,9	6,4	3,6	5,4	1,2	11,5	8,3	4,6	14,4	9,1	13,6	9,9	25,2	23,4	17,7	36,7	57,5
Czech Republic	3,0	2,0	3,0	2,8	4,7	4,0	3,6	7,8	6,3	5,0	2,4	11,8	6,0	8,5	10,4	8,7	47,8
Poland	3,5	2,9	1,4	2,3	3,0	3,1	9,7	4,3	3,2	8,5	8,7	7,5	7,9	18,8	19,3	26,2	31,5
United kingdom	3,6	1,0	2,2	1,4	3,0	4,7	2,3	5,7	3,7	3,3	7,1	10,7	9,4	6,3	18,0	35,3	31,6
EU-28	4,1	4,4	4,7	4,7	5,2	6,9	7,2	7,1	7,5	7,4	8,2	8,7	10,0	12,6	18,3	26,3	33,8
Croatia	4,5	7,2	3,2	5,7	6,2	4,9	5,6	5,9	5,3	7,8	7,4	7,0	9,1	8,9	35,9	29,2	32,3
Greece	5,4	7,2	0,6	4,1	9,2	8,9	18,9	10,1	1,8	21,2	14,1	8,6	30,5	15,8	54,2	41,3	57,5
Estonia	6,7	2,5	1,6	1,6	6,1	0,9	3,4	3,7	2,4	3,0	4,9	9,7	1,4	4,1	16,2	10,3	27,4
Lithuania	7,8	2,3	0,4	2,8	6,3	5,3	5,0	6,6	13,0	5,8	9,9	12,0	25,6	18,8	17,8	19,2	50,1
Serbia	9,7	7,9	8,2	6,2	15,1	13,8	10,6	10,9	13,8	15,0	7,9	20,9	15,6	20,9	48,5	39,7	61,4
Slovakia	9,8	10,4	6,6	7,6	12,9	9,1	12,0	11,0	14,0	9,1	15,3	13,9	7,8	11,0	10,8	15,5	45,3
Latvia	10,0	11,0	11,7	8,7	8,2	8,1	10,3	16,4	24,5	7,6	11,3	23,4	18,2	16,2	31,6	27,6	57,7
Romania	14,8	24,8	28,0	42,4	21,6	36,7	33,2	55,5	26,6	30,3	40,1	45,3	15,4	60,1	29,3	61,4	67,3
Hungary	22,8	15,5	7,8	13,7	22,0	18,2	15,4	17,0	27,2	15,2	30,6	31,1	12,5	20,9	36,2	51,1	52,9
Bulgaria	40,2	43,2	49,0	38,4	42,4	26,9	32,3	52,0	36,2	42,5	41,4	30,2	40,2	52,3	43,9	54,6	72,1

Source: Guio et al., 2018.

2.3. Micro- and macro-level determinants of child deprivation

In the existing literature on (material) deprivation determinants (as documented for the whole population), a distinction is drawn between so-called "micro-level" and "macro-level" determinants. The micro-level determinants are socio-economic characteristics measured at individual or household level that have a relationship with deprivation.[30] By contrast, the macro-level determinants look at macro-variables such as GDP, unemployment, inequality, welfare state regime etc. to account for differences in deprivation between countries (see, for example Kenworthy et al., 2011). Recently, multilevel studies have combined the micro-level and macro-level approaches, by jointly considering individual and country characteristics in pooled data settings (see Kim et al., 2010; Chzhen & Bradshaw, 2012; Nelson, 2012; Whelan & Maître, 2012; Whelan & Maître, 2013; Israel & Spannagel, 2013; Bárcena-Martín et al., 2014; Chzhen, 2014; Visser et al., 2014; Saltkjel & Malmberg-Heimonen, 2017; Bárcena-Martín et al., 2017b; Verbunt & Guio, 2019).

2.3.1. Micro-level determinants

It is well documented that demographic and socioeconomic characteristics of households influence child income poverty and deprivation (see for example Tárki, 2011). Both social stratification – the social stratum to which the household belongs – and resources are at play, and the relation between the social stratum and the resources as joint determinants of deprivation is probably much more complex than a reduced form empirical model can account for: the social stratum influences not only the level of resources a household commands, but also their use. To specify an empirical model, notwithstanding this difficulty, we distinguish three sets of household-level variables that can explain children's deprivation:

1) the longer-term command over resources;
2) needs related to health and housing;
3) the size and composition of the household.

Deprivation emerges in the confrontation between available resources and needs. As will become clear, the distinction between variables captured under set 1) and variables grouped under sets 2) and 3) is largely (but not fully) a distinction between "resources" and "needs". However, important factors that influence both the household's command over resources and its needs are not available in our micro dataset (EU-SILC). This holds, for instance, for the household's consumption of in-kind benefits for which we use as "proxy" the national social

[30] For an extensive review of the micro-level determinants of (material) deprivation, see Perry (2002) and Boarini and Mira d'Ercole (2006).

spending in-kind in the multilevel models. Yet, some relevant elements are missing in both the single level and multilevel models: in-kind support from family/friends, as well as a direct measure of wealth. It is also important to highlight that the national social spending in-kind that we use is only a crude measure. Indeed, when using this aggregate we also miss important relevant elements: what is the proportion of the benefits that goes to children, what proportion goes to poor/deprived children, what are the quality and affordability of services?

First, children's material well-being depends on how much the household can consume, which, in turn, depends on its "command over resources". Although current (disposable) household income is usually used as a proxy for "command over resources", the association between current income and deprivation is far from perfect. This imperfect link is documented extensively in the literature (see among others Whelan et al., 2001; Whelan & Maître, 2006; 2007; Berthoud & Bryan, 2011; Fusco et al., 2011; Nolan & Whelan, 2011; Verbunt & Guio, 2019). It can be explained by difficulties in measuring income (as is notably the case for self-employed people) and deprivation, and by the fact that households with equal resources may have different needs and face different costs. But, importantly, it can also be explained by the fact that current income is only one element in a household's command over resources. A household's command over resources is also determined by its previous, current and future income, its wealth and its ability to borrow.

We use three variables, available in EU-SILC, which can plausibly serve as proxies for the household's *longer-term* command over resources (in addition to its current income), its *wealth* and its *ability* to overcome short-term financial difficulties: current educational attainment, current (quasi-)joblessness and migrant status. Borrowing from economic jargon, these indicators can be related to the household's *permanent* income, its *wealth* and its ability to overcome *liquidity* constraints.[31] *Ceteris paribus* (for a given level of current income and other household characteristics), a higher level of education can indeed be expected to correlate statistically with: i) a stronger position on the labour market, hence less vulnerability with regard to adverse income shocks (e.g. income shocks because of unemployment or precarious employment); ii) parents that were higher educated and therefore richer, which implies more important bequests and thus wealth; iii) easier access to financial institutions to overcome liquidity constraints; iv) for younger people, a higher future return on human capital. *Ceteris paribus,* if someone in the household was born outside the EU, this correlates statistically with similar social factors: a more vulnerable position on the labour market, less inherited wealth,

[31] The extent to which one needs additional "social stratification" indicators to gauge an individual's or a household's permanent income, over and above its current income, is a moot question; see Kim et al. (2018) and Brady et al. (2018) for recent explorations of this issue. Here, we start from the theoretical expectation that education, joblessness and migrant status do play a role.

and more difficult access to financial institutions.[32] *Ceteris paribus,* (quasi-)joblessness at the household level is likely to signal a precarious position on the labour market for all working age household members, which is a predictor of future unemployment risks and, in addition, may hamper access to financial institutions to overcome liquidity constraints. Given its availability in EU-SILC, we are able to add a measure of the household's debt burden, which directly influences its longer-term command over resources, in addition to the three proxies just mentioned.

To sum up, in order to proxy as well as possible the longer-term command over resources at the household level, we use six variables:

A. The yearly (disposable) non-equivalised income of households expressed, in purchasing power standards (PPS) per 1000 (*Household income*).[33][34] Both the logarithm and linear forms of the income variable are introduced in the regressions. The best regression fit was obtained with the non-logarithm form of the variable. We use non-equivalised income, because the size and composition of the household enter separately in group 3) of our explanatory variables (see below).

B. The educational attainment of the highest educated parent (operationalised by three dummies: *Low education* (no education, primary education or a lower secondary education), *Medium education* (upper secondary or post-secondary non-tertiary education) and *High education* (tertiary education used as the reference category).

C. The (quasi-)jobless status of the household (*Jobless*), based on the household's work intensity, with a threshold set at 20%. These are households where on average the adults (aged 18-59, excluding students) work 20% or less of their total work potential during the past year. This indicator covers the population aged 0-59 (i.e. also children).

[32] On the impact of migrant status on (material) deprivation, see de Neubourg et al. (2012).
[33] The disposable income of a household is obtained by summing up all monetary incomes received from any source by any member of the household or the household itself and then deducting taxes and social contributions paid by the household.
[34] On the basis of Purchasing Power Parities (PPP), Purchasing Power Standards (PPS) convert the amounts expressed in a national currency to an artificial common currency that equalises the purchasing power of different national currencies (including for those countries that share a common currency). It should be noted that PPS can be considered to be an imperfect tool to measure price differences in relation to deprivation. Reference budgets, priced baskets of goods and services that are needed for households in given countries, regions or cities to achieve a given standard of living, are a theoretically sound alternative. However, reference budgets are at this moment not yet available for all countries in the dataset.

D. A dummy measuring whether one household member was born outside the EU (*Migrant*).[35][36]

E. The debt burden of the household (*Debt burden*), which equals one if payment of debts from hire purchases or loans other than mortgage or loan connected with the dwelling are considered as a heavy financial burden to the household.

F. The presence of self-employed people in the household (*Self-employment*), a dummy variable which we include to take into account difficulties in measuring income for this sub-population.

Secondly, children living in households with the same resources but different needs may experience very different standards of living. Needs increase the level of resources necessary for a household to maintain its standard of living. Needs notably depend on health, tenure status, and the housing situation (see among others Whelan et al., 2004; Fusco et al., 2011; Verbunt & Guio, 2019).[37] So, we introduce three variables to proxy the household's needs (and related costs):

A. The self-reported health status variable (*Bad health*), which has a value of one if at least one person in the household reports having bad or very bad health.[38]

B. A tenure dummy (*Rent*), which has a value of one if the household rents its dwelling on the private market or with a social (free or reduced) tariff, as compared to owning its own house.[39]

C. Two housing burden dummies, which measure if households' housing costs, including mortgage repayment (instalment and interest) or rent, insurance and service charges

[35] For the three non-EU countries covered in the paper (Iceland, Serbia and Switzerland), a child is considered as migrant if at least one member of its household was born in a country which is neither the country of residence nor an EU country.

[36] An additional variable that captures whether someone in the household is born in another EU country was co-regressed with the non-EU migrant dummy as a robustness check. The coefficient of this variable was small, but significant. The BIC indicates that the original model (i.e., the model that does not include this variable) should be preferred over the model that includes this variables.

[37] Childcare costs were included in the model (using as a proxy based on childcare attendance). However, the variable was missing for a large share of the sample of children and had no significant impact on child deprivation for the rest of the sample. A variable on childcare cost burden was collected in the EU-SILC ad-hoc module on public services in 2016, and should be more appropriate to test the impact of childcare costs on child deprivation when it becomes available.

[38] We tested "limitation in daily activity" and "suffering from a chronic condition" as alternatives for the bad health variable. The bad health specification had the best fit to the data.

[39] We introduced separate dummies for private market renting, renting with a free or reduced tariff and owning a house with a mortgage. The coefficients of both the market and social renting variables gave very similar results, while the owning a house with a mortgage variable was insignificant.

(sewage removal, refuse removal, regular maintenance, repairs and other charges) are a heavy (*Heavy housing burden*) or a light housing burden (*Light housing burden*), with no housing burden as the category of reference.

Thirdly, we include three socio-demographic variables related to the household size and composition:

A. The total number of dependent children (i.e. all children aged 0-17 and dependent students aged between 18-24) in the household (*Number of dependent children*), instead of implicitly adjusting the household income for its size and composition with an equivalence scale (as is done for the calculation of income poverty).[40]

B. The age of the oldest child in the household among those children aged 1-15 (*Age of oldest child*), in order to test whether the composition of the deprivation basket induces a systematic bias in favour of younger/older children, as would be the case if some of the items are less relevant for some age groups.

C. A dummy indicating if children live in a single-parent household (*Single parent*). A priori, we expect this variable to be related both to the longer-term command on resources and the needs of the household. From a permanent income perspective, a single parent household is more vulnerable (it has fewer possibilities for employment risk pooling across adults in the household than households with more than one adult). From a needs perspective, single parents face fixed costs (housing, childcare costs, etc.) which generally represent a higher share of their household resources than households with more than one adult (remember that we do not equivalise household incomes). (They also face more difficulties in reconciling working life and family life and therefore are more likely to opt for part-time employment or inactivity; inactivity or a very low level of activity is however already taken into account by the variable on (quasi-)joblessness.)

These three sets of household-level variables are used in the single level models (for each country), as well as at the micro level of the multilevel model (for the pooled dataset). All summary statistics can be found in Table 2.A1 in Appendix 2. Table 2.A2 in Appendix 2 presents the correlation coefficients between these variables.

[40] In order to reflect differences in a household's size and composition, the total (net) household income is usually divided by the number of "equivalent adults", using the so-called OECD-modified scale, which gives a weight to all members of the household (1 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 14). The resulting figure, the equivalised disposable income, is attributed equally to each member of the household (adults and children).

2.3.2. Combining micro- and macro-level determinants

In multilevel models, household-level risk factors are complemented by country-level variables. The selection of explanatory variables included in these models needs careful consideration: depending on the research question one wants to answer, it may be appropriate or inappropriate to include certain variables in the model.

Table 2.A3 in Appendix 2 summarises the results obtained with multilevel models in existing research on (child) deprivation. Multilevel models typically include explanatory variables which correlate with the average level of household income in the country, most often GDP per capita; they often also include aggregate measures of social spending. With the exception of Bárcena-Martín and co-authors (2014) and Whelan and Maître (2012), the papers we found do not include household income at the individual level, whilst they include individual household variables related to education, socio-economic status and employment. This choice of variables at the micro level raises questions: the most plausible argument to include variables related to education, status and employment at the micro level, is that these variables correlate with the household's "longer-term command over resources", as explained above. However, current income certainly also correlates with the household's "longer-term command over resources"; presumably it is even the best proxy for a household's longer-term command over resources (see Kim et al., 2018 and related literature). If the research objective is to explain child deprivation across Europe, we do not see good reasons for leaving out the best proxy for "longer-term command over resources" when it is available in the dataset. In fact, models excluding individual household income at the micro level but including national GDP per capita and social transfers at the macro level, are bound to mix up direct and indirect impacts of such variables. This is not to say that excluding individual household income in a multilevel model examining deprivation is always wrong. For instance, if the research question focuses on the relationship of cash transfers with material deprivation across countries, given their level of GDP per capita and given household *needs* measured at the micro level, one might want to exclude household income at the micro level, in order to gauge the relationship of cash transfers with deprivation.[41] But we feel uncomfortable with models that include all kinds of variables that determine households' longer-term command over resources except household income, and then add the level of cash transfers as explanatory variable.

However, once household income is included at the micro level, the inclusion of macro-variables that directly influence individual household incomes – such as GDP per capita or cash transfers – needs careful consideration. A priori, we expect that only macro-variables

[41] We thank Brian Nolan for extensive discussion on this issue, which is not to say that he would agree with our conclusion.

without direct impact on individual incomes have an impact on between-country differences in deprivation, when individual incomes are accounted for at the micro level. A prime example of such a macro-variable is spending on *in-kind* social benefits: receipt of in-kind benefits is not included in individual household incomes. If a variable has a significant relationship with deprivation when it is included at both the macro and micro levels, such a result is *prima facie* counterintuitive and deserves further interpretation. We return to this when we discuss our results.

To test whether social transfers have a significant association with child deprivation, we mobilise a large number of indicators that capture differences in social spending across the 31 countries analysed, in terms of spending size (total, cash and in-kind), targeting on families/children, pro-poorness and adequacy:

A. Social welfare generosity is operationalised by several variables. A first measure expresses total social spending as a percentage of GDP and is derived from the Eurostat European System of integrated Social Protection Statistics (ESSPROS) database (*Total social benefits, % of GDP*). In addition, following the approach presented in Chapter 1, we also distinguish between in-cash (*Cash social benefits, % of GDP*) and in-kind (*In-kind social benefits, % of GDP*) social spending. Social spending covers sickness/healthcare, disability, family/children, unemployment, pension, survivor, housing and all not elsewhere classified social exclusion benefits.[42] These variables measure the generosity of the welfare state in the country, as a proportion of the GDP. Alternatively, we also use household-level variables that measure the level of net social benefits received by households with children (any benefit, not just family-related benefits), and are directly derived from the EU-SILC micro-data. This is the average equivalised social transfer computed per child (*Cash social benefits, in PPS per child*). Lacking additional information in EU-SILC on the distribution of in-kind benefits in PPS, we use in-kind social benefits derived from the ESSPROS database and expressed in PPS per head (*In-kind social benefits, in PPS per head*). Total social spending sums up both in-cash and in-kind social benefits (T*otal social spending, in PPS per head*).

B. We evaluate the relationship of social spending geared to families and children with child deprivation. We use the ESSPROS average family transfer expressed as a proportion of GDP, covering both in-kind and in-cash benefits (F*amily social spending benefits, % of GDP*) and the average gross equivalised family benefits per child based

[42] It might seem counterintuitive to include pensions and survivor benefits in this concept when explaining differences in child deprivation. However (see for example Diris et al., 2017), pensions constitute an important share of household income for non-elderly individuals in some countries (mainly those where intergenerational households are more prevalent).

on EU-SILC micro-data (*Family cash social benefits, PPS per child*).[43] One should remember that cash-transfers are already included in individual household income whilst in-kind transfers are not. Hence, if we obtain a significant coefficient for a macro-variable including cash-transfers to the target population, the interpretation is not straightforward (see above).

C. The pro-poorness of in-cash social benefits is an important aspect of the redistributive system. The question of the optimal degree of universalism and targeting is still open to debate. We measure the degree of targeting by the share of transfers that is distributed to the lowest five deciles of the pre-transfer household income distribution of children (*Pro-poorness bottom 50*).[44] The countries with the highest share of transfers (more than 75%) going to the bottom 50% of the distribution are the Czech Republic, Denmark, Greece, Iceland, Ireland, Malta, Poland, Portugal and the UK (see Table 2.A1 in Appendix 2). Again, significant coefficients for such a variable require careful interpretation, since individual incomes of poor households in our dataset already include these transfers. A first descriptive analysis indicates that the negative relationship between targeting (*Pro-poorness bottom 50*) and size (as measured by social transfers in % of GDP or per head) is not confirmed by our data (see the correlations between the country-level variables in Table 2.A4 in Appendix 2).

D. Nelson (2012) argues against analysing the relationship of social transfers with child deprivaiton via an expenditure-based approach, as we proposed above. Expenditure-data mix information on system generosity with information on the business cycle and the composition of the population. Also, these data refer to gross public spending (in ESSPROS data, and to a certain extent also in EU-SILC data), and do not account for national differences in taxation.[45] Furthermore, by looking at the national average of social spending per head, the expenditure approach cannot account for variations in treatment of families by household composition or social situation. These are the main reasons why some authors opt for a “household-type” approach (rather than an expenditure approach): it makes it possible to overcome these drawbacks and better measure cross-country differences in social transfers (Nelson, 2012; Chzhen, 2014). Household-types simulate the level of benefits and taxes for standardised household types across countries, instead of averaging actual expenditure data. Whilst it has

[43] We computed additional variables that consider the level of family benefits expressed as a proportion of total social spending (ESSPROS) and as a proportion of household income (EU-SILC micro-data). Both variables were found to have a statistically insignificant relationship with child deprivation and explained little about between-country differences in child deprivation.
[44] Following Marx et al., 2013; Diris et al., 2017.
[45] In EU-SILC, the amount of the various social transfers received by people/households are gross amounts except for the total amount of pensions received by the household and for the total amount of transfers received (with and without pensions) for which both gross and net figures are available.

advantages, this approach has also limitations. One of the limitations, especially for comparative analyses, is the difficulty to propose a representative set of "household types" for the various countries considered (Bárcena-Martín et al., 2017b; Bárcena-Martín et al., 2018). Still, the "household type" approach is an interesting alternative for measuring the adequacy of minimum income schemes. In this chapter, the indicator used is the minimum income benefit (for the type under review) expressed as a percentage of national median household income (*Adequacy of minimum income benefit schemes*). We focus on one type: a married couple with two children, eligible for cash housing assistance.[46] The data are derived from the OECD database.

After considering income at household level, we introduce GDP per capita expressed in Purchasing Power Standards (*GDP per capita*) to reflect general differences in standard of living. GDP per capita varies extensively across the 31 countries analysed and ranges from 10,100 (Serbia) and 12,800 PPS (Bulgaria) to 74,500 PPS (Luxembourg).

Even though we control for low work intensity at household level (see above for the definition of the "(quasi-)jobless" indicator), we also introduce the unemployment rate to account for the possible effect of the business cycle on the size and pro-poorness of social benefits. The definition of the unemployment rate is the standard definition of the International Labour Office (ILO) – i.e. the number of people unemployed (ILO concept) as a percentage of the active population; it is derived from the Eurostat database (*Unemployment rate*).

All summary statistics of the country-level variables can also be found in Table 2.A1 in Appendix 2. As explained above, most of the papers using multilevel approaches test crossed effect between micro- and macro-variables. A cross-level interaction allows the coefficient of a household determinant to vary with a variable defined at the country-level. These interactions are also investigated.

2.4 The model and the estimation strategy

We use an unweighted count of child deprivation items (ranging from 0 to 17) as the dependent variable in our model. This has the advantage of using all the information on the number of deprivations suffered by children, without reducing it to a binary variable (i.e. the deprivation

[46] We tested the sensitivity of our results to choice of the "standard" family type. Tests were made with married couples with two children not eligible for cash housing assistance, single-parent households with two children eligible for cash housing assistance and single-parent households with two children not eligible for cash housing assistance). Altering the reference family had no impact on our results.

rate). Our reference population covers children aged between 1 and 15 years, i.e. the age group for which the information is collected.

The dependent variable displays a large degree of over-dispersion. Over-dispersion in count data occurs when the variance is larger than its mean. It is therefore recommended to use a negative binomial model, as this technique weakens the highly restrictive assumption made in the traditional Poisson model that the variance is equal to the mean. Instead, the negative binomial model estimates an additional random parameter that takes the unobserved heterogeneity into account. The estimate of the dispersion parameter is significantly greater than zero in all models, indicating that the dependent variable is indeed over-dispersed and that the negative binomial models are the most suitable models.

We run both single level and multilevel negative binomial models to investigate the within- and between-country determinants of child deprivation. The single level models investigate the relationship of the household-level variables with child deprivation. The main advantage of estimating single level models for each country is that all the estimated (individual/household-level) coefficients are country-specific and, hence, give a more precise estimate of the explanatory power of the model *within countries.* Multilevel models are particularly appropriate to study nested data designs, where respondents are organised within more than one level. In our study, individuals (*i*) are nested within countries (*j*). They are useful to account for unobservable differences in the dependent variable *between countries*. The differences in the composition of the population in terms of household-level risk factors may not fully explain the between-country differences in the risk of child deprivation. Country-level variables are therefore included in the model to better understand the relationship with child deprivation of variables not fully captured at the household level. Formally, the model is given by the following formula:

$$E[y_{ij}|[x_{hij}, z_{cj}, U_j] = \mu_{ij}$$

$$\log(\mu_{ij}) = \beta_0 + \sum_{h=1}^{H} \beta_h x_{hij} + \sum_{c=1}^{C} \beta_c z_{cj} + U_j$$

$$\mu_{ij} = e^{\beta_0 + \sum_{h=1}^{H} \beta_h x_{hij} + \sum_{c=1}^{C} \beta_c z_{cj} + U_j}$$

$$Var(y_{ij}|\mu_{ij}) = \mu_{ij} + v\mu_{ij}$$

with:

$E(y_{ij})$ is the expected number of deprivation items for individual *i* (*i=1,..., N*) living in country *j* (*j=1,…,J*)
μ_{ij} is the conditional mean of the dependent variable for individual *i* (*i=1,..., N*) living in country *j* (*j=1,…,J*)
β_0 is the overall intercept

x_{hij} is the value of the h^{th} ($h = 1,\ldots,H$) independent variable defined at the household level for individual i ($i=1,\ldots, N$) living in country j ($j=1,\ldots,J$)
β_h is the coefficient of the h^{th} ($h = 1,\ldots,H$) independent variable defined at the household level
z_{cj} is the c^{th} ($c = 1,\ldots,C$) independent variable defined at the country level for country j ($j=1,\ldots,J$)
β_c is the coefficient of the c^{th} ($c = 1,\ldots,C$) independent variable defined at the country level
U_j is the error term for country j ($j=1,\ldots,J$), $\sim$N(0, σ^2)
v is an over-dispersion parameter

We calculate pseudo R^2-measures to assess the overall explanatory power of the employed models. In the single level models, we use the McFadden pseudo R^2-measure. This measure is based on the likelihood value, and higher values indicate a better fit of the model to the data. Following the approach presented in Section 1.4.2, we define a measure of explained between-country variance in the multilevel models as the difference between the variance in random intercept values of the empty multilevel model and the variance in random intercept values of the models that include independent variables. We then apply Shapley decompositions on the pseudo R^2-measures to establish and compare the relative explanatory power of the independent variables (Shapley, 1953; see Section 1.4.3 for a description of the method). The Shapley approach calculates the exact contribution of each independent variable to the total R^2-value. The method has been used to decompose the goodness-of-fit measure in both linear and logistic regression models (Deutsch & Silber, 2006; Verbunt & Guio, 2019).

2.5. Results

2.5.1. National single level model

We ran negative binomial models at the country level. Table 2.2 reveals a considerable *cross-country variation in the McFadden pseudo R^2-measure* (see column 1). This means that the effectiveness of the household-level variables differs strongly across countries, which is a first interesting result. This model is the most effective in explaining child deprivation in countries with the lowest share of child deprivation (Austria, Belgium, Denmark, Netherlands and Sweden). Conversely, the countries where the single level model has a lower explanatory power are Bulgaria, Cyprus, Estonia, Italy, Latvia, Lithuania, Malta, Poland, Portugal, Romania, Serbia and Slovakia. The specific situation of Greece and Hungary should be stressed: these countries have very high levels of child deprivation but their R^2 is at the level of the weighted average of the 31 countries (Hungary) or higher (Greece). In countries where the single level model has a lower explanatory power, differences in socio-economic characteristics of households play a (much) smaller role in explaining the number of deprivations suffered by children. In several of these countries, this may be because the general standard of living is low and all children have, as a consequence, a greater likelihood of being (more) deprived.

In terms of relative share of explanatory power, Table 2.2 and Figure 2.1 show that the group of variables related to resources (income, presence of self-employed people in the household, education, (quasi-)joblessness, debt burden and migration) make, on average, a relative contribution of 55% to the fit. The variables related to needs (housing cost burden, bad health and tenure status ["rent" variable]) represent 38%. The other socio-demographic variables (household structure and size) contribute to around 7%. Figure 2.1 clearly illustrates that the explanatory power of the different variables differs between countries. In the richest countries, the explanatory power of the variables related to needs is larger. In countries with the highest proportion of child deprivation, the explanatory power of resources variables is generally greater.

The relationship of individual household income with child deprivation is significant in all 31 countries (see Table 2.3 for the detailed results). With an average contribution of 25% to the fit (from 7% in Slovakia to 36% in Cyprus, 37% in Portugal and 50% in Greece; see Table 2.2), it is the most important variable related to resources.

The educational level of the parents is also strongly associated with child deprivation, even when income, labour market attachment and other household-related demographic differences are taken into account. This confirms our expectation that educational attainment is a good proxy for the longer-term command over resources, independently from other proxies of command over resources. It makes an average contribution of 15% to the fit and is the third most important variable across the dataset (after income and housing cost burden). The education variables are significant in all models tested and in all countries (with the exception of lower education in Sweden and medium education in Denmark and Luxembourg). The association is the strongest in Bulgaria, Hungary and Romania (27-37%) as well as, to a much lesser extent, Poland, Lithuania, Slovakia, Portugal and Malta (20-22%). These are all countries with (very) high child deprivation levels. A plausible explanation for this diverging effect across countries, which does not contradict our theoretical expectation, is that higher education is more scarce in these countries and thus more valuable on the labour market.

Living in a (quasi-)jobless household is positively related with child deprivation in the majority of countries, even when household income is controlled for (see also Fusco et al. (2011) and De Graaf-Zijl & Nolan (2011) for similar results). The impact is, however, not significant in Austria, Czech Republic, Denmark, Hungary, Iceland, Lithuania, Luxembourg, Netherlands, Poland and Hungary (Table 2.3). The contribution of (quasi-)joblessness to the fit is the highest in Serbia, Ireland, Spain, Croatia, Malta and Slovakia. The average contribution is 6%.

The other variables related to households' longer-term command on resources have a more limited association with child deprivation (i.e. self-employment, migrant, debt burden). For similar income levels, households with self-employed member(s) tend to suffer from a lower number of deprivations: in all but two countries the impact is significant and negative; the exception is Switzerland where the figure is positive and high (0.39) and France where it is not significant. This confirms previous results (see also Fusco et al., 2011; Berthoud & Bryan, 2011) and may be partly explained by the difficulty of measuring self-employment income in surveys such as EU-SILC or by the challenge of discriminating between personal and professional assets and costs for the self-employed. There are, however, many countries where the coefficient of self-employment is close to zero or negative, but not significant. Migration has the largest relative contribution to the fit measures in Denmark, Sweden and Switzerland: 7-12%, as opposed to 3% for the average. Households with a high debt burden also have a higher deprivation risk (this explains 6% of the fit, on average, across the 31 countries analysed). The share of the fit is the highest (10-15%) in richest countries such as Denmark, Iceland, Sweden and Switzerland.

As expected, households with higher costs face a higher child deprivation risk. The variables related to the housing burden appear to have a strong association with child deprivation intensity in most countries: it explains more than 20% of the fit in almost all countries and as much as 43% in Slovenia (average contribution to the fit: 27%). Children living in households renting their dwelling tend to suffer more from deprivation than those owning it in all countries, except in Bulgaria, Estonia, Romania, Serbia and Slovakia, where the difference by tenure status is not significant. This variable explains a large share of the fit in Austria, Belgium, Denmark, Germany, Luxembourg, Netherlands, Sweden, Switzerland (12-18%) and in the UK (26%). The average fit is 7%. Finally, households in which at least one adult suffers from health problems also face higher risks of child deprivation (except in Bulgaria and Lithuania), which is in line with results shown in other studies (Fusco et al., 2011). This is explained by the burden of additional healthcare costs of having a household member with (very) bad health.

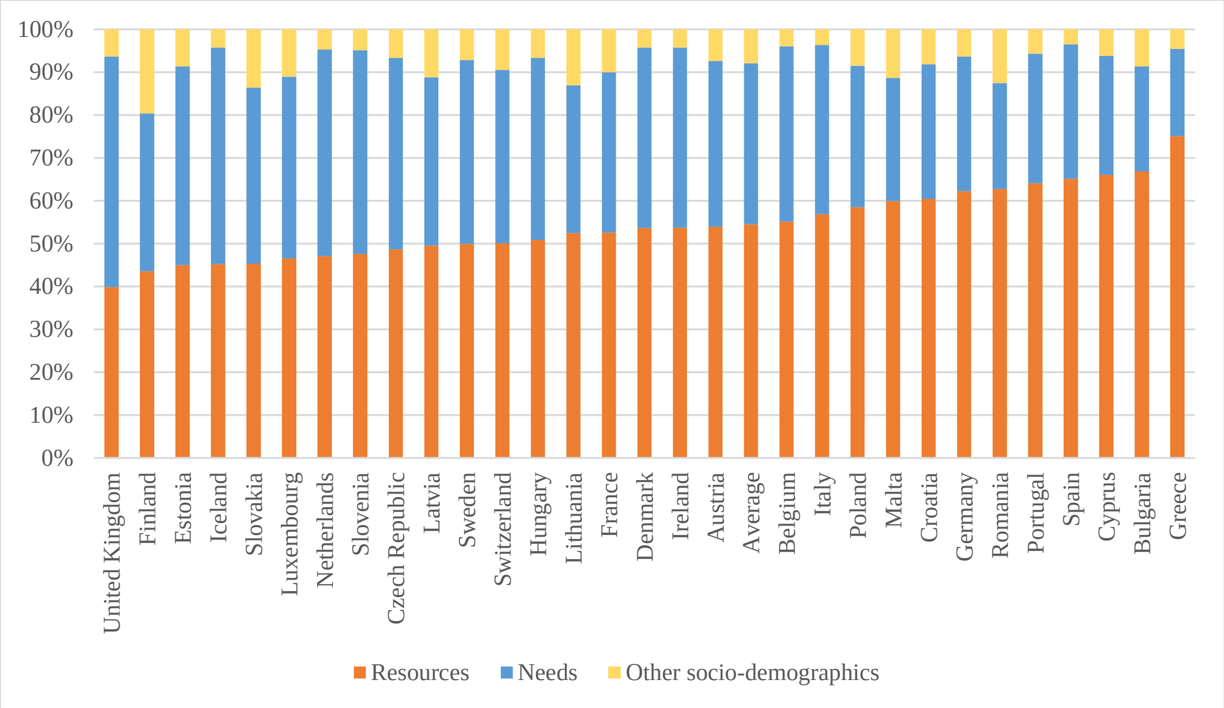
Among the socio-demographic variables included in the model, the number of children is positively related to child deprivation in all countries. Living in a single-parent household increases the risk of child deprivation in many countries (22 out of 31). In the countries where it is not, this can be interpreted as the fact that it is not living in single-parent households *per se* that increases the child deprivation intensity, but the associated characteristics of these households in terms of low income and low labour market attachment. The age of the oldest child has no significant relationship with the child deprivation risk in two thirds of the countries studied. This is an important result as it indirectly confirms that the composition of the 17 deprivation-item basket proposed by Guio and co-authors (2018) does not lead to systematic differences between age groups.

Table 2.2 – Shapley decompositions of the within-country pseudo R^2-measure

		Resources					Needs			Other socio-demograhpics
	R^2	**Income**	**Education**	**Quasi-joblessness**	**Debt burden**	**Migrant**	**Housing burden**	**Bad health**	**Rent**	
Belgium	0.23	28,2% (1)	11,8% (4)	8,4% (5)	4,5% (6)	2,3% (9)	21,2% (2)	4% (8)	15,7% (3)	4% (7)
Bulgaria	0.07	22,2% (2)	37,3% (1)	6,4% (5)	0,8% (8)	0,1% (9)	22% (3)	1,7% (6)	0,8% (7)	8,7% (4)
Czech Republic	0.20	20,5% (2)	16,2% (3)	8% (5)	3,8% (8)	0,1% (9)	31,8% (1)	4,1% (7)	8,8% (4)	6,7% (6)
Denmark	0.24	25,2% (1)	4% (7)	3,9% (8)	11,9% (4)	8,7% (5)	25,1% (2)	3% (9)	14% (3)	4,3% (6)
Germany	0.18	31,5% (1)	15,5% (3)	9,1% (5)	5,4% (7)	0,7% (9)	16,7% (2)	4,7% (8)	10% (4)	6,4% (6)
Estonia	0.14	19,3% (2)	11,1% (3)	9,5% (4)	3,9% (6)	1,1% (8)	42,3% (1)	2,9% (7)	1,1% (9)	8,7% (5)
Ireland	0.18	28,4% (2)	8,8% (4)	11,9% (3)	4,3% (6)	0,3% (9)	30,5% (1)	3,9% (8)	7,6% (5)	4,3% (7)
Greece	0.19	50,3% (1)	13,1% (3)	6% (4)	1,3% (9)	4,3% (6)	16,1% (2)	2,8% (7)	1,4% (8)	4,6% (5)
Spain	0.20	29% (1)	17,2% (3)	10,6% (4)	3,7% (7)	4,6% (5)	25,5% (2)	1,7% (9)	4,2% (6)	3,5% (8)
France	0.17	23,7% (2)	15,3% (3)	5% (6)	3,9% (8)	4,6% (7)	25,9% (1)	2,9% (9)	8,6% (5)	10% (4)
Croatia	0.15	26,9% (1)	18,8% (3)	12,8% (4)	1,9% (8)	1,5% (9)	21,6% (2)	5,4% (6)	2% (7)	8,9% (5)
Italy	0.14	26,8% (2)	15,6% (3)	5,3% (5)	4,3% (7)	4,8% (6)	30,1% (1)	2,7% (9)	6,7% (4)	3,7% (8)
Cyprus	0.13	35,6% (1)	16,2% (3)	5,6% (6)	6,7% (4)	1,9% (9)	20,9% (2)	3,4% (8)	3,5% (7)	6,2% (5)
Latvia	0.14	25% (2)	15,8% (3)	4,8% (5)	3,8% (6)	0,1% (9)	34,3% (1)	2,1% (8)	2,8% (7)	11,2% (4)
Lithuania	0.14	23,5% (2)	21,3% (3)	4% (5)	1,8% (7)	1,9% (6)	32,3% (1)	1,1% (9)	1,2% (8)	13,1% (4)
Luxembourg	0.20	22,8% (2)	9,9% (5)	1,8% (9)	8,4% (6)	3,6% (8)	24,7% (1)	3,8% (7)	13,9% (3)	11,1% (4)
Hungary	0.17	18,6% (3)	27,4% (2)	3,8% (5)	1% (8)	0,1% (9)	37,3% (1)	2,8% (6)	2,3% (7)	6,7% (4)
Malta	0.15	20,1% (2)	19,7% (3)	11,6% (4)	8,3% (6)	0,2% (9)	21,7% (1)	2,1% (8)	4,9% (7)	11,4% (5)
The Netherlands	0.25	22,3% (2)	8,4% (4)	5,1% (6)	6,8% (5)	4,5% (8)	29,3% (1)	2,3% (9)	16,7% (3)	4,7% (7)
Austria	0.23	17,4% (3)	17,6% (2)	4% (8)	8,9% (5)	6% (7)	22,6% (1)	4% (9)	12,1% (4)	7,4% (6)
Poland	0.13	29,6% (1)	22,3% (3)	3,2% (6)	3% (7)	0,3% (9)	24,9% (2)	3% (8)	5,1% (5)	8,5% (4)
Portugal	0.17	37,2% (1)	19,8% (3)	5,1% (6)	1,6% (8)	0,3% (9)	21,8% (2)	2,6% (7)	5,8% (4)	5,7% (5)
Romania	0.09	30,1% (1)	26,8% (2)	2,8% (5)	2,7% (6)	0,3% (9)	22,4% (3)	2% (7)	0,3% (8)	12,6% (4)
Slovenia	0.17	16,9% (2)	16,3% (3)	3,9% (6)	7% (4)	3,5% (7)	43,3% (1)	1,8% (9)	2,4% (8)	4,9% (5)
Slovakia	0.14	7,2% (5)	20% (2)	13,3% (4)	4,6% (6)	0,2% (9)	37,1% (1)	1,5% (8)	2,6% (7)	13,6% (3)
Finland	0.17	18,3% (3)	7,7% (5)	8,9% (4)	6,5% (7)	2,1% (8)	29,6% (1)	0,6% (9)	6,6% (6)	19,6% (2)
Sweden	0.28	13,6% (4)	4% (8)	6,5% (7)	14% (3)	11,8% (5)	21,5% (1)	3,2% (9)	18,2% (2)	7,2% (6)
United Kingdom	0.19	15% (3)	7,9% (5)	8,7% (4)	7,5% (6)	0,7% (9)	23,7% (2)	3,8% (8)	26,3% (1)	6,4% (7)
Iceland	0.16	14,4% (4)	12,2% (5)	3,2% (8)	15,1% (3)	0,3% (9)	29,2% (1)	16% (2)	5,3% (6)	4,3% (7)
Serbia	0.13	31,9% (1)	17,1% (3)	10,9% (4)	0,5% (7)	0,1% (9)	23,9% (2)	7,2% (6)	0,3% (8)	8,2% (5)
Switzerland	0.20	18,4% (2)	9% (6)	5,6% (8)	9,9% (4)	7,1% (7)	21,3% (1)	1,4% (9)	17,7% (3)	9,5% (5)
Average	0.17	25.3% (1)	15.3% (3)	6.9% (6)	4.9% (7)	2.7% (9)	24.7% (2)	3.1% (8)	10% (4)	7% (5)

Source: EU-SILC (2014) cross-sectional data, authors' computation. *NB:* The income column includes the relative contribution of the household disposable income variable and the self-employment dummy. For Croatia, the "light housing burden" variable has been dropped, as the Shapley decomposition model did not converge when this variable is included. *Reading note*: The R^2 captures the relative fit of the (full) model to the data. The percentages reflect the relative contribution to the fit and the number between brackets ranks the variables according to their respective relative contribution (e.g. for Belgium, income and self-employment have the highest relative contribution to the fit (1), i.e. 28,2% to the R^2-measure of 0.23).

Figure 2.1 – Shapley decompositions of the within-country pseudo R^2-measures



Source: EU-SILC 2014 cross-sectional data, authors' computation. *NB:* "Resources" refers to income, self-employment, low and medium education, (quasi-)joblessness, debt burden and migration; "Needs" refers to light and heavy housing cost burden, rent and bad health; "Other socio-demographics" refers to number of dependent children, age of the oldest child, single parent.

Table 2.3 – Single level negative binomial model

		Resources							Needs				Other socio-demographics		
Country	**Intercept**	**Household income**	**Low education**	**Medium education**	**(Quasi-) jobless**	**Self-employment**	**Debt burden**	**Migrant**	**Heavy housing burden**	**Light housing burden**	**Bad health**	**Rent**	**Number of dependent children**	**Single parent**	**Age of oldest child**
Belgium	-0.2934	-0.0001***	0.5582***	0.3364***	0.2649***	-0.5986***	0.5497***	0.0046	1.5538***	0.7538***	0.3504***	0.7013***	0.029	0.2258***	-0.0142
Bulgaria	0.9403***	-0.0001***	0.7345***	0.3395***	0.1331**	-0.1736***	0.1375**	-0.0922	0.7595***	0.3546**	0.0801	0.0005	0.0041	0.1158	0.0244***
Czech Republic	-0.5801**	-0.0002***	0.9064***	0.5112***	0.086	-0.2469***	0.3204***	0.3518*	1.5299***	0.6606***	0.3321***	0.3648***	0.0811***	0.1972***	-0.0107*
Denmark	-1.2799***	-0.0001***	0.5404***	0.0504	-0.1335	-0.5449***	1.1392***	0.7624***	1.6162***	1.1928***	0.4008**	0.9339***	-0.0626	0.2154	0.0253*
Germany	-0.9912***	-0.0001***	0.9486***	0.5119***	0.6238***	-0.2738*	0.5777***	0.1995**	1.2815***	0.5561***	0.5807***	0.5677***	0.0833**	0.3078***	-0.0049
Estonia	-1.382***	-0.0001***	0.5481***	0.2768***	0.5406***	-0.4163***	0.419***	0.1699**	1.9254***	1.0666***	0.237***	-0.07	0.0353	0.3684***	0.0265***
Ireland	-0.6408***	-0.0002***	0.339***	0.1798***	0.2373***	-0.3254***	0.2902***	0.0049	1.9681***	1.2807***	0.5288***	0.2791***	-0.0233	0.1112***	-0.0017
Greece	0.8189***	-0.0001***	0.3755***	0.1781***	0.1048***	-0.0939***	0.0964***	0.1776***	0.9293***	0.5203**	0.2981***	0.1081***	0.0472***	0.1338**	0.0028
Spain	-0.5108**	-0.0001***	0.5756***	0.3957***	0.442***	-0.1505***	0.448***	0.3259***	1.2697***	0.1664	0.2251***	0.24***	0.0467***	0.066	0.0076**
France	-0.5168***	-0.0001***	0.6332***	0.3905***	0.2235***	0.01	0.3781***	0.3299***	1.164***	0.71***	0.3098***	0.344***	0.089***	0.2667***	0.0096*
Croatia	-23.6173***	-0.0002***	0.9207***	0.4176***	0.4551***	-0.1635*	0.2218***	0.1625**	24.0614***	23.1233	0.3335***	0.3527***	0.0939***	-0.163	0.0044
Italy	0.1116	-0.0001***	0.6864***	0.2191***	0.2158***	-0.2077***	0.4973***	0.3809***	0.6938***	-0.4688**	0.4857***	0.3692***	0.0746***	-0.0158	0.0035
Cyprus	0.0202	-0.0001***	0.3697***	0.1677***	0.1899***	-0.034	0.2848***	0.1525***	1.2895***	0.4985***	0.3278***	0.1542***	-0.0106	0.3755***	0.0135***
Latvia	-0.1542	-0.0001***	0.6017***	0.2827***	0.1481**	-0.2177***	0.2731***	0.1007	1.3495***	0.7091***	0.1841***	0.1223***	0.1118***	0.0906	0.0144***
Lithuania	-0.9646***	-0.0001***	0.8792***	0.4643***	0.0714	-0.4225***	0.2672***	0.4799***	1.7587***	1.133***	0.0943	0.1708**	0.13***	0.1155	0.0042
Luxembourg	-1.7437***	-0.0001***	0.3623***	0.1219	-0.1286	-0.3858*	0.6929***	0.4037***	1.5178***	0.5754**	0.5629***	0.6549***	0.009	0.8042***	0.0058
Hungary	-0.5097***	-0.0002***	1.0159***	0.5985***	-0.0212	-0.6102***	0.1136***	-0.1384	2.0151***	1.2331***	0.1543***	0.25***	0.0404***	0.2127***	-0.0017
Malta	-0.4359*	-0.0001***	0.5236***	0.1848**	0.3472***	-0.1432*	0.636***	0.1987***	1.0945***	0.4071**	0.5662***	0.1504**	0.1435***	0.266***	-0.0034
The Netherlands	-0.8299***	-0.0001***	0.5395***	0.2234***	0.0587	-0.0355	0.7384***	0.5932***	1.7179***	1.0258***	0.6247***	0.7235***	-0.0492*	0.4331***	-0.0082
Austria	-1.52***	-0.0001***	1.1523***	0.5769***	0.1478	-0.4813***	0.9784***	0.2211***	1.4519***	0.668***	0.3637***	0.6205***	0.066*	0.2845***	0.0051
Poland	-0.3773**	-0.0002***	1.0793***	0.6337***	0.076	-0.4437***	0.3795***	0.6914***	0.9752***	0.0262	0.2113***	0.3569***	0.1073***	0.3239***	0.0037
Portugal	0.261*	-0.0001***	0.5541***	0.2571***	0.1008**	-0.4336***	0.1884***	0.1799***	1.1159***	0.5653***	0.1639***	0.183***	0.0268	0.0312	0.0091**
Romania	1.1457***	-0.0003***	0.5131***	0.3385***	0.1211*	-0.0396	0.2779***	-15.2373	0.7842***	0.3786***	0.1684***	0.0965	0.0486***	0.3355***	0.0059
Slovenia	-1.1679***	-0.0001***	0.7046***	0.3442***	0.1563**	-0.2937***	0.437***	0.2404***	1.8831***	0.8745***	0.4594***	0.1622***	0.1164***	0.1743**	-0.0024
Slovakia	-1.5961***	-0.0001***	0.8941***	0.4741***	0.552***	-0.2366***	0.206***	1.2629	2.02***	1.0698***	0.2235***	0.2067***	0.1554***	0.2993**	0.0071
Finland	-1.4217***	-0.0001***	0.5983***	0.2891***	0.6315***	-0.1197*	0.5583***	0.4719***	1.5859***	0.828***	0.3424**	0.4078***	0.0927***	0.2259***	0.0004
Sweden	-2.6208***	-0.0001***	0.0472	0.4236***	0.6224***	-0.3495*	1.3778***	0.804***	1.6201***	1.1335***	0.7127***	0.8543***	0.0747*	0.2699*	0.0226
United Kingdom	-0.9145***	-0.0001***	0.3394***	0.1905***	0.2885***	-0.0731	0.3892***	0.0976**	1.0651***	0.4919***	0.4128***	0.8403***	-0.0141	0.1425***	0.0153***
Iceland	-0.5677**	-0.0001***	0.5673***	0.2272***	0.0616	-0.0887	0.5411***	-0.2607*	1.067***	0.326**	1.0468***	0.2701***	-0.0149	0.0038	0.014
Serbia	0.4812*	-0.0003***	0.6203***	0.2509***	0.2514***	-0.3455***	0.1104***	-0.0064	0.6683***	-0.1271	0.2625***	0.1341***	0.0394***	0.141**	0.0165***
Switzerland	-2.8659***	-0.0001***	0.5984***	0.3126***	0.902***	0.3948**	0.7305***	0.55***	1.7356***	0.9975***	0.3889*	0.926***	0.1328***	0.7218***	-0.0001

Source: EU-SILC 2014 cross-sectional data, authors' computation.

2.5.2 European multilevel model

We add a multilevel structure to investigate the *between-country* differences in child deprivation across the 31 countries analysed. We start with an empty random intercept model (M1, Table 2.4) and gradually introduce variables. First, the household-level variables are added (M2, Table 2.4). Next, we use a series of models containing one institutional variable, with the aim of comparing their between-country explanatory strengths (M3-M12, Table 2.4). In the next set of models, we introduce GDP per capita levels and the unemployment rate to assess which institutional variables remain significant after controlling for macroeconomic circumstances (M13-M22, Table 2.4). We then investigate the relationship of social spending, in terms of spending size and pro-poorness of cash transfers, with child deprivation when household income is not regressed (M23-M26, Table 2.4). Finally, we assess cross-level interactions between the household-level variables and GDP per capita (Table 2.5). The estimated residuals at the country-level are given in Table 2.A5 in Appendix 2.

2.5.2.1. M1-M2: Empty and household-level model

The variance in random intercepts (0.70) of the empty model (M1, Table 2.4) indicates that significant differences in child deprivation exist between the 31 countries covered.

The household-level variables are introduced in M2 (Table 2.4). The sign and magnitude of the coefficients are in line with the results from the single level analysis. The household-level variables explain a large share of the original unobserved between-country differences of the empty model (71%). Most of the between-country explanatory power of the household-level variables is driven by household income: it explains 57 per cent of the original variation in random intercepts. The other household-level variables (i.e.; cross-country *compositional differences* in education, (quasi-)joblessness, needs (and related costs), socio-demographics etc.) play a much smaller role: they account for only 19 per cent of the unobserved between-country differences.

2.5.2.2. M3-M12: Assessing the association of institutional variables

Models 3 to 12 each add one institutional variable to M2. All ten institutional variables have a statistically significant negative relationship with child deprivation intensity, when they are introduced separately. The purpose of the current set of models is to assess whether social spending explains between-country differences, once differences in household determinants are taken into account. Several conclusions can be drawn.

The Shapley decompositions reveal that *in-kind social benefits* have a greater between-country explanatory power than *cash transfers*. This is to be expected: in-kind social transfers are not included at the micro level, whilst cash transfers are included in household income. This result holds when transfers are expressed as a percentage of GDP and in PPS per head/child. In-kind benefits expressed as a percentage of GDP or in PPS per head explain, respectively, 28 and 35 per cent of the unobserved between-country differences (M5 and M8). The corresponding figures for cash benefits are 8 and 23 per cent. This shows that the provision of in-kind services freely (or at a reduced rate) is a crucial driver. It allows households to spend their resources on other goods and necessities (see Aaberge et al., 2017). However, one must not conclude that cash-transfers are, policy-wise, less important: in our model, their role is more limited, given the fact that we control for individual household incomes.

The model further indicates that *social spending targeted at families* is negatively associated with child deprivation. Specifically, social spending devoted only to children and families explains 15% (% GDP, M9) and 19% (in PPS per head, M10) of the between-country differences. Whilst it is to be expected that in-kind transfers targeted at families have a negative relationship with child deprivation, even when household incomes are included at the micro level, it is difficult to explain why cash transfers targeted at families would have this result. However, both measures of family targeting (in % of GDP or in PPS) are highly correlated with GDP (see Table 2.A4 in Appendix 2). The next round of models control for such differences between countries and test whether the impact of pro-families' transfers is still significant (Models M19-M20). Next, the *pro-poorness of cash transfers* is also negatively associated with child deprivation intensity, even if it explains only a minor part of the between-country differences in child deprivation (9%, M11). Variables that capture the size of social spending are comparatively much more effective in explaining between-country differences in child deprivation.

The final institutional concept we regressed is the *adequacy of minimum income* to attain the poverty threshold. This variable explains a non-negligible amount of between-country differences in child deprivation intensity (16%, M12).

Without controlling for GDP per capita, social benefits expressed in PPS per child/head explain between-country differences in child deprivation intensity more effectively than the social benefit concepts expressed as a percentage of GDP. This is easily explained: the latter concept captures the relative size of social benefits within the economy, whereas the former also captures differences in absolute living standards. The next round of models (M13-M22) will take these differences into account.

2.5.2.3. M13-M22: The role of GDP

In models M13-M22, we introduce GDP per capita levels and the unemployment rate to assess whether social benefits remain significant after controlling for macroeconomic circumstances. These models show that *in-kind social benefits* (in % of GDP [M15] and in PPS per head [M18]), *pro-poorness of social transfers* (M21) and *the proportion in GDP of total social benefits* (which regroups in-kind and in-cash transfers [M13]) have a significant negative relationship with the intensity of child deprivation. Family benefits [M19-M20] and cash transfers (in % of GDP [M14] and in PPS [M17]) as well as the total social benefits (in-kind plus in-cash) in PPS per child [M16] and measures of adequacy of minimum income safety nets [M22] are not significant once differences in GDP are taken into account.

By looking at the explanatory power of the significant variables, we can conclude that:

A. In-kind services explain 21% (% GDP, [M15]) to 24% (PPS/head, [M18]) of between-country differences, once GDP is included in the model (as against 28% [M5] and 35% [M8] when it was not). This means that in-kind benefits remain a key variable to explain differences in child deprivation between the 31 pooled countries, even when differences in economic development (GDP per capita) are taken into account.

B. The global generosity of the welfare state (total transfers in % of GDP, [M13]) accounts for 16% of between-country differences and is mainly driven by social transfers in-kind, as social transfers in-cash have an insignificant relationship with child deprivation once the level of GDP per capita is controlled for. Some of the conclusions of previous papers on the relationship of cash social transfers with child deprivation thus have to be nuanced in the light of this result. Papers which combine both in-cash and in-kind transfers into a single global variable have highlighted a significant effect of transfers on child deprivation (once household income and GDP per capita are controlled for) which might not have been significant if only in-cash transfers had been included in the model (Bárcena-Martín et al., 2017b). Of course, this result does not imply that cash benefits have no relationship at all with deprivation. Instead, the cushioning effect of cash transfers is likely captured by household income, which consists of both market income and cash transfers. We will test this assumption in the next set of models [M23-M25] by excluding household income from the model.

C. Pro-poorness of social transfers explains 7% of between-country differences.

The models with social benefits expressed as a percentage of GDP provide a slightly better explanation of between-country differences *as a whole* than models with social benefit size expressed in PPS per head/child when differences in living standards are controlled for (through GDP per capita), but the difference is negligible. Relative indicators (in proportion of GDP) provide information on the way a country prioritises on social transfers, whereas transfers expressed in PPS provide information on the level of such transfers.

GDP per capita is an important predictor of child deprivation intensity and explains 14 to 20% of the total unobserved between-country differences, depending on the social spending concept that is co-regressed. The unemployment rate coefficient is insignificant and explains only 5 to 8% of the original unobserved country differences. In the interpretation of the latter result, it is important to stress that household (quasi-)joblessness is already regressed at the individual level and that the inclusion of the national unemployment rate mainly aims at accounting for the possible effect of the business cycle on the size and pro-poorness of social benefits.

The fact that GDP per capita has a negative association with child deprivation, while individual household income and other micro-drivers are controlled for, is not expected a priori and deserves further interpretation. Why should children with similar household socio-economic background and household income be better protected against deprivation if they live in more prosperous countries?

One reason could be that countries with higher GDP per capita provide more in-kind benefits, which would reduce deprivation for given income levels. We tested this and our results indicate that this is partly true. When in-kind transfers are included in PPS per head, GDP is not significant due to the high correlation between in-kind transfers level (PPS) and GDP [M18]. However, GDP per capita remains significant *after controlling for in-kind benefits in % of GDP* [M15]. This result implies that GDP per capita may also capture the effects of some "hidden" contextual variables which cannot be included in the model with the available data, such as the average household wealth and the size of gifts between households. One may also conjecture (though this hypothesis would need further examination) that richer countries have features that lead to less volatility of incomes, notably within the working-age population and at the bottom end of the income distribution: a larger public sector and better functioning automatic stabilisers in their welfare edifice reduce this volatility.[47] In other words, it seems plausible to argue that these contextual variables increase households' "permanent income", notably within the working-age population and at the bottom end of the income distribution, and therefore reduce child deprivation. Another possible reason might be that GDP per capita is a proxy for "qualitative" differences, such as the effectiveness of public support, notably the quality of

[47] A rough proxy for "volatility" might be the size of cash benefits in relation to GDP, as cash benefits tend to reduce volatility of incomes. We should immediately concede that we do not find convincing evidence for this hypothesis: cash benefits (as a % of GDP) have a negative, but weakly insignificant (p=0.11) relationship with child deprivation after controlling for GDP per capita [M14].

public social services. Richest countries are expected to provide public services of better quality (education, childcare, public transport systems, etc.), which should increase permanent income and/or decrease household needs and related costs in the most effective way.

To sum up, it may be the case that GDP is a proxy for the overall "level of social development" of societies, which can only be partially measured by existing data: individual household income and the other micro-determinants of child deprivation are insufficient to measure the overall, societal "level of development" which has a favourable (statistically negative) relationship with the intensity and individual risk of deprivation.

2.5.2.3. M23-M26: Sensitivity to household disposable income

Models 23 to 25 confirm the *cushioning effect of cash transfers through individual household income*. These models replicate models M13-15, except that household disposable income is no longer included and the pro-poorness of cash transfers is included. They show that all social spending concepts (total, cash, in-kind) have a statistically significant negative relationship with child deprivation intensity after controlling for the unemployment rate, GDP per capita, household-level risk factors (with the exception of individual household income) and the pro-poorness of cash social benefits. This is a very important result regarding the results of previous papers on the relationship of social transfers with deprivation, as this confirms our expectation that only social transfers not included in household income have a relationship with child deprivation.

2.5.2.4. Cross-level interactions

Several multilevel deprivation studies have pointed out that the association of variables at the household level with child deprivation should not be understood independently from variables at the country level. The general consensus in these studies is that the coefficients of certain risk factors at the individual level are influenced by countries' level of affluence or welfare state generosity (Nelson, 2012; Bárcena-Martín et al., 2014; Visser et al., 2014; Saltkjel & Malmberg-Heimonen, 2017). We examine this relationship by introducing a series of cross-level interactions between GDP per capita and the household-level variables. We also add random slopes to the household-level variables to ensure that the coefficients of the cross-level interactions with GDP per capita are not influenced by other effects.[48] All random slopes, with

[48] A random slope allows the relationship between the explanatory variable and an independent variable at the household-level to be different for each country by adding a random term to the coefficient of the household-level variable. The covariance between the random intercepts and the random slopes were not estimated for

the exception of the age of the oldest child, are statistically different from zero. This confirms our findings from the single level analysis that the relationship of the household-level variables with child deprivation differs across countries. The results of our cross-level interactions are shown in Table 2.5. Specifically, we find that GDP per capita levels mitigate the association for the household-level variables that relate to households' resources, while they increase the relationship for variables that capture households' needs.

Variables that capture or directly influence households' ability to generate *resources* on the labour market have a more moderate relationship with child deprivation intensity in the more affluent countries, except variables with regard to debt burden and migration background (see below). A positive cross-level interaction between GDP per capita and household income indicates that the negative association of household income becomes smaller when GDP per capita increases. So, household income has a larger effect in less affluent countries. In addition, the negative cross-level interaction between the low and medium education dummies and GDP per capita indicates that the negative relationship of low education with deprivation is smaller in the most affluent countries, i.e. children in low-educated households are better protected from deprivation in the more affluent countries. Whelan and Maître (2012) already showed for the whole population that the negative relationship with deprivation of lacking educational qualifications increases as GDP declines. However, in contrast to their results, in our model the interaction effects do not explain away the impact of GDP per capita as an independent variable. These results imply that the variables in our model that aim to capture households' command on resources have a relatively stronger association with child deprivation in countries with a low standard of living than in countries with a high standard of living. Finally, while the coefficient of (quasi-)joblessness varies across countries (i.e. the random slope is significant), it does not depend on GDP per capita.

The results indicate that the deprivation-increasing (i.e., statistically positive) relationship of variables related to household *needs* (such as having a heavy housing cost burden, renting one's dwelling or having at least one household member struggling with bad health) with child deprivation increases if GDP per capita increases. The cross-level interaction with the light housing burden dummy is positive, but not significant. These results confirm the single level analysis, in which variables that measure household needs/costs contribute more to the fit in richer countries. As argued in the previous section, a plausible interpretation for this result is that households living in more affluent countries also face relatively higher personal costs related to housing and health.

computational reasons. We also conducted a robustness check of a model that does not include random slopes. The results indicate that none of the significant cross-level interactions lose their significancy or change sign. Two insignificant relationships (i.e. slight housing burden, number of children) become significant once the random slopes are dropped.

The coefficient of being a single parent or having someone in the household with a migration background is larger in the more affluent countries. The cross-level interaction between GDP per capita and the number of children living in the household and the age of the oldest child is insignificant.

Table 2.4 – Multilevel negative binomial model (M1-M2, empty and household-level model)

	M1			M2		
	Coeff.	p>z	Shapley R²	Coeff.	p>z	Shapley R²
Household-level variables						
Household income						0.57
Household income				-0.03	0.00	
Other						0.14
Self-employment				-0.19	0.00	
(Quasi-)joblessness				0.32	0.00	
Low education				0.75	0.00	
Medium education				0.41	0.00	
Bad health				0.35	0.00	
Heavy housing burden				1.51	0.00	
Light housing burden				0.75	0.00	
Rent				0.33	0.00	
Debt burden				0.41	0.00	
Number of dependent children				0.14	0.00	
Single parent				0.07	0.00	
Age of oldest child				0.01	0.00	
Migrant				0.30	0.00	
Constant	0.34	0.03		-1.11	0.00	
Random Estimates						
Random intercept	0.70	0.00		0.20	0.00	
Explained between-country variance				0.71		
Over-dispersion parameter	1.91	0.00		0.66	0.00	
Model information						
N of observations	88,901			88,901		
N of countries	31			31		

Source: EU-SILC 2014 cross-sectional data, authors' computation.

Table 2.4 – Multilevel negative binomial model (M3-M7, assessing the association of institutional variables)

	M3			M4			M5			M6			M7		
	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2
Household-level variables															
Household income			0.45			0.51			0.42			0.37			0.41
Other			0.13			0.15			0.1			0.13			0.15
Country-level variables															
Total social benefits, % of GDP	-0.04	0.00	0.20												
Cash social benefits, % of GDP				-0.04	0.04	0.08									
In-kind social benefits, % of GDP							-0.09	0.00	0.28						
All social benefits, in PPS per child										-0.12	0.00	0.31			
Cash social benefits, in PPS per child													-0.19	0.00	0.23
Random Estimates															
Random intercept	0.15	0.00		0.18	0.00		0.14	0.00		0.14	0.00		0.15	0.00	
Explained between-country variance	0.78			0.75			0.80			0.81			0.78		
Over-dispersion parameter	0.66	0.00		0.66	0.00		0.66	0.00		0.66	0.00		0.66	0.00	
Model information															
N of observations	88,901			88,901			88,901			88,901			88,901		
N of countries	31			31			31			31			31		

Source: EU-SILC 2014 cross-sectional data, authors' computation. *NB*: The coefficients of the household-level variables (with the exception of income) are not shown and are very similar to the coefficients in M2.

Table 2.4 – Multilevel negative binomial model (M8-M12, assessing the association of institutional variables, continued)

	M8			M9			M10			M11			M12		
	Coeff.	p>z	Shapley R²	Coeff.	p>z	Shapley R²	Coeff.	p>z	Shapley R²	Coeff.	p>z	Shapley R²	Coeff.	p>z	Shapley R²
Household-level variables															
Household income			0.35			0.48			0.44			0.52			0.51
Other			0.12			0.12			0.14			0.14			0.07
Country-level variables															
In-kind social benefits, in PPS per head	-0.24	0.00	0.35												
Family cash social benefits, % of GDP				-0.20	0.04	0.15									
Family cash social benefits, PPS per head							-0.24	0.00	0.19						
Pro-poorness bottom 50										-0.02	0.02	0.09			
Adequacy of minimum-income													-0.01	0.03	0.16
Random Estimates															
Random intercept	0.13	0.00		0.18	0.00		0.16	0.00		0.17	0.00		0.18	0.00	
Explained between-country variance	0.82			0.75			0.77			0.75			0.75		
Over-dispersion parameter	0.66	0.00		0.66	0.00		0.66	0.00		0.66	0.00		0.71	0.00	
Model information															
N of observations	88,901			88,901			88,901			88,901			88,901		
N of countries	31			31			31			31			29		

Source: EU-SILC 2014 cross-sectional data, authors' computation. *NB*: The coefficients of the household-level variables (with the exception of income) are not shown and are very similar to the coefficients in M2.

Table 2.4 – Multilevel negative binomial model (M13-M17, the role of GDP)

	M13			M14			M15			M16			M17		
	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2
Household-level variables															
Household income			0.30			0.33			0.30			0.29			0.32
Other			0.12			0.14			0.09			0.12			0.13
Country-level variables															
GDP per capita	-0.01	0.03	0.18	-0.02	0.01	0.19	-0.02	0.01	0.17	-0.01	0.60	0.14	-0.02	0.12	0.15
Unemployment rate	0.00	0.74	0.07	0.01	0.57	0.08	-0.01	0.58	0.06	-0.01	0.65	0.05	0.00	0.88	0.06
Total social benefits, % of GDP	-0.03	0.03	0.16												
Cash social benefits, % of GDP				-0.03	0.11	0.07									
In-kind social benefits, % of GDP							-0.06	0.01	0.21						
All social benefits, in PPS per child										-0.10	0.13	0.20			
Cash social benefits, in PPS per child													-0.06	0.59	0.14
Random Estimates															
Random intercept	0.12	0.00		0.13	0.00		0.12	0.00		0.13	0.00		0.14	0.00	
Explained between-country variance	0.83			0.81			0.83			0.81			0.80		
Over-dispersion parameter	0.66	0.00		0.66	0.00		0.66	0.00		0.66	0.00		0.66	0.00	
Model information															
N of observations		88,901			88,901			88,901			88,901			88,901	
N of countries		31			31			31			31			31	

Source: EU-SILC 2014 cross-sectional data, authors' computation. *NB*: The coefficients of the household-level variables (with the exception of income) are not shown and are very similar to the coefficients in M2.

Table 2.4 – Multilevel negative binomial model (M18-M22, the role of GDP, continued)

	M18			M19			M20			M21			M22		
	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2
Household-level variables															
Household income			0.28			0.35			0.34			0.34			0.36
Other			0.10			0.12			0.13			0.13			0.10
Country-level variables															
GDP per capita	0.00	0.71	0.14	-0.02	0.01	0.18	-0.02	0.05	0.17	-0.02	0.00	0.20	-0.02	0.00	0.19
Unemployment rate	-0.01	0.52	0.05	0.00	0.92	0.06	0.00	0.90	0.06	0.00	0.99	0.07	-0.01	0.61	0.05
In-kind social benefits, in PPS per head	-0.23	0.03	0.24												
Family cash social benefits, % of GDP				-0.02	0.86	0.09									
Family cash social benefits, PPS per head							-0.03	0.83	0.11						
Pro-poorness bottom 50										-0.02	0.05	0.07			
Adequacy of minimum-income													-0.01	0.32	0.10
Random Estimates															
Random intercept	0.13	0.00		0.14	0.00		0.14	0.00		0.13	0.00		0.14	0.00	
Explained between-country variance	0.82			0.80			0.80			0.82			0.80		
Over-dispersion parameter	0.66	0.00		0.66	0.00		0.66	0.00		0.66	0.00		0.71	0.00	
Model information															
N of observations	88,901			88,901			88,901			88,901			88,901		
N of countries	31			31			31			31			31		

Source: EU-SILC 2014 cross-sectional data, authors' computation. *NB*: The coefficients of the household-level variables (with the exception of income) are not shown and are very similar to the coefficients in M2.

Table 2.4 – Multilevel negative binomial model (M23-M25, sensitivity to household disposable income)

	M23			M24			M25		
	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2	Coeff.	p>z	Shapley R^2
Household-level variables									
Household income									
Other			0.17			0.20			0.13
Country-level variables									
GDP per capita	-0.03	0.00	0.26	-0.03	0.00	0.28	-0.03	0.00	0.26
Unemployment rate	0.01	0.40	0.09	0.02	0.24	0.11	0.00	0.81	0.07
Total social benefits, % of GDP	-0.04	0.01	0.19						
Cash social benefits, % of GDP				-0.05	0.03	0.09			
In-kind social benefits, % of GDP							-0.09	0.01	0.25
Pro-poorness bottom 50	-0.02	0.16	0.07	-0.02	0.05	0.08	-0.02	0.18	0.07
Random Estimates									
Random intercept	0.15	0.00		0.17	0.00		0.16	0.00	
Explained between-country variance	0.78			0.77			0.78		
Over-dispersion parameter	0.83	0.00		0.83	0.00		0.83	0.00	
Model information									
N of observations	88,901			88,901			88,901		
N of countries	31			31			31		

Source: EU-SILC 2014 cross-sectional data, authors' computation. *NB*: The coefficients of the household-level variables (with the exception of income) are not shown and are very similar to the coefficients in M2.

Table 2.5 – Negative binomial model with cross-level interactions

	Main effects		**Interaction with GDP per capita**		**Random estimates**	
	Coeff.	p>z	Coeff.	p>z	Coeff.	p>z
Household-level variables						
Household income						
Household income	-0.04	0	0.003	0.03	0.00008	0
Other						
Self-employment	-0.21	0.02	0.003	0.93	0.02	0
Quasi-joblessness	0.25	0	0.02	0.58	0.03	0
Low education	0.94	0	-0.08	0.02	0.04	0
Medium education	0.53	0	-0.05	0.03	0.01	0
Debt burden	0.1	0.28	0.13	0	0.04	0
Bad health	0.18	0	0.07	0	0.01	0.04
Heavy housing burden	1.18	0	0.1	0.05	0.06	0
Light housing burden	0.49	0	0.08	0.13	0.06	0
Rent	-0.08	0.38	0.16	0	0.04	0
Number of dependent children	0.11	0	0.01	0.17	0	0
Single parent	-0.1	1.1	0.07	0	0.01	0.02
Age of oldest child	0.01	0	-0.0004	0.77	0.00002	0.16
Migrant	0.12	0.27	0.06	0.09	0.04	0
Constant	0.37	0.71			0.28	0
Country-level variables						
GDP per capita	-0.39	0				
Unemployment rate	0.03	0.2				
Total social benefits, % of GDP	-0.02	0.38				
Pro-poorness (bottom 50)	0	0.85				
Model information						
Over-dispersion parameter	0.55	0				
N of observations	88,901					
N of countries	31					

Source: EU-SILC 2014 cross-sectional data, authors' computation. *NB*: GDP per capita is expressed in PPS per 10,000, instead of in PPS per 1,000.

2.6. Conclusion

This paper is the first one that analyses the determinants of child deprivation in 31 European countries, using the scale officially adopted in March 2018 to measure child-specific deprivation at EU level. Our analyses show that the factors which are important in explaining child deprivation within countries are not necessarily the same as those explaining variation between countries. They demonstrate that both single and multilevel models are useful and complementary to explain child deprivation in the 31 countries analysed (all 28 EU countries as well as Iceland, Serbia and Switzerland).

In regard to *within-country* differences in child deprivation, the single level model is the most effective in explaining child deprivation in countries with the lowest share of child deprivation (Austria, Belgium, Denmark, Netherlands and Sweden). Conversely, the countries where the single level model has a lower explanatory power are Bulgaria, Cyprus, Estonia, Italy, Latvia, Lithuania, Malta, Poland, Portugal, Romania, Serbia and Slovakia. In several of these countries, this may be because the general standard of living is low and all children have, as a consequence, a greater likelihood of being (more) deprived.

In all countries analysed, the results confirm the combined relationship of variables related to the “longer-term command over resources” (current household income, parents’ education, household labour market attachment, burden of debts, migration status) and variables indicating household needs (costs related to housing, tenure status and bad health) with child deprivation. The three most powerful predictors are: housing cost burden, household income and educational level of parents. However, our results also clearly illustrate that the explanatory power of the household-level variables differs between countries. In the richest countries, the explanatory power of the variables related to household needs is the largest, whereas in the most deprived countries, the explanatory power of resources is generally greater (with the exception of debt and migration). This means that countries not only differ in terms of socio-economic composition (as stated in most papers explaining differences in deprivation between countries), but also in terms of the association of each variable with the child deprivation risk, i.e. household income, (quasi-)joblessness, housing cost burden do not have the same relationship with child deprivation across countries. Our results highlight that the age of the oldest child has no significant relation with the child deprivation in two thirds of the countries studied. This is an important result as it indirectly confirms that the composition of the deprivation basket does not lead to systematic differences between age groups.

In regard to *between-country* differences, we ran a large number of multilevel models and compared them systematically, to identify those results which remain robust to alternative specifications (i.e. total, cash and in-kind social spending as a % of GDP and in PPS per head/child, total and cash social spending on families and children as a % of GDP and in PPS per head/child, pro-poorness of social spending, adequacy of minimum income benefit schemes; with and without controlling for GDP per capita or household income).

Our results indicate that all social spending concepts have a statistically significant negative relationship with child deprivation (i.e. they reduce it), when GDP per capita is omitted. However, once GDP per capita and the household-level variables (including household income) is controlled for, only the level of in-kind social benefits provided and the pro-poorness of social transfers have a significant negative relationship with child deprivation. This confirms our expectation that only social transfers not included in household income at micro-level play a role in predicting child deprivation. The between-country explanatory power of the pro-poorness of cash transfers is limited, whereas in-kind social benefits levels is a crucial variable. We further showed that the impact of cash benefits operates mainly through household income (i.e. aggregated cash transfer levels are only significant when household income is omitted from the model). This explains and also qualifies the conclusions of papers which have analysed the relationship of social transfers with differences in deprivation in the EU, using multilevel models but without controlling for individual household income. This should not lead to the conclusion that cash transfers are unrelated to child deprivation; what our model shows is, quite logically, that cash transfers don't have a statistically significant association independently from the distribution of household income at the micro level.

GDP per capita is an important predictor of child deprivation in nearly all model specifications, even when household income is co-regressed at the individual level. In total, GDP per capita and individual household income explain, respectively, between 14-20% and 28-36% of the between-country differences in child deprivation (depending on the other country-level variables that are included in the model). The observation that GDP per capita is negatively associated with child deprivation, while individual household income and other micro-drivers are controlled for, is not expected a priori. It seems that GDP per capita correlates with "hidden" contextual factors, which are not available from our dataset. The following factors come to mind: household wealth, between-households support in-kind, the quality and affordability of education, childcare, healthcare and public transport systems. In other words, GDP is a proxy for the "level of social development" of societies, and child deprivation correlates negatively with the "level of social development", so conceived. An additional hypothesis to explain this result, is that the notion of “affordability” changes with the average level of incomes; we cannot pursue this hypothesis in the context of this paper, but it needs further research.

Finally, crossed-effects in multilevel models also indicate that the association of certain individual risk factors is influenced by countries’ level of affluence. We find that GDP per capita levels mitigates the association of household-level variables that relate to households’ resources (except for debt and migration status, which we construe as components of “longer-term resources”) with child deprivation, while they increase the relationship of variables that capture households’ needs. These results confirm the findings from the single level analysis and illustrate the importance of looking at national drivers of child deprivation.

Appendix 2.

Table 2.A1 – Summary statistics

Country	Income poverty rate	Average number of items lacked	Low education	High Education	(Quasi-) jobless	Debt burden	Self employment	Migrant	Bad health	Rent	Heavy housing burden	Slight housing burden	Average N of dependent children	Average age of oldest child	Single parent
Belgium	0,20	1,24	0,13	0,34	0,14	0,15	0,13	0,22	0,10	0,29	0,39	0,34	2,43	9,42	0,16
Bulgaria	0,28	7,53	0,30	0,43	0,16	0,10	0,13	0,01	0,16	0,23	0,48	0,48	2,04	10,03	0,07
Czech Republic	0,11	1,24	0,04	0,66	0,08	0,08	0,21	0,02	0,06	0,23	0,31	0,63	2,12	9,37	0,12
Denmark	0,04	0,38	0,03	0,30	0,03	0,04	0,14	0,10	0,03	0,17	0,13	0,37	2,31	10,40	0,07
Germany	0,11	0,76	0,04	0,43	0,07	0,08	0,06	0,15	0,05	0,39	0,22	0,62	2,13	9,85	0,14
Estonia	0,26	1,18	0,11	0,46	0,08	0,07	0,11	0,12	0,09	0,20	0,28	0,54	2,28	9,61	0,08
Ireland	0,13	1,79	0,12	0,27	0,23	0,18	0,13	0,10	0,04	0,36	0,48	0,45	2,60	9,92	0,15
Greece	0,24	3,04	0,15	0,42	0,13	0,20	0,34	0,12	0,07	0,26	0,53	0,45	2,12	9,60	0,05
Spain	0,26	2,04	0,30	0,22	0,16	0,16	0,19	0,15	0,07	0,22	0,61	0,37	2,05	9,51	0,08
France	0,15	1,14	0,09	0,41	0,09	0,14	0,12	0,13	0,08	0,35	0,34	0,25	2,38	10,11	0,14
Croatia	0,18	1,81	0,12	0,68	0,14	0,31	0,14	0,19	0,24	0,15	0,70	0,28	2,26	10,19	0,03
Italy	0,17	1,59	0,19	0,52	0,08	0,12	0,28	0,12	0,07	0,28	0,61	0,38	1,98	9,60	0,09
Cyprus	0,14	2,47	0,08	0,39	0,08	0,50	0,11	0,20	0,04	0,22	0,82	0,15	2,29	9,86	0,06
Latvia	0,25	3,21	0,10	0,47	0,10	0,10	0,11	0,12	0,15	0,25	0,41	0,48	2,20	9,73	0,15
Lithuania	0,22	2,21	0,06	0,48	0,09	0,05	0,13	0,08	0,10	0,13	0,35	0,56	2,08	10,33	0,12
Luxembourg	0,11	0,67	0,25	0,38	0,05	0,21	0,08	0,23	0,10	0,29	0,44	0,43	2,21	9,52	0,13
Hungary	0,19	4,32	0,21	0,53	0,15	0,14	0,05	0,00	0,16	0,15	0,44	0,49	2,39	10,33	0,11
Malta	0,21	1,76	0,45	0,25	0,13	0,08	0,14	0,13	0,02	0,17	0,60	0,32	2,11	9,89	0,09
The Netherlands	0,05	0,43	0,06	0,33	0,04	0,02	0,15	0,07	0,02	0,13	0,10	0,43	2,37	9,88	0,10
Austria	0,09	0,71	0,08	0,45	0,07	0,06	0,17	0,20	0,09	0,39	0,18	0,60	2,22	9,98	0,14
Poland	0,21	1,83	0,07	0,60	0,06	0,11	0,25	0,01	0,14	0,20	0,66	0,31	2,18	9,79	0,06
Portugal	0,26	2,72	0,49	0,26	0,12	0,11	0,14	0,12	0,13	0,28	0,47	0,46	1,92	9,97	0,11
Romania	0,29	6,27	0,22	0,60	0,05	0,06	0,25	0,00	0,10	0,05	0,43	0,53	2,16	10,80	0,04
Slovenia	0,14	0,95	0,06	0,47	0,04	0,17	0,16	0,16	0,04	0,28	0,35	0,57	2,19	9,55	0,05
Slovakia	0,15	2,32	0,07	0,58	0,09	0,16	0,19	0,00	0,13	0,16	0,39	0,53	2,35	9,55	0,05
Finland	0,09	0,42	0,02	0,32	0,04	0,06	0,30	0,04	0,02	0,14	0,25	0,58	2,73	9,98	0,08
Sweden	0,10	0,28	0,05	0,34	0,04	0,05	0,13	0,18	0,03	0,23	0,07	0,37	2,45	9,41	0,09
The United Kingdom	0,11	1,50	0,19	0,30	0,16	0,17	0,12	0,16	0,06	0,44	0,38	0,44	2,29	9,43	0,23
Iceland	0,06	0,61	0,10	0,25	0,03	0,17	0,19	0,06	0,04	0,20	0,32	0,55	2,50	10,45	0,10
Serbia	0,26	3,55	0,20	0,59	0,19	0,25	0,23	0,11	0,32	0,22	0,75	0,24	2,29	9,85	0,03
Switzerland	0,05	0,32	0,04	0,34	0,03	0,06	0,11	0,18	0,02	0,42	0,28	0,61	2,27	9,77	0,08
Average	0,16	1,74	0,14	0,42	0,10	0,12	0,16	0,12	0,08	0,29	0,41	0,42	2,21	9,80	0,12

Source: EU-SILC (2014) cross-sectional data, authors' computation.

Table 2.A1 – Summary statistics (continued)

Country	Total social benefits (% of GDP)	In-kind social benefits (% of GDP)	In-cash social benefits (% of GDP)	Total social benefits (in PPS per head)	Cash social benefits (in PPS per child)	In-kind social benefit (in PPS per head)	Family cash social benefits (% of GDP)	Average gross cash benefits (1000 PPS per child)	Pro-poorness of cash social benefits (bottom 50, child population)	Adequacy of minimum income (% of median income)	GDP per capita (1000 PPS)	Unemployment rate (% of working age population)
Belgium	29,00	9,20	19,80	6,35	3,43	2,92	2,20	1,94	71,33	38,08	33,00	8,50
Bulgaria	17,90	5,60	12,30	2,07	1,30	0,77	1,90	0,45	55,49	20,37	12,80	11,40
Czech Republic	19,10	6,30	12,80	3,07	1,53	1,54	1,70	0,71	77,12	41,61	23,80	6,10
Denmark	32,20	12,70	19,50	6,17	2,69	3,48	3,60	1,26	79,64	63,15	35,10	6,60
Germany	27,80	10,60	17,20	7,44	3,69	3,75	3,10	2,56	66,27	54,09	34,60	5,00
Estonia	15,00	4,60	10,40	2,88	1,91	0,96	1,60	1,33	55,99	35,35	20,90	7,40
Ireland	19,40	7,30	12,10	6,95	4,65	2,31	2,50	2,46	76,96	64,12	37,70	11,30
Greece	25,50	5,00	20,50	2,04	1,08	0,96	1,10	0,35	76,15	7,88	19,40	26,50
Spain	24,90	7,60	17,30	3,60	1,91	1,69	1,30	0,07	70,29	22,78	24,70	24,50
France	32,20	11,70	20,50	6,76	3,32	3,43	2,50	1,72	71,82	38,73	29,60	10,30
Croatia	21,20	7,00	14,20	2,49	1,37	1,12	1,50	0,58	70,68	33,03	16,10	17,20
Italy	28,90	7,10	21,80	3,45	1,63	1,82	1,60	0,54	64,55	0,00	26,60	12,70
Cyprus	22,10	3,40	18,70	3,24	2,49	0,75	1,40	0,95	64,47		22,40	16,10
Latvia	14,30	4,00	10,30	2,03	1,36	0,67	1,30	0,70	60,51	41,86	17,50	10,80
Lithuania	14,00	4,60	9,40	2,45	1,50	0,95	1,10	0,62	59,41	40,91	20,70	10,70
Luxembourg	22,40	6,90	15,50	9,76	5,23	4,53	3,50	3,75	62,52	49,16	74,50	6,00
Hungary	19,70	6,30	13,40	3,23	2,10	1,13	2,30	1,50	64,06	23,93	18,70	7,70
Malta	18,80	6,50	12,30	3,32	1,96	1,35	1,20	0,92	77,86	35,94	24,90	5,80
The Netherlands	28,90	10,20	18,70	5,36	2,12	3,24	0,90	1,04	74,08	50,07	36,00	7,40
Austria	29,20	8,90	20,30	7,39	4,43	2,96	2,80	2,63	66,54	49,93	35,70	5,60
Poland	18,50	4,30	14,20	2,17	1,29	0,88	1,40	0,34	75,19	43,35	18,60	9,00
Portugal	25,50	6,80	18,70	2,81	1,49	1,32	1,20	0,29	76,84	29,33	21,20	14,10
Romania	14,40	4,30	10,10	1,29	0,62	0,67	1,20	0,25	59,31	23,45	15,30	6,80
Slovenia	23,70	7,60	16,10	3,77	2,10	1,67	1,90	1,45	70,71	41,73	22,80	9,70
Slovakia	17,90	6,10	11,80	2,80	1,53	1,27	1,70	0,79	65,87	28,62	21,30	13,20
Finland	31,10	11,80	19,30	6,85	3,62	3,24	3,20	2,25	71,28	48,34	30,50	8,70
Sweden	29,00	13,50	15,50	7,68	3,96	3,71	3,10	2,20	70,70	41,98	34,10	7,90
The United Kingdom	27,10	10,30	16,80	6,31	3,45	2,86	2,80	1,91	84,71	56,86	29,90	6,10
Iceland	23,70	11,10	12,60	6,36	2,96	3,41	2,70	1,07	78,51	52,10	32,50	5,00
Serbia	22,80	6,00	16,80	1,64	1,04	0,60	1,20	0,26	62,18		10,10	19,20
Switzerland	24,40	8,00	16,40	6,85	3,62	3,23	1,50	1,77	71,45	40,19	45,00	4,90
Average	26,04	8,64	17,41	4,97	2,57	2,40	2,10	1,28	71,27	36,58	27,58	10,43

Source: EU-SILC 2014 cross-sectional data, authors' computation.

Table 2.A2 – Correlation coefficients between household-level variables

	Average number of deprivation	Household income	Jobless	Self-employed	Debt burden	Low education	Medium educcation	Rent	Heavy housing burden	Light housing burden	Age of the oldest child	Bad health	Number of dependent children	Migrant	Single parent
Average number of deprivation	1	-0.34	0.31	-0.04	0.17	0.3	0.09	0.2	0.37	-0.21	0.09	0.18	0.11	0.1	0.11
Household income	-0.34	1	-0.22	0.02	0.08	-0.21	-0.22	-0.18	-0.26	0.07	0.04	-0.09	0.08	-0.03	-0.18
Jobless	0.31	-0.22	1	-0.13	0.04	0.27	0	0.24	0.17	0.12	0.02	0.16	0.08	0.07	0.27
Self-employed	-0.04	0.02	-0.13	1	-0.01	-0.01	0.01	-0.13	-0.01	0.01	0.04	0	0.02	-0.04	-0.11
Debt burden	0.17	0.08	0.04	-0.01	1	0.04	0.06	0.07	0.25	-0.16	0.02	0.07	0.02	0.03	0.04
Low education	0.3	-0.21	0.27	-0.01	0.04	1	-0.35	0.17	0.19	-0.11	0.08	0.13	0.1	0.09	0.1
Medium educcation	0.09	-0.22	0	0.01	0.06	-0.35	1	0.06	0.09	-0.02	0.07	0.05	-0.03	-0.03	0.03
Rent	0.2	-0.18	0.24	-0.13	0.07	0.17	0.06	1	0.1	-0.08	-0.08	0.03	0.01	0.19	0.26
Heavy housing burden	0.37	-0.26	0.17	-0.01	0.25	0.19	0.09	0.1	1	-0.71	0.05	0.13	0.03	0.08	0.06
Light housing burden	-0.21	0.07	0.12	0.01	-0.16	-0.11	-0.02	-0.08	-0.71	1	-0.03	-0.07	-0.03	-0.05	-0.04
Age of the oldest child	0.09	0.04	0.02	0.04	0.02	0.08	0.07	-0.08	0.05	-0.03	1	0.07	0.34	0.01	0.06
Bad health	0.18	-0.09	0.16	0	0.07	0.13	0.05	0.03	0.13	-0.07	0.07	1	0.01	0.01	-0.01
Number of dependent children	0.11	0.08	0.08	0.02	0.02	0.1	-0.03	0.01	0.03	-0.03	0.34	0.01	1	0.1	-0.07
Migrant	0.1	-0.03	0.07	-0.04	0.03	0.09	-0.03	0.19	0.08	-0.05	0.01	0.01	0.1	1	0.01
Single parent	0.11	-0.18	0.27	-0.11	0.04	0.1	0.03	0.26	0.06	-0.04	0.06	-0.01	-0.07	-0.01	1

Source: EU-SILC 2014 cross-sectional data, authors' computation.

Table 2.A3 – Literature review of multilevel (material) deprivation studies

Micro-/Macro-Determinants	Sample and Econometrics	Deprivation definition and determinants	Main Findings
Nelson (2012)	Data: EU-SILC (2008), 26 European countries, cross-sectional Unit of analysis: Individual (below 65 years of age) Model: Multilevel logistic model Dependent variable: Material deprivation	Deprivation Index: Standard EU definition Determinants: Micro (female, age dummies, single person, lone parent, two-parent family, primary education, unemployed, non-EU migrant) and macro (type-case social assistance benefits, GDP per capita, activity rate, unemployment rate, long-term unemployment rate, educational expenditure, active labour market policy (ALMP) expenditure, public service expenditure, non-means-tested benefit expenditure)	All individual determinants which are normally related to material deprivation have a substantial and significant effect. Household income is not taken up as a variable in the model. Social assistance benefits are negatively associated with material deprivation. After controlling for social benefits, GDP per capita, the activity rate, the unemployment rate and the long-term unemployment rate are significant, while non-means-tested benefit expenditure, ALMP, education expenditure and public services expenditures are not significant. Looking at effects of cross-level interactions, the author finds that social assistance benefits reduce the influence of four individual-level variables on material deprivation (i.e. single person, lone parent, unemployed, primary education).
Whelan and Maître (2012)	Data: EU-SILC (2009), 28 European countries, cross-sectional Unit of analysis: Individual (household reference person) Model: Multilevel linear model Dependent variable: Basic deprivation	Deprivation Index: Basic Deprivation which comprises items relating to enforced absence of a meal, clothes, a leisure activity, a holiday, a meal with meat or a vegetarian alternative, adequate home heating, shoes. Determinants: Micro (logarithm of **household income**, professional occupation, education (pre-primary, primary, lower secondary, higher education), age, gender, marital status, immigrant, number of children, lone parent, employment status, tenure) and macro (logarithm of Gross National Disposable Income per head (GNDH), welfare regime dummies and Gini)	All individual determinants which are normally related to material deprivation have a substantial and significant effect. Household reference person's socioeconomic variables were related to basic deprivation and account for substantial proportions of both within-country and between-country variance. The addition of macro-economic factors to the model contributed relatively little to the explanatory power and only GNDH was significant. The welfare regime dummies add little in terms of variance explanation. Further, there is a set of significant interactions between micro variables and GNDH: the impact of the micro variables is contingent on the level of aggregated income in society.
Chzhen and Bradshaw (2012)	Data: EU-SILC (2009), 24 European countries, cross-sectional	Deprivation Index: Standard EU definition Determinants: Micro (gender of lone parent, number of children, age of youngest child, marital status,	All individual determinants which are normally related to material deprivation have a substantial and significant effect. Household income is not taken up as a variable in the model.

	Unit of analysis: Individual, **children living in lone parent families** Model: Multilevel logistic model Dependent variable: Material Deprivation	education, economic activity) and macro (logarithm of GDP per capita, logarithm of social transfers)	The effect of transfers is negatively associated with material deprivation, but only when the differences in GDP per capita are not controlled for. Once the variation in country wealth is taken into account, the effect of social transfers disappears.
Visser et al. (2014)	Data: European Social Survey (ESS), 25 European countries, cross-sectional Unit of analysis: Individual Model: Multilevel linear model Dependent variable: Economic Deprivation	Deprivation Index: Confirmatory factor analysis on three variables measured on an ordinal scale (0-6): 'I have had to manage on a lower household income', 'I have had to draw on my savings or get into debt to cover ordinary living expenses' and 'I have had to cut back on holidays or new household equipment'. Determinants: Micro (national income position (quartiles), job status, employment status, marital status, number of children, urbanization, parental education, age, ethnicity) and macro (unemployment rate, GDP per capita, relative changes in the percentage of unemployment people and GDP, total social spending expenditure)	All individual determinants which are normally related to material deprivation have a substantial and significant effect. Household income is not taken up as a variable in the model. Macroeconomic circumstances and social protection expenditures show a significant impact on deprivation, after controlling for the individual level variables. Various crossed effects between micro- and macro-variables are found: the impact of the relative national income position on material deprivation varies according to the economic circumstances and the generosity of the welfare state. The paper also shows that adverse economic circumstances affect the deprivation-reducing impact of social transfers (country-level interaction).
Bárcena-Martin et al. (2014)	Data: EU-SILC (2007), 28 countries, cross-sectional Unit of analysis: Individual (household reference person) Model: Multilevel linear model Dependent variable: Material Deprivation	Deprivation Index: Linear index , weighted by frequency weights Determinants: Micro (female, young, old, tertiary education, working, tenure status, **household** income, household structure variables) and macro (long-term unemployment rate, S80/S20, GDP per capita, total social spending expenditure)	All individual determinants which are normally related to material deprivation have a substantial and significant effect. A (jointly) significant impact of social policy generosity, inequality and GDP is found. The introduction of country-specific factors reduces the proportion of total variance due to between-country differences in deprivation by 72.7 per cent, while individual-level variables reduce this proportion by only 9.4 per cent. Cross-level interactions show that social policy generosity, higher GDP and lower inequalities decrease the effect of the individual-level variables on material deprivation.
Chzhen (2014)	Data: EU-SILC (2008-2012), 31 European countries, cross-sectional Unit of analysis: Individual level, **child population** Model: Multilevel logistic model	Deprivation Index: Standard EU definition Determinants: Micro (low work intensity, lone parent, large family, migrant, owner-occupier, one adult works in public sector, age of youngest child, highest level of education) and macro (Minimum income	All individual determinants which are normally related to material deprivation have a substantial and significant effect. Total social spending and the unemployment rate reduces material deprivation for children. The negative effect of the minimum income protection scheme indicator was

	Dependent variable: Severe child deprivation	protection scheme, total social spending, unemployment rate)	statistically significant only when other country-level characteristics were not accounted for. Income, measured both at the individual and country-level, is not included in the model.
Bárcena-Martın et al. (2017)	Data: EU-SILC (2009), 27 European countries, cross-sectional Unit of analysis: Individual, **child population** Model: Multilevel linear model Dependent variable: Child deprivation	Deprivation Index: Linear index with frequency weights based on 14 specific items included in the child-specific module of the EU-SILC 2009. Determinants: Micro (age of the child, work intensity, lone parent, urban area, owner, chronic illness or condition, female household reference person (HRP), tertiary education HRP, young HRP, immigrant HRP) and macro (GDP per capita, long unemployment rate, s80s20, social spending expenditure functions)	Child deprivation is significantly related to household characteristics and to country-level determinants. The latter explain more than half of the cross-national variation in child deprivation levels, once the micro-level determinants have been controlled for. GDP per capita and inequality has a statistically significant association with child material deprivation in all model specifications. A strong and negative relationship between social protection as a share of the GDP and child deprivation is found. Some benefit functions targeted at children do not have the intended negative impact on child deprivation, while other functions not explicitly targeted at children appear to be effective in reducing child deprivation. Household income and cross-level interactions are not regressed.
Saltkjel and Malmberg-Heimonen (2017)	Data: EU-SILC (2009), 27 European countries, cross-sectional Unit of analysis: Individual, **child population** Model: Multilevel linear model Dependent variable: Material Deprivation	Deprivation Index: Standard EU definition Determinants: Micro (gender, age, country of birth, marital status, limiting longstanding illness, self-defined economic status, education level) and macro (Social protection expenditure in PPS per head, divided by the inverse of the employment rate)	All individual determinants which are normally related to material deprivation have a substantial and significant effect. Welfare generosity is related to a lower risk of material deprivation among disadvantaged groups, when assessing a combination of the main effects of welfare generosity and the group-specific effects. Income, measured both at the individual and country-level, is not included in the model.

Note: Extension of the literature review of Bárcena-Martin et al. (2014) (online appendix).

Table 2.A4 – Correlation coefficients between country-level variables

	GDP per capita	Median income	Total social benefits, % of GDP	In-kind social benefits, % of GDP	Cash social benefits, % of GDP	Total social spending, in PPS per head	In-kind social benefits, in PPS per head	Cash social benefits, in PPS per child	Family cash social benefits, % of GDP	Family social benefits, PPS per head	Pro-poorness of cash social benefits (bottom 50)	Adequacy of minimum-income benefit	Unemployment rate
GDP per capita	1	0.85	0.39	0.45	0.26	0.88	0.84	0.84	0.61	0.81	0.16	0.49	-0.43
Median income	0.85	1	0.67	0.7	0.49	0.92	0.91	0.85	0.65	0.76	0.31	0.56	-0.46
Total social benefits, % of GDP	0.39	0.67	1	0.81	0.9	0.62	0.7	0.5	0.51	0.38	0.44	0.19	-0.03
In-kind social benefits, % of GDP	0.45	0.7	0.81	1	0.47	0.75	0.84	0.6	0.71	0.5	0.47	0.5	-0.36
Cash social benefits, % of GDP	0.26	0.49	0.9	0.47	1	0.38	0.43	0.3	0.24	0.19	0.3	-0.07	0.23
Total social spending, in PPS per head	0.88	0.92	0.62	0.75	0.38	1	0.96	0.96	0.81	0.9	0.27	0.63	-0.49
In-kind social benefits, in PPS per head	0.84	0.91	0.7	0.84	0.43	0.96	1	0.84	0.77	0.78	0.33	0.59	-0.5
Cash social benefits, in PPS per child	0.84	0.85	0.5	0.6	0.3	0.96	0.84	1	0.78	0.94	0.18	0.63	-0.45
Family social spending benefits, % of GDP	0.61	0.65	0.51	0.71	0.24	0.81	0.77	0.78	1	0.79	0.18	0.57	-0.48
Family social benefits, PPS per head	0.81	0.76	0.38	0.5	0.19	0.9	0.78	0.94	0.79	1	0.03	0.59	-0.54
Pro-poorness bottom 50	0.16	0.31	0.44	0.47	0.3	0.27	0.33	0.18	0.18	0.03	1	0.38	-0.06
Adequacy of minimum income benefit	0.49	0.56	0.19	0.5	-0.07	0.63	0.59	0.63	0.57	0.59	0.38	1	-0.59
Unemployment rate	-0.43	-0.46	-0.03	-0.36	0.23	-0.49	-0.5	-0.45	-0.48	-0.54	-0.06	-0.59	1

Source: EU-SILC 2014 cross-sectional data, authors' computation.

Table 2.A5 – Country-level residual estimates in the multilevel negative binomial model

Country	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
Belgium	-0.12	-0.23	0.01	-0.04	-0.08	-0.01	-0.04	-0.02	-0.18	-0.06	-0.18	-0.22	0.03
Bulgaria	1.68	1.28	1.05	1.13	1.10	1.00	1.06	0.98	1.26	1.09	0.95	1.08	0.91
Czech Republic	-0.12	-0.21	-0.39	-0.34	-0.33	-0.37	-0.38	-0.33	-0.27	-0.34	-0.03	-0.16	-0.36
Denmark	-1.31	-0.19	0.18	-0.02	0.27	0.01	-0.14	0.15	0.14	-0.18	0.05	0.13	0.20
Germany	-0.61	-0.18	0.01	-0.11	0.09	0.16	0.06	0.22	0.04	0.13	-0.26	0.02	0.08
Estonia	-0.17	-0.23	-0.57	-0.46	-0.50	-0.41	-0.33	-0.48	-0.30	-0.21	-0.55	-0.25	-0.55
Ireland	0.24	0.28	0.12	0.13	0.26	0.57	0.70	0.35	0.39	0.57	0.46	0.61	0.32
Greece	0.78	0.40	0.50	0.62	0.17	0.12	0.15	0.15	0.23	0.19	0.56	0.05	0.28
Spain	0.38	-0.09	-0.02	-0.02	-0.09	-0.19	-0.19	-0.17	-0.22	-0.36	-0.07	-0.26	-0.14
France	-0.20	0.11	0.48	0.32	0.48	0.37	0.28	0.44	0.21	0.22	0.17	0.13	0.40
Croatia	0.26	-0.56	-0.64	-0.62	-0.61	-0.78	-0.76	-0.77	-0.65	-0.71	-0.53	-0.62	-0.81
Italy	0.13	-0.12	0.11	0.15	-0.17	-0.24	-0.27	-0.17	-0.20	-0.29	-0.24	-0.57	0.02
Cyprus	0.57	0.48	0.43	0.61	0.10	0.34	0.49	0.17	0.37	0.41	0.36	.	0.35
Latvia	0.83	0.51	0.13	0.27	0.18	0.22	0.30	0.18	0.37	0.38	0.29	0.56	0.10
Lithuania	0.46	0.32	-0.06	0.05	0.05	0.09	0.15	0.07	0.15	0.17	0.08	0.37	-0.04
Luxembourg	-0.73	-0.54	-0.58	-0.55	-0.60	0.06	-0.02	0.05	-0.24	0.04	-0.70	-0.40	0.14
Hungary	1.13	0.54	0.39	0.44	0.43	0.40	0.48	0.33	0.61	0.60	0.41	0.38	0.33
Malta	0.23	-0.02	-0.21	-0.17	-0.12	-0.16	-0.11	-0.18	-0.18	-0.10	0.18	-0.03	-0.17
The Netherlands	-1.17	-0.27	-0.04	-0.14	-0.04	-0.17	-0.33	0.01	-0.48	-0.32	-0.16	-0.11	0.03
Austria	-0.68	-0.30	-0.05	-0.10	-0.18	0.03	0.07	-0.08	-0.14	0.02	-0.37	-0.15	0.01
Poland	0.27	-0.14	-0.33	-0.20	-0.43	-0.40	-0.35	-0.41	-0.25	-0.35	0.00	-0.06	-0.39
Portugal	0.67	0.15	0.24	0.28	0.08	-0.04	-0.03	-0.02	0.00	-0.08	0.32	0.06	0.11
Romania	1.50	1.08	0.71	0.84	0.79	0.72	0.74	0.76	0.93	0.85	0.84	0.92	0.67
Slovenia	-0.39	-0.26	-0.24	-0.24	-0.26	-0.34	-0.33	-0.35	-0.28	-0.21	-0.23	-0.21	-0.31
Slovakia	0.51	0.19	-0.03	0.02	0.06	0.00	0.02	0.01	0.14	0.08	0.10	0.08	-0.06
Finland	-1.21	-0.42	-0.09	-0.25	-0.04	-0.14	-0.19	-0.13	-0.17	-0.18	-0.37	-0.28	-0.13
Sweden	-1.60	-0.73	-0.49	-0.74	-0.21	-0.36	-0.44	-0.34	-0.51	-0.51	-0.70	-0.69	-0.46
United Kingdom	0.07	-0.01	0.16	0.04	0.24	0.21	0.19	0.19	0.16	0.15	0.36	0.23	0.16
Iceland	-0.82	-0.16	-0.14	-0.29	0.15	0.06	-0.06	0.16	-0.02	-0.20	0.05	0.03	-0.05
Serbia	0.93	-0.02	-0.04	0.03	-0.17	-0.35	-0.29	-0.36	-0.18	-0.26	-0.20	.	-0.33
Switzerland	-1.48	-0.67	-0.62	-0.63	-0.63	-0.39	-0.44	-0.39	-0.76	-0.54	-0.62	-0.63	-0.35

Source: EU-SILC 2014 cross-sectional data, authors' computation.

Table 2.A5 – Country-level residual estimates in the multilevel negative binomial model (continued)

Country	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26
Belgium	0.02	-0.05	-0.02	-0.08	-0.02	-0.11	-0.10	-0.09	-0.14	0.00	0.73	0.83	1.05
Bulgaria	0.92	0.93	0.97	0.98	0.95	0.99	0.98	0.76	0.90	0.78	-0.53	-0.65	-0.62
Czech Republic	-0.33	-0.38	-0.39	-0.33	-0.37	-0.29	-0.30	-0.14	-0.29	-0.19	-0.05	0.03	0.01
Denmark	0.10	0.23	0.00	-0.05	0.13	0.00	-0.04	0.13	0.09	0.27	-0.69	-0.67	-0.57
Germany	0.04	0.09	0.11	0.00	0.19	-0.02	-0.01	-0.10	0.02	0.04	0.18	0.27	0.18
Estonia	-0.48	-0.54	-0.44	-0.37	-0.52	-0.37	-0.35	-0.58	-0.39	-0.71	0.57	0.43	0.50
Ireland	0.33	0.43	0.59	0.58	0.39	0.51	0.52	0.61	0.67	0.36	-0.16	0.12	-0.14
Greece	0.30	0.23	0.23	0.23	0.27	0.25	0.25	0.38	0.20	0.50	-0.28	-0.04	-0.02
Spain	-0.19	-0.02	-0.10	-0.13	-0.05	-0.13	-0.15	-0.12	-0.11	-0.19	-0.34	-0.22	-0.09
France	0.31	0.41	0.34	0.20	0.43	0.17	0.17	0.19	0.16	0.39	-0.82	-0.65	-0.64
Croatia	-0.84	-0.72	-0.77	-0.78	-0.74	-0.78	-0.78	-0.74	-0.75	-0.77	0.46	0.43	0.39
Italy	0.05	-0.15	-0.21	-0.17	-0.15	-0.14	-0.15	-0.22	-0.35	-0.07	-0.05	0.09	-0.06
Cyprus	0.45	0.17	0.37	0.42	0.22	0.38	0.39	0.30	.	-0.09	-0.09	-0.20	-0.39
Latvia	0.17	0.12	0.22	0.29	0.17	0.30	0.30	0.17	0.34	0.03	-0.10	-0.18	-0.21
Lithuania	0.01	0.03	0.09	0.16	0.06	0.17	0.18	0.02	0.22	-0.05	0.35	0.28	-0.58
Luxembourg	0.26	0.11	0.21	0.37	0.15	0.43	0.43	0.20	0.40	0.25	0.07	0.10	0.13
Hungary	0.35	0.31	0.36	0.38	0.28	0.37	0.38	0.29	0.26	0.45	0.40	0.55	0.12
Malta	-0.13	-0.16	-0.17	-0.10	-0.22	-0.09	-0.09	0.08	-0.11	-0.20	0.10	0.07	-0.04
The Netherlands	0.00	0.01	-0.15	-0.15	0.00	-0.12	-0.12	-0.03	-0.05	0.13	-0.09	-0.36	-0.33
Austria	0.03	-0.12	-0.01	-0.06	-0.11	-0.12	-0.11	-0.20	-0.11	-0.10	0.00	-0.13	0.01
Poland	-0.31	-0.49	-0.41	-0.34	-0.44	-0.32	-0.33	-0.19	-0.28	-0.25	0.40	0.21	0.24
Portugal	0.12	0.03	-0.02	0.00	0.00	0.02	0.01	0.17	0.01	0.33	0.89	0.87	0.71
Romania	0.73	0.66	0.69	0.78	0.70	0.82	0.82	0.68	0.73	0.85	-0.40	-0.36	-0.04
Slovenia	-0.32	-0.34	-0.36	-0.36	-0.37	-0.35	-0.34	-0.32	-0.33	-0.41	-0.08	0.08	-0.03
Slovakia	-0.06	0.02	0.01	0.05	0.02	0.07	0.07	0.02	0.04	-0.06	0.43	0.41	0.16
Finland	-0.23	-0.11	-0.18	-0.30	-0.15	-0.33	-0.33	-0.32	-0.30	-0.21	0.27	0.35	0.48
Sweden	-0.60	-0.27	-0.40	-0.54	-0.35	-0.58	-0.58	-0.58	-0.60	-0.47	-0.41	-0.45	-0.24
United Kingdom	0.11	0.18	0.16	0.09	0.15	0.06	0.06	0.31	0.13	0.41	0.43	0.41	0.16
Iceland	-0.13	0.11	0.02	-0.06	0.12	-0.05	-0.08	0.09	-0.01	-0.08	-0.04	-0.32	0.11
Serbia	-0.34	-0.33	-0.34	-0.37	-0.33	-0.37	-0.37	-0.47	.	-0.34	-0.69	-0.25	-0.49
Switzerland	-0.31	-0.41	-0.37	-0.32	-0.38	-0.33	-0.32	-0.31	-0.35	-0.59	0.05	-0.31	0.43

Source: EU-SILC 2014 cross-sectional data, authors' computation.

Chapter 3

Geometric composite indicators with compromise Benefit-of-the-Doubt weights

(This chapter is published in *European Journal of Operational Research*; co-author: Nicky Rogge)

3.1. Introduction

The flexibility and the optimistic stance in the determination of the weights is often praised as the most important advantage of the Benefit-of-the-Doubt (BoD) method in the construction of composite indicators (CIs).[49] In a setting in which objective knowledge on the true policy weights is usually lacking or incomplete, the BoD-model derives for each country the set of optimal weights from the observed sub-indicator values themselves. More in particular, the BoD-model defines importance weights for each country such that the impact of sub-indicators of relative strength is maximized and the impact of sub-indicators of relative weakness is minimized in the composite value. This quality explains much of the appeal of the BoD-model: in what is usually a sensitive evaluation environment, disappointed countries can no longer blame a low CI-score on damaging or unfair weights. Any other weighting scheme than the one specified by the BoD-model would worsen the CI-score. The increasingly popular BoD-model has by now become an established method to construct CIs in various contexts.[50]

In recent years, several interesting extensions of the BoD-model have been proposed in the literature. One such extension is the so-called "pessimistic" version of the BoD-model that relates to the minimum efficiency concept introduced by Zhu (2004) in the DEA-context. The conceptual starting point of the pessimistic counterpart is opposite to the one of the traditional, 'optimistic' BoD-model (Zhou et al., 2007; Rogge, 2012). Specifically, the pessimistic BoD-model assesses the policy performance of countries under a 'worst-case' evaluation scenario, in which weights for the sub-indicators are defined such that the CI-value of each country is

[49] The BoD-technique is inspired by (the multiplier formulation of) Data Envelopment Analysis (DEA) (Charnes et al., 1978), an efficiency measurement technique popular in the Management Science and Operations Research literature.

[50] Examples include human development (Blancard & Hoarau, 2013), environmental performance (Rogge, 2012), measuring active ageing (Amado et al., 2016), measuring local police effectiveness (Verschelde & Rogge, 2012), measuring quality in health care (Shwartz et al., 2016).

minimized relative to the other countries. In the definition of the weights this entails assigning relatively high (low) weights to sub-indicators on which the evaluated country performs relatively weakly (strongly) as compared to the other countries in the sample.

Other interesting extensions are the multiplicative versions of the BoD-model which compute CIs by a multiplicative aggregation function. Multiplicative CIs do not imply, contrary to their linear equivalents, full compensability which makes that a poor performance on one sub-indicator cannot be fully compensated by sufficiently high values on other sub-indicators (see, e.g. Ebert & Welsch, 2004). In addition, multiplicative CIs penalize inequality among sub-indicators (Nardo et al., 2008, p. 33). In the BoD-literature, there are, broadly speaking, two streams of multiplicative BoD-models.[51] A first stream of 'direct' multiplicative BoD-models combine both multiplicative aggregation and BoD-weighting in one computation step (Zhou et al., 2010; Blancas et al., 2012; Tofallis, 2014). More recently, Van Puyenbroeck and Rogge (2017) proposed an alternative, 'indirect' approach to construct multiplicative BoD-based CIs. It concerns a two-step procedure in which, in a first step, importance weights of the different sub-indicators are estimated using the original BoD-model and, in a second step, BoD-derived importance weights are used in the construction of the CIs as geometric mean quantity indices. It is argued by Van Puyenbroeck and Rogge (2017) that this 'indirect' multiplicative BoD-model avoids some of the disadvantages in the 'direct' multiplicative BoD-models.[52]

This chapter contributes to the literature in three ways. Firstly, the chapter extends the indirect multiplicative BoD-model of Van Puyenbroeck and Rogge (2017) by combining optimistic and pessimistic BoD-weighting so as to obtain a comprehensive view on countries' policy performances.[53] A second contribution of the chapter is more innovative and relates to the 'optimistic' and 'pessimistic' BoD-models that are used in the extension of the framework of Van Puyenbroeck and Rogge (2017). In particular, as existing versions of the optimistic and pessimistic BoD-models may cause implausible results in the identification of a country's sub-indicators of comparative strength and weakness, we advocate new versions of optimistic and pessimistic BoD-weighting. These alternative BoD-weighting models are different in the sense that the benchmark is a fixed, hypothetical country. It is shown that by evaluating each country

---

[51] Both 'direct' (Giambona & Vassallo, 2014) and 'indirect' (Van Puyenbroeck & Rogge, 2017) versions of the multiplicative BoD-model have been applied on European social inclusion data in recent studies.

[52] The 'disadvantages' of the direct multiplicative BoD-models include the lack of commensurability (as in the model of Zhou et al., 2010) and the presence of a scaling factor as in the model of Tofallis (2014) (for a more elaborate discussion of the disadvantages of the direct multiplicative BoD-model, see Van Puyenbroeck & Rogge, 2017).

[53] Note that the idea of combining 'optimistic' and 'pessimistic' weighting scenarios in the indirect BoD-CI framework so as to obtain a comprehensive view on countries' policy performances was suggested by Van Puyenbroeck and Rogge (2017) in the concluding section of their paper.

in the sample set relative to a fixed benchmark, sub-indicators can no longer be simultaneously identified as a comparative strength and weakness. A third original contribution is the development of a measure for the degree of unbalance in a country's policy portfolio mix. It concerns a ratio of geometric CI under optimistic weighting and the geometric CI under pessimistic weighting that can be further decomposed using a multiplicative Bortkiewicz (Bortkiewicz, 1923) decomposition to explain for (changes in) the degree of unbalance in a country's policy portfolio mix.

This chapter is structured as follows. Section 3.2 briefly describes Van Puyenbroeck and Rogge's (2017) 'indirect' framework for multiplicative BoD-based CI-construction. This section also illustrates some important issues with the existing optimistic and pessimistic BoD-models in the identification of countries' comparative strengths and weaknesses. To resolve for these issues, we present alternative versions of the optimistic and pessimistic BoD-weighting models and implement them into the 'indirect' framework to derive a compromise geometric CI. In Section 3.3, we adjust Van Puyenbroeck and Rogge's (2017) inter-temporal geometric CI-framework, so as to incorporate both optimistic and pessimistic BoD-based weights. Section 3.4 develops a measure for the degree of unbalance in a country's policy portfolio mix. This section also shows how this measure can be decomposed using a multiplicative Bortkiewicz decomposition to explain for (changes in) the degree of unbalance in a country's policy portfolio mix. Section 3.5 concludes.

Throughout, we illustrate our findings with the commonly agreed EU indicators (period 2008-2013) from the overarching portfolio of social protection and social inclusion objectives as endorsed by the Heads of State and Government in the Europe 2020-strategy and employed by Social OMC (Social Protection Committee, 2015). Specifically, the nine overarching commonly agreed EU social inclusion indicators are: *(i) at risk of poverty or social exclusion rate, (ii) relative median poverty risk gap, (iii) income quintile ratio (S80/S20), (iv) early school leavers, (v) aggregate replacement ratio, (vi) at-risk-of-poverty rate anchored at a fixed moment in time (2008), (vii) employment rate of older workers, (viii) in work at-risk-of poverty rate,* and *(ix) activity rate.* For all except three indicators (i.e. the aggregate replacement ratio, the employment rate of older workers and the activity rate), higher values represent worse social inclusion performances. To put all indicators on a common basis so that all measure social inclusion, the other six indicators are transformed by taking the inverse of the regular indicator.

3.2. Compromise geometric mean composite indicators

The indirect CI-framework of Van Puyenbroeck and Rogge (2017) computes multiplicative CIs as geometric mean quantity index numbers using BoD-derived sub-indicator importance weights by a two-step procedure. In a first step, importance weights for the different sub-indicators are estimated using the original (linear) BoD-model. In a second step, (normalized) country sub-indicator values are weighted and geometrically aggregated using the BoD-based importance weights as obtained in the first step. Formally,

$$CI_i = \prod_{r=1}^{s} \left(\frac{y_{r,i}}{y_{r,B}}\right)^{\omega_{r,i}} \qquad (1)$$

where $y_{r,i}$ in the numerator is the performance of the i^{th} country ($i = 1,\ldots,n$) on the r^{th} social inclusion sub-indicator ($r = 1,\ldots,s$). In the denominator, there are the baseline sub-indicator values, $y_{r,B}$, relative to which the performances of country i are compared (in our application below, the baseline performance values are equal to the (population-weighted aggregate) sub-indicator values for the EU-27 countries, i.e. $y_{r,B} = y_{r,EU27}$).[54] The sub-indicator exponents $\omega_{r,i}$ defines how much the r^{th} sub-indicator contributes to the aggregate CI, with $\sum_{r=1}^{s} \omega_{r,i} = 1$. The sub-indicator exponent values $\omega_{r,i}$ indicate the percentage change in the CI_i-value as result of a 1% increase in $\frac{y_{r,i}}{y_{r,B}}$. Note that the multiplicative CIs as in (1) are tailor-made per country to compare the evaluated country itself with some base performance observation. As such, they are bilateral in nature and, hence, should be interpreted in relative terms. For example, a CI_i -value of 1.1 indicates that the policy performance of the country i on the whole outperforms the baseline policy performance by 10%. A CI_i -value of 0.8, on the other hand, indicates that the policy performance of the country i on the whole underperforms the baseline policy performance by 20%.

In their original work, Van Puyenbroeck and Rogge (2017) use the traditional optimistic (linear) BoD-model to estimate the exponents $\omega_{r,i}$ of the geometric CI (see Appendix 1 for a detailed description of the model). Specifically, Van Puyenbroeck and Rogge (2017) use the “pie-shares” of the traditional optimistic BoD-model which designate the relative importance of each

---

54 As noted by Van Puyenbroeck and Rogge (2017), the choice of a specific set of base performance indicators $y_{r,B}$ is largely arbitrary. Depending on the evaluation context, base performance values other than the sample average of each sub-indicator can be specified (e.g., median, maximum, etc.). Within the EU social policy setting, benchmarking performances to the EU27-average was endorsed by the European Commission in its yearly Joint Employment Report.

sub-indicator within the linear CI. However, as noted above, Van Puyenbroeck and Rogge (2017) suggested to combine 'optimistic' and 'pessimistic' weighting scenarios in their indirect BoD-CI framework so as to obtain a comprehensive view on countries' policy performances. To derive 'pessimistic' sub-indicator importance weights, one could resort to the pessimistic version of the BoD-model as proposed by Zhou, Ang and Poh (2007) (ZAP, henceforth) or Athanassoglou's (2016) worst-case equivalent of the optimistic BoD-model (for a description of these models, see Appendix 1). However, as we will illustrate below, whereas the traditional optimistic and the two pessimistic versions of the BoD-models can be effective tools for benchmarking countries' performances under respectively optimistic or pessimistic evaluation scenarios in separate performance evaluations, they seem less suitable to be combined into one overall CI-framework that aims at simultaneously identifying countries' sub-indicators of comparative strength and sub-indicators of comparative weakness. The reason is the potential implausible result that sub-indicators are simultaneously identified as a country's comparative strengths by the optimistic BoD-model and comparative weaknesses by the pessimistic BoD-model, or vice versa.

To illustrate this counterintuitive result, consider the simple setting with three countries (i = 1,2,3) and two sub-indicators (r = 1,2) described in Table 3.1. The optimistic BoD-model and Athanassoglou's pessimistic BoD-model are able to correctly identify the sub-indicators of comparative strength ($\omega_{1,1}^{trad}$ = 1.00 and $\omega_{2,2}^{trad}$ = 1.00) and comparative weakness ($\omega_{2,1}^{Atha}$ = 1.00 and $\omega_{1,2}^{Atha}$ = 1.00) for countries 1 and 2. The ZAP-model, on the other hand, gives a high weight to the sub-indicators on which the respective countries perform strongest (for country 1 the sub-indicator 1 with $\omega_{1,1}^{ZAP}$ = 0.53 and for country 2 the sub-indicator 2 with $\omega_{2,2}^{ZAP}$ = 0.94), hence falsely indicating this sub-indicator as relative weakness. Country 3, with strong performances on both sub-indicators, is, - again - illogically, awarded a high weight to sub-indicator 1 in both the optimistic BoD-model ($\omega_{1,3}^{trad}$ = 0.64) and Athanassoglou's pessimistic BoD-model ($\omega_{1,3}^{Atha}$ = 1). This example clearly shows that an integrated CI-framework using combinations of existing versions of the optimistic and pessimistic BoD-weighting models may cause implausible results.

Table 3.1 – Example of counterintuitive weight setting

		Performance $y_{r,i}$		$\omega_{r,i}^{trad}$		$\omega_{r,i}^{ZAP}$		$\omega_{r,i}^{Atha}$	
		r = 1	*r = 2*	*r = 1*	*r = 2*	*r = 1*	*r = 2*	*r = 1*	*r = 2*
country	*i = 1*	100	1	1.00	0.00	0.53	0.47	0.00	1.00
	i = 2	10	2	0.00	1.00	0.05	0.95	1.00	0.00
	i = 3	80	1.8	0.64	0.36	0.33	0.67	1.00	0.00

The implausible results in the identification of countries comparative strengths and weaknesses are due to the fact that the traditional BoD-models identify different benchmarks to which performances are evaluated. To solve for this inconvenience, we propose an alternative version of the optimistic and pessimistic BoD-weighting model in which a fixed country is used as the benchmark. However, in a fixed benchmark setting, the weights are sensitive to the variability of the sub-indicators. We therefore apply a max-min normalisation such that the normalized sub-indicator values $y_{\mathrm{r,i}}^{n}$ fully reflect countries' comparative performances. Specifically, the normalization of the observed sub-indicator values is as following:

$$y_{\mathrm{r,i}}^{n} = \frac{y_{r,i} - y_r^{min}}{y_r^{max} - y_r^{min}} + 1 \qquad (2)$$

with y_r^{min} and y_r^{max} respectively the lowest and highest value on the sub-indicator *r* observed among the countries in the sample set. In the interpretation of the normalized sub-indicator values, a $y_{\mathrm{r,i}}^{n}$-value of one (i.e., the lowest possible value) indicates that country *i* is the worst performing country on the *r^th^* sub-indicator, whereas the maximum possible score of two indicates that country *i* is the best performing country on the *r^th^* sub-indicator.[55]

The benchmark in our alternative weighting model is a fixed, hypothetical country that constitutes the maximum values of the normalized sub-indicators (i.e., $\max_i(y_{r,i}^n) = 2, \forall\, r$). We consider such an ideal country as a natural reference point to evaluate countries' comparative strengths and weaknesses and, consequently, to determine the optimistic and pessimistic weights of sub-indicators within the geometric CI. Specifically, our version of the optimistic BoD-weighting model as in (3) awards relatively high (low) importance weights to sub-

[55] In order to avoid infeasibilities in our model, it was necessary to add a non-negative factor to $\frac{y_{r,i}-y_r^{min}}{y_r^{max}-y_r^{min}}$ such that $\min(y_{r,i}^n) > 0$. The size of the factor, 1 in our case, does not influence the sub-indicator exponent shares $\omega_{r,i}$ derived from the model.

indicators on which the evaluated country is closely situated to the hypothetical ideal country, whereas our version of the pessimistic BoD-weighting model as in (4) gives relatively high (low) importance weights to sub-indicators on which the evaluated country is more distant from the hypothetical best practice country. Formally,

$$CI_i^{opt\,weight} = \max_{w_{r,i}^{+}} \sum_{r=1}^{s} w_{r,i}^{+}\, y_{r,i}^{n} \qquad (3)$$

$$s.t.: \sum_{r=1}^{s} w_{r,i}^{+}\ \max_i(y_{r,i}^{n}) = 1$$

$$L_r < \frac{w_{r,i}^{+} y_{r,i}^{n}}{\sum_{r=1}^{s} w_{r,i}^{+}\, y_{r,i}^{n}} < U_r$$

$$w_{r,i}^{+} \geq 0$$

$$CI_i^{pes\,weight} = \min_{w_{r,i}^{-}} \sum_{r=1}^{s} w_{r,i}^{-}\, y_{r,i}^{n} \qquad (4)$$

$$s.t.: \sum_{r=1}^{s} w_{r,i}^{-}\ \max_i(y_{r,i}^{n}) = 1$$

$$L_r < \frac{w_{r,i}^{-} y_{r,i}^{n}}{\sum_{r=1}^{s} w_{r,i}^{-}\, y_{r,i}^{n}} < U_r$$

$$w_{r,i}^{-} \geq 0$$

with $CI_i^{opt\,weight}$ and $CI_i^{pes\,weight}$ the (linear) BoD-based CIs for the evaluated country *i* and $w_{r,i}^{+}$ and $w_{r,i}^{-}$ the optimistic and pessimistic weights for the evaluated country *i* on the r^{th} sub-indicator as computed by our alternative versions of the optimistic and pessimistic (linear) BoD-model.

In the computations below, in order to exclude (or at least minimize the prevalence of) unappealing weighting scenarios in which CIs only comprise a minority (or even just one) of the sub-indicators, weight restrictions are added in both models. In particular, we imposed that (BoD-estimated) optimistic and pessimistic sub-indicator shares should be at least 2% (Lower bound, L_r= 0.02) and at maximum 35% (Upper bound, U_r= 0.35) (see, e.g. Cherchye et al. (2007) and Wong & Beasley (1990) for more on weight restrictions).[56] In addition, to mitigate

[56] Ideally, upper and lower bound values on weight restrictions should be specified by a panel of experts and/or stakeholders. In the current illustrative application we lack such expert/stakeholder information. Nevertheless, we still defined a lower and upper weight bound value so as to avoid unappealing low and high BoD-weights.

the sensitivity of the optimistic and pessimistic evaluation outcomes (i.e., CI-scores and importance weights) to the influences of outliers and measurement errors, the optimistic and pessimistic models as in (3) and (4) were robustified, using insights after Cazals, Florens and Simar (2002) (i.e., robust order-m model specification with 5,000 computations rounds in each of which a subsample of 20 randomly selected peer countries is considered).

Following the reasoning of the 'indirect' approach of Van Puyenbroeck and Rogge (2017), in the first step, we use the country-specific sub-indicator importance weights as obtained endogenously by the models (3) and (4) to estimate the (robust) optimistic and pessimistic sub-indicator exponents $\omega_{r,i}^{+}$ and $\omega_{r,i}^{-}$. The sub-indicator importance shares ("pie-shares") of our alternative optimistic and pessimistic models as in (3) and (4) are defined as the products of the normalized sub-indicator values $y_{r,i}^{n}$ and the corresponding BoD-weights (respectively, $w_{r,i}^{+}$ and $w_{r,i}^{-}$) divided by the value of the objective function (respectively, $\sum_{r=1}^{s} w_{r,i}^{+} y_{r,i}^{n}$ and $\sum_{r=1}^{s} w_{r,i}^{-} y_{r,i}^{n}$). In the second step, the computed sub-indicator exponents are employed in the construction of an optimistic and pessimistic geometric CI_i^{+} and CI_i^{-}. It is important to stress that the (max-min) normalized sub-indicators $y_{r,i}^{n}$ are solely used in the determination of the optimistic $\omega_{r,i}^{+}$ and pessimistic $\omega_{r,i}^{-}$ sub-indicator exponents in models (3) and (4)). The optimistic and pessimistic geometric CI_i^{+} and CI_i^{-} in (5a) and (5b) use the original sub-indicator values $y_{r,i}$ and $y_{r,B}$:

$$CI_i^{+} = \prod_{r=1}^{s} \left(\frac{y_{r,i}}{y_{r,B}}\right)^{\omega_{r,i}^{+}} \quad (5a) \quad \text{and} \quad CI_i^{-} = \prod_{r=1}^{s} \left(\frac{y_{r,i}}{y_{r,B}}\right)^{\omega_{r,i}^{-}} \quad (5b)$$

The corresponding optimistic $\omega_{r,i}^{+}$ and pessimistic $\omega_{r,i}^{-}$ sub-indicator exponents can be found, respectively, in Tables 3.2 and 3.3. While not delving too deeply into the exact optimistic and pessimistic sub-indicator exponents, it is important to briefly note that valuable insights can be retrieved from both tables. For instance, one readily notices that the sub-indicator exponents as determined by the optimistic and pessimistic versions of the BoD-model can be quite diverse without violating the additionally imposed weight restrictions. This illustrates that the Benefit-of-the-Doubt in weighting still plays in the evaluations of the countries' social inclusion policy performances, however, only within the a priori confines. Another interesting observation is that binding weight restrictions reveal the sub-indicators of comparative strength and weakness. More in particular, in the interpretation of the optimistic sub-indicator exponents $\omega_{r,i}^{+}$, upwardly binding weight restrictions identify social inclusion sub-indicators of comparative strength. The opposite holds for the interpretation of the pessimistic sub-indicator exponents $\omega_{r,i}^{-}$, i.e.

upwardly binding weight restrictions revealing social inclusion sub-indicators of comparative weakness.

Table 3.2 – Optimistic weights

	Ind1		*Ind2*		*Ind3*		*Ind4*		*Ind5*		*Ind6*		*Ind7*		*Ind8*		*Ind9*	
Country	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*
EU27	0.10	0.10	0.03	0.03	0.03	0.02	0.02	0.02	0.35	0.31	0.03	0.03	0.12	0.17	0.02	0.02	0.30	0.29
Belgium	0.02	0.04	0.34	0.13	0.22	0.34	0.02	0.02	0.02	0.02	0.02	0.07	0.02	0.02	0.32	0.34	0.02	0.02
Bulgaria	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.04	0.05	0.02	0.02	0.14	0.33	0.31	0.26	0.31	0.26	0.11
Czech Republic	0.08	0.29	0.02	0.09	0.31	0.28	0.07	0.10	0.02	0.02	0.26	0.10	0.02	0.02	0.21	0.09	0.02	0.02
Denmark	0.10	0.17	0.02	0.02	0.21	0.11	0.02	0.02	0.02	0.02	0.05	0.09	0.11	0.21	0.13	0.03	0.34	0.34
Germany	0.09	0.13	0.02	0.06	0.02	0.02	0.02	0.02	0.28	0.03	0.02	0.02	0.18	0.35	0.02	0.02	0.34	0.35
Estonia	0.11	0.08	0.04	0.11	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.35	0.35	0.06	0.02	0.35	0.35
Ireland	0.02	0.02	0.32	0.35	0.06	0.03	0.02	0.04	0.02	0.12	0.02	0.02	0.25	0.05	0.10	0.35	0.18	0.02
Greece	0.08	0.03	0.03	0.02	0.02	0.02	0.04	0.32	0.35	0.35	0.02	0.02	0.27	0.02	0.02	0.02	0.18	0.19
Spain	0.10	0.13	0.02	0.02	0.02	0.02	0.02	0.02	0.34	0.35	0.02	0.03	0.12	0.06	0.02	0.02	0.35	0.35
France	0.11	0.24	0.35	0.34	0.02	0.02	0.02	0.02	0.34	0.29	0.10	0.03	0.02	0.02	0.02	0.02	0.02	0.02
Italy	0.28	0.28	0.05	0.02	0.15	0.03	0.02	0.02	0.35	0.35	0.02	0.04	0.02	0.21	0.09	0.03	0.02	0.02
Cyprus	0.02	0.02	0.35	0.35	0.04	0.04	0.02	0.08	0.02	0.02	0.02	0.02	0.21	0.11	0.09	0.02	0.23	0.34
Latvia	0.02	0.02	0.02	0.02	0.02	0.02	0.13	0.15	0.02	0.02	0.02	0.02	0.35	0.35	0.07	0.05	0.35	0.35
Lithuania	0.02	0.02	0.02	0.02	0.02	0.02	0.35	0.32	0.04	0.02	0.02	0.02	0.35	0.28	0.02	0.02	0.15	0.28
Luxembourg	0.35	0.14	0.18	0.26	0.02	0.02	0.02	0.15	0.35	0.35	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Hungary	0.02	0.02	0.12	0.11	0.33	0.35	0.02	0.02	0.35	0.35	0.07	0.04	0.02	0.02	0.06	0.08	0.02	0.02
Malta	0.29	0.03	0.03	0.33	0.23	0.35	0.02	0.02	0.02	0.02	0.03	0.08	0.02	0.02	0.34	0.14	0.02	0.02
Netherlands	0.31	0.19	0.20	0.17	0.02	0.25	0.02	0.02	0.02	0.02	0.13	0.02	0.02	0.04	0.12	0.04	0.16	0.24
Austria	0.04	0.28	0.03	0.02	0.19	0.18	0.03	0.06	0.34	0.12	0.02	0.04	0.02	0.02	0.02	0.02	0.31	0.26
Poland	0.02	0.02	0.17	0.02	0.02	0.02	0.35	0.32	0.35	0.32	0.02	0.23	0.02	0.02	0.02	0.02	0.02	0.02
Portugal	0.02	0.06	0.02	0.02	0.02	0.02	0.02	0.02	0.29	0.35	0.02	0.02	0.25	0.14	0.02	0.02	0.34	0.35
Romania	0.02	0.02	0.02	0.02	0.02	0.02	0.15	0.03	0.35	0.35	0.02	0.28	0.35	0.24	0.02	0.02	0.05	0.02
Slovenia	0.08	0.13	0.02	0.06	0.35	0.35	0.34	0.35	0.04	0.02	0.03	0.02	0.02	0.02	0.09	0.02	0.03	0.02
Slovakia	0.02	0.08	0.03	0.02	0.35	0.35	0.27	0.10	0.02	0.02	0.24	0.35	0.02	0.02	0.03	0.04	0.02	0.02
Finland	0.07	0.08	0.31	0.31	0.12	0.18	0.02	0.02	0.02	0.02	0.02	0.06	0.12	0.03	0.18	0.27	0.13	0.02
Sweden	0.33	0.12	0.02	0.02	0.12	0.07	0.02	0.02	0.02	0.02	0.03	0.07	0.31	0.33	0.02	0.02	0.13	0.33
United Kingdom	0.09	0.02	0.05	0.22	0.02	0.02	0.02	0.02	0.08	0.02	0.02	0.02	0.35	0.31	0.02	0.02	0.35	0.35

Source: Authors' calculations from Eurostat. *Note*: *Ind1*= at risk of poverty or social exclusion rate, *Ind2*= relative median poverty risk gap, *Ind3*= income quintile ratio (S80/S20), *Ind4*= early school leavers, *Ind5*= aggregate replacement ratio, *Ind6*= at-risk-of-poverty rate anchored at a fixed moment in time (2008), *Ind7*= employment rate of older workers, *Ind8*=in-work at risk of poverty rate, *Ind9*=activity rate.

Table 3.3 – Pessimistic weights

	Ind1		*Ind2*		*Ind3*		*Ind4*		*Ind5*		*Ind6*		*Ind7*		*Ind8*		*Ind9*	
Country	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*	*2008*	*2013*
EU27	0.04	0.03	0.08	0.06	0.04	0.03	0.32	0.30	0.02	0.02	0.21	0.21	0.05	0.02	0.22	0.31	0.02	0.02
Belgium	0.02	0.02	0.02	0.02	0.02	0.02	0.25	0.14	0.08	0.34	0.04	0.02	0.35	0.21	0.02	0.02	0.19	0.21
Bulgaria	0.35	0.35	0.18	0.16	0.23	0.31	0.03	0.02	0.04	0.08	0.11	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Czech Republic	0.02	0.02	0.08	0.02	0.02	0.02	0.02	0.03	0.23	0.35	0.02	0.02	0.28	0.30	0.02	0.02	0.31	0.21
Denmark	0.02	0.02	0.10	0.31	0.02	0.02	0.35	0.16	0.35	0.35	0.07	0.05	0.03	0.02	0.04	0.04	0.02	0.02
Germany	0.02	0.02	0.29	0.02	0.09	0.02	0.29	0.25	0.02	0.08	0.14	0.24	0.02	0.02	0.10	0.33	0.02	0.02
Estonia	0.02	0.02	0.02	0.02	0.03	0.25	0.20	0.05	0.34	0.35	0.33	0.24	0.02	0.02	0.02	0.03	0.02	0.02
Ireland	0.16	0.33	0.02	0.02	0.02	0.02	0.29	0.08	0.17	0.02	0.23	0.35	0.02	0.02	0.07	0.02	0.02	0.14
Greece	0.05	0.09	0.08	0.35	0.13	0.31	0.05	0.02	0.02	0.02	0.28	0.15	0.02	0.02	0.34	0.02	0.03	0.02
Spain	0.02	0.02	0.14	0.29	0.02	0.23	0.34	0.33	0.02	0.02	0.18	0.03	0.02	0.02	0.24	0.04	0.02	0.02
France	0.02	0.02	0.02	0.02	0.03	0.02	0.33	0.23	0.02	0.02	0.02	0.05	0.35	0.28	0.10	0.30	0.11	0.06
Italy	0.02	0.02	0.03	0.07	0.02	0.03	0.25	0.24	0.02	0.02	0.09	0.11	0.31	0.03	0.03	0.14	0.22	0.34
Cyprus	0.05	0.08	0.02	0.02	0.02	0.02	0.32	0.04	0.35	0.31	0.18	0.30	0.02	0.02	0.02	0.19	0.02	0.02
Latvia	0.10	0.13	0.18	0.05	0.31	0.31	0.02	0.02	0.23	0.17	0.10	0.26	0.02	0.02	0.02	0.02	0.02	0.02
Lithuania	0.06	0.12	0.22	0.02	0.25	0.35	0.02	0.02	0.05	0.11	0.29	0.28	0.02	0.02	0.07	0.07	0.02	0.02
Luxembourg	0.02	0.02	0.02	0.02	0.02	0.02	0.20	0.02	0.02	0.02	0.02	0.19	0.35	0.32	0.26	0.34	0.09	0.05
Hungary	0.17	0.19	0.02	0.02	0.02	0.02	0.04	0.06	0.02	0.02	0.02	0.02	0.35	0.30	0.02	0.02	0.34	0.34
Malta	0.02	0.02	0.02	0.02	0.02	0.02	0.22	0.33	0.02	0.06	0.02	0.02	0.35	0.28	0.02	0.02	0.31	0.23
Netherlands	0.02	0.02	0.02	0.02	0.05	0.02	0.35	0.34	0.22	0.31	0.02	0.17	0.21	0.06	0.09	0.03	0.02	0.02
Austria	0.02	0.02	0.03	0.05	0.02	0.02	0.10	0.09	0.02	0.02	0.11	0.10	0.35	0.35	0.33	0.33	0.02	0.02
Poland	0.09	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.34	0.29	0.29	0.32	0.18	0.27
Portugal	0.02	0.02	0.02	0.07	0.25	0.26	0.35	0.34	0.02	0.02	0.02	0.09	0.02	0.02	0.28	0.16	0.02	0.02
Romania	0.21	0.18	0.35	0.20	0.14	0.31	0.02	0.04	0.02	0.02	0.06	0.02	0.02	0.02	0.14	0.17	0.03	0.03
Slovenia	0.02	0.02	0.27	0.02	0.02	0.02	0.02	0.02	0.06	0.13	0.07	0.18	0.35	0.35	0.03	0.15	0.16	0.11
Slovakia	0.06	0.02	0.02	0.31	0.02	0.02	0.02	0.02	0.12	0.04	0.02	0.02	0.35	0.34	0.09	0.02	0.30	0.21
Finland	0.02	0.02	0.02	0.02	0.02	0.02	0.31	0.35	0.35	0.35	0.20	0.09	0.04	0.09	0.02	0.02	0.02	0.04
Sweden	0.02	0.02	0.05	0.08	0.02	0.02	0.18	0.12	0.32	0.35	0.06	0.04	0.02	0.02	0.31	0.33	0.02	0.02
United Kingdom	0.02	0.02	0.02	0.02	0.20	0.02	0.30	0.31	0.03	0.05	0.30	0.31	0.02	0.02	0.08	0.23	0.02	0.02

Source: Authors' calculations from Eurostat. *Note*: *Ind1*= at risk of poverty or social exclusion rate, *Ind2*= relative median poverty risk gap, *Ind3*= income quintile ratio (S80/S20), *Ind4*= early school leavers, *Ind5*= aggregate replacement ratio, *Ind6*= at-risk-of-poverty rate anchored at a fixed moment in time (2008), *Ind7*= employment rate of older workers, *Ind8*=in-work at risk of poverty rate, *Ind9*=activity rate.

To obtain a comprehensive view on countries' policy performances, the optimistic and pessimistic geometric CIs are combined into one overall CI as in (6). The result is a geometric CI that is calculated under compromise weighting, with the compromise weights being defined as the (arithmetic) average of optimistic and pessimistic BoD-derived sub-indicator importance weights:[57]

$$CI_i = \sqrt{CI_i^+ \times CI_i^-} \qquad (6)$$

$$CI_i = \prod_{r=1}^{s} \left(\frac{y_{r,i}}{y_{r,B}}\right)^{\frac{\omega_{r,i}{}^{+} + \omega_{r,i}{}^{-}}{2}}$$

The compromise CI-scores (and rankings) to evaluate the social inclusion policy performance of the EU Member States for the years 2008 and 2013 are listed in Table 3.4. This table reveals some interesting information. Firstly, a comparison of the different CI-scores (and CI-rankings) shows that several Member States perform more or less *consistently* strongly/poorly on the different sub-indicators of social inclusion. More in particular, the comparison indicates several sub-groups of Member States. A first group contains Member States with a social democratic welfare state (i.e. *Finland, Sweden, Netherlands*) which are evaluated as excellent performers under both optimistic and pessimistic weighting in both 2008 and 2013. This suggests that these countries have strong and highly balanced social inclusion portfolios. A second group of EU Member States consists of some Central and Eastern European countries (i.e., *Czech Republic*, *Slovakia*, *Slovenia*) that obtain rather high compromise *CI*-scores, but predominantly because of strong social inclusion performance evaluations under optimistic weighting (i.e., high CI^+-scores and ranks). This suggests that these countries have a rather strong, yet, slightly less cohesive social inclusion policy. The third group consists of some "old" EU Member States (i.e. *Germany, Austria, Denmark, France*) that have a more or less balanced (yet overall relatively strong) social inclusion portfolio. Countries in the fourth group (i.e. *Belgium, Poland*) perform reasonably well on social inclusion sub-indicators in which they hold a comparative

[57] Zhou et al. (2010) also combined optimistic and pessimistic (BoD) weights in an overall CI based on multiplicative aggregation with a control parameter. In fact, a control parameter α (with $0 \leq \alpha \leq 1$) can also be introduced in our 'indirect' multiplicative BoD framework to determine the importance of the optimistic and pessimistic weights in the composite value:

$$CI_i = \prod_{r=1}^{s} \left(\frac{y_{r,i}}{y_{r,B}}\right)^{\alpha * \omega_{r,i}{}^{+} + (1-\alpha) * \omega_{r,i}{}^{-}}$$

However, their approach is different in that a (max-min) normalisation is used to obtain an overall CI, which, unlike our approach, does not allow for a logical inter-temporal decomposition, nor a decomposition of the degree of unbalance.

advantage (relatively high CI^{+}-scores and ranks), but have a mediocre performance on the social inclusion sub-indicators on which they have a comparative social disadvantage (mediocre CI^{-}-scores and ranks). The countries in the fifth group are *United Kingdom*, *Estonia,* and *Cyprus* that have an overall balanced, but mediocre performance on the set of social inclusion sub-indicators (mediocre CI^{+}- and CI^{-}-scores and ranks). The countries in the sixth group (i.e *Malta, Lithuania*, *Hungary*) on the other hand show some signs of a mediocre, slightly unbalanced social inclusion portfolio mix. Finally, there is a seventh group of countries (some Southern European countries, i.e. *Italy*, *Greece, Portugal*, and some of the new Eastern European Member States, i.e. *Romania, Bulgaria*, *Latvia*) that perform rather consistently poor to very poor on the different sub-indicators.

Table 3.4 – CI-scores (ranks) under the different weighting schemes

Country	$CI_{i,2008}$	$CI^{+}_{i,2008}$	$CI^{-}_{i,2008}$	$CI_{i,2013}$	$CI^{+}_{i,2013}$	$CI^{-}_{i,2013}$
EU27	1.00 (19)	1.00 (23)	1.00 (11)	1.00 (17)	1.00 (24)	1.00 (9)
Belgium	1.15 (9)	1.36 (7)	0.96 (13)	1.17 (8)	1.47 (5)	0.93 (13)
Bulgaria	0.83 (25)	0.99 (24)	0.70 (25)	0.83 (26)	1.01 (23)	0.68 (27)
Czech Republic	1.37 (1)	1.75 (2)	1.07 (7)	1.33 (2)	1.67 (4)	1.07 (7)
Denmark	1.19 (7)	1.31 (9)	1.08 (6)	1.14 (9)	1.22 (13)	1.06 (8)
Germany	1.10 (11)	1.09 (19)	1.11 (4)	1.13 (11)	1.16 (16)	1.10 (4)
Estonia	1.00 (18)	1.15 (17)	0.87 (17)	0.98 (19)	1.11 (20)	0.87 (18)
Ireland	1.13 (10)	1.17 (16)	1.10 (5)	1.12 (13)	1.43 (7)	0.88 (16)
Greece	0.85 (24)	0.95 (27)	0.77 (22)	0.84 (25)	1.02 (22)	0.69 (26)
Spain	0.81 (26)	0.97 (25)	0.68 (26)	0.82 (27)	0.98 (25)	0.69 (25)
France	1.17 (8)	1.28 (12)	1.06 (8)	1.18 (7)	1.26 (11)	1.10 (5)
Italy	0.89 (22)	0.96 (26)	0.82 (20)	0.86 (24)	0.91 (28)	0.81 (21)
Cyprus	1.07 (14)	1.22 (15)	0.93 (14)	1.00 (18)	1.13 (18)	0.88 (15)
Latvia	0.86 (23)	1.05 (21)	0.70 (24)	0.88 (22)	1.04 (21)	0.74 (24)
Lithuania	1.05 (15)	1.29 (10)	0.85 (18)	1.01 (16)	1.23 (12)	0.83 (20)
Luxembourg	1.08 (13)	1.28 (11)	0.92 (15)	1.08 (14)	1.32 (10)	0.88 (17)
Hungary	1.02 (17)	1.25 (14)	0.84 (19)	0.98 (20)	1.13 (17)	0.85 (19)
Malta	0.96 (21)	1.27 (13)	0.73 (23)	0.96 (21)	1.21 (14)	0.76 (23)
Netherlands	1.34 (3)	1.46 (5)	1.23 (2)	1.29 (4)	1.35 (8)	1.23 (1)
Austria	1.04 (16)	1.08 (20)	1.01 (10)	1.13 (10)	1.18 (15)	1.08 (6)
Poland	1.10 (12)	1.51 (4)	0.80 (21)	1.13 (12)	1.43 (6)	0.89 (14)
Portugal	0.80 (27)	1.00 (22)	0.64 (27)	0.86 (23)	0.97 (26)	0.77 (22)
Romania	0.77 (28)	0.92 (28)	0.64 (28)	0.79 (28)	0.92 (27)	0.68 (28)
Slovenia	1.32 (4)	1.77 (1)	0.99 (12)	1.27 (5)	1.71 (1)	0.94 (12)
Slovakia	1.29 (5)	1.63 (3)	1.01 (9)	1.29 (3)	1.69 (2)	0.99 (10)
Finland	1.25 (6)	1.34 (8)	1.16 (3)	1.38 (1)	1.67 (3)	1.14 (3)
Sweden	1.34 (2)	1.46 (6)	1.24 (1)	1.25 (6)	1.34 (9)	1.16 (2)
United Kingdom	1.00 (20)	1.10 (18)	0.91 (16)	1.04 (15)	1.12 (19)	0.97 (11)

Source: Authors' calculations from Eurostat.

3.3. Inter-temporal decomposition: accounting for changes in social inclusion

In this section, we extend Van Puyenbroeck and Rogge's (2017) geometric measure of country performance change over time and its tripartite decomposition according to the above developed compromise indirect BoD-model. Because we shift to an inter-temporal framework, our notation will be extended accordingly ($y_{r,i}^{t}$ resp. $y_{r,i}^{t+1}$, $\omega_{r,i,t}^{+}$ resp. $\omega_{r,i,t+1}^{+}$, and $\omega_{r,i,t}^{-}$ *resp.* $\omega_{r,i,t+1}^{-}$). Evidently, it may well be the case that the base performance standard also changes over time, as is for example the case in our application (i.e., EU27 sub-indicator values change over time). We consequently represent base performance data by $y_{r,B}^{t}$ and $y_{r,B}^{t+1}$. As a measure of Member State performance change (PC_i) we use the ratio of the respective CIs for period *t+1* and a preceding period *t*. Under the optimistic and pessimistic BoD weighting scenario, this yields respectively (7a) and (7b):

$$PC_i^{+} = \frac{CI_{i,t+1}^{+}}{CI_{i,t}^{+}} = \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\right)^{\omega_{r,i,t+1}^{+}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\omega_{r,i,t}^{+}}} \qquad (7a)$$

$$PC_i^{-} = \frac{CI_{i,t+1}^{-}}{CI_{i,t}^{-}} = \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\right)^{\omega_{r,i,t+1}^{-}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\omega_{r,i,t}^{-}}} \qquad (7b)$$

The interpretation is clear: a PC_i^{+}or PC_i^{-}-value larger (smaller) than one indicates that the country's global performance on the set of sub-indicators has generally improved (deteriorated) under optimistic (pessimistic) weighting during the period of study. A PC_i^{+}or PC_i^{-}-value equal to one suggests that overall there was a status quo in country *i*'s policy performance under optimistic (pessimistic) weighting.

As in the static framework, the compromise version of the geometric performance change measure is defined as the geometric mean of its optimistic and pessimistic counterparts:

$$PC_i = \sqrt{PC_i^{+} \times PC_i^{-}} \qquad (8)$$

Obviously, this ("Fisher ideal") metric preserves the intuitive interpretation of the constituting measures, i.e. a PC_i-value above (below) one indicating performance progress (regress) of country *i* between *t* and *t+1* under compromise weighting; and a PC_i-value equal to one indicating overall a status quo in country's *i* policy performance.

The tripartite decomposition of the compromise geometric performance change measure PC_i is straightforward. However, given the presence of both an optimistic and pessimistic weighting scenario, the tripartite decomposition adjusts into 2 x 3 types of changes.[58] Formally,

$$PC_i = \prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,i}^{t}}\right)^{\frac{\omega_{r,i,t+1}^{+}+\omega_{r,i,t}^{+}}{4}} \times \prod_{r=1}^{s}\left(\frac{y_{r,B}^{t}}{y_{r,B}^{t+1}}\right)^{\frac{\omega_{r,i,t+1}^{+}+\omega_{r,i,t}^{+}}{4}} \times \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t+1}^{+}}{4}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t}^{+}}{4}}}$$

$$\times \prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,i}^{t}}\right)^{\frac{\omega_{r,i,t+1}^{-}+\omega_{r,i,t}^{-}}{4}} \times \prod_{r=1}^{s}\left(\frac{y_{r,B}^{t}}{y_{r,B}^{t+1}}\right)^{\frac{\omega_{r,i,t+1}^{-}+\omega_{r,i,t}^{-}}{4}} \times \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t+1}^{-}}{4}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t}^{-}}{4}}} \qquad (9)$$

or, in short,

$$PC_i = \sqrt{(\Delta OWN_i^{+} \times \Delta BP_i^{+} \times \Delta W_i^{+}) \times (\Delta OWN_i^{-} \times \Delta BP_i^{-} \times \Delta W_i^{-})} \qquad (10)$$

The components 'ΔOWN_i^{+}' and 'ΔBP_i^{+}' respectively measure country *i*'s own performance change and the base performance change over period *t* and *t+1* under optimistic weighting, thereby averaging out the effect of changes in the optimistic sub-indicator exponents during that period. Straightforwardly, the components 'ΔOWN_i^{-}' and 'ΔBP_i^{-}' measure the same change factors yet under the pessimistic BoD-based weighting scenario. In the interpretation of the component values, ΔOWN_i^{+} and ΔOWN_i^{-} values greater (smaller) than one indicate that the evaluated country *i* generally improved on the set of sub-indicators during the period of study. Values for ΔBP_i^{+} and ΔBP_i^{-} greater (smaller) than one indicate that in general the base performance declined (advanced) between *t* and *t+1*. The combinations of the two aforementioned change components, respectively $\Delta OWN_i^{+} \times \Delta BP_i^{+}$ and $\Delta OWN_i^{-} \times \Delta BP_i^{-}$, provide an answer to the question whether the progress (regress) realized by the evaluated country *i* on the set of sub-indicators is larger or smaller than the overall progress (regress) in

[58] The full decomposition is given in Appendix 3.2.

the baseline performance. The components 'ΔW_i^+' and 'ΔW_i^-' measure the changes in respectively the optimistic and pessimistic BoD-estimated sub-indicator exponents. Values of ΔW_i^+ and ΔW_i^- larger (smaller) than one suggest that for the evaluated country *i* the weighting system has changed such that policy sub-indicators in which it holds respectively a comparative advantage/disadvantage are rewarded at the end of the period more (less) generously than at the start of the period.

Table 3.5 shows for each EU Member State the decomposition of the performance change in social inclusion policy for the period 2008-2013. The compromise PC_i-values indicate an overall improvement in social inclusion for twelve countries (i.e., $PC_i > 1$), a decline for thirteen countries (i.e., $PC_i < 1$) and almost no change for two countries (i.e., $PC_i = 1$ for Bulgaria and Malta). Note that there is quite some variation among countries in terms of the levels of social inclusion performance change. Among the countries that realized progress, Finland showed the strongest improvement in social inclusion (PC_i-value = 1.11), whereas Belgium (PC_i = 1.02), Spain (PC_i = 1.01), and France (PC_i = 1.01), for instance, realized only very slight improvements. A similar remark holds for the countries with PC_i-values below one. Sweden showed the strongest decline with a PC_i-value of 0.93. However, the majority of the countries that saw a decline in their social comparative position have values that are close to one, indicating only minor deterioration in social inclusion over time (examples are Estonia, Ireland, Luxembourg). The minor changes in social inclusion observed in the majority of the EU Member States is unsurprising given the rather short time span (2008-2013) and the presence of compromise weighting in the computation of the CI_i s underlying the PC_i.

Table 3.5 – Decomposition of EU Member States' social inclusion performance change

Country	PC_i	PC_i^+	ΔOWN_i^+	ΔBP_i^+	ΔW_i^+	PC_i^-	ΔOWN_i^-	ΔBP_i^-	ΔW_i^-
EU27	1.00	1.00	1.02	0.98	1.00	1.00	1.02	0.98	1.00
Belgium	1.02	1.04	1.02	1.01	1.00	0.98	1.04	0.96	0.98
Bulgaria	1.00	1.01	1.02	0.99	1.00	0.99	0.99	1.01	0.99
Czech Republic	0.97	0.98	1.01	1.01	0.96	1.00	1.03	0.97	0.99
Denmark	0.96	0.97	0.98	0.99	0.99	0.99	1.04	0.97	0.98
Germany	1.03	1.03	1.03	0.98	1.02	1.00	1.01	0.99	1.00
Estonia	0.98	0.98	1.00	0.98	1.00	1.00	1.02	0.99	0.99
Ireland	0.99	1.11	1.05	1.00	1.05	0.89	0.95	1.00	0.95
Greece	0.99	1.04	1.04	0.96	1.05	0.95	0.89	1.02	1.05
Spain	1.01	1.00	1.03	0.98	1.00	1.01	1.02	0.98	1.01
France	1.01	0.99	0.99	1.00	1.00	1.02	1.04	0.96	1.02
Italy	0.97	0.97	1.00	0.98	0.99	1.00	1.01	0.97	1.01
Cyprus	0.94	0.96	0.96	1.00	1.00	0.98	1.01	0.98	0.99
Latvia	1.02	1.00	1.03	0.97	1.00	1.03	1.05	1.00	0.98
Lithuania	0.96	0.97	1.04	0.95	0.99	0.99	0.98	1.02	0.99
Luxembourg	0.99	1.02	1.02	0.99	1.00	0.98	1.03	0.98	0.97
Hungary	0.96	0.95	0.96	0.99	1.00	1.01	1.02	0.98	1.01
Malta	1.00	0.98	0.99	1.01	0.98	1.02	1.08	0.96	0.98
Netherlands	0.96	0.96	0.99	1.01	0.96	1.00	1.05	0.95	1.00
Austria	1.09	1.05	1.03	0.99	1.03	1.04	1.06	0.98	1.00
Poland	1.03	0.97	1.01	0.96	1.00	1.06	1.06	0.99	1.00
Portugal	1.08	0.98	1.01	0.97	1.00	1.10	1.12	0.97	1.01
Romania	1.03	1.00	1.04	0.97	0.99	1.03	1.02	1.01	1.00
Slovenia	0.96	0.99	1.02	0.97	1.00	0.97	0.97	0.99	1.02
Slovakia	1.01	1.02	1.04	1.00	0.98	0.99	1.01	0.98	1.00
Finland	1.11	1.12	1.06	1.01	1.04	0.99	1.05	0.96	0.99
Sweden	0.93	0.96	1.00	0.99	0.97	0.97	1.01	0.98	0.98
United Kingdom	1.04	1.01	1.02	0.98	1.01	1.03	1.05	0.99	1.00

Source: Authors' calculations from Eurostat.

Note: $PC_i = PC_i^+ \times PC_i^- = \left(\Delta OWN_i^+ \times \Delta OWN_i^-\right) \times \left(\Delta BP_i^+ \times \Delta BP_i^-\right) \times \left(\Delta W_i^+ \times \Delta W_i^-\right)$.

Indeed, the performance change measure depends considerably on the implemented weighting system. The results in Table 3.5 show that for 12 Member States the estimated performance change is opposite under optimistic and pessimistic BoD-based weighting (i.e., $PC_i^+ > 1$ and $PC_i^- < 1$, or vice versa). This finding suggests a change in the degree of unbalance in the social inclusion portfolio mix. Examples are Ireland and Greece who saw a deterioration on the social inclusion sub-indicators of comparative weakness, causing PC_i^--values to be below one (i.e., PC_i^- = 0.89 and 0.95, respectively) and improvements on the social inclusion indicators of comparative strength, causing PC_i^+-values to be higher than one (i.e., PC_i^+= 1.11 and 1.04, respectively). Moreover, the compromise PC_i -values of 0.99 for both Ireland and Greece suggest that these improvements and deteriorations more or less cancelled each other out. The results in Table 3.5 also indicate that other countries, such as Poland and Portugal evolved towards a more balanced performance in their social portfolio mix, performing better on the social inclusion sub-indicators in which they hold a social comparative disadvantage and worse on the social inclusion sub-indicators in which they hold a social comparative advantage. In the next section we apply a multiplicative Bortkiewicz decomposition to account for such changes in the degree of unbalance in Member States' social inclusion portfolio mix.

In the remainder of this section, we analyse the tripartite decomposition of the estimated compromise PC_i according to Van Puyenbroeck and Rogge (2017). The results in Table 3.5 reveal some specific patterns in the social inclusion developments within the EU. The compromise ΔOWN_i components (i.e., $\Delta OWN_i = \Delta OWN_i^+ \times \Delta OWN_i^-$) are greater than one for the majority of the EU Member States.[59] This result implies that, globally, most Member States improved their performance on the set of social inclusion policy indicators over the period 2008-2013. The underlying ΔOWN_i^+ and ΔOWN_i^- reveal a more nuanced picture, though. The ΔOWN_i^+-values indicate that for 6 Member States (most notably Cyprus and Hungary with ΔOWN_i^+ = 0.96) social inclusion deteriorated under optimistic BoD-weighting. The ΔOWN_i^--values show that for 5 Member States social inclusion got worse under pessimistic BoD-weighting. Particularly the sharp fall in performance on the weakest social inclusion sub-indicators as observed for crisis-hit Greece and Ireland (ΔOWN_i^- = 0.89 for Greece and ΔOWN_i^- = 0.95 for Ireland) seems worrisome.

As to the ΔBP_i-component, results in Table 3.5 show that BP_i^+ and ΔBP_i^- are smaller than one for the majority of the EU Member States. This suggests that for the most of the EU Member

[59] The ΔOWN_i -component corresponds with the first method of analysing employment and social developments and levels in the Joint Employment Report that was outlined by the European Commission and Council in March 2014. It provides a "synthesized historical overview", by summarizing the change in social inclusion in a certain year, as compared with earlier periods in time (Council of the European Union, 2014, p. 51).

States the change in the baseline social inclusion performance was beneficial, i.e. the changes in the baseline social inclusion performance making it easier for these countries to realize performance progress as compared to the baseline performance. A more detailed screening of the change in EU27 social inclusion sub-indicator data shows that considerable decline occurred in four out of the nine sub-indicators (‘AROPE’, ‘relative median poverty risk’, ‘fixed at-risk-of-poverty rate’, and ‘in work at-risk-of-poverty rate’). Countries with high BoD-based importance weights on these social inclusion sub-indicators (e.g., Belgium, Czech Republic, Malta, the Netherlands, and Finland under optimistic BoD-weighting and, e.g., Greece, Bulgaria, Lithuania, and Romania under pessimistic BoD-weighting) have, consequently, ΔBP_i-values greater than one (as for these countries the negative change in EU27-values on these indicators is beneficial to their PC_i).

The combinations $\Delta OWN_i \times \Delta BP_i$ yield further useful insights.[60] Some Member States were able to strengthen their comparative position due to improvements in the social inclusion sub-indicators of comparative strength or weakness relative to the baseline performance EU27. Finland is an example of the former: it improved its comparative position in social inclusion relative to the baseline performance EU27 (i.e., Finland; $\Delta OWN_i^+ \times \Delta BP_i^+ \times \Delta OWN_i^- \times \Delta BP_i^- = 1.07$) due to additional gains on the social dimensions on which it performs strongly. Portugal ($\Delta OWN_i^+ \times \Delta BP_i^+ \times \Delta OWN_i^- \times \Delta BP_i^- = 1.07$) is an example of a country that was able to improve its social inclusion policy performance vis-à-vis the baseline performance due to progress in its social comparative weaknesses. The combinations $\Delta OWN_i \times \Delta BP_i$ under both optimistic and pessimistic weighting also identify Member States that saw a deterioration in social inclusion relative to the EU-baseline due to a worsening of their performance on social inclusion indicators of comparative strength or weakness. Examples are Hungary and Greece with $\Delta OWN_i^+ \times \Delta BP_i^+ \times \Delta OWN_i^- \times \Delta BP_i^-$-values of 0.95 and 0.90, respectively. For most EU Member States, however, improvements in the set of social inclusion sub-indicators approximately equalled the improvements in the EU baseline performance (i.e., $\Delta OWN_i \times \Delta BP_i \approx 1$). This suggests that for these Member States there was no or only minor convergence/divergence.

Finally, as to the measured change in the EU Member States social inclusion performance that is due to the policy weight changes (ΔW_i-component), correlational analysis suggests a strong

[60] This combination corresponds with the third method of analysing employment and social developments and levels in the Joint Employment Report as outlined by the European Commission and Council in March 2014. Specifically, these combinations point out the synthesized “dynamics of socio-economic convergence/divergence” by summarizing the change in the social inclusion policy performance of each Member State between consecutive periods relative to the change at the EU-level (Council of the European Union, 2014, p. 51).

positive correlation (i.e., correlation of 0.63) between the compromise BoD sub-indicator importance weights as in 2008 and 2013. This suggests that, on average, Member States did not undergo serious shifts in the social comparative (dis)advantages over this period. This observation is also confirmed by the many ΔW_i^+ and ΔW_i^--values that are close or equal to one in Table 3.5. Conversely, the ΔW_i^+ and ΔW_i^--values in Table 3.5 also show that some Member States did undergo considerable shifts in their social inclusion policy weights. Examples of such countries are the Czech Republic (ΔW_i^+ = 0.96), the Netherlands (ΔW_i^+ = 0.96), and Sweden (ΔW_i^+ = 0.97). However, of all EU Member States, Greece and Ireland underwent the most drastic policy weight changes. Greece ($\Delta W_i^+ = 1.05$ and $\Delta W_i^- = 1.05$) saw an expansion of its social comparative disadvantage in the sub-indicators 'relative median poverty risk gap' and 'the income inequality ratio', where previously the social comparative disadvantages were predominantly situated in the sub-indicators 'in work at-risk-of poverty rate' and 'fixed at-risk-of-poverty rate'. At the same time, Greece developed a social comparative advantage in the sub-indicator 'early school leavers' at the expense of the social inclusion sub-indicator 'employment rate of older workers'. Ireland also underwent drastic shifts in its social comparative (dis)advantages, but the effects of both ΔW_i^+ and ΔW_i^--components completely cancelled each other out ($\Delta W_i^+ \times \Delta W_i^- = 1.05 \times 0.95 = 1$).

3.4. Decomposing the degree of unbalance in country performance

We consider the ratio of geometric CI_i under optimistic weighting and the geometric CI_i under pessimistic weighting, CI_i^+/CI_i^-, as a measure of the degree of unbalance in countries' social inclusion policy performances. We develop a tripartite decomposition of this ratio to account for (changes in) the degree of unbalance. The decomposition is of the Bortkiewicz-type (1923) and inspired by recent work of Armknecht and Silver (2014, p. 7). It basically involves a decomposition of the logarithm of the ratio into three aspects:

$$\ln(CI_i^+) - ln(CI_i^-) = R_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right), \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right) * CV_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) * \sigma_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right) \quad (11)$$

or, alternatively:

$$\frac{CI_i^+}{CI_i^-} = e^{R_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right), \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right) * CV_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) * \sigma_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)} \quad (12)$$

This decomposition shows that the ratio between CI_i^+ and CI_i^- depends on three factors:[61]

(i) The first factor, $R_i^{\omega_{r,i}^-}$, measures the correlation between the ratio $(\omega_{r,i}^+/\omega_{r,i}^-)$ and $ln(y_{r,i}/y_{r,B})$, weighted by the pessimistic BoD-based sub-indicator importance weights $\omega_{r,i}^-$. In essence, a positive correlation implies that strong performances on the sub-indicators (high $ln(y_{r,i}/y_{r,B})$-values) are correlated with high $\omega_{r,i}^+$-values, while poor country performances (low $ln(y_{r,i}/y_{r,B})$-values) are associated with high $\omega_{r,i}^-$-values. In the interpretation of this factor, a positive (negative) correlation implies that the geometric CI_i developed using the optimistic BoD-based sub-indicator importance weights $\omega_{r,i}^+$ is greater (smaller) than or equal to the geometric CI_i calculated using the pessimistic BoD-based sub-indicator importance weights $\omega_{r,i}^-$. A perfect positive correlation means that the sub-indicator importance weights as obtained from the optimistic BoD-model (3) are maximizing the geometric CI_i^+, while the sub-indicator importance weights as specified by the pessimistic BoD-model (4) are minimizing the geometric CI_i^-. In essence, this first factor measures the extent to which the optimistic (pessimistic) BoD-derived importance weights are maximizing (minimizing) the CI_i^+-value (CI_i^--value).

(ii) The second factor, $CV_i^{\omega_{r,i}^-}$, represents the coefficient of variation of the ratio of the optimistic and pessimistic BoD-based indicator importance weights, $(\omega_{r,i}^+/\omega_{r,i}^-)$, weighted by the respective pessimistic BoD-based sub-indicator importance weights $\omega_{r,i}^-$ and equals $\sqrt{\sum_{r=1}^{s}\frac{(\omega_{r,i}^+)^2}{\omega_{r,i}^-} - \sum_{r=1}^{s}\omega_{r,i}^-}$. Given that we imposed the optimistic and pessimistic BoD-based importance weights to be situated between a lower bound of 2% (L_r = 0.02) and an upper bound of 35% (U_r = 0.35), the maximal attainable $CV_i^{\omega_{r,i}^-}$-value equals 3.6.[62] The final magnitude of this factor depends on two aspects. Firstly, the more dissimilar (similar) the optimistic and

---

[61] The full decomposition is given in Appendix 3. Note that the decomposition is also equivalent to:
$\ln(CI_i^+) - ln(CI_i^-) = R_i^{\omega_{r,i}^+}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right), \left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right)\right) * CV_i^{\omega_{r,i}^+}\left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right) * \sigma_i^{\omega_{r,i}^+}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)$
and $\frac{CI_i^+}{CI_i^-} = e^{R_i^{\omega_{r,i}^+}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right)\right) * CV_i^{\omega_{r,i}^+}\left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right) * \sigma_i^{\omega_{r,i}^+}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)}$. The interpretation of the three factors can be readily adjusted to this equivalent format.

[62] The $CV_i^{\omega_{r,i}^-}$ attains its maximal value when the highest possible $\left(\omega_{r,i}^+\right)^2$ values in the nominator correspond with the lowest possible $\omega_{r,i}^-$ values in the denominator. Specifically, with L_r =0.02 and U_r = 0.35 the weighted coefficient of variation equals $\sqrt{2 * \left(\frac{U_r^2}{L_r}\right) + 3 * \left(\frac{L_r^2}{L_r}\right) + 2 * \left(\frac{L_r^2}{U_r}\right) + \left(\frac{L_r^2}{(1-2U_r-5L_r)}\right) + \left(\frac{(1-2U_r-5L_r)^2}{L_r}\right) - 1} \approx 3.6$.

pessimistic BoD-derived importance weights are, the larger (smaller) is the gap between CI_i^+ and CI_i^-. In fact, if the optimistic $\omega_{r,i}^+$ and pessimistic $\omega_{r,i}^-$ BoD-derived importance weights are completely similar (that is, the optimistic and pessimistic weighting scheme are equivalent), then the $CV_i^{\omega_{r,i}^-}$-component is equal to zero and, hence, no difference between CI_i^+ and CI_i^- exists. Secondly, the $CV_i^{\omega_{r,i}^-}$-component also depends on the degree of concentration ('variation') of the sub-indicator importance weights. With the degree of concentration we imply the extent to which the importance weights diverge from equal weighting. A higher $CV_i^{\omega_{r,i}^-}$ denotes more concentrated optimistic ($\omega_{r,i}^+$) and pessimistic ($\omega_{r,i}^-$) importance weights. The more concentrated the optimistic importance weights, the lower weight will be attached to sub-indicators of comparative weakness and, hence, the higher the CI_i^+-value (provided that $R_i^{\omega_{r,i}^-}$ is positive). To explain this effect, it is important to stress that the multiplicative aggregator functions penalize inequality among the sub-indicators. In contrast, more concentrated pessimistic weights imply that predominantly sub-indicator of comparative weakness are considered in the construction of the CI, and hence, resulting in a lower CI_i^--value (provided that $R_i^{\omega_{r,i}^-}$ is positive). Logically, when imposing more strict weight restrictions (e.g., L_r = 0.05 and U_r = 0.25 instead of L_r = 0.02 and U_r = 0.35), importance weights become automatically less concentrated and the maximum possible gap between CI_i^+ and CI_i^- becomes smaller. In fact, in the extreme case of equal weighting, the coefficient of variation equals zero and hence, no difference between CI_i^+ and CI_i^- can exist.

(iii) The third factor, $\sigma_i^{\omega_{r,i}^-}$, measures the standard deviation of $ln(y_{r,i}/y_{r,B})$ weighted by the pessimistic BoD-derived importance weights $\omega_{r,i}^-$. This third factor captures the intensity of the differences in the evaluated Member State's performances on the set of sub-indicators vis-à-vis the baseline sub-indicator performance values. The idea is that if a country shows a strong variation in its performances on the set of sub-indicators, the aggregation of these performances are more likely to diverge under different weighting schemes.

The decomposition as in (12) is used to measure and decompose the degree of unbalance in the EU Member States' social inclusion policy performances in 2008 and 2013. The results are displayed in Table 3.6. This table shows several interesting patterns. Firstly, both in 2008 and 2013, the sign of the first factor $R_i^{\omega_{r,i}^-}$ is positive for all except one of the EU Member States.[63]

[63] Germany is the sole Member State with a negative $R_i^{\omega_{r,i}^-}$-value ($R_i^{\omega_{r,i}^-}$ = - 0.05 in 2008) implying that the geometric CI_i^--value is (slightly) higher than the CI_i^+-value. This suggests that for Germany the optimistic (pessimistic) BoD-derived importance weights fail to maximize (minimize) the geometric CI_i^+ (CI_i^-).

This causes the geometric CI_i^+ to be greater than the geometric CI_i^-. For twenty-one Member States $R_i^{\omega_{\bar{r},i}}$-values are higher than 0.30 (seven Member States with $R_i^{\omega_{\bar{r},i}}$ greater than 0.60 and fourteen Member States with $R_i^{\omega_{\bar{r},i}}$ between 0.30-0.60). This indicates that for these twenty-one Member States our alternative versions of the optimistic and pessimistic BoD-models as in (3) and (4) are relatively effective in, respectively, maximizing and minimizing the geometric CI_i^+ and CI_i^-. However, for the other six Member States (with positive $R_i^{\omega_{\bar{r},i}}$-values lower than 0.30), assigning importance weights to the social inclusion sub-indicators other than the ones specified by our optimistic (pessimistic) BoD-models, could have resulted in significantly higher CI_i^+-values (lower CI_i^--values). A second interesting pattern is that the $CV_i^{\omega_{\bar{r},i}}$-values for all Member States are closely situated to the maximum attainable standard deviation value of 3.6. This result is not a surprise, as our alternative versions of the optimistic and pessimistic BoD-model as in (3) and (4) were specifically designed to avoid that the optimistic and pessimistic models would indicate the same sub-indicator simultaneously as comparative strength and comparative weakness. Of all EU Member States, the Netherlands has the lowest $CV_i^{\omega_{\bar{r},i}}$-values (i.e., $CV_{i,2008}^{\omega_{\bar{r},i}}$= 2.84 and $CV_{i,2013}^{\omega_{\bar{r},i}}$= 2.93). The reason for these rather low $CV_i^{\omega_{\bar{r},i}}$-values is that the social inclusion sub-indicators of comparative strength receive more or less equal BoD-derived importance weights. In other words, the Netherlands' optimistic weights are relatively unconcentrated. Finally, the $\sigma_i^{\omega_{\bar{r},i}}$-values reveal that the differences between the evaluated Member State's performances on the set of sub-indicators vis-à-vis the baseline sub-indicator performance values can be significant.[64] Examples of Member States with significant $\sigma_i^{\omega_{\bar{r},i}}$-values are Spain (values of 0.29 and 0.23) and Slovenia (values of 0.29 and 0.31). These countries show diverging performances on social dimensions on which they perform comparatively poorly.

The tripartite decomposition as in (12) also yields useful information concerning the change in the degree of unbalance in Member States' social inclusion policy portfolio mix. In particular, by looking at the change in the gap between the geometric CI_i^+ and CI_i^-, we can analyse whether the degree of unbalance in the social inclusion policy performance changed over time. Results in Table 3.6 show that for most EU Member States the gap between the geometric CI_i^+ and CI_i^- did change significantly during the period 2008-2013. However, there are some exceptions. The

---

[64] Note that for the baseline performance, EU27, it holds that for all r sub-indicators $(y_{r,B}/y_{r,B})$ is equal to one. As a consequence, the $\sigma_{ln(y_{r,i}/y_{r,B})}^{\omega_{\bar{r},i}}$-factor equals zero and the geometric CI_i^+ and CI_i^- are perfectly equal.

differences in CI_i^+/CI_i^--values in 2008 and 2013 for countries like Belgium, Ireland, Greece and Finland suggest an increasing unbalance in the way that these countries perform on the set of social inclusion sub-indicators. For Belgium (change from $CI_{i,2008}^+/CI_{i,2008}^-$ = 1.42 to $CI_{i,2013}^+/CI_{i,2013}^-$ = 1.58), for instance, this increase in unbalance was mainly due to a change in the sub-indicator importance weights which became more effective in maximizing (minimizing) the CI_i^+ (CI_i^-)-value ($R_{i,2008}^{\omega_{r,i}^-}$= 0.46 and $R_{i,2013}^{\omega_{r,i}^-}$= 0.74). On the other hand, Member States like Hungary, Poland and Portugal became increasingly more balanced in their social inclusion performances. Portugal, for instance, moved from a highly unbalanced social inclusion policy performance $CI_{i,2008}^+/CI_{i,2008}^-$= 1.55) to a moderately balanced social inclusion policy portfolio mix ($CI_{i,2013}^+/CI_{i,2013}^-$= 1.26). This change was mainly because of realized balancing in its performance vis-à-vis the EU27 baseline performance on the social inclusion sub-indicators on which it performs poorly ($\sigma_{i,2008}^{\omega_{r,i}^-}$= 0.32 and $\sigma_{i,2013}^{\omega_{r,i}^-}$= 0.15). In other words, the smaller variation in its performances resulted in a smaller gap between the optimistic CI_i^+ and pessimistic CI_i^-.

Table 3.6 – Multiplicative Bortkiewicz decomposition

Country	$\ln\left(\frac{CI^+_{i,2008}}{CI^-_{i,2008}}\right)$	$R^{\bar{\omega}_{r,i}}_{i,2008}$	$CV^{\bar{\omega}_{r,i}}_{i,2008}$	$\sigma^{\bar{\omega}_{r,i}}_{i,2008}$	$\ln\left(\frac{CI^+_{i,2013}}{CI^-_{i,2013}}\right)$	$R^{\bar{\omega}_{r,i}}_{i,2013}$	$CV^{\bar{\omega}_{r,i}}_{i,2013}$	$\sigma^{\bar{\omega}_{r,i}}_{i,2013}$
EU27	0.00	-	3.12	0.00	0.00	-	3.08	0.00
Belgium	0.35	0.46	3.49	0.22	0.46	0.74	3.44	0.18
Bulgaria	0.34	0.48	3.29	0.22	0.39	0.52	3.21	0.23
Czech Republic	0.49	0.72	3.12	0.22	0.45	0.65	2.93	0.24
Denmark	0.19	0.31	2.89	0.22	0.14	0.19	3.00	0.25
Germany	-0.02	-0.05	3.30	0.10	0.05	0.20	3.50	0.08
Estonia	0.28	0.48	3.48	0.16	0.24	0.46	3.49	0.15
Ireland	0.06	0.13	2.99	0.14	0.48	0.60	3.46	0.23
Greece	0.21	0.37	3.14	0.18	0.40	0.49	3.50	0.23
Spain	0.36	0.36	3.46	0.29	0.35	0.44	3.51	0.23
France	0.18	0.28	3.45	0.19	0.14	0.29	3.44	0.14
Italy	0.17	0.55	3.14	0.10	0.11	0.37	3.25	0.09
Cyprus	0.28	0.39	3.22	0.22	0.24	0.52	3.43	0.14
Latvia	0.40	0.84	3.50	0.14	0.34	0.50	3.51	0.20
Lithuania	0.42	0.88	3.51	0.14	0.39	0.73	3.46	0.16
Luxembourg	0.33	0.51	3.55	0.18	0.40	0.71	3.27	0.17
Hungary	0.40	0.56	3.41	0.21	0.29	0.56	3.47	0.15
Malta	0.56	0.67	3.43	0.24	0.47	0.58	3.42	0.24
Netherlands	0.17	0.36	2.84	0.17	0.09	0.16	2.93	0.19
Austria	0.07	0.13	3.42	0.16	0.08	0.15	2.96	0.19
Poland	0.64	0.79	3.58	0.23	0.47	0.81	3.48	0.17
Portugal	0.44	0.39	3.49	0.32	0.23	0.44	3.50	0.15
Romania	0.36	0.56	3.48	0.18	0.30	0.47	3.43	0.19
Slovenia	0.58	0.59	3.38	0.29	0.60	0.56	3.51	0.31
Slovakia	0.48	0.65	3.41	0.22	0.53	0.78	3.47	0.19
Finland	0.15	0.22	2.76	0.25	0.38	0.47	3.09	0.26
Sweden	0.16	0.20	3.29	0.25	0.14	0.18	3.32	0.24
United Kingdom	0.19	0.77	3.46	0.07	0.15	0.47	3.51	0.09

Source: Authors' calculations from Eurostat. *Note*: $\ln\left(\frac{CI^+_{i,t}}{CI^-_{i,t}}\right) = R^{\bar{\omega}_{r,i}}_{i,t} \times CV^{\bar{\omega}_{r,i}}_{i,t} \times \sigma^{\bar{\omega}_{r,i}}_{i,t}$.

3.5. Discussion

In this chapter, we extended Van Puyenbroeck and Rogge's (2017) 'indirect' multiplicative BoD-approach in two important ways. On a methodological level, following the suggestion made by the authors, we combined optimistic and pessimistic BoD-weighting into their indirect CI-framework. In doing so, we use alternative versions of the optimistic and pessimistic BoD weighting models as the combination of the existing optimistic and pessimistic BoD-models may yield implausible results in the joint identification of strengths and weaknesses in countries' policy performances (i.e., sub-indicators can simultaneously be identified as a country's comparative strength and comparative weakness). Whereas previous versions of the optimistic and pessimistic BoD-models used country-specific benchmarks, our alternative BoD-weighting models assess country performances towards a fixed, hypothetical, best-case benchmark. The idea is that our models yield optimistic and pessimistic BoD-based importance weights that more plausibly reflect countries comparative strengths and weaknesses. In fact, a post hoc analysis of the optimistic and pessimistic importance weights showed that the weights derived from our alternative optimistic (pessimistic) BoD-model were reasonably effective in maximizing (minimizing) the geometric CI-value for the majority of the countries included in the dataset. This illustrates that the Benefit-of-the-Doubt principle still plays in our alternative weighting models. However, at the same time, it became evident that awarding high (low) importance weights to sub-indicators in which countries realize a comparative advantage (disadvantage) relative to the other countries does not automatically result in the maximum (minimum) possible attainable CI-value, which is guaranteed in the 'direct' multiplicative BoD-models (Zhou et al., 2010; Tofallis, 2014). Clearly, this result stresses the need for further research on a 'direct' multiplicative BoD-model that deals with the issues of commensurability (Zhou et al., 2010) and the presence of a scaling factor (Tofallis, 2014).

On a conceptual level, we showed that by integrating the two different weighting sets into one framework, we obtain synthesized, yet detailed information that can be useful for monitoring specific policy performance developments. We adjusted the inter-temporal version of Van Puyenbroeck and Rogge's (2017) 'indirect' geometric CI-framework and its tripartite decomposition in change components (e.g., divergence or convergence towards the baseline performance) accordingly. The adjusted inter-temporal version provides a more comprehensive and nuanced view on changes in countries' policy performances. In addition, to explain for (changes in) the degree of unbalance in countries' policy performances, we proposed a multiplicative Bortkiewicz decomposition of the ratio of countries' geometric CI-values under different weighting schemes.

We illustrated the relevance of the geometric composite indicator with compromise BoD-weighting to support decision/policy making using EU social inclusion data. While our empirical results primarily serve to illustrate the proposed method, they are interesting in their own right. The observation that there is a broadly shared EU-wide concern to combat poverty and social inclusion while, at the same time, the different traditions and instruments to achieve this goal are, under the subsidiarity principle, still largely to be situated at the level of national social policy, marks a setting in which performance benchmarking fits uneasily with the idea of some ‘imposed’ policy priority weighting scheme. Both the optimistic and the pessimistic BoD-models interpret comparative performances as a revealed evidence of policy priorities: the optimistic model stipulates that the weights should explicitly reflect these policy priorities, whereas the pessimistic model attempts to incentivize policy makers to prioritize on the dimensions on which the country lags behind. Hence the potential value added of a compromising BoD-based weighting method within the EU social policy framework. As to the empirical results, the recurrent finding in the CI-ranking that Romania, Bulgaria, Spain, Greece and Latvia are among the worst performing member states, may be a valuable trigger for further action by the national states concerned, by national poverty agencies, NGO’s, etc. Conversely, we identified EU Member States such as Sweden, the Czech Republic, the Netherlands and Finland as good performing countries, thereby effectively recognizing that different social policy models might overall lead to (comparatively) good results.

Of course, depending on the policy context and monitoring objectives, many other (BoD-derived) weighting methods can be readily implemented within Van Puyenbroeck and Rogge’s (2017) indirect multiplicative BoD-framework. For instance, one could resort to the so-called conditional BoD-model to determine the sub-indicator importance shares (Verschelde & Rogge, 2012). The main idea of such a conditional BoD-weighting model is that differences between entities’ exogenous policy environments are taken into account in the assessment of its comparative strengths (or weaknesses). Further robustness and sensitive analysis can be integrated in the indirect multiplicative BoD-model by comparing entities’ performances based on (different sets of) conditional and unconditional weights. In fact, the multiplicative Bortkiewicz decomposition proposed in this chapter of the ratio of an unconditional BoD-CI and the conditional BoD-CI (or other unconditional BoD-CIs) may prove useful in explaining the differences between both indexes. We leave the integration of such alternative BoD-weighting models within the geometric index number framework as a scope for further research and applications.

Appendix 3.

Appendix 3.1. Traditional optimistic and pessimistic BoD-models

A key feature of the traditional optimistic BoD-model is that it chooses for each country the weighting scheme such that the country is evaluated optimally vis-à-vis the other countries. This implies that the sub-indicators on which the evaluated country is relatively close (further away) to the strongest performer(s) in the sample obtain a relatively high (low) importance weight. Formally this involves solving the following linear programming model:

$$CI_i^{trad} = \max_{w_{r,i}^{trad}} \sum_{r=1}^{s} w_{r,i}^{trad} y_{r,i} \quad (A1)$$

$$s.t.$$

$$w_{r,i}^{trad} y_{r,j} \le 1 \quad (j = 1, \dots, i, \dots, n)$$

$$w_{r,i}^{trad} \ge 0 \quad (r = 1, \dots, s)$$

with CI_i^{trad} the CI for evaluated country *i* as computed by traditional, 'optimistic' BoD-model, $y_{r,i}$ the performance of the evaluated country *i* on the r^{th} sub-indicator ($r = 1,\dots,s$), $y_{r,j}$ the performances of the country *j* ($j = 1,\dots,n$) on the r^{th} sub-indicator ($r = 1,\dots,s$), $w_{r,i}^{trad}$ the optimal weights for the evaluated country *i* on the sub-indicator *r* as computed by the traditional 'optimistic' version of the BoD-model. It are these optimal BoD-weights that are used by Van Puyenbroeck and Rogge (2017) to derive the sub-indicator exponents $\omega_{r,i}$ for the geometric CI as in (1). Specifically, Van Puyenbroeck and Rogge (2017) use the "pie-shares" of the traditional linear BoD-model which designate the relative importance of each sub-indicator within the CI. More formally, it concerns the products of the original sub-indicator values $y_{r,i}$ and the corresponding optimistic BoD-weights $w_{r,i}^{trad}$ divided by the value of the objective function $\sum_{r=1}^{s} w_{r,i}^{trad}\, y_{r,i}$.

The pessimistic BoD-model of Zhou, Ang and Poh (2007) (ZAP) evaluates how close each country is to the worst performing country in the sample under the least optimistic evaluation conditions, i.e. by defining the sub-indicator weights such that the CI-value is minimized.[65] More precisely, the ZAP-model assigns relatively high (low) weights to the sub-indicators on

65 In the literature, the ZAP-model has been applied in the context of the Environmental Performance Index (Rogge, 2012), the Human Development Index (Hatefi & Torabi, 2010) and a gender wellbeing index (Domínguez-Serrano & Blancas, 2011), amongst others.

which the evaluated country is the closest (furthest away) to the weakest performer(s) in the sample. Formally,

$$CI_i^{ZAP} = \min_{w_{r,i}^{ZAP}} \sum_{r=1}^{s} w_{r,i}^{ZAP} y_{r,i} \qquad (A2)$$
$$s.t.$$
$$w_{r,i}^{ZAP} y_{r,i} \geq 1 \quad (j = 1, \dots, n)$$
$$w_{r,i}^{ZAP} \geq 0 \qquad (r = 1, \dots, s)$$

with CI_i^{ZAP} and $w_{r,i}^{ZAP}$ the ZAP-versions of respectively the CI and the pessimistic weights on the sub-indicator *r (r = 1,…,s)* for the evaluated country *i*.

Athanassoglou (2016) criticized the ZAP-model on the grounds that it may fail to capture the essence of a country's worst-case relative performance, precisely because it evaluates a country's performances in terms of the distance to the worst performer(s). Athanassoglou (2016) argued that the pessimistic model should rather evaluate countries in terms of their distance from the observed best performance under the least optimistic evaluation conditions. Specifically,

$$CI_i^{Ath} = \min_{i} \min_{w_{r,i}^{Ath}} \sum_{r=1}^{s} w_{r,i}^{Ath} y_{r,i} \qquad (A3)$$
$$s.t.$$
$$\sum_{r=1}^{s} w_{r,i}^{Ath} y_{r,i} = 1 \ (1\ contraint;\ inner\ minimization)$$
$$w_{r,i}^{Ath} \geq 0 \qquad (r = 1, \dots, s;\ inner\ minimization)$$

with CI_i^{Ath} and $w_{r,i}^{Ath}$ as Athanassoglou's worst-case equivalent of the optimistic BoD model in a strict mathematical sense of respectively the CI and the weights on the sub-indicator *r* (*r = 1,…,s*) for the evaluated country *i*.[66]

66 The solution amounts to solving *n* linear programs (the inner minimization of (A3)) and choosing the benchmark country for which the optimal solution which is the smallest (the outer minimization of (A3)).

Appendix 3.2. Inter-temporal decomposition

The PC_i-values under optimistic (7a) and pessimistic (7b) weighting are decomposed into three types of changes. By rearranging we get:

$$PC_i^+ = \frac{CI_{i,t+1}^+}{CI_{i,t}^+} = \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\right)^{\omega_{r,i,t+1}^+}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\omega_{r,i,t}^+}} \qquad (7a)$$

$$PC_i^+ = \prod_{r=1}^{s}\left(\frac{\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}}{\frac{y_{r,i}^{t}}{y_{r,B}^{t}}}\right)^{\frac{\omega_{r,i,t+1}^+ + \omega_{r,i,t}^+}{2}} \times \prod_{r=1}^{s}\left\{\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\right) * \left(\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)\right\}^{\frac{\omega_{r,i,t+1}^+ - \omega_{r,i,t}^+}{2}}$$

$$PC_i^- = \frac{CI_{i,t+1}^-}{CI_{i,t}^-} = \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\right)^{\omega_{r,i,t+1}^-}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\omega_{r,i,t}^-}} \qquad (7b)$$

$$PC_i^- = \prod_{r=1}^{s}\left(\frac{\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}}{\frac{y_{r,i}^{t}}{y_{r,B}^{t}}}\right)^{\frac{\omega_{r,i,t+1}^- + \omega_{r,i,t}^-}{2}} \times \prod_{r=1}^{s}\left\{\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\right) * \left(\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)\right\}^{\frac{\omega_{r,i,t+1}^- - \omega_{r,i,t}^-}{2}}$$

Which leads to:

$$PC_i^+ = \prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,i}^{t}}\right)^{\frac{\omega_{r,i,t+1}^+ + \omega_{r,i,t}^+}{2}} \times \prod_{r=1}^{s}\left(\frac{y_{r,B}^{t}}{y_{r,B}^{t+1}}\right)^{\frac{\omega_{r,i,t+1}^+ + \omega_{r,i,t}^+}{2}} \times \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\,\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t+1}^+}{2}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\,\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t}^+}{2}}}$$

$$PC_i^+ = \Delta OWN_i^+ \times \Delta BP_i^+ \times \Delta W_i^+$$

$$PC_i^- = \prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,i}^{t}}\right)^{\frac{\omega_{r,i,t+1}^- + \omega_{r,i,t}^-}{2}} \times \prod_{r=1}^{s}\left(\frac{y_{r,B}^{t}}{y_{r,B}^{t+1}}\right)^{\frac{\omega_{r,i,t+1}^- + \omega_{r,i,t}^-}{2}} \times \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t+1}^-}{2}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t}^-}{2}}}$$

$$PC_i^- = \Delta OWN_i^- \times \Delta BP_i^- \times \Delta W_i^-$$

Since the compromise version of the geometric performance change (PC_i) measure is defined as the geometric mean of its optimistic (PC_i^+) and pessimistic (PC_i^-) counterparts, we obtain:

$$PC_i = \sqrt{PC_i^+ \times PC_i^-} \quad (8)$$

$$PC_i = \prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,i}^{t}}\right)^{\frac{\omega_{r,i,t+1}^+ + \omega_{r,i,t}^+}{4}} \times \prod_{r=1}^{s}\left(\frac{y_{r,B}^{t}}{y_{r,B}^{t+1}}\right)^{\frac{\omega_{r,i,t+1}^+ + \omega_{r,i,t}^+}{4}} \times \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t+1}^+}{4}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t}^+}{4}}}$$

$$\times \prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,i}^{t}}\right)^{\frac{\omega_{r,i,t+1}^- + \omega_{r,i,t}^-}{4}} \times \prod_{r=1}^{s}\left(\frac{y_{r,B}^{t}}{y_{r,B}^{t+1}}\right)^{\frac{\omega_{r,i,t+1}^- + \omega_{r,i,t}^-}{4}} \times \frac{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t+1}^-}{4}}}{\prod_{r=1}^{s}\left(\frac{y_{r,i}^{t+1}}{y_{r,B}^{t+1}}\frac{y_{r,i}^{t}}{y_{r,B}^{t}}\right)^{\frac{\omega_{r,i,t}^-}{4}}} \quad (9)$$

or, in short:

$$PC_i = \sqrt{(\Delta OWN_i^+ \times \Delta BP_i^+ \times \Delta W_i^+) \times (\Delta OWN_i^- \times \Delta BP_i^- \times \Delta W_i^-)} \quad (10)$$

Appendix 3.3. Multiplicative Bortkiewicz (1924) decomposition

The degree of unbalance is given by:

$$\frac{CI_i^+}{CI_i^-} = \prod_{r=1}^{s} \left(\frac{y_{r,i}}{y_{r,B}}\right)^{\omega_{r,i}^+ - \omega_{r,i}^-}$$

Taking the logarithm on both sides gives:

$$\ln(CI_i^+) - \ln(CI_i^-) = \sum_{r=1}^{s} \omega_{r,i}^+ \times ln\left(\frac{y_{r,i}}{y_{r,B}}\right) - \sum_{r=1}^{s} \omega_{r,i}^- \times ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$$

$$\ln(CI_i^+) - ln(CI_i^-) = \sum_{r=1}^{s} \omega_{r,i}^- \times \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) \times ln\left(\frac{y_{r,i}}{y_{r,B}}\right) - \sum_{r=1}^{s} \omega_{r,i}^- \times ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$$

The weighted ($\omega_{r,i}^-$) covariance cov_i of the ratio ($\omega_{r,i}^+/\omega_{r,i}^-$) and $ln(y_{r,i}/y_{r,B})$ is defined as:

$$cov_i^{\omega_{r,i}^-}\left(\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right), ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right) = R_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right), \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right) \times \sigma_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) \times \sigma_i^{\omega_{r,i}^-}\left(\frac{y_{r,i}}{y_{r,B}}\right)$$

Which is equivalent to:

$$cov_i^{\omega_{r,i}^-}\left(\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right), ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right) = \sum_{r=1}^{s} \omega_{r,i}^- \times \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) \times ln\left(\frac{y_{r,i}}{y_{r,B}}\right) - \sum_{r=1}^{s} \omega_{r,i}^- \times ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$$

And with the weighted ($\omega_{r,i}^-$) mean of the ratio ($\omega_{r,i}^+/\omega_{r,i}^-$) equal to one, ($\sum_{r=1}^{s}\left(\omega_{r,i}^- \times \frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) = 1$), it is also equivalent to:

$$cov_i^{\omega_{r,i}^-}\left(\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right), ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right) = R_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right), \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right) \times CV_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) \times \sigma_i^{\omega_{r,i}^-}\left(\frac{y_{r,i}}{y_{r,B}}\right)$$

Finally, we obtain:

$$\ln(CI_i^+) - ln(CI_i^-) = R_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right), \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right) \times CV_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) \times \sigma_i^{\omega_{r,i}^-}\left(\frac{y_{r,i}}{y_{r,B}}\right) \tag{11}$$

$$\frac{CI_i^+}{CI_i^-} = e^{R_i^{\omega_{r,i}^-}\left(ln\left(\frac{y_{r,i}}{y_{r,B}}\right), \left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right) \times CV_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right) \times \sigma_i^{\omega_{r,i}^-}\left(\frac{y_{r,i}}{y_{r,B}}\right)} \tag{12}$$

Chapter 4

Non-parametric estimation and evaluation of capability sets

4.1. Introduction

Well-being is now widely recognized by academics and policy makers as a multidimensional phenomenon going well beyond income alone. Although few people would argue that income is not an essential element of well-being, there is a growing consensus that other dimensions (such as health, housing and employment) should also be taken into account in well-being evaluations (Stiglitz et al., 2009). The capability approach has been particularly influential on this shifting focus from narrow economic indicators to broader considerations of human well-being since its development in the 80s by Nobel prize laureate Amartya Sen. The capability approach is a broad normative framework that is used for the evaluation and assessment of individual well-being and social arrangements (Sen, 1993, p. 30). It is used in development studies, welfare economics, social policy, political philosophy with both abstract and theoretical contributions, as well as applied and empirical studies (Robeyns, 2018).

One of the main achievements of the capability approach, in addition to its advocacy towards a multidimensional perspective on well-being and poverty measurement, is that it broadens the informational basis of welfare economics to include opportunity freedoms. Opportunity freedom is measured by the capability sets of individuals. A capability set is the set of all the potential feasible functioning vectors from which an individual can choose. A functioning vector is a description of the state of "being and doings" on various well-being dimensions (Sen, 1985, p. 10). The functioning bundle that an individual ultimately chooses from the capability set is called the observed/achieved functioning bundle. The achieved functionings are thus the various things that an individual does or is while leading his/her life (Sen, 1993, p. 31). The quality of life of an individual is assessed in terms of his/her capability set, as it represents the freedom to achieve valuable functionings (Nussbaum & Sen, 1993, p. 31). According to the capability approach, there is not only an instrumental aspect to being able to choose more (i.e. more options can make better alternatives available). More freedom to lead different types of life is also intrinsically valuable

(Sen, 1988, p. 290). Nussbaum and Sen (1993, p. 39) argue that “this view is contrary to the one typically assumed in standard consumer theory, in which the contribution of a set of feasible choices is judged exclusively by the value of the best element available. Even the removal of all the elements of a feasible set other than the chosen best element is seen as no real loss, since the freedom to choose does not matter in itself”. The theory thus argues that capabilities are a richer metric to measure quality of life than achieved functionings. Sen’s famous example of fasting and starving illustrates this argument well. An individual that is malnourished because he/she is fasting (i.e. he/she can be well-nourished, but the individual chooses not to) should be considered better off than an individual that is malnourished because he/she is starving (i.e. he/she simply cannot be well-nourished because he/she does not have access to decent food).

A major challenge in the operationalization of the capability approach is the unobservability of individuals’ capability sets. Datasets on well-being typically gather information on the achieved functionings of individuals and not on their capabilities. There are two types of methodologies to arrive at capability measures in the current literature. The first collects and analyses primary capability data through surveys in which respondents are asked directly to evaluate their capabilities on certain domains (e.g. the survey question “I am able to express my political views” in Anand et al., 2009). Krishnakumar and Chávez-Juárez (2016) warn that such self-reported evaluations of life situations are subjective and that this subjective nature should be taken into account while drawing conclusions based on such datasets. The second one is the so-called latent variable approach and uses Structural Equation Models (SEM) in which the capabilities of an individual are assumed to be unobservable, with each observed outcome (i.e. the achieved functionings) representing a partial manifestation of it (Krishnakumar, 2007). SEM consist of a measurement part and a structural part. The structural part is made up of equations that link exogenous ‘causes’ in the form of economic, social, political, institutional, cultural and other factors with latent capabilities. The measurement part links the latent capabilities and other exogenous variables that take into account individual preferences and characteristics with the observed functioning achievements. The estimates of these models are used to construct (scalar based) capability measures (Krishnakumar & Ballon, 2008; Ballon & Krishnakumar, 2008; Bhattacharya & Banerjee, 2012; Addabbo et al., 2014).

This chapter proposes an alternative data-driven and non-parametric approach to *estimate* capability sets. Specifically, we propose to estimate group capability sets by enveloping the achieved functioning bundles of group members (Charnes et al., 1978). This approach amounts to reasoning in terms of representative agents, where all group members are of the same type of the representative agent. Groups are defined on the basis of exogenous characteristics that influence the group's ability to generate capabilities. Individuals that share the same exogenous characteristics are assumed to be able to generate the same capabilities. We realize that this assumption disregards the fact that, formally, the unit of reference in the capability approach is the individual. However, it is common practice in empirical capability approach studies to use information on (group-averaged) functioning achievements to make claims about the opportunities of groups. Social indicators (e.g. employment rate) or the UNDP's Human Development Index are well-known examples of such an approach. Our approach similarly builds on the idea that the observed functioning achievements of group members entail information on the capabilities of a group. In addition, the representative agent assumption is mitigated in that not all group members are assumed to share the same capabilities. Specifically, a robust version of the non-parametric estimation model, in which the influence of individuals with outlying functioning achievements on the estimation of capability sets is reduced, is presented. The basic idea is that outliers and extreme points in the cloud of observed functioning bundles deviate from the group's 'true' capabilities and should not (always) be used in the estimation of a group's capability set (Daraio & Simar, 2007).

Our approach differs from the SEM methodology in several ways. First, our approach is closely related to the original conceptualization of the capability approach in that capability *sets* are estimated. The latent variable approach disregards the set framework and calculates capability *scalars* in each dimension. Second, the SEM approach is parametric and makes commitments to the form of the equations or to the distributions of the error terms. The approach presented in this chapter is non-parametric and makes minimal assumptions on the structure of the group capability set we wish to estimate. That is, capability sets are assumed to be compact, convex and comprehensive, which are standard assumptions in the capability set literature (Xu, 2002; Farina et al., 2004; Patanaik & Xu, 2007). Finally, SEM require information on exogenous factors that influences the relationship between latent capabilities and the actual functioning achievements (e.g. preferences). In the present approach, the observed functioning bundles are only used to estimate the capability sets of groups. We do not consider the manner in which individuals choose functioning bundles from their capability sets.

Next to presenting a methodology to estimate capability sets, we also provide a framework to non-parametrically *evaluate* the estimated group capability sets by quantifying the extent to which groups differ in their ability to generate capabilities. We find inspiration in the contribution of Muellbauer (1987).[67] Muellbauer suggests to evaluate sets by measuring the distance of their boundary from the origin along a particular ray or multiple rays that have a connection to the distribution of the population.[68] Building on these ideas, we propose two metrics to evaluate sets. The first one evaluates sets on the basis of a single ray that goes through a functioning bundle that represents the average functioning achievement in the sample. This measure is transitive and allows for a complete multilateral ranking and comparison of the estimated capability sets. However, using a single ray to evaluate sets may not fully gauge the diversity of options desirable for freedom offered by capability sets. We therefore present a second set evaluation metric that calculates the average distance between the estimated group capability frontiers along multiple rays that go through the observed functioning bundles of group members. We believe such an averaging approach is appealing, as it not only fully exploits all the information that is available (i.e. the observed functioning bundles), it is also inherently 'democratic' in the sense that each individual matters in the evaluation. However, this metric can only be used for bilateral comparisons of capability sets.

The proposed framework can be used to make relative group-level well-being assessments. For policy makers such an aggregated level is a relevant level of analysis, as it allows to identify and assess the situation of certain societal groups of interest (i.e. typically those who are lagging behind) and set up appropriate policy interventions. In this regard, the presented approach can help to "assess the institutional support for the individual", which is one of the key ambitions of the capability approach (Douglas & Ney, 1998, p. 72). We illustrate our method on the EU-SILC cross-sectional 2013 dataset, considering four well-being functioning dimensions (i.e. household income, material well-being, housing quality and health). We estimate country-specific capability sets of 32 European countries and apply the multilateral evaluation tool to compare the country-specific capability sets with a pan-European one. In addition, we apply both the bilateral and multilateral set evaluation metric to benchmark the capability sets of the French and the Germans.

[67] The theoretical literature on the ranking of (capability) sets is now sizable. Some approaches focus on counting the quantity of options available to sets (Pattanaik & Xu, 1990), while other approaches involve a preference ordering over elements of the sets. For surveys, see Barberà et al., 2004; Foster, 2011.

[68] The idea of using distance measures to evaluate capability sets has more recently been echoed in theoretical contributions of Gaertner and Xu (2006, 2008, 2011), Gaertner (2012) and Farina and co-authors (2004), who provide an axiomatic characterization of the ranking induced by the (Euclidian) distance between the boundary of capability sets and a minimal reference functioning bundle. In the concluding section, we offer suggestions on how some of these theoretical contributions can be integrated in our empirical measurement framework.

The aim of this exercise is to study the extent to which both metrics lead to a different evaluation outcome.

This chapter is structured as follows. Section 4.2 presents the methodology and consists of three subsections. We first discuss how group capability sets are defined and non-parametrically estimated. Next, we present the non-parametric capability set evaluation framework. Finally, we present the robust version of our estimation model which mitigates the influence of outlying observations on the estimation of capability sets. Section 4.3 presents the data. Section 4.4 illustrates our approach empirically. Section 4.5 concludes and presents possible avenues for follow-up research.

4.2. Methodology

4.2.1. Estimating capability sets

In any application, the capability sets of individuals are unknown. Typically, the only information that is available to the analyst is the sample of functioning observations. We use a model to empirically reconstruct capability sets from the achieved functioning bundles of group members. A key assumption behind our approach is that all group members have the same capabilities, i.e. they share a capability set, the "group capability set". In other words, if a group member can achieve a certain functioning bundle, then it is assumed that another group member should also be able to achieve this functioning bundle. This approach amounts to reasoning in terms of representative agents, where all group members are of the same type of the representative agent and share exogenous characteristics that influence the group's ability to generate capabilities. The representative agent assumption is a standard technique in economics, but also in the literature on the measurement of multidimensional well-being. That is, all studies in this literature build on the assumption that the well-being function of individuals or countries can be portrayed by representative agents. For instance, Fleurbaey and co-authors (2009) use a hedonic life satisfaction regression to estimate group-specific (i.e. regression-averaged) multidimensional well-being functions. The Human Development Index takes the idea of a representative agent a step further and assumes that all countries have the same underlying multi-dimensional well-being function.

Let us introduce some notation. We consider the case in which functioning observations belong to G (mutually exclusive) groups $g_1, g_2, ..., g_G$. Let Ψ^{g_j} denote true group capability set of group g_j (with *j = 1,...G*) we wish to estimate. Let $\mathbb{R}_{++}$ denote the set of all positive numbers, $\mathbb{R}^r_{++}$ be the *r*-fold Cartesian product of $\mathbb{R}_{++}$. We define $f_i^{g_j} = \left(f_{i1}^{g_j}, ..., f_{ir}^{g_j} \right) \in \mathbb{R}^r_{++}$ as an achieved functioning vector of individual i belonging to group g_j with $i = 1,...,\# g_j$. For all f_i, f_j we write $f_i \geq f_j$ when $f_{ik} \geq f_{jk}$ for all k $\left(k = 1,...,r\right)$, $f_i > f_j$ when $f_i \geq f_j$ and $f_i \neq f_j$ for at least one k $\left(k = 1,...,r\right)$, and $f_i \gg f_j$ when $f_i > f_j$ for all k $\left(k = 1,...,r\right)$.

We assume that the 'true' group capability set Ψ^{g_j} is compact, convex and comprehensive:

A1 Compactness: Ψ^{g_j} is a closed and bounded set.

A2 Convexity: given any two functioning vectors of the capability set, any linear combination of these two functioning vectors is also an element of the capability set. Formally,

if $f \in \Psi^{g_j}$ and $f^{'} \in \Psi^{g_j}$ then $\lambda f + \left(1-\lambda\right) f^{'} \in \Psi^{g_j}$ for any $\lambda \in \left[0,1\right]$.

A3 Comprehensiveness: if a functioning vector is an element of the capability set, then functioning vectors with less achievements are also elements of that set. Formally,

if $f \in \Psi^{g_j}$ and $f \leq f^{'}$ then $f^{'} \in \Psi^{g_j}$.

These assumptions are standard in microeconomic theory and in the capability set literature (Xu, 2002; Farina et al., 2004; Patanaik & Xu, 2007). They are also rather minimal, with the aim to let the data reveal as much as possible about the underlying model, rather than making strong, untested assumptions that have potential to influence the results of the estimation in a misleading and perhaps large way (Simar & Wilson, 2008).

We further assume that all the observed functioning bundles of group members are included in the group capability set:

> **A4 Inclusion of all observations**: the observed functioning vectors of all group members belong to the capability set of the respective group. Formally,
>
> $$f_i^{g_j} \in \Psi^{g_j}, \forall i,\ i = 1,\ldots,\#g_j\ .$$

This assumption is crucial to empirically reconstruct a group capability set on the basis of the observed functioning achievements of group members. Of course, the more homogeneous the group (i.e. the stricter the group definition), the more realistic the assumption is that a group capability set can be constructed from the observed functioning bundles of group members. Put differently, the more exogenous characteristics that individuals share that are influential on their ability to generate capabilities (e.g. gender, age, educational background, IQ, the socio-economic background of parents, etc.), the more realistic it becomes to assume that if an individual can achieve a certain functioning bundle, then another individual with the same exogenous characteristics should be able to achieve this functioning bundle as well. In most empirical settings, the realism of our approach will thus depend on the availability of good group identification variables that can be used to define more or less homogeneous groups in the dataset. Assumption A4 implies, in addition, that the observed functioning bundles are measured without error. Obviously, both the presence of a rich set of exogenous characteristics that can be used to define groups and the absence of measurement error are rather strong requirements for most practical-empirical settings. Therefore, while we maintain assumption A4 to simplify our theoretical exposition, we will relax it in Section 4.2.3.

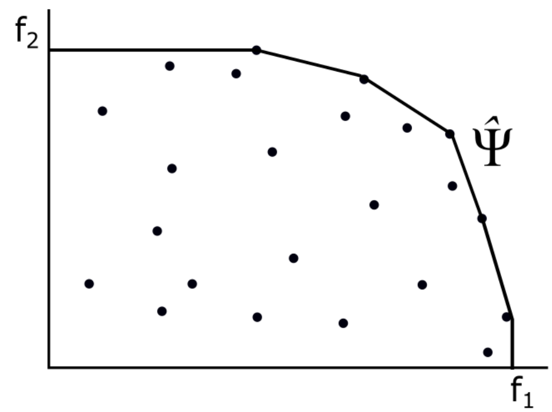
The estimator of the true capability set Ψ^{g_j}, $\hat{\Psi}^{g_j}$, satisfying assumptions A1-A4, is the intersection of all possible sets that satisfy assumptions A1-A4. It is the set of all functioning bundles f that can be stated in the following form (Charnes et al., 1978):[69]

$$\hat{\Psi}^{g_j} = \left\{ f \mid \sum_{i=1}^{\# g_j} \lambda_i f_i^{g_j} \geq f, \sum_{i=1}^{\# g_j} \lambda_i \leq 1, \lambda_i \geq 0, i = 1,\dots,\# g_j \right\} \quad (1)$$

Figure 4.1 illustrates how a group capability set, satisfying assumptions A1-A4, is estimated by enveloping the observed functioning achievements of group members (reflected by the dots) on two functioning dimensions. The capability set estimator is the smallest free disposable (A2) convex (A3) set covering all observed functioning dimensions (A4). Compactness (A1) ensures that the capability set is closed. The frontier of the group capability set denoted by the bold line reflects the group capability set frontier. The functioning bundles on the group capability set frontier represent the best possible functioning achievements within the group capability set. Formally, a functioning bundle f belongs to the boundary of $\hat{\Psi}^{g_j}$ when $f \in \hat{\Psi}^{g_j}$ and there is not another functioning bundle $f' \in \hat{\Psi}^{g_j}$ such that $f' \gg f$.

[69] Assumptions A2-A3 are standard in the non-parametric frontier estimation literature, but can be relaxed. If one drops the assumption of convexity (A2), one arrives at the so-called Free Disposable Hull (FDH) set. The FDH set is operationalized by imposing binary constraints associated on the λ_i variables, i.e. $\lambda_i \in \{0,1\}$. Formally, $\hat{\Psi}^{g_j} = \left\{ f \mid \sum_{i=1}^{\# g_j} \lambda_i f_i^{g_j} \geq f, \sum_{i=1}^{\# g_j} \lambda_i \leq 1, \lambda_i \in \{0,1\}, i = 1,\dots,\# g_j \right\}$. The assumption of comprehensiveness can be replaced by the star-shaped assumption, as assumed in Gaertner and Xu (2006, 2008, 2011). Sets are star-shaped if $f \in \hat{\Psi}^{g_j}$ and $tf \in \hat{\Psi}^{g_j}$ for all $0 \leq t \leq 1$. The star-shaped assumption is operationalized by setting equalities (instead of inequalities) for the functioning equations ($\sum_{i=1}^{\# g_j} \lambda_i f_i^{g_j} = f_i$). Formally, $\hat{\Psi}^{g_j} = \left\{ f \mid \sum_{i=1}^{\# g_j} \lambda_i f_i^{g_j} = f, \sum_{i=1}^{\# g_j} \lambda_i \leq 1, \lambda_i \geq 0, i = 1,\dots,\# g_j \right\}$.

Figure 4.1 – Constructing group capability sets from observed functioning bundles



It is important to note that while we assume individuals can freely choose any functioning bundle from their group's capability set, we do not consider the manner in which they do so. Observed functioning bundles below the group capability set frontier can be interpreted in two ways. First, it can be an endogenous choice of the individual to choose a functioning bundle in the capability set that lies below the frontier. This interpretation is fully consistent with the capability approach: the individual could choose a functioning bundle with higher achievements from the capability set, but opts, - for whichever reason - , not to do so. Individuals do not necessarily choose within their capability set the functioning vector that would give them the highest level of individual well-being. Returning to Sen's famous fasting-starving example, the individual is capable to be well-nourished, but chooses not to because he/she is fasting. Second, bundles below the group capability set frontier can also be interpreted as an empirical misspecification of the estimated frontier. That is, the estimated frontier is misspecified in the sense that it does not reflect the individual's 'true' capabilities and the individual is, in fact, not capable of attaining the functioning bundles on the estimated frontier. Returning to Sen's famous fasting-starving example, the individual is simply not capable to be well-nourished, even though its group capability set would suggest otherwise. As discussed earlier, the realism of the estimated group capability set hinges strongly on the way groups are defined. The most strict group definition would be to say that "groups" are individual-specific. In this case, the estimated capability set of an individual i ($\hat{\Psi}^i$) reduces to the set of functioning bundles that are smaller than or equal to the individual's achieved functioning bundle ($\hat{\Psi}^i = \{f \mid f_i \geq f\}$). This is our model's (rather conservative) estimate of an individual's capability set.

4.2.2. Evaluating capability sets

The previous section presented a non-parametric methodology to estimate a group capability set from the observed functioning bundles of group members. In this section, we present an appropriate methodology to quantify the extent by which groups differ in their ability to generate capabilities. The aim of the evaluation exercise is to assess which estimated group capability set is the largest. We follow Muellbauer's (1987) suggestion to evaluate two sets by comparing distances between their boundaries along a ray.

We use the distance between a reference functioning bundle f_o and the boundary of the group capability set Ψ^{g_j} along the ray that goes through f_o as a tool to evaluate capability sets. Formally, this measure is defined as:

$$\theta_o^{g_j} = \sup\{\theta \mid \theta f_o \in \Psi^{g_j}\} \quad (2)$$

where $\theta_o^{g_j}$ gives the factor by which the functioning achievements in bundle f_o need to be multiplied in order for f_o to be situated on the frontier of the capability set Ψ^{g_j}. If f_o lies below (above) the boundary of Ψ^{g_j}, then $\theta_o^{g_j} > 1$ ($\theta_o^{g_j} < 1$). If f_o lies on the boundary of Ψ^{g_j}, then $\theta_o^{g_j} = 1$.

Since the true group capability set Ψ^{g_j} is unknown, we replace the true group capability set Ψ^{g_j} with its estimator as defined in (1) in equation (2) to estimate the distance measure $\theta_o^{g_j}$. Formally,

$$\hat{\theta}_o^{g_j} = \sup\left\{\theta \mid \theta f_o \in \hat{\Psi}^{g_j}\right\} \quad (3)$$

The distance of reference functioning bundle f_o towards the estimated group capability set frontier of group g_j ($\hat{\Psi}^{g_j}$), $\hat{\theta}_o^{g_j}$, can be calculated by the following linear programme:

$$
\begin{aligned}
&\hat{\theta}_o^{g_j} = \underset{\lambda_i,\theta}{Max}\ \theta \\
&\sum_{i=1}^{\# g_j} \lambda_i f_i^{g_j} - \theta f_o \geq 0 && (r \text{ times, once for every } k=1,\dots,r), \\
&\sum_{i=1}^{\# g_j} \lambda_i \leq 1, && (4) \\
&\lambda_i \geq 0 && (\# g_j \text{ times, once for every } i=1,\dots,\# g_j)
\end{aligned}
$$

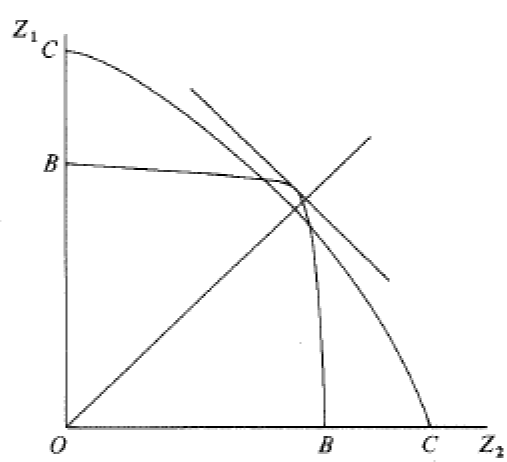
Let $SE_o\left(\hat{\Psi}^{g_1}, \hat{\Psi}^{g_2}\right)$ denote a metric that evaluates the estimated capability sets of group g_1 ($\hat{\Psi}^{g_1}$) and g_2 ($\hat{\Psi}^{g_2}$) along a ray that goes through f_o. It is defined as the ratio of the distance of a reference bundle f_o to the boundary of $\hat{\Psi}^{g_1}$ along the ray that goes through f_o ($\hat{\theta}_o^{g_1}$) and distance of the same reference bundle f_o to the boundary of $\hat{\Psi}^{g_2}$ along the same ray that goes through f_o ($\hat{\theta}_o^{g_2}$). Formally,

$$SE_o\left(\hat{\Psi}^{g_1}, \hat{\Psi}^{g_2}\right) = \frac{\hat{\theta}_o^{g_1}}{\hat{\theta}_o^{g_2}} \qquad (5)$$

If $SE_o\left(\hat{\Psi}^{g_1}, \hat{\Psi}^{g_2}\right)$ is equal to one, then the reference functioning bundle f_o is equally far removed from the estimated boundaries of both frontiers (along the ray that goes through the reference functioning bundle f_o). It can then be concluded that the boundaries of both frontiers coincide along the ray that goes through f_o and that both groups are able to generate an equal amount of functioning achievements. If $SE_o\left(\hat{\Psi}^{g_1}, \hat{\Psi}^{g_2}\right)$ differs from one, then there exists a gap between both estimated group capability set frontiers along the ray that goes through f_o. If $SE_o\left(\hat{\Psi}^{g_1}, \hat{\Psi}^{g_2}\right)$ is smaller (larger) than one, then the distance between the origin and the intersection with the group capability set frontier of $\hat{\Psi}^{g_2}$ along the ray that goes through f_o is larger (smaller) than the distance between the origin and the intersection with the group capability set frontier of $\hat{\Psi}^{g_1}$ along the same ray. The intersection point of the group capability set frontier of $\hat{\Psi}^{g_2}$ ($\hat{\Psi}^{g_1}$) along the ray that goes through f_o then lies outside the group capability set of $\hat{\Psi}^{g_1}$ ($\hat{\Psi}^{g_2}$). It can then be concluded that group g_2 (g_1) is able to generate more capabilities than group g_1 (g_2). The $SE_o\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)$ metric should be interpreted in relative terms. For example, a $SE_o\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)$ value of 0.8 (1.25) indicates that the extent of the capability set $\hat{\Psi}^{g_2}$, as measured through the ray that goes through f_o, is 25 per cent larger (smaller) than the capability set of $\hat{\Psi}^{g_1}$. Put differently, group g_2 (g_1) is capable of generating functioning achievements that are 25 per cent higher than the functioning achievements that group g_1 (g_2) is capable of generating (as evaluated by the ray that goes through f_o). Note that it is important to be specific about the order of comparison, as it follows from (3) that $SE_o\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)=SE_o\left(\hat{\Psi}^{g_2},\hat{\Psi}^{g_1}\right)^{-1}$.

Given that an endless variety of rays present themselves, the selection of the ray that is used to evaluate sets is arbitrary. There is a priori no information available on why a certain ray should be preferred over another. The evaluation outcome depends on the specific ray that is employed. With the ray shown in Figure 4.2, set OBB is better than set OCC, while most would argue that set OCC offers more freedom. A ray from the origin going through B or C would classify set OCC as better off.

Figure 4.2 – Evaluating sets



Source: Muellbauer (1987, p. 51)

Muellbauer's (1987) suggests to use rays that are based on the actual distribution in the population to evaluate sets. He further suggested to employ two kinds of set evaluation metrics. The first one evaluates sets on the basis of a single ray that represents the "average or representative value". In our framework, this could be the ray that goes through a functioning bundle that represents the average functioning achievement across groups. Let f_μ denote the average bundle $\left(f_{\mu 1},...,f_{\mu r}\right)$.[70] Thus,

$$SE_\mu\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)=\frac{\hat{\theta}_\mu^{g_1}}{\hat{\theta}_\mu^{g_2}} \qquad (6)$$

gives the set multilateral evaluation metric based on a single, average ray. It is important to note that the set evaluation metric in (4) is transitive (i.e. $SE_\mu\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)\times SE_\mu\left(\hat{\Psi}^{g_2},\hat{\Psi}^{g_3}\right)=SE_\mu\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_3}\right)$) and allows to evaluate all the estimated group capability sets simultaneously. That is, a complete multilateral ranking and comparison of the estimated group capability sets can be obtained.

Muellbauer (1987) also suggests to evaluates sets on the basis of a distribution of rays that represents the actual distribution in the population. Thus, instead of evaluating the gap between the frontiers of two sets along a ray that goes through a single point of reference as in (6) and (5), one can also consider multiple rays that go through the observed functioning bundles of all the individuals in the sample. Let $SE_i\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)$ represent the distance between the boundaries of the capability sets of group g_1 and g_2 along the ray that goes through the functioning bundle of individual i. We define the bilateral group capability set evaluation metric as the geometric average of the $SE_i\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)$ measures across all individuals that are members of the group capability sets under evaluation:

[70] With $f_{\mu k}=\frac{1}{\#g_1+...+\#g_G}\left(\sum_{i\in g_1} f_{ik}^{g_1}+...+\sum_{i\in g_G} f_{ik}^{g_G}\right)$ as the average functioning achievement on functioning dimension k $(k=1,...,r)$ across G groups.

$$\overline{SE}\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)=\left[\prod_{i\in g_1} SE_i\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)\times\prod_{i\in g_2} SE_i\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)\right]^{\frac{1}{\#g_1+\#g_2}} \qquad (7)$$

$$\overline{SE}\left(\hat{\Psi}^{g_1},\hat{\Psi}^{g_2}\right)=\left[\prod_{i\in g_1}\left(\frac{\hat{\theta}_i^{g_1}}{\hat{\theta}_i^{g_2}}\right)\times\prod_{i\in g_2}\left(\frac{\hat{\theta}_i^{g_1}}{\hat{\theta}_i^{g_2}}\right)\right]^{\frac{1}{\#g_1+\#g_2}}$$

We believe such an averaging approach is appealing, as it fully exploits all the information that is available (i.e. the observed functioning bundles). It also 'democratic' in the sense that each individual not only matters in the estimation of group capability sets, but also in their evaluation. It is clear that the set evaluation metric in (7) pays more attention to the diversity in the options that are desirable for freedom than the metric presented in (6). Diversity comes at a cost, however. The set evaluation in (7) is, in essence, a bilateral measure and, as such, it cannot be used to evaluate all the estimated group capability sets simultaneously. In the empirical section of the chapter (Section 4.4), we will compare the extent by which the multilateral measure in (6) and the bilateral measure in (7) result in different evaluation outcomes.

4.2.3. Robustification

As discussed in Section 4.2.1, in order for assumption 4 'Inclusion of all observations' (A4) to be realistic, groups are to be sufficiently homogeneous and all functioning achievements are measured without error. Both conditions may be problematic in many empirical applications. Specifically, outlying observations can be a source of misspecification as they may disproportionately shift the estimated group capability sets, which may, in turn, result in biased evaluations of the capability sets. If the estimated group capability set is dominated by a couple of extreme observations and the majority of the individuals in the group have significantly lower functioning achievements, then the basic premise behind our approach that all group members have access to same capabilities becomes difficult to defend. Instead, it could be argued that these outlying individuals have capabilities that deviate from the group's true capabilities. This may be due measurement errors or exogenous effects that are not accounted for in the empirical group definition. An alternative group definition with better group identification variables might give a better fit of the estimated group capability set to the true group capability set. The issue of outliers becomes particularly pressing for datasets with large numbers of observations. To give a crude example, if all group members in the dataset are used to construct a group capability set, then it is very likely that the resulting group capability set frontier will be (solely) determined by a millionaire who has "everything" on every possible functioning dimension. It would of course be unrealistic to assume that every group member can become a millionaire, regardless of the exogenous characteristics that the group members share with the millionaire.

We therefore present an alternative estimator that estimates a robust capability set, as opposed to the estimator in (1) that envelops all the data (after insight of Cazals et al., 2002; Daraio & Simar, 2005). We use the so-called locally convex order-*m* robust frontier estimator (Daraio & Simar, 2007). The robust capability set estimator replaces the goal of estimating the absolute highest functioning bundles that the group is capable of achieving with the idea of estimating functioning bundles that are more realistically achievable for group members. Let $\Psi_m^{g_j}$ be the 'true' robust group capability set of group g_j. We still assume that $\Psi_m^{g_j}$ is compact (A1), convex (A2) and comprehensive (A3), but we no longer assume it includes all the observations (A4). The robust frontier estimator estimates an empirical capability set $\hat{\Psi}_m^{g_j}$ from a strict subset of *m* observations that is drawn (randomly and with replacement) from the full set of observations. We then calculate the distance of a reference functioning bundle towards the estimated frontier. Let the resulting distance estimate of the reference functioning f_o bundle to this frontier be denoted as $\hat{\theta}_{o,b}^{m,g_j}$. Then, this estimation procedure is repeated a large number of times (B times, in casu $B = 2{,}000$) and we average these B estimates. The obtained averages are called robust order-*m* distance measures $\hat{\theta}_o^{m,g_j}$ (i.e. $\hat{\theta}_o^{m,g_j} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{o,b}^{m,g_j}$).

Some observations may not be included in the estimated robust group capability set $\hat{\Psi}_m^{g_j}$. These individual observations represent functioning achievements that are higher than what the robust group capability set would indicate as achievable, and can thus be considered as super-achievers. The probability of observing bundles above the order-*m* frontier is a function of the size of the subsample *m* that is drawn from the original sample. If the size of the subsample *m* increases, then, on average across Monte Carlo draws, the probability of observing functioning achievement bundles in the robust group capability set increases.[71] The parameter *m* serves as a trimming value,

[71] An interesting feature of the robust group capability set estimator is that it converges to the full group capability set estimator when the size of the subsample *m* goes to infinity (i.e. $\hat{\Psi}_m^{g_j} \rightarrow \hat{\Psi}^{g_j}$ as $m \rightarrow \infty$).

which allows us to tune the percentage of super-achievers. We follow the non-parametric frontier estimation literature and fix *m* at its value for which the marginal decrease in the fraction of super-achiever observations becomes sufficiently small (for more technical details, see Daraio & Simar, 2005). However, to test the sensitivity of the results to the size of *m*, we also estimate and evaluate group capability sets under several *m*-values (see Appendix 4).

The statistical properties of the locally convex order-*m* frontier estimator have still to be investigated, but Daraio and Simar (2007) conjecture that they share the same properties as the original (non-convex) order-*m* estimators under the appropriate convexity assumptions on $\hat{\Psi}_m$. Order-*m* estimators are consistent estimators of the true capability set if the assumptions on the capability sets (A1-A3) are true.[72] That is, for a fixed *m*, if the sample size increases, the order-*m* distance estimator $\hat{\theta}_o^{m,g_j}$ will converge to the true, but unknown distance θ_o^{m,g_j}. Order-*m* estimators have a fast rate of convergence (i.e. at a rate of $\sqrt{n}$). This implies that the set evaluation metrics presented in (6) and (7) are also a consistent estimators of the 'true' distance between two capability set frontiers (provided that the assumptions made on the group capability sets are true).

4.3. Data

The proposed method is applied on the EU-SILC 2013 cross-sectional dataset, which includes an ad-hoc module with questions on subjective well-being. The European Commission uses this dataset to monitor Member States' performances on social inclusion. The original dataset contains

[72] In addition, two statistical assumptions (SA) on the sampling process should also hold (Simar and Wilson, 2008). These are:

> **SA1 Random sample**. The achieved functioning bundles f^{g_j} are realization of identically and independently distributed random variables on the convex attainable set of Ψ^{g_j}.
>
> **SA2 Positiveness**. The probability density of f^{g_j} is completely positive on the boundary of the capability set and is continuous in any direction toward the interior of Ψ^{g_j}.

The first statistical assumption is common in most empirical studies and states that the observation are considered as random draws from the population. The realism of this assumption ultimately depends on the quality of the survey data that is available to the researcher. We used appropriate sample weights in the Monte Carlo subsampling procedure (individuals with a higher weight have a higher probability to be included in the subsample that is used to estimate the robust group capability sets) and in the computation of the bilateral evaluation metric to ensure the validity of this assumption. The second statistical assumption says that the probability of observing functioning bundles in an open neighborhood of the frontier is strictly positive. This assumption implies that *some* individuals will want to maximize their functioning achievements and be as close as possible to the group capability set frontier. We believe this assumption to be both theoretically and empirically realistic.

614,788 observations. The analysis is conducted on a dataset of 336,979 individuals.[73] We define groups on the basis of country membership. This has several advantages. First, country membership is, disregarding the issue of migration which concerns only a small share of the population, truly exogenous to individuals. Second, this approach ensures that individuals in the same household belong to the same group. This is an important feature, as some functioning dimensions are defined at the household-level (such as household income, see below). Third, the large group sample sizes ensure a consistent estimation of the group capability frontiers. However, we recognize at the same time that, ideally, more and better group identification variables should be employed to maintain the realism behind our key assumption that all group members share the same capabilities. The empirical analysis should therefore merely be considered as an illustration of our approach.

Four well-being functioning dimensions are considered: household equivalised disposable income, health, housing quality and material well-being. The health, housing quality and material well-being dimensions are constructed from hedonic regressions. We conduct separate regressions on a pooled dataset and on a reduced dataset with only France and Germany for, respectively, the multilateral European and the bilateral French-German evaluation exercise. It is to be expected that the regressions on the reduced dataset for France and Germany delivers estimates that are more precise and more in line with French-German specificities. In addition, as some variables were missing for some countries in the pooled setting, more variables and/or variables of a better quality were used in the regressions for France and Germany.

The yearly disposable equivalised income of households is obtained by summing up all monetary incomes received from any source by any member of the household or the household itself and then deducting taxes and social contributions paid by the household. Household disposable incomes are expressed in PPS and are divided by an equivilisation scale to take differences in household size into account.[74] Given that household sample surveys are often expected to perform poorly in the tails of the distributions, a winsorizing procedure for the lowest 0.5% and the highest 1% income observations is adopted, replacing those extreme values with the values of trimming thresholds. This approach also has the advantage that there are no negative and zero incomes at the lower end of the distribution.

[73] The individuals that were excluded had missing observations on variables that were used in hedonic regressions. These hedonic regressions were employed to construct the functioning well-being dimensions (see below).
[74] The OECD-modified equivilisation scale is used, which gives a weight of one to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 14.

The SILC dataset includes an ordinal variable that measures individuals' self-assessed health. While this health indicator provides a global assessment of health, this variable is not suitable for our purposes. The reasons are twofold. First, it is generally recommended to work with cardinal indicators in non-parametric frontier estimation.[75] Second, the subjectivity of the indicator significantly limits its interpersonal comparability. For this reason, a cardinal indicator based on the predicted values of an ordered probit regression of self-assessed health status on five objective health dummies (i.e., suffering from a chronic illness, being very limited in daily activities, being limited in daily activities, having unmet medical needs, having unmet dental needs), two socio-demographic variables (i.e., gender and age dummies) and four socio-economic variables (i.e., (the logarithm of) household disposable income, sickness and disability transfer dependency, an unemployment dummy, highest obtained educational degree dummies) is constructed.[76] This procedure imposes cardinality and the constructed measure is as close as possible to the self-reported health status, while ensuring that individuals with the same observed characteristics obtain the same health measure (Decancq & Lugo, 2009).

It is important to control for unobserved individual heterogeneity as these may influence aspirations, expectations and psychological adaptation to ill-health, which may, in turn, influence the reported health status (Schokkaert, 2007; Lokshin & Ravallion, 2008). Panel data can capture such unobserved individual heterogeneity by including individual-specific and time-invariant factors in the hedonic regression. Panel data with information on subjective well-being is unfortunately not available in the EU-SILC dataset. A second best solution, in the case of a lack of panel data, is to use information on personality traits (Xara & Schokkaert, 2017). Unfortunately, EU-SILC does not contain this information either. We therefore follow Xara and Schokkaert (2017) and use information about self-rated affects or emotions available in EU-SILC 2013, as a sort of third best solution to control for individual-specific time-invariant characteristics. More specially, we include four self-rated affects variables: respondents had to indicate whether they felt "very nervous", "down in the dumps", "calm and peaceful", and "downhearted or depressed" over the past four weeks. Responses are recorded as a categorical variable measured on a 5-point scale ranging from 1 ("All of the time") to 5 ("None of the time"). We have reversed the scores of the "calm and peaceful" variable, so that higher values indicate that the respondent has felt this

[75] Some non-parametric frontier estimation models allow for ordinal data (e.g. Chen et al., 2017). However, these models give a cardinal interpretation to ordinal scores and are therefore less appealing.
[76] Sickness and disability transfer dependency is defined as the share of gross sickness and disability benefits within gross household income. For the French and Germany regression, sickness and disability transfer dependency is defined as the share of net sickness and disability benefits within disposable household income.

emotion more frequently. We further include country dummies and a dummy indicating whether the individual is married as additional control variables. The results are shown in Table 4.1 and are in line with other subjective health regressions on the EU-SILC dataset (Madden, 2011; Aristei & Bracalente, 2014; Coveney et al., 2018; Xara & Schokkaert, 2017). Following Van Doorslaer and Jones (2002), the predicted values are normalized by subtracting the smallest individual prediction from the predicted values.

Table 4.1 – Hedonic self-assessed health regression (Ordered Probit)

	Pooled dataset			France and Germany		
	Coef.	Std. Err.	Sign.	Coef.	Std. Err.	Sign.
Predictors						
Chronic illness	-0.82	0.01	***	-0.687	0.017	***
Very limited in daily activities	-1.82	0.01	***	-1.930	0.030	***
Limited in daily activities	-0.87	0.01	***	-0.990	0.020	***
Unmet medical needs	-0.23	0.01	***	-0.235	0.029	***
Unmet dental needs	-0.14	0.01	***	-0.191	0.028	***
Male	0.01	0.00	***	-0.036	0.013	***
Age 25-35	-0.35	0.01	***	-0.313	0.028	***
Age 35-45	-0.65	0.01	***	-0.630	0.028	***
Age 45-55	-0.88	0.01	***	-0.803	0.028	***
Age 55-65	-1.08	0.01	***	-0.974	0.029	***
Age 65-70	-1.19	0.01	***	-1.064	0.036	***
Age 70-75	-1.28	0.01	***	-1.149	0.035	***
Age 75-80	-1.33	0.01	***	-1.209	0.040	***
Age >80	-1.35	0.01	***	-1.281	0.039	***
No education	-0.59	0.03	***	-0.416	0.125	***
Primary education	-0.37	0.01	***	-0.239	0.029	***
Lower secondary education	-0.24	0.01	***	-0.210	0.024	***
Upper secondary education	-0.18	0.01	***	-0.172	0.015	***
Postsecondary (non-tertiary) educaiton	-0.15	0.01	***	-0.156	0.032	***
Unemployed	-0.03	0.01	***	-0.184	0.029	***
Equivalised household disposable income (log)	0.04	0.00	***	0.048	0.009	***
Transfer dependency	-0.26	0.01	***	-0.571	0.063	***
Control variables						
Feeling nervous	0.05	0.00	***	0.060	0.008	***
Feeling down in the dumps	0.09	0.00	***	0.122	0.009	***
Feeling calm and peaceful	0.10	0.00	***	0.114	0.009	***
Feeling downhearted or depressed	0.09	0.00	***	0.122	0.009	***
Married	0.00	0.00		-0.025	0.014	*
Country dummies	(coefficients not shown)			(coefficient not shown)		
Model information						
N	350 533			31 961		
R^2	0.304			0.300		

Source: EU-SILC (2013) cross-sectional data, author's computation. *Note*: *=significant at 10% level, **=significant at 5% level, ***=significant at 1% level.

The housing quality index is derived from a hedonic price equation, following Decancq and co-authors (2015) and Xara and Schokkaert (2017). (Imputed) Rent (per 1000 PPS) is regressed on a series of housing characteristics such as the dwelling type (i.e. detached house, semi-detached house, smaller apartment building, large apartment building), the number of rooms in the dwelling, the availability of a bath or shower, the availability of a toilet, not having a leaking roof or rot, having a sufficiently warm dwelling, having enough light in the dwelling and the number of home appliances (i.e., phone, television and washing machine). Neighbourhood effects are captured by five variables: three dummies measuring whether the respondent does not suffer from (1) noise from neighbours or from the street, (2) pollution, grime or other environment problems, and (3) crime, violence or vandalism in the area and two variables capturing the average household satisfaction (on a scale of 0-10) with green areas in the neighborhood and the average household satisfaction with the living area.[77] We account for regional price differences by including regional dummies and dummies indicating the degree of urbanization (i.e. urban, suburban, rural). The results are shown in Table 4.2. Housing quality is defined as the predicted housing price value excluding the price differences control variables). To correct for household size, we substitute 'equivalized rooms' for 'rooms' (following Decancq et al., 2015).

[77] We regress independent variables of a subjective nature by lack of more objective characteristics in the dataset. By taking the average household satisfaction, the subjectivity of the indicator is partially mitigated.

Table 4.2 – Hedonic housing quality regression (OLS)

	Pooled dataset			**France and Germany**		
	Coef.	Std. Err.	Sign.	Coef.	Std. Err.	Sign.
Indepedent variables						
Semi detached house	-0.020	0.001	***	-0.015	0.004	***
Small appartment building	-0.038	0.001	***	-0.068	0.004	***
Large appartment building	-0.020	0.001	***	-0.064	0.005	***
Number of rooms	0.066	0.000	***	0.093	0.001	***
Bath or shower	0.001	0.004		-0.011	0.029	
Toilet	-0.004	0.004		0.043	0.026	*
No rot	0.014	0.001	***	0.023	0.004	***
Warm dwelling	0.022	0.001	***	0.047	0.005	***
Enough light	-0.001	0.001		-0.007	0.005	
Number of home appliances	-0.007	0.002	***	-0.002	0.011	
No noise	0.006	0.001	***	-0.008	0.004	**
No environmental problems	0.003	0.001	**	0.015	0.004	***
No crime, violence or vandalism	0.010	0.001	***	0.030	0.004	***
Average household satisfaction with green areas	0.001	0.000	***	0.002	0.001	**
Average household satisfaction with the living area	0.009	0.000	***	0.011	0.001	***
Control variables						
Urban	0.109	0.001	***	0.073	0.004	***
Sub-urban	0.053	0.001	***	0.038	0.003	***
Regional dummies	(coefficients not shown)			(coefficients not shown)		
Model information						
N	503 655			50 228		
R^2	0.55			0.2578		

Source: EU-SILC (2013) cross-sectional data, author's computation. *Note*: *=significant at 10% level, **=significant at 5% level, ***=significant at 1% level. Note: in the pooled regression gross imputed rent is used for non-market tenants as the dependent variable (except for Germany, for which net imputed rent is used as gross imputed rent is missing). In the France-Germany regression net imputed is used for non-market tenants as the dependent variable.

The material standard of living index is constructed on the basis of a hedonic subjective regression. The variable "Satisfaction with financial situation (0-10)" is taken as the dependent variable in an ordered probit regression model. We take up 15 dummies that directly capture whether the respondent has a comfortable material life. Six items are measured at the household-level. These are: (1) the household has a computer, (2) the household has a car, (3-5) the household has no arrears in payments (of mortgage, utility bills, installments), (6) the household is able to go on an annual holiday, (7) the household eats meat/chicken/fish daily, (8) the household is capable to face unexpected expenses, and (9) the household replaces worn-out furniture. Five other items are measured at the personal-level. These are: (10) the individual replaces worn-out clothing by new (not second-hand) ones, (11) the individual gets together with friends/family for a drink/meal, (12) the individual regularly participates in a leisure activity, (13) the individual spends a small amount of money on him/herself, (14) the individual has internet for personal use.[78] In addition, the analysis in Chapter 2 of this dissertation showed that variables capturing households' longer-term command over resources and needs can explain the (objective) material well-being situation of households. We thus also include the (logarithm of) equivalised household disposable income, an unemployment dummy, dummies capturing the highest obtained educational degree, the health index and household quality index to 'indirectly' capture the material well-being situation of households.

To account for unobserved individual heterogeneity, we again include the four self-rated affects variables (i.e. feeling very nervous, feeling down in the dumps, feeling calm and peaceful, and feeling downhearted or depressed), three socio-demographic control dummies (i.e. gender, age dummies, martial dummy) and country dummies. In addition, nine dummies are introduced in the model that measure whether the household or individual does not have an item due to other reasons than affordability (i.e. not owning the item due to affordability is the reference category). These are: (1) the household does not have a computer, (2) the household does not have a car, (3) the household does not replace worn-out furniture, (4) the individual does not replace worn-out clothing by new (not second-hand) ones, (5) the individual has at least two pairs of properly fitting of shoes, (6) the individual does not get together with friends/family for a drink/meal, (7) the individual does not participate regularly in a leisure activity, (8) the individual does not spend a small amount of money on him/herself and (9) the individual does not has internet for personal use. We include these dummies to capture potential aspirational effects. A positive coefficient of not owning an item due to other reasons than affordability (with not owning the item due to affordability as the reference) indicates that individuals have lower aspirations. The results are shown in Table 4.3 and are in line with the expectations. We use the estimated coefficients to

[78] The material well-being dummies measured at the personal level are unfortunately missing in some countries.

calculate the predicted values of the regression model. The predicted values are normalized by subtracting the predicted values by the smallest individual prediction.

Table 4.3 – Hedonic material well-being regression (Ordered Probit)

	Pooled dataset			**France and Germany**		
	Coef.	Std. Err.	Sign.	Coef.	Std. Err.	Sign.
Indepedent variables						
Computer	0.169	0.009	***	0.057	0.045	
Car	0.126	0.007	***	0.087	0.029	***
No arrears (mortgage)	0.189	0.011	***	0.239	0.041	***
No arrears (utility bills)	0.138	0.007	***	0.049	0.036	
No arrears (installments)	0.179	0.012	***	0.203	0.054	***
Annual holiday	0.472	0.005	***	0.338	0.019	***
Eat meat/chicken/fish daily	0.215	0.007	***	0.092	0.026	***
Able to face unexpected expenses	0.530	0.005	***	0.496	0.017	***
Replace worn-out furniture				0.333	0.020	***
Replace worn-out clothing				0.212	0.026	***
Friends and family				0.328	0.026	***
Leisure activities				0.201	0.025	***
Money for selfspending				0.325	0.024	***
Internet				0.068	0.039	*
Equivalised household disposable income (log)	0.151	0.002	***	0.161	0.008	***
Unemployed	-0.346	0.007	***	-0.163	0.028	***
No education	-0.052	0.023	**	0.115	0.116	
Primary education	-0.086	0.008	***	-0.016	0.028	
Lower secondary education	-0.012	0.006	**	-0.033	0.023	
Upper secondary education	-0.067	0.005	***	-0.088	0.014	***
Postsecondary (non-tertiary) education	-0.089	0.010	***	-0.135	0.029	***
Health index	0.099	0.002	***	0.082	0.008	***
Housing quality index	1.865	0.029	***	1.018	0.060	***
Control variables						
Feeling nervous	0.052	0.002	***	0.029	0.008	***
Feeling down in the dumps	0.074	0.003	***	0.086	0.009	***
Feeling calm and peaceful	0.092	0.002	***	0.053	0.008	***
Feeling downhearted or depressed	0.113	0.003	***	0.092	0.009	***
Male	-0.078	0.004	***	-0.111	0.012	***

Age 25-35	-0.219	0.007	***	-0.176	0.025	***
Age 35-45	-0.261	0.008	***	-0.185	0.026	***
Age 45-55	-0.252	0.008	***	-0.165	0.027	***
Age 55-65	-0.197	0.009	***	-0.136	0.029	***
Age 65-70	-0.102	0.011	***	-0.070	0.036	*
Age 70-75	-0.033	0.011	***	0.063	0.036	*
Age 75-80	0.017	0.013		0.071	0.042	*
Age >80	0.128	0.013	***	0.125	0.044	***
Married	0.173	0.004	***	0.190	0.014	***
No computer (other reason than affordability)	0.129	0.010	***	0.042	0.046	
No car (other reason than affordability)	0.175	0.008	***	0.241	0.034	***
Does not replace worn-out furniture (other reason than affordability)				0.253	0.026	***
Does not replace worn-out clothing (other reason than affordability)				0.168	0.040	***
No friends and family (other reason than affordability)				0.314	0.029	***
No leisure activities (other reason than affordability)				0.144	0.027	***
No money selfspending (other reason than affordability)				0.135	0.035	***
No internet (other reason than affordability)				0.026	0.039	
Country dummies	(coefficients not shown)			(coefficient not shown)		
Model information						
N	336,843			30,344		
R^2	0.13			0.13		

Source: EU-SILC (2013) cross-sectional data, author's computation. *Note*: *=significant at 10% level, **=significant at 5% level, ***=significant at 1% level.

Table 4.4 gives the mean functioning achievements and sample sizes for the countries analysed. Only small (average) differences in health exist across European countries. The differences in housing quality and material well-being are moderate. For instance, the average housing quality achievement of Ireland, the country that performs the best on the housing quality index, is 55 per cent higher than the average housing quality achievement of Serbia, the country that performs the worst on the housing quality index. There exist very large differences across countries on the household income dimension. Average household income ranges from 5,018 PPS (Romania) to 31,711 PPS (Switzerland).

Table 4.4 – Summary statistics

Country	Income	Health	Housing quality	Material well-being	N
Pooled regression					
Austria	22884	11.97	0.34	4.24	9692
Belgium	21063	12.04	0.39	4.30	9605
Bulgaria	7804	12.11	0.27	3.29	7659
Switzerland	31711	12.04	0.35	4.45	11502
Cyprus	20240	12.12	0.35	3.80	9862
Czech Republic	13004	11.96	0.32	3.87	10906
Germany	21343	11.94	0.34	4.19	16357
Denmark	21203	12.03	0.37	4.28	4345
Estonia	11093	11.83	0.32	3.81	9320
Greece	11005	12.00	0.28	3.56	14297
Spain	17585	12.06	0.35	3.94	24077
Finland	20087	11.68	0.37	4.22	10190
France	22059	11.84	0.36	4.16	14195
Croatia	9194	11.67	0.29	3.37	6126
Hungary	9068	11.89	0.28	3.35	16737
Ireland	20217	12.05	0.40	3.98	5789
Iceland	20003	12.06	0.34	4.24	2669
Italy	17975	11.77	0.29	3.79	22263
Lithouania	9587	12.00	0.32	3.67	7458
Luxemburg	30854	12.04	0.37	4.39	5438
Latvia	8329	11.70	0.30	3.41	9701
Malta	17080	12.04	0.36	4.07	6649
Netherlands	20979	11.92	0.38	4.30	9272
Norway	28458	12.09	0.39	4.55	5322
Poland	10674	11.89	0.29	3.62	20999
Portugal	13171	11.83	0.32	3.76	9048
Romania	5018	12.07	0.28	3.36	13948
Serbia	6453	11.99	0.26	3.29	10174
Sweden	20322	12.03	0.35	4.33	5445
Slovenia	15043	11.78	0.35	3.88	6804
Slovakia	11237	12.01	0.29	3.77	10567
The United Kingdom	20562	12.03	0.37	4.25	10427
Pan-European	17503	11.94	0.33	3.96	336843
French-German regression					
Germany	21512	3.39	0.40	4.13	14115
France	22361	3.32	0.44	4.25	16229

Source: EU-SILC (2013) cross-sectional data, author's computation.

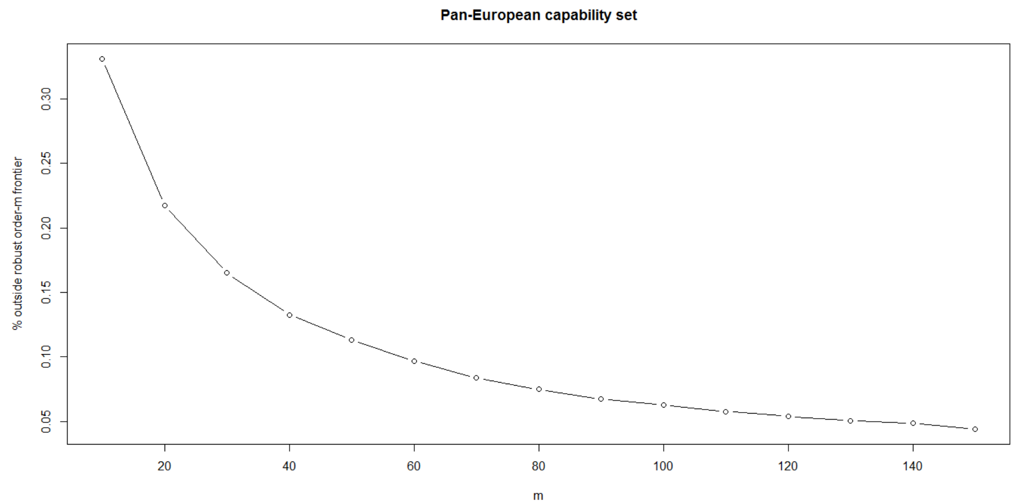
4.4. Empirical illustration

We conduct two empirical analyses to illustrate our approach. We first use the multilateral set evaluation metric to benchmark a pan-European capability set with 32 country-specific capability sets. The former set is constructed from all the observations (across all 32 countries) in the dataset (i.e. no exogenous characteristics are used to define the group), whereas the latter country-specific capability sets are constructed from (country) subsets of the dataset. This allows us to quantify the extent by which Europeans from different countries differ in their ability of generating capabilities. We use the ray that goes through a functioning bundle that represents the average pan-European functioning achievement to evaluate the capability sets. We then proceed with applying both the bilateral and multilateral set evaluation metric on the estimated capability sets of the French and the Germans. This is to assess the extent to which both measures result in diverging evaluation outcomes.

4.4.1. Multilateral evaluation

Figure 4.3 plots the percentage of Europeans that lie outside the pan-European capability set frontier for different values of m. As expected, the plot shows that this percentage decreases as m increases. The m-parameter is set at 70, as for this value an "elbow" effect can be observed in the plot, indicating the start of a more or less robust percentage of individuals that lie outside of the frontier. An m-value of 70 corresponds to a pan-European group capability set that includes 91.6 per cent of the observed functioning bundles.

Figure 4.3 – The relationship between *m* and the percentage of observations that lie outside of the estimated pan-European capability set



Source: EU-SILC (2013) cross-sectional data, author's computation.

Table 4.5 compares the pan-European capability set with the 32 country-specific European capability sets using the multilateral set evaluation metric $SE_{\mu}\left(\hat{\Psi}^{i}_{70},\hat{\Psi}^{j}_{70}\right)$ in (6) (with the pan-European capability set as $\hat{\Psi}^{i}_{70}=\hat{\Psi}^{EU}_{70}$ and the country-specific capability sets as $\hat{\Psi}^{j}_{70}$ with *j* = *1*,…, *32*) and gives the percentage of observations that lie outside of the estimated capability sets. Before discussing the between-country differences in capability sets, we first discuss the variation in outlying observations across European countries. When there are large differences across countries in the percentage of observations that lie outside of the estimated capability sets, the distributional interpretation of these estimated robust capability set differs and one should be cautious with interpreting the results. However, in our illustration, the average difference between the percentage of individuals that lie outside their own country's capability set and the percentage of individuals that lie outside the pan-European capability set is only 1.1 per cent. Thus, the estimated group capability sets have, on average across European countries, a similar distributional interpretation. Lithuania and Slovakia have, respectively, the highest and lowest percentage of observations not included in their estimated capability sets, with 11,9 and 7,98 per cent of outlying observations (versus 8.4 per cent for the pan-European capability set). Thus, even for these countries, the deviation is rather limited and a more or less consistent distributional interpretation is ensured.

Table 4.5 – Multilateral set evaluation

Country	Set evaluation	% outside frontier	Country	Set evaluation	% outside frontier
Belgium	0.974 (5)	9.11%	Hungary	1.291 (30)	9.08%
Bulgaria	1.353 (31)	9.19%	Malta	0.999 (14)	8.18%
Czechia	1.107 (23)	9.01%	The Netherlands	0.979 (7)	10.03%
Denmark	0.972 (4)	10.12%	Autstria	0.998 (12)	9.44%
Germany	0.999 (13)	8.79%	Poland	1.171 (25)	8.87%
Estonia	1.101 (22)	8.19%	Portugal	1.056 (20)	8.95%
Ireland	0.966 (3)	10.68%	Romania	1.803 (33)	11.27%
Greece	1.207 (26)	9.83%	Slovenia	1.037 (19)	9.40%
Spain	1.009 (16)	9.56%	Slovakia	1.269 (29)	11.90%
France	0.981 (8)	9.19%	Finland	0.987 (10)	11.83%
Croatia	1.267 (28)	9.13%	Sweden	0.985 (9)	9.39%
Italy	1.07 (21)	9.96%	United Kingdom	0.976 (6)	8.43%
Cyprus	1.011 (17)	9.69%	Iceland	1.015 (18)	9.73%
Latvia	1.219 (27)	8.28%	Norway	0.955 (2)	9.54%
Lithouania	1.14 (24)	7.98%	Switzerland	0.987 (11)	10.29%
Luxembourg	0.951 (1)	9.62%	Serbia	1.448 (32)	8.96%
Pan-European	1 (15)	8.38%			

Source: EU-SILC (2013) cross-sectional data, author's computation. *Note*: the number between brackets ranks the country in terms of the set evaluation outcome. The order of comparison of the set evaluation metric $SE_\mu\left(\hat{\Psi}^i_{70},\hat{\Psi}^j_{70}\right)$ is as follows: the pan-European capability set ($\hat{\Psi}^i_{70}=\hat{\Psi}^{EU}_{70}$) is evaluated against the country-specific capability sets $\hat{\Psi}^j_{70}$ (*j=1*,…,*32*).

We now discuss the set evaluation outcomes. The results indicate that strong differences in capability sets exist across European countries. The people of Luxembourg and Norway are the best off. The extent of their capability sets as measured through the pan-European average ray is, respectively, 5.2 and 4.7 per cent larger than the capability set of the average European. The people of Romania and Serbia are the worst off. The pan-European capability set is, respectively, 80.3 and 44.8 per cent larger than the capability sets of the Romanians and Serbs. Note that because the set evaluation metric is transitive, it is possible to directly compare the estimated country-specific capability sets. For instance, the Luxembourgish are capable of achieving functionings that are

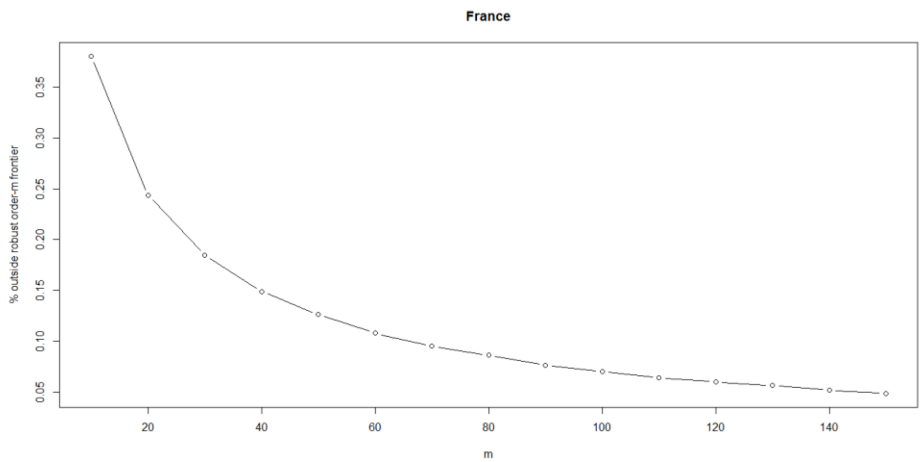
89.6 larger than the functionings that the Romanians are capable of achieving, i.e. $SE_{\mu}\left(\hat{\Psi}_{70}^{Romania},\hat{\Psi}_{70}^{Luxembourg}\right)=SE_{\mu}\left(\hat{\Psi}_{70}^{EU},\hat{\Psi}_{70}^{Luxembourg}\right)\times SE_{\mu}\left(\hat{\Psi}_{70}^{EU},\hat{\Psi}_{70}^{Romania}\right)^{-1}$; $0.527=(0.951)\times(1.803)^{-1}$.

Roughly four clusters of countries can be established from the capability set evaluation outcomes. Individuals in the first group of countries have capability sets that are larger than the pan-European one. Norway, Luxembourg, Ireland, Denmark, Belgium, the United Kingdom, the Netherlands, France, Finland, Sweden and Switzerland belong to this group. Individuals in the next group of countries have capabilities that correspond approximately to the capabilities of the average European. The capability set evaluation metric of these countries is close to one. These are: Austria, Germany, Malta, Spain, Cyprus and Iceland. Note that most of the Western and Northern European countries belong to the first and second group. The third cluster of countries consists of Slovenia, Portugal, Italy, Estonia and Czechia. Individuals living in these countries have smaller capability sets as compared to the capability set of the average European. Specifically, the capability set of the average European is between 3.7 (Slovenia) and 10.7 per cent (Czechia) larger. Individuals in the last group of countries have capabilities that are even more below the European average. These countries are mostly situated in Central and Eastern Europe: Poland, Greece, Slovakia, Latvia, Hungary, Bulgaria, Romania and Serbia. These large disparities in capabilities across European countries indicate that country membership is an identification variable with significant discriminatory power in a pan-European setting.

4.4.2. Comparison of the multilateral and bilateral evaluation outcomes

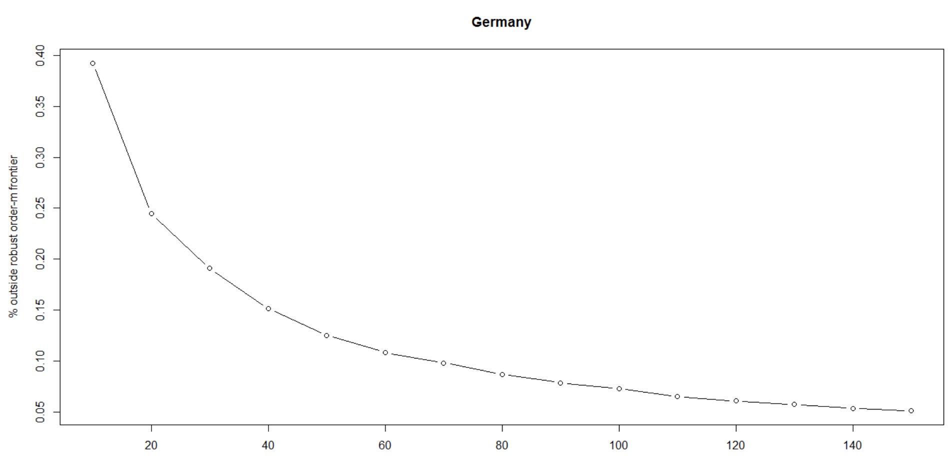
In this section, we compare the capability sets of the French and the Germans, using both the multilateral evaluation metric $SE_{\mu}\left(\hat{\Psi}_{m}^{Germany},\hat{\Psi}_{m}^{France}\right)$ as in (6) and the bilateral evaluation metric $\overline{SE}\left(\hat{\Psi}_{m}^{Germany},\hat{\Psi}_{m}^{France}\right)$ as in (7). It is important to re-emphasize that the data on the functioning dimensions that were used to estimate the French and German capability sets differ from the data in the previous section, as both datasets are derived from different hedonic regressions. Figures 4.4-4.5 plot the percentage of individuals that lie outside their own robust group capability set frontiers for different values of *m* for France and Germany, respectively. The plots indicate that a highly similar percentage of individuals lie outside the estimated French and German robust group capability sets for the same values of *m*. This facilitates the interpretation of the set evaluation outcomes across both countries. We again set the *m*-values at 70, corresponding with group capability sets that include approximately 90 per cent of the observed functioning bundles (90.5 per cent for France and 90.2 per cent for Germany).

Figure 4.4 – The relationship between *m* and the percentage of observations outside the estimated group capability set (France)



Source: EU-SILC (2013) cross-sectional data, author's computation.

Figure 4.5 – The relationship between *m* and the percentage of observations outside the estimated group capability set (Germany)



Source: EU-SILC (2013) cross-sectional data, author's computation.

Table 4.6 gives the outcome of the bilateral and multilateral set evaluation exercises. The results indicate that the Germans are slightly better off than the French, though the differences are almost negligible. Specifically, the German capability set is 1 per cent larger than the French capability set, as measured through the bilateral, "multiple rays" evaluation metric. When the multilateral, "average ray" evaluation metric is employed, the Germans have a capability set that is 0.6 per cent larger than the capability set of the French.[79] Thus, taking diversity into account by evaluating sets along multiple rays increases the estimated French-Germans differences in capabilities, though the differences remain almost negligible. The evaluation outcomes under the multilateral and bilateral evaluation metrics are highly similar.

Table 4.6 – Multilateral and bilateral set evaluation for France and Germany

$\overline{\overline{SE}}\left(\hat{\Psi}_{70}^{Germany},\hat{\Psi}_{70}^{France}\right)$	$SE_{\mu}\left(\hat{\Psi}_{70}^{Germany},\hat{\Psi}_{70}^{France}\right)$	**% outside estimated set (France)**	**% outside estimated set (Germany)**
1.010	1.006	9.51%	9.83%

Source: EU-SILC (2013) cross-sectional data, author's computation.

That being said, the fact that both metrics lead to similar evaluation outcomes does not imply that the choice for a particular ray does not have an impact on the comparisons of sets. Table 4.7 gives the minimal value, the first quartile, the third quartile and the maximal value of the distances between the boundaries of the estimated capability sets along the rays that go through the functioning bundles of the French and the Germans. The resulting distance measures vary between 0.989 and 1.09. Thus, depending on the ray that is used to evaluate sets, one obtains diverging results. This stresses the fact that is important to motivate the choice for a particular ray used in the capability set evaluation. It also shows that our averaging approach not only fully gauges the diversity in options that are available to sets, but also leads to a robust evaluation.

[79] This result diverges from the multilateral set evaluation framework presented in the previous section, where the estimated German capability set was smaller than the French one. To interpret this result, it is important to note that the French-German capability sets are estimated from different datasets. In addition, different rays were employed to evaluate the estimated sets (i.e. a pan-European average ray versus country-specific French-German rays).

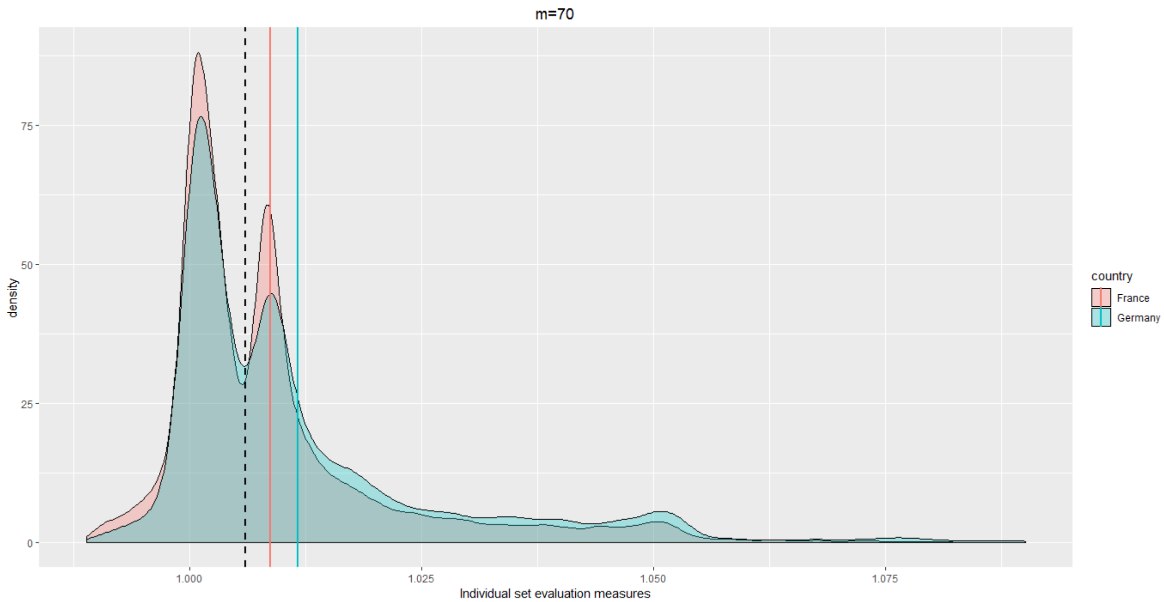
Table 4.7 – Distances between frontiers along multiple rays for France and Germany

$\underset{i \in g_1, g_2}{\text{Min}} \left(\frac{\hat{\theta}_i^{70,g_1}}{\hat{\theta}_i^{70,g_2}} \right)$	$\underset{i \in g_1, g_2}{\text{Q}_1} \left(\frac{\hat{\theta}_i^{70,g_1}}{\hat{\theta}_i^{70,g_2}} \right)$	$\underset{i \in g_1, g_2}{\text{Q}_3} \left(\frac{\hat{\theta}_i^{70,g_1}}{\hat{\theta}_i^{70,g_2}} \right)$	$\underset{i \in g_1, g_2}{\text{Max}} \left(\frac{\hat{\theta}_i^{70,g_1}}{\hat{\theta}_i^{70,g_2}} \right)$
0.989	1.001	1.013	1.090

Source: Calculations from EU-SILC (2013). *Note*: g_1 represents Germany and g_2 represents France.

Figure 4.6 gives the kernel distributions of the distances between the boundaries of the estimated capability sets along the rays that go through the functioning bundles of the French and the Germans. The set evaluation outcomes of the rays that go through the functioning bundles of the French and Germans are reflected, respectively, in red and blue. The dotted line, the solid red line and the solid blue line give, respectively, the outcome of the multilateral evaluation exercise, the outcome of the bilateral evaluation metric that only uses rays that go through the observed functioning bundles of the French, and the outcome of the bilateral evaluation metric that only uses rays that go through the observed functioning bundles of the Germans. The capability set evaluation is relatively more favourable towards a certain group if one uses the rays that go through the observed functioning bundles of the individuals that belong to this group. As can be deduced from the red and blue lines, the outcome of the bilateral evaluation metric that only uses rays that go through the observed functioning bundles of the French is comparatively more favourable to the French, as compared to the outcome of the bilateral evaluation metric that only uses rays that go through the observed functioning bundles of the Germans.

Figure 4.6 – Kernel densities of individual set evaluation measures for *m* = 70



Source: EU-SILC (2013) cross-sectional data, author's computation.

The sensitivity of the results in both sections to the size of *m* are presented in Appendix 4. An important conclusion of this sensitivity analysis is that changing the *m*-parameter has virtually no influence on the rankings induced by the set evaluation metrics.

4.5. Conclusion

While the theoretical literature on the ranking of capability sets has advanced considerably in recent decades, empirical applications of this literature are virtually nonexistent due the fact that capability sets of individuals cannot be observed. This has been a longstanding difficulty in the operationalization of the capability approach. The first novelty of this chapter is that a method is proposed to non-parametrically estimate capability sets on the basis of observed functioning bundles of group members. A key assumption behind our approach is that group members have the same capabilities, i.e. they share a capability set. We used a non-parametric frontier estimator to estimate these group capability sets by enveloping the observed functioning bundles of group members (Charnes et al., 1978). Our approach is data-driven and makes only minimal assumptions on the capability sets we which to estimate (i.e. we consider capability sets that are compact, convex and comprehensive). The estimator is also robust to the influence of outliers (Daraio & Simar, 2007). We are aware that the realism of our approach hinges strongly on the availability of good identification variables that can be used to define groups. Good identification variables are exogenous characteristics that strongly influence an individual's ability to generate capabilities. If groups are defined on the basis of such variables, then it becomes more realistic to assume that highly similar individuals could attain the same functioning achievements. The identification power of the variables that are used to define groups can be empirically assessed by comparing differences in the estimated group capability sets. If large between-group differences in capabilities exist, then this provides an indication that the identification variables that are used to construct groups are influential in determining the capabilities of a group.

We presented a tangible methodology to compare the estimated capability sets. Specifically, we operationalized Muellbauer's (1987) suggestion to evaluate sets by estimating the distance(s) between the boundaries along rays that have a connection to the distribution of the population. We presented two metrics to evaluate sets. The first one evaluates capability sets along a ray that represents the average functioning achievement across groups. This metric is transitive and allows for a complete multilateral ranking and comparison of the estimated capability sets. The second metric evaluates capability sets by averaging the distances between their boundaries along multiple rays that go through the observed functioning bundles of group members. This metric takes into account the diversity in options offered by sets, but can only be used for bilateral comparisons.

We illustrated our approach on the EU-SILC 2013 cross-sectional dataset. We considered four well-being functioning dimensions in the analysis: household disposable income, material living well-being, housing quality and health. We conducted two series of analyses. The first analysis uses the multilateral set evaluation metric to benchmark the capability sets of European countries vis-à-vis a pan-European one, for which no identification variables were used in the definition of the group. The results indicate that there exist strong differences in capability sets across European countries. More in particular, we found strong differences in capability sets between Western and Northern European countries and Southern, Central and Eastern European countries. In the second analysis, we compared the capability sets of the French and Germans using both the bilateral and multilateral set evaluation outcome. The results from the bilateral evaluation metric indicate that the evaluation outcome (strongly) depends on the selection of the ray that is used to evaluate sets. This shows that it is important to motivate the choice for a particular ray in the evaluation of sets. However, averaging the distance measures along the rays that go through the observed functioning bundles resulted in an evaluation outcome that was highly similar to the one obtained through the multilateral "single, average ray" evaluation metric. This is an indication that the multilateral evaluation metric may be preferable over the bilateral evaluation metric in some empirical settings, given that the bilateral evaluation metric is computationally heavy to calculate for empirical settings with large amount of observations and that it is unable to deliver a complete multilateral ranking and comparison of the estimated capability sets.

We compared country-specific capability sets in our illustration. This approach is useful if one wants to obtain a Bird's-eye view on differences in capability sets between-countries. However, such average measures miss the inequalities between-groups (i.e. inequalities within countries). A more comprehensive approach would be to consider inequalities between-countries jointly with inequalities within-countries. Archetypical sociological characteristics that come to mind in defining groups are generation, gender, migration background, social stratification etc. As pointed out in the influential Stiglitz-Sen-Fitoussi (2009) report, accounting for differences in capabilities between groups is necessary to fill the gap between country-wide estimates and people's feelings about their own conditions. The approach presented in this chapter can be employed to do exactly this. One could benchmark the capability sets of groups against a 'neutral group' that doesn't use any characteristics in its definition (similar to the pan-European group in our illustration) or a group that is considered the best off. Another potentially interesting avenue for future research is the extension of the framework presented in this chapter to an inter-temporal setting. This would allow to account for changes in groups' estimated capabilities over time.

Using the distance from the origin along a particular ray coincides with evaluating sets by their best element according to Leontief preferences. One can think of several other measures that can be employed to compare the estimated capability sets. The idea of using distance measures to evaluate capability sets has more recently been echoed in several theoretical contributions (Gaertner & Xu, 2006, 2008, 2011; Geartner, 2012; Farina et al., 2004). We leave the integration of such alternative evaluation frameworks within our setting as an interesting avenue for further research. However, we would like to point that the "directed cone" set ranking measure of Gaertner and Xu (2011) can be readily empirically operationalized through the so-called directional Data Analysis Envelopment (DEA) model. The authors provide an axiomatic characterization of the ranking induced by the maximal Euclidian distance from a reference functioning bundle/point of orientation to the boundary of the capability set within a directed cone (which can be shrunk in a single direction). One needs only to make minor adjustments to the framework proposed in this chapter to arrive at their setting (i.e. replacing the assumption of comprehensiveness upheld in this paper with the star-shaped upheld in theirs, applying a unity-normalization to ensure a measurement scale common to all functionings, ensure that the directed cone reduces to a line segment or multiple line segments).

Appendix 4.

This appendix discusses the sensitivity of the results to the size of *m*. Table 4.A1 gives the results from the multilateral set evaluation outcome in the multilateral setting (Section 4.4.1) for different values of *m*. Table 4.A2 gives the outcome of the bilateral and multilateral set evaluation exercises (Section 4.4.2) for different values of *m*. Table 4.A3 gives the minimal value, the first quartile, the third quartile and the maximal value of the distances between the boundaries of the estimated capability sets along the rays that go through the functioning bundles of the French and the Germans (Section 4.4.2) for different values of *m*. Figures 4.A1-4.A3 give the kernel distributions of the distances between the boundaries of the estimated capability sets along the rays that go through the functioning bundles of the French and the Germans (Section 4.4.2) for different values of *m*.

Table 4.A1 – Multilateral set evaluation (for different *m*-values)

Country	***m* = 20**		***m* = 40**		***m* = 70**		***m* = 150**	
	Set evaluation	**% outside frontier**	**Set evaluation**	**% outside frontier**	**Set evaluation**	**% outside frontier**	**Set evaluation**	**% outside frontier**
Belgium	0.946 (5)	24.16%	0.96 (5)	14.33%	0.974 (5)	9.11%	0.98 (4)	5.04%
Bulgaria	1.535 (31)	23.73%	1.418 (31)	13.96%	1.353 (31)	9.19%	1.274 (31)	4.93%
Czechia	1.165 (22)	23.33%	1.13 (22)	13.69%	1.107 (23)	9.01%	1.073 (23)	4.77%
Denmark	0.941 (4)	26.61%	0.957 (4)	15.85%	0.972 (4)	10.12%	0.981 (5)	5.40%
Germany	0.994 (13)	22.73%	0.994 (13)	13.91%	0.999 (13)	8.79%	0.998 (12)	4.55%
Estonia	1.171 (23)	21.73%	1.136 (23)	12.81%	1.101 (22)	8.19%	1.071 (22)	3.85%
Ireland	0.939 (3)	26.28%	0.956 (3)	15.88%	0.966 (3)	10.68%	0.972 (3)	6.09%
Greece	1.29 (26)	25.51%	1.243 (26)	15.27%	1.207 (26)	9.83%	1.156 (27)	4.97%
Spain	1.008 (16)	25.15%	1.007 (16)	14.88%	1.009 (16)	9.56%	1.006 (17)	5.23%
France	0.961 (8)	24.33%	0.974 (8)	14.59%	0.981 (8)	9.19%	0.987 (8)	4.95%
Croatia	1.36 (29)	22.78%	1.29 (28)	13.94%	1.267 (28)	9.13%	1.21 (28)	5.00%
Italy	1.108 (21)	25.00%	1.081 (21)	15.47%	1.07 (21)	9.96%	1.047 (21)	5.25%
Cyprus	1.01 (17)	25.60%	1.01 (17)	15.07%	1.011 (17)	9.69%	1.006 (16)	5.19%
Latvia	1.356 (28)	21.69%	1.262 (27)	13.11%	1.219 (27)	8.28%	1.151 (26)	4.30%
Lithouania	1.242 (24)	21.93%	1.175 (24)	13.13%	1.14 (24)	7.98%	1.097 (24)	3.85%
Luxembourg	0.924 (2)	22.71%	0.936 (1)	14.40%	0.951 (1)	9.62%	0.958 (1)	5.27%
Hungary	1.408 (30)	24.01%	1.338 (30)	13.95%	1.291 (30)	9.08%	1.225 (30)	4.74%
Malta	0.998 (14)	20.98%	0.996 (14)	12.70%	0.999 (14)	8.18%	1 (15)	4.42%
The Netherlands	0.955 (7)	26.12%	0.967 (7)	15.84%	0.979 (7)	10.03%	0.986 (7)	5.49%
Autstria	0.99 (12)	23.60%	0.989 (12)	14.53%	0.998 (12)	9.44%	1 (13)	4.99%
Poland	1.272 (25)	22.62%	1.213 (25)	13.67%	1.171 (25)	8.87%	1.122 (25)	4.63%
Portugal	1.1 (20)	22.97%	1.069 (20)	13.77%	1.056 (20)	8.95%	1.034 (20)	4.82%
Romania	2.105 (33)	27.64%	1.931 (33)	16.85%	1.803 (33)	11.27%	1.645 (33)	5.96%
Slovenia	1.057 (19)	23.34%	1.043 (19)	14.65%	1.037 (19)	9.40%	1.026 (19)	5.02%
Slovakia	1.335 (27)	28.79%	1.295 (29)	17.50%	1.269 (29)	11.90%	1.221 (29)	6.33%
Finland	0.969 (9)	29.33%	0.979 (10)	17.90%	0.987 (10)	11.83%	0.994 (11)	6.78%
Sweden	0.969 (10)	25.14%	0.975 (9)	14.68%	0.985 (9)	9.39%	0.991 (10)	5.05%
United Kingdom	0.951 (6)	22.00%	0.963 (6)	13.45%	0.976 (6)	8.43%	0.983 (6)	4.51%
Iceland	1.018 (18)	24.97%	1.017 (18)	15.37%	1.015 (18)	9.73%	1.015 (18)	5.03%
Norway	0.915 (1)	25.16%	0.938 (2)	14.99%	0.955 (2)	9.54%	0.964 (2)	5.07%
Switzerland	0.972 (11)	25.40%	0.98 (11)	15.73%	0.987 (11)	10.29%	0.988 (9)	5.61%
Serbia	1.65 (32)	23.12%	1.529 (32)	13.66%	1.448 (32)	8.96%	1.372 (32)	4.63%
Pan-European	1 (15)	21.74%	1 (15)	13.25%	1 (15)	8.38%	1 (14)	4.40%

Source: Calculations from EU-SILC (2013). Note: the number between brackets ranks the country in terms of the set evaluation outcome. The order of comparison of the set evaluation metric $SE_{\mu}\left(\hat{\Psi}^{i}_{m},\hat{\Psi}^{j}_{m}\right)$ is as follows: the pan-European capability set ($\hat{\Psi}^{i}_{m}=\hat{\Psi}^{EU}_{m}$) is evaluated against the country-specific capability sets $\hat{\Psi}^{j}_{m}$ (j = 1,…,32).

Table 4.A2 – Multilateral and bilateral set evaluation for France and Germany (for different *m*-values)

	Bilateral	Multilateral	**% outside estimated set (France)**	**% outside estimated set (Germany)**
m=20	1.016	1.010	24.33%	24.47%
m=40	1.012	1.007	14.90%	15.14%
m=70	1.010	1.006	9.51%	9.83%
m=150	1.008	1.004	4.85%	5.14%

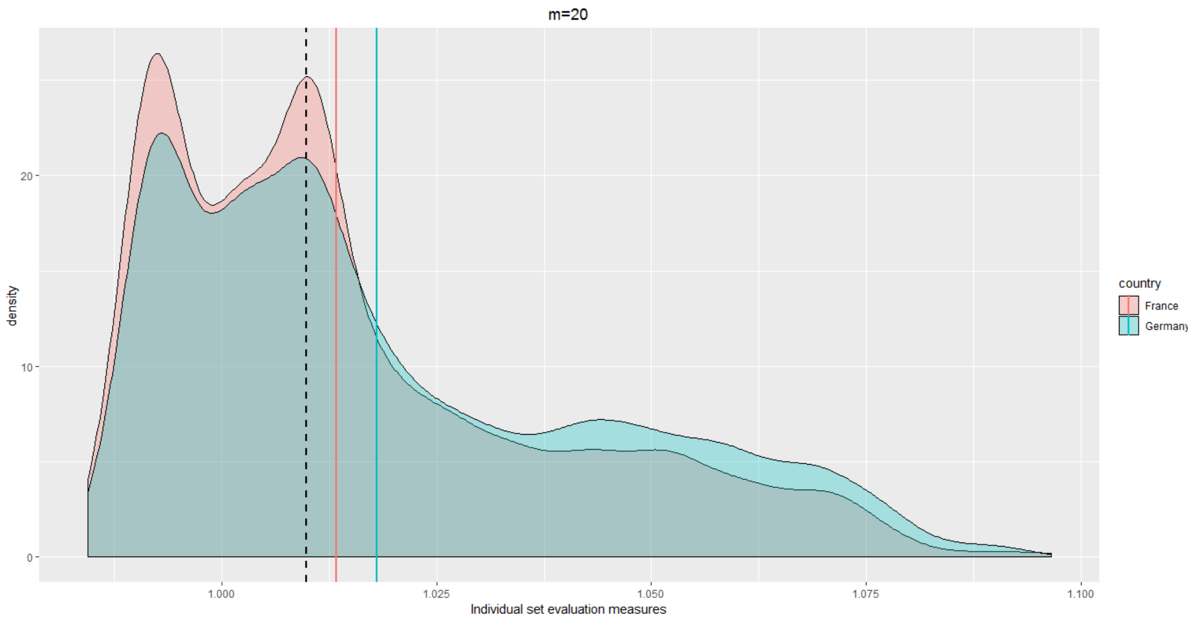
Source: EU-SILC (2013) cross-sectional data, author's computation.

Table 4.A3 – Distances between frontiers along multiple rays for France and Germany (for different *m*-values)

	$\underset{i\in g_1,g_2}{\text{Min}}\left(\frac{\hat{\theta}_i^{m,g_1}}{\hat{\theta}_i^{m,g_2}}\right)$	$\underset{i\in g_1,g_2}{\text{Q}_1}\left(\frac{\hat{\theta}_i^{m,g_1}}{\hat{\theta}_i^{m,g_2}}\right)$	$\underset{i\in g_1,g_2}{\text{Q}_3}\left(\frac{\hat{\theta}_i^{m,g_1}}{\hat{\theta}_i^{m,g_2}}\right)$	$\underset{i\in g_1,g_2}{\text{Max}}\left(\frac{\hat{\theta}_i^{m,g_1}}{\hat{\theta}_i^{m,g_2}}\right)$
m = 20	0.984	0.999	1.033	1.097
m = 40	0.987	1.000	1.019	1.095
m = 70	0.989	1.001	1.013	1.090
m = 150	0.985	1.002	1.012	1.089

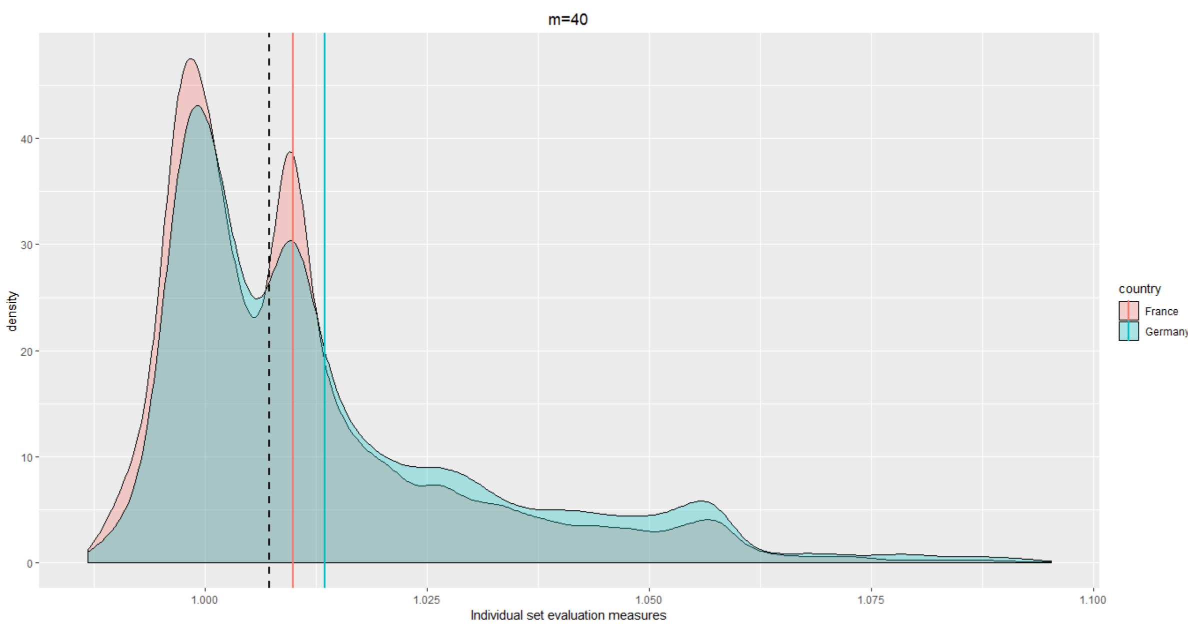
Source: EU-SILC (2013) cross-sectional data, author's computation. *Note*: g_1 represents Germany and g_2 represents France.

Figure 4.A1 – Kernel densities of individual set evaluation measures for *m* = 20



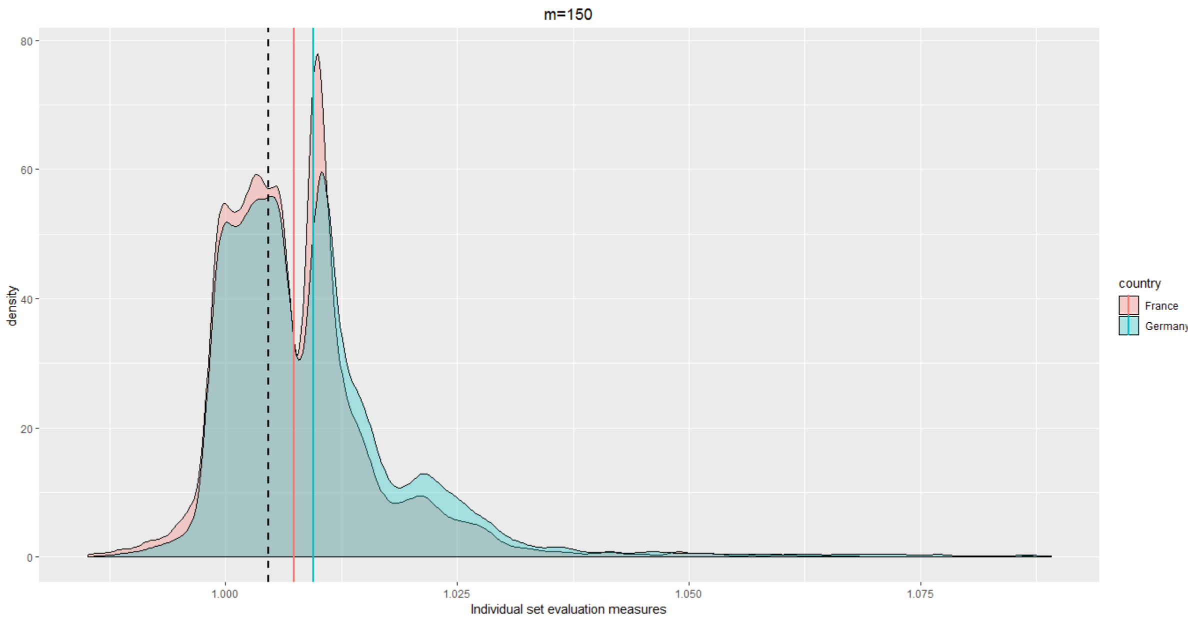
Source: EU-SILC (2013) cross-sectional data, author's computation.

Figure 4.A2 – Kernel densities of individual set evaluation measures for *m* = 40



Source: EU-SILC (2013) cross-sectional data, author's computation.

Figure 4.A3 – Kernel densities of individual set evaluation measures for m = 150



Source: EU-SILC (2013) cross-sectional data, author's computation.

The following conclusions can be derived:

The average difference between the percentage of individuals that lie outside their own country's capability set and the percentage of individuals that lie outside the pan-European capability set is around 2.6, 1.4, 1.1 and 0.6 percent for, respectively, m = 20, m = 40, m = 70, m =150. Thus, the deviation in outlying observation from the pan-European average decreases as m increases. This result is easily explained, as the percentage of outlying observations also becomes smaller as m increases.

The estimated between-country differences in capability sets become smaller for higher sizes of m. This result holds for the bilateral and multilateral set evaluation metrics (Table 4.A1 and 4.A2), but also for the individual set evaluation measures (Table 4.A3 and Figures 4.A1-4.A3). When the initial (i.e. for small sizes of m) between-country differences are large, the impact of increasing m on the set evaluation outcome is also large. For instance, the estimated differences in the Hungarian and pan-European capability set under m = 20 (1.408) are much larger than under m = 150 (1.225) (Table 4.A1). The impact is small when the initial differences are small. See, for instance, Malta, where the set evaluation metric is 0.998 and 1 for, respectively, $m = 20$ and $m = 150$ (Table 4.A1).

Overall, adjusting the size of m has virtually no impact on the rankings induced by the set evaluation (for both the multilateral and bilateral illustration). This is a key result and shows that the set evaluation outcomes are insensitive to the size of m.

General conclusion

Beyond GDP

> *"What we measure affects what we do. If we focus only on material wellbeing – on, say, the production of goods, rather than on health, education, and the environment – we become distorted in the same way that these measures are distorted; we become more materialistic."* - Joseph Stiglitz

The broader research objective pursued in this dissertation, viz., to contribute to a better measurement and understanding of social inclusion and social exclusion in the European Union, can be situated in the "Beyond GDP" movement. GDP, Gross Domestic Product, is the most widely used measure of economic activity. While GDP mainly measures market production, it often serves as a generic indicator of countries' societal progress as well. The perceived equivalence between economic production and societal success has been a subject of scrutiny for many decades. The main scope of moving beyond GDP is to complement GDP for measuring progress, wealth and well-being of nations. The "beyond GDP" movement started in the 1970s and it enjoyed an exceptional boom over the last decade (Fleurbaey & Blanchet, 2013, p. 1). The UNDP's Human Development Index, (a composite index of GDP per capita, life expectancy and education) that ranks countries into four tiers of human development since 1990, is the first major initiative to complement GDP per capita at the global international stage. In 2007, the European Commission, the European Parliament, Club of Rome, OECD and WWF hosted a conference titled "Beyond GDP". The consensus was to widen measures of economic growth and come up with measures that can inform policy making. Another key international initiative in the "Beyond GDP" movement was the Commission on the Measurement of Economic Performance and Social Progress, generally referred to as the Stiglitz-Sen-Fitoussi Commission after the surnames of its leaders. The Commission examined how the wealth and social progress of a nation could be measured, without relying on the unidimensional gross domestic product (GDP) measure. The final report by the Stiglitz-Sen-Fitoussi Commission was published in September 2009.

The Stiglitz-Sen-Fitoussi report outlines three main critiques on GDP as a measure of economic and social performance. The first line of critique is flagshipped under the term "classical GDP issues" and pertain to the fact that GDP does not adequately capture material living standards (e.g. GDP does not measure non-market activities, the difficulty of capturing quality change in the metric, etc.). The second line of critique on GDP is related to the fact that quality of life or well-being is something inherently multidimensional that goes beyond income. The final line of critique deals with the fact that sustainability issues are poorly addressed in a metric that only captures current economic production (e.g. externalities of production on the environment, the degradation of natural resources, etc.).

This dissertation exclusively focussed on the Report's second critique on GDP, i.e. its limited ability to measure quality of life. The main message is that a multidimensional definition has to be used when assessing well-being. The Report identified the following key dimensions that are missed by conventional income measures: (i) material living standards (income, consumption and wealth), (ii) health, (iii) education, (iv) personal activities including work, (v) political voice and governance, (vi) social connections and relationships, (vii) environment (present and future conditions), (viii) insecurity of an economic as well as a physical nature. All these dimensions shape people's well-being and should, ideally, be considered simultaneously.

In all the chapters of this dissertation, such a multidimensional perspective on the measurement of social inclusion (exclusion) is upheld. In Chapter 1, the drivers of income poverty and material deprivation, which represent two conceptually distinct ways of measuring social exclusion, were investigated jointly. In Chapter 2, the drivers of a new EU child deprivation indicator, which is a multidimensional construct consisting of 17 sub-indicators, covering both material and social aspects of deprivation, were studied. In Chapter 3, a composite indicator model was proposed to aggregate performances on sub-indicators. The approach was illustrated on nine commonly agreed EU indicators (period 2008–2013) from the overarching portfolio of social protection and social inclusion. In Chapter 4, a method was presented to operationalize the measurement and evaluation of capability sets. In its illustration, four well-being functioning dimensions were considered: income, health, housing quality and material living conditions. In this regard, it is important to note that the capability approach is inherently multidimensional and has, in fact, since its development innervated the growing acceptance of the multidimensional nature of well-being and deprivation within economics.

The Stiglitz-Sen-Fitoussi report (2009) issued five other recommendations, in addition to its advocacy for a shift towards a multidimensional perspective, to ameliorate the measurement of well-being. I believe that there exists a considerable overlap between four out of five of these *recommendations* and the research objectives pursued in this dissertation.[80] These are:

1) The assessment of links between various quality-of-life dimensions to ameliorate policies (Chapter 1 and Chapter 2).
2) Quality of life depends on people's objective conditions and capabilities (Chapter 4).
3) The development of single summary measures of quality-of-life (Chapter 3 and Chapter 4).
4) Inequalities should be addressed in a comprehensive way (Chapter 3 and Chapter 4).

These recommendations will therefore be used as a flagship to position the *contributions* of the different chapters within the literature and within the broader of research objective pursued in this dissertation of going beyond GDP as a metric to measure societal progress. Finally, the limitations of the chapter will be discussed and suggestions for *further research* will be presented.

The assessment of links between various quality-of-life dimensions to ameliorate policies

The report *recommends* that "it is critical to address questions about *how developments in one domain of quality of life affect other domains*, and how developments in all the various fields are related to *income*. This is important because the consequences for quality of life of having *multiple disadvantages far exceed the sum of their individual effects*. Developing measures of these cumulative effects requires information on the "joint distribution" of the most salient features of quality of life across everyone in a country through dedicated surveys. Steps in this direction could also be taken by including in all surveys some standard questions that allow *classifying respondents based on a limited set of characteristics*. When designing *policies* in specific fields, indicators pertaining to different quality-of-life dimensions should be considered jointly, to address the interactions between dimensions and the needs of people who are disadvantaged in several domains" (Stiglitz et al., 2009, p. 59).

[80] The fifth recommendation pertains to further use and development of subjective well-being data. Stiglitz and co-authors (2009, p. 58) argue that "quantitative measures of these subjective aspects hold the promise of delivering not just a good measure of quality of life per se, but also a better understanding of its determinants, reaching beyond people's income and material conditions." Throughout this dissertation, social inclusion (exclusion) is measured through objective measures. However, I would like to point out that subjective well-being measures (of health and satisfaction with the financial situation) were used in hedonic regressions to construct cardinal and objective functioning indicators in Chapter 4.

The *contribution* of Chapter 1 and Chapter 2 to the literature jointly lie in the research objectives and in the accompanying methodological strategy. I believe both go hand in hand with the abovementioned recommendation. The objective in both chapters was to obtain a better conceptual understanding of key EU social exclusion indicators. The methodological strategy to obtain a better understanding of these indicators was dual: the micro-drivers and macro-drivers of social exclusion were considered jointly in both a single level and multilevel regression framework.

The micro-drivers were investigated by regressing the relationship of certain individual or household-level characteristics with observed social exclusion outcomes of individuals in the *single level models*. In Chapter 1, variables that captured households' longer-term command over resources (i.e. education and work intensity of the household), households' needs (i.e. suffering from a bad health, renting) and socio-demographic variables (i.e. household structure and composition, young household health, migration background of the household) were considered. In Chapter 2, the household-level variables were similarly grouped into three clusters, i.e. variables related to the longer-term command over resources (i.e. household disposable income, parental education attainment, quasi-joblessness, migration status, self-employment), variables capturing needs (i.e. bad health, tenure status, suffering from a housing burden) and socio-demographic variables (i.e. age of the oldest child, single parenthood, number of dependent children). One of the key contributions of Chapter 2 was that it is explicitly argued why these have variables have a relationship with deprivation. In most papers on deprivation (including Chapter 1), the expectation that such "social stratification" variables are related to deprivation is taken for granted without further argument.

The methodological strategy of decomposing within-country and between-country explained variance measures in terms of the contribution of the regressed variables allowed to deliver a nuanced picture on several issues that are relevant for policy makers. Regarding within-country differences, the following key results were obtained. First, the comparison of the explanatory power of the model across countries provides an answer to questions on the effectiveness of the variables that are available in the EU-SILC dataset in explaining within-country differences in social exclusion (i.e. *how much do/don't we know?*). In Chapter 1, the model as a whole was able to explain a large share of the original unobserved within-country differences in the risk of consistent poverty (59 per cent, on average, across countries), while the explanatory power of the employed model was much smaller for the other categories of the dependent variable (36 per cent for the risk of ‘income poverty only’ and 32 per cent for ‘material deprivation only’, on average across countries). In Chapter 2, the model was the most effective in countries with the highest share of child deprivation. The explanatory power was low in countries with a high occurrence of child deprivation.

Second, the comparison of the relative contributions of the different independent variables to the overall explained within-country fit measures allowed the comparison of the extent to which the regressed socio-economic characteristics differ in their explanatory power (i.e. *which variables explain how much of what?*). In Chapter 1, the estimation results revealed that a household’s short term ability to generate resources on the labour market, as measured by its work intensity, is the most strongly related to income poverty, whereas a household’s long term ability to generate resources on the labour market, as measured by the educational attainment of its members, was found to be more strongly related to material deprivation. It was further found that variables that capture costs play an important role in explaining the risk of material deprivation, but are rather unimportant for the risk of income poverty. The socio-demographic variables were found to have a moderate, yet non-negligible relationship with both income poverty and severe material deprivation. In Chapter 2, the results confirm the combined relationship of variables related to the longer-term command over resources and variables indicating household needs with child deprivation. On average across countries, variables related to resources and variables related to needs were estimated to have, respectively, a relative contribution of 55 and 38 per cent to the fit measure (the socio-demographic variables have a remaining contribution of 7 per cent to the fit measure), respectively. The most important variables were household income, the housing cost burden dummies and the educational level of parents with, respectively, a relative share of 25.3, 24.7 and 15.3 per cent of the explained fit.

Third, the comparison of the explanatory power of the different variables across countries allowed for the comparison of national specificities in the predictors of social exclusion (i.e. *how do the results differ across countries?*). In Chapter 1, a large cross-country variation in the within-country explained variance measure was found in the ‘material deprivation only’ category, with a high (low) effectiveness of the model in Western and Northern (Southern, Central and Eastern) European countries. Results also showed that a household’s work intensity and education have a large relative contribution to the explained within-country variance measures in Central and Eastern European countries, whereas their explanatory power is much more limited in Western and Northern European countries. In addition, it was found that households’ socio-demographic characteristics and costs have a much stronger relationship with social exclusion in Western and Northern European countries than in most Central and Eastern European countries. In Chapter 2, it was similarly shown that the explanatory power of the household-level variables differs strongly across countries. In the richest countries, the explanatory power of variables related to household needs is the largest, whereas in the most deprived countries, the explanatory power of resources is generally greater (with the exception of debt and migration). This means that countries not only differ in terms of socio-economic composition (as stated in most papers explaining differences in deprivation between countries), but also in terms of the association of each variable with the child deprivation risk. These patterns were confirmed in a multilevel model that included cross-level interactions with GDP per capita and the household-level variables.

The single level models were complemented with *multilevel models* with the aim of investigating the explanatory power of country-level variables on unobserved between-country differences in social exclusion. Country-level variables of a macroeconomic nature and of an institutional nature were jointly regressed. Both chapters consider institutional variables that are new to the multilevel social exclusion literature (i.e. the in-kind versus in-cash dichotomy within social spending, the pro-poorness of cash transfers). However, given the exclusive and specific focus on child deprivation in Chapter 2, more institutional variables (social spending levels expressed as a percentage of GDP per capita or in PPS head, family social spending levels, adequacy of minimum income schemes) were regressed in this chapter. Another key difference between Chapter 1 and Chapter 2 is that household income is not included as an independent variable at the micro-level, whereas it was in Chapter 2. As argued in more detail in Chapter 2, the inclusion of household income has important consequences on the interpretation of the relationship between the country-level variables and child deprivation.

The models in Chapter 1 and Chapter 2 were found to be quite effective in explaining *between-country differences* in deprivation, while the explanatory power of the regressed variables are more limited for the risk of income poverty. Specifically, in Chapter 1, around 80 per cent of the between-country differences in the risk of material deprivation ('only' and 'consistent poverty'), and about half of the between-country differences in the risk income poverty 'only' are explained by the model. In Chapter 2, around 80 per cent of the between-country differences in child deprivation were explained by the model.

Several *significant relationships* of country-level variables with the risk of social exclusion were uncovered. First, it was shown that it is, in contrast to the usual approach in the multilevel poverty and deprivation literature, important to distinguish between cash and in-kind social spending. Specifications which combine both cash and in-kind transfers into a single global variable may result in (in)significant relationships, which can become no longer (in)significant if cash or in-kind transfer levels are regressed separately. For instance, in Chapter 1, it was found that cash benefits are negatively associated with the risk of 'income poverty only', while in-kind benefit levels are negatively associated with all categories of the dependent variable. In Chapter 2, in-kind transfers were found to be a crucial explanatory variable of child deprivation. Cash transfers level were only significant once household income was omitted from the model at the micro-level. It was argued that the cushioning effect of cash transfers operates through household income (which consist both of market income and social transfers). Second, in Chapter 1, it was shown that living in a country in which social benefits are strongly targeted towards the poor significantly reduces the risk of income poverty and the risk of material deprivation (but only for those individuals that also have a low income). In Chapter 2, it was found that the pro-poorness of cash transfers is negatively related to child deprivation, even after controlling for household income at the micro-level and GDP per capita at the country-level, though its explanatory power was found to be rather limited. Third, variables capturing aggregated levels of affluence were crucial in explaining the risk of deprivation. In both chapters and in all specifications, they were the most important country-level variable. In Chapter 1, median income levels explained almost 40 per cent of between-country differences in material deprivation. In Chapter 2, GDP per capita and household income levels included at the micro-level explained around half of the between-country differences. The significant, negative relationship of GDP per capita, after controlling for individual household income, with child deprivation was given considerable attention. It was argued that GDP per capita correlates with "hidden" contextual factors, which are not available in the EU-SILC dataset and can be flagshipped under the term "level of social development" of societies.

I now discuss the *limitations* of Chapter 1 and Chapter 2 and propose some suggestions for further research that can amend some of these limitations. As pointed out in the conclusion of Chapter 1, one should be reluctant to draw policy conclusions from a cross-sectional data setting. Significant relationships in a cross-sectional setting do not necessary imply causality. Think of, for instance, the positive relationship between having children (as compared to having none) and the risk of income poverty and/or material deprivation in Chapter 1. It was argued that this family type suffers from a higher risk of social exclusion because less resources are available for consumption.

However, this relationship may also reversed: households that have a higher social exclusion risk (because they belong to a lower social strata), may very well (opt to) have more children. The determination of the risk of social exclusion could be improved by moving towards a *dynamic model*. The main advantage of panel data is that time and country fixed effects can be controlled for. In this regard it is important to note that, while there are multiple macro-level panel studies that account for income poverty rates, no macro-level panel studies that account for material deprivation rates exist, to the best of my knowledge, in the current literature. At the micro-level, panel studies found interesting relationships between current income, permanent income, deprivation and the manner in which resources are accumulated and eroded (see, for instance, Berthoud & Bryan, 2011). I believe a macro-level panel perspective on material deprivation is an equally interesting research setup. In particular, it would be interesting to verify how changes in the level of affluence of countries or countries' social spending levels result in changes in material deprivation rates. However, as pointed out by Brady and co-authors (2009), macro-level studies may suffer from a black-box problem of causal inference because, unlike the multilevel studies, micro-level mechanisms are unobserved. Dynamic multilevel studies take this issue into account by adding a time dimension to the multilevel setup (for a first dynamic logistic multilevel study on income poverty, see Bosco & Poggi, 2016; for a first dynamic probit multilevel study on child poverty, see Bárcena-Martín et al., 2017a). Though such models are econometrically complex, they offer in a sense "the best of both worlds" by combining both household-level and country-level variables in a model that includes a time component. As argued in the conclusion of Chapter 1, in a dynamic multilevel context, it could prove interesting to analyse how changes in both country-level (e.g. economic crisis, changes in the pro-poorness of social spending) and household-level (e.g. moving into unemployment, divorce) variables over time affect the probability of moving in and out of material hardship and/or income poverty.

Another limitation lies in the *choice of models* that were used to predict the risk of social exclusion. In Chapter 1, a multinomial logistic regression model was employed. This model has been used to study the differences in the risk of income poverty and severe material deprivation in single level settings (Fusco et al., 2011). In Chapter 2, a negative binomial regression model was employed. The multilevel versions of both models are new to the social exclusion literature. They are appealing, as they take, respectively, the nominal and count nature (and the issue of over-dispersion) of the dependent variable into account. A downside lies in the fact that they are computationally heavy to estimate and that sampling weights could not be included in the estimation in the software programmes that were employed to estimate the results (i.e. both Stata and MLwiN). Hopefully, in the future this feature will become available for researchers in the abovementioned programmes. The main advantage of the multinomial logistic regression model is that it allows to account for the accumulation of disadvantages in terms of income poverty and severe material deprivation. However, a key limitation of this model is that it relies on the assumption of independence from irrelevant alternatives. This assumption states that the odds of occurrence of one nominal outcome over another does not depend on the presence or absence of other "irrelevant" alternatives (e.g. the relative probabilities of being ‘income poor only’ or ‘materially deprived only’ do not change if being ‘consistently poor’ is added as an additional outcome). A way to mitigate this rather restrictive assumption is to allow for correlation between the error terms across (social exclusion) outcome categories. An alternative model that presents itself is a (multilevel) bivariate probit model. This model allows, similar to the multinomial logistic regression model, for the joint estimation of the outcomes of income poverty and severe material deprivation. However, the bivariate probit model differs from the multinomial model in that it uses a latent variable structure to estimate two correlated outcomes jointly, instead of estimating the nominal outcome categories separately. In such a setting, it would be interesting to analyse the impact of the regressed variables on the estimated residuals of income poverty and severe material deprivation outcomes and their joint covariance (following a similar strategy pursued in Chapter 1). In Chapter 2, a negative binomial multilevel regression model was employed. The negative binomial model takes both the count nature and the issue of over-dispersion of the dependent variable into account. However, a zero inflated model presents itself as an interesting alternative. Zero inflated Poisson models, which can be either of a traditional Poisson nature or of a negative binomial nature, are designed to account for frequent zero-valued observations, as Poison or negative binominal models may underestimate these observations.[81] The appropriateness of the zero inflated models were tested in the single level setting as a robustness check. It was found that zero inflated models perform slightly better, but the gain in fit was not major. The zero inflated models were very difficult to estimate in the multilevel setting. The comparison of both types of models are left as a scope for further research.[82]

Next, Chapter 2 showed that the relationship of GDP per capita, or rather the “*overall level of societal development*” that it captures, with material deprivation needs further elaboration. It was

[81] Zero-inflated model use a two-step estimator. The first step consists of a logistic model modelling the probability of a zero versus a positive value. The second step models the count using either a Poisson or a negative binominal model.
[82] Another alternative count regression model is the hurdle model. See Notten and Guio (2018) on why this model might not be appropriate for studying deprivation.

argued that GDP per capita correlates with "hidden" contextual factors, which can be household wealth, between-households support in-kind, the quality and affordability of education, childcare, healthcare and public transport systems. The disentanglement of these factors and their relationship with deprivation needs further research. Some of these factors can be investigated by, for instance, testing the significance and explanatory power of some country-level variables that were not regressed in Chapter 2 (e.g. in-kind (non-social) public services such as education, public transport). That being said, it is important to be transparent about the limitations of the multilevel framework. One such limitations is that the multilevel were not appropriate to regress a large amount of country-level variables, because of the presence of only a limited amount of observations at the country-level. Other relationships are more difficult to investigate due to data limitations (e.g. how can one measure the quality of public services?). Finally, I would like to point out that new data initiatives on the development of the joint distributions of income, consumption and wealth are promising in this regard, and may help to provide answers to some of the open research questions and hypotheses presented in Chapter 2.

Finally, it could prove interesting to re-conduct the analysis in Chapter 1 on more recent editions of the EU-SILC dataset. This is particularly relevant as the European Union adopted *a new indicator of "material and social deprivation"* in March 2017. This measure was developed by Guio and co-authors (2012, 2017) and covers the entire population of the 28 EU Member States. It includes 13 deprivation items and replaces the 9-item "standard" material deprivation index adopted in 2009. It would be interesting to analyse whether the results in Chapter 1 hold for the new deprivation index. It could also prove interesting to investigate whether the insignificant relationship of cash transfers with the risk of deprivation, as found in Chapter 1, still holds if pensions are included in the cash transfer concept.

Quality of life depends on people's objective conditions and capabilities

The Stiglitz-Sen-Fitoussi report *recommends* the Capability Approach as "useful in thinking about how to measure quality of life" (Stiglitz et al., 2009, p. 42). The authors further state that "the information relevant to valuing quality of life goes beyond people's self-reports and perceptions to include measures of their functionings and freedoms. In effect, what really matters are the capabilities of people, that is, the extent of their opportunity set and of their freedom to choose among this set, the life they value" (Stiglitz et al., 2009, p. 15).

The main *contribution* of Chapter 4 is that a method is proposed to non-parametrically reconstruct group capability sets on the basis of group members' observed functioning bundles. The unobservability of capability sets has been a longstanding difficulty in the empirical operationalisation of the capability approach. The key assumption behind the approach is that group members have the same capabilities, i.e. they share a capability set, a group capability set. I recognize that this assumption is rather strong and that it is contestable, both on a conceptual and an empirical level. That being said, it is common practice in economics to make simplifying assumptions about the world. As nicely formulated by Sen, there is always a trade-off between relevance and usability of well-being measures:

> There are two major challenges in developing an appropriate approach to the evaluation of the standard of living. First, it must meet the motivation that makes us interested in the concept … doing justice to the richness of the idea. It is an idea with far-reaching relevance, and we cannot just redefine it in some convenient but arbitrary way. Second, the approach must nevertheless be practical in the sense of being usable for actual assessments ... This imposes restrictions on the kinds of information that can be required and the techniques of evaluation that may be used.
>
> These two considerations – relevance and usability – pull us, to some extent, in different directions. Relevance may demand that we take on board the inherent complexities of the idea of well-being as fully as possible, whereas usability may suggest that we try to shun complexities if we reasonably can. Relevance wants us to be ambitious; usability urges restraint. This is, of course, a rather common conflict in economics, and while we have to face the conflict squarely, we must not make heavy weather of it (Sen, 1987, p. 20).

I believe the proposed methodology is flexible in that researchers themselves can decide on the 'right' balance between relevance and usability. The proposed approach gains relevance by employing good identification variables (that are known to determine individuals' capabilities) in the definition of groups. If groups are defined on a lot of such characteristics, then the assumption that the estimated group capability set truly corresponds with a group member's actual capability set becomes realistic. Think, for instance, of a group capability set that is estimated from the observed functioning bundles of identical twins. However, relevance comes at a cost in terms of usability. By making groups more homogeneous (by applying a stricter group definition), one also increases the total amount of groups under evaluations and reduces the average number of group members across groups. Policy maker are most often only interested in well-being comparisons of some key sociological groups (e.g. based on ethnicity, gender, social stratification). It is not really relevant for policy makers to know whether identical twins Emma and Maria are better off than identical twins Kevin and Jonathan.

But what are good identification variables? In Chapter 4 it was argued that "the identification power of the variables that are used to define groups can be empirically assessed by comparing differences in the estimated group capability sets. If large between-group differences in capabilities exist, then this provides an indication that the identification variables that are used to construct groups are influential in determining the capabilities a group can generate." In the illustration, groups are defined on the basis of country membership. Large disparities in capabilities across European countries, notably between Western and Northern European countries on the one hand and Southern, Central and Eastern European countries on the other hand, were found. Country membership can thus be considered as a strong identification variable in a pan-European setting.

There exists considerable scope for *further research* on the relationship between exogenous characteristics of individuals and the extent of their capability sets. I believe the so-called conditional order-*m* Data Envelopment Analysis model is particularly appropriate to do this within the non-parametric frontier estimation framework (Bădin et al., 2000; Bădin, et al., 2002). In the framework presented in Chapter 4, the identification variables define groups in a rather radical way. Group capability sets are constructed from the observed functioning bundles of individuals, but only if these individuals strictly satisfy the identification criteria (e.g. you are a group member if you are young and female). In reality, such exogenous characteristics influence the ability of individuals to generate capabilities, but in a much more smooth way (e.g. being young and female has an influence on your ability to generate capabilities, but are these characteristics strong enough to box someone into a group?). The basic idea behind the conditional order-*m* DEA is that in the estimation of an individual's capability set, observed functioning bundles of individuals that share similar exogenous characteristics are sampled from the full dataset. If one then similarly estimates an individual's capability set from the observed functioning bundle of individuals that do not necessary share similar exogenous characteristics (i.e., i.i.d. drawn), one could estimate the distance between a capability set frontier that is conditional on certain exogenous characteristics and a capability set frontier that is unconditional on exogenous characteristics. A next step could be to non-parametrically regress the impact of these exogenous variables to explain differences in capability sets that are shaped by exogenous variables and capability sets that are not. This would give policy makers and researcher useful insights on the influence of exogenous variables on individuals' abilities to generate capabilities. Finally, I believe it would be interesting to compare the results from these non-parametric regressions with results obtained from parametric Structural Equation Models that are currently being used in the literature to estimate capabilities.

The development of single summary measures of quality-of-life

Another *recommendation* in the Stiglitz-Sen-Fitoussi (2009, p. 59) report is to "construct different scalar indexes that aggregate across quality-of-life dimensions." Even though the authors are critical about the perceived status of summary measures as the "holy grail of all efforts to go beyond conventional economic measures", they conclude that there "are strong demands to develop a single summary measure" (Stiglitz et al., 2009, p. 217 and p. 59). The authors further emphasize that this does not imply that researchers should be aspired to come up with a single, "one size fits all" aggregate measure. As mentioned in the introduction, aggregating the various aspects of well-being dimensions cannot be accomplished without value judgments that are necessarily controversial. Methods of aggregation crucially depend on the question one wishes to answer.

Chapter 3 and Chapter 4 *contribute* to the literature by presenting methodologies that embrace the multidimensional nature of social exclusion (inclusion) in its measurement. Both approaches share some similarities. First, the objectives of both techniques are similar in spirit in that they can be employed to make bilateral well-being comparison, i.e. comparing aggregated well-being (social inclusion, capability) levels of one entity versus another one. Both methods are illustrated on European social inclusion data, where countries' performances are compared to the European average. Second, both methods are inspired by the "Benefit-of-the-Doubt" composite indicator framework. In Chapter 3, the multiplier version of an alternative version of the traditional "Benefit-of-the-Doubt" model is used to compute shadow price weights of social inclusion indicators within a geometric composite index. In Chapter 4, the envelopment version of the "Benefit-of-the-Doubt" model is used to estimate group capability sets. The fundamental difference lies in the fact that Chapter 3 is concerned with creating a composite indicator of aggregated social inclusion performances. Chapter 4 uses individual-level social inclusion data to estimate group capability sets and deviates from the composite indicator approach in that relative performances in social inclusion are measured by the distance between the boundaries of two capability sets along a ray (multiple rays). Chapter 4 is, in fact, the first study that uses the "Benefit-of-the-Doubt" method on social inclusion data at the micro-level.

The report further mentions three limits on synthetic indices that aggregate across domains. First, by retaining the notion of a "representative agent", macro-level composite indicators cannot track the accumulation of disadvantages by certain subgroups. This issue can be taken into account within both frameworks and will be discussed in the following Section "inequalities should be addressed in a comprehensive way". A second limit is related to the choice of weights for various domains. It is argued that the "lack of diversity of viewpoints about the relative importance of various dimensions" is seen as a drawback in the traditional composite indicator approach (Stiglitz et al., 2009, p. 209). In Chapter 3, this limitation is taken into account by deriving both optimistic and pessimistic country-specific "Benefit-of-the-doubt"-derived shadow price weights. As argued in the introduction, the optimistic country-specific "Benefit-of-the-doubt" derived weights align well with the subsidiarity principle in the European Union social policy setting, whereas the pessimistic country-specific "Benefit-of-the-doubt" derived weights are more appealing from a normative (i.e., social justice) point of view. In Chapter 4, the arbitrariness in the evaluation exercise lies in the selection of the ray that is used to compare the capability sets. I believe this limit is taken into account by using a ray (multiple rays) that has (have) a connection to the distribution of the population in the evaluation of the capability sets. The bilateral evaluation framework is particularly appropriate to consider multiple viewpoint in the evaluation. A third limit mentioned in the Stiglitz-Sen-Fitoussi has to do with the interpretation of changes in these aggregate indicators. For instance, a concern raised regarding the HDI is that as time passes and the HDI is updated year to year, its movements have tended to be dominated by changes in the GDP component, at least for those developed countries (such as France and the United States) whose performance in the health and education domains is close to the top. The inter-temporal decomposition of the geometric mean quantity index number framework presented in Van Puyenbroeck and Rogge (2017) and further extended to a dual-weighting setting in Chapter 3 allows to account for such changes over time.

I believe there is scope for *further research* to take the last abovementioned issue in the report, i.e. the lack of interpretation of changes in the evaluation metric, into account within the framework presented in Chapter 4. As described in the conclusion of the fourth chapter, interesting research setups lie in the assessment of changes in groups' capability sets. I now present a proposal of such an inter-temporal capability evaluation framework. In an inter-temporal framework, the notation needs to be extended accordingly. In particular, I distinguish between an estimated capability set $\hat{\Psi}^{g_j,t}$ pertaining to period *t* and an estimated capability set $\hat{\Psi}^{g_j,t+1}$ pertaining to period *t+1*, the capability set evaluation metric $SE_{\mu_t}\left(\hat{\Psi}^{g_1,t},\hat{\Psi}^{g_2,t}\right)$ using the average ray μ_t pertaining to period *t* and the capability set evaluation metric $SE_{\mu_t}^{t+1}\left(\hat{\Psi}^{g_1,t+1},\hat{\Psi}^{g_2,t+1}\right)$ using the average ray μ_{t+1} pertaining to period *t+1*. Following the inter-temporal perspective upheld in Chapter 3, a natural measure of capability set change (*CC*) is based on the ratio of the capability set evaluation metric $SE_{\mu_t}^{t+1}\left(\hat{\Psi}^{g_1,t+1},\hat{\Psi}^{g_2,t+1}\right)$ and $SE_{\mu_t}\left(\hat{\Psi}^{g_1,t},\hat{\Psi}^{g_2,t}\right)$:

$$CC\left(SE_{\mu_t}^{t}\left(\hat{\Psi}^{g_1,t},\hat{\Psi}^{g_2,t}\right),SE_{\mu_t}^{t+1}\left(\hat{\Psi}^{g_1,t+1},\hat{\Psi}^{g_2,t+1}\right)\right)=\frac{SE_{\mu_{t+1}}^{t+1}\left(\hat{\Psi}^{g_1,t+1},\hat{\Psi}^{g_2,t+1}\right)}{SE_{\mu_t}^{t}\left(\hat{\Psi}^{g_1,t},\hat{\Psi}^{g_2,t}\right)}$$

$$CC\left(SE_{\mu_t}^{t}\left(\hat{\Psi}^{g_1,t},\hat{\Psi}^{g_2,t}\right),SE_{\mu_t}^{t+1}\left(\hat{\Psi}^{g_1,t+1},\hat{\Psi}^{g_2,t+1}\right)\right)=\frac{\dfrac{\hat{\theta}_{\mu_{t+1}}^{g_1,t+1}}{\hat{\theta}_{\mu_{t+1}}^{g_2,t+1}}}{\dfrac{\hat{\theta}_{\mu_t}^{g_1,t}}{\hat{\theta}_{\mu_t}^{g_2,t}}}$$

The capability change metric $CC\left(SE_{\mu_t}^{t},SE_{\mu_{t+1}}^{t+1}\right)$ quantifies the extent in capability set changes between *t* and *t+1*, with *CC*-values smaller (larger) than one reflecting a relative enlargement (contraction) of g_2's capability set vis-à-vis g_1's capability set. Finding inspiration in the inter-temporal framework of Van Puyenbroeck and Rogge (2017), this measure can be further decomposed in a capability set change component of the first group ΔCC^{g_1}, a capability set change component of the second group ΔCC^{g_2} and an evaluation change component related to the ray that was employed to evaluate the sets ΔEC:

$$CC\left(SE_{\mu_t}^{t}\left(\hat{\Psi}^{g_1,t},\hat{\Psi}^{g_2,t}\right),SE_{\mu_t}^{t+1}\left(\hat{\Psi}^{g_1,t+1},\hat{\Psi}^{g_2,t+1}\right)\right)=\sqrt{\left(\frac{\hat{\theta}_{\mu_t}^{g_1,t+1}}{\hat{\theta}_{\mu_t}^{g_1,t}}\right)\times\left(\frac{\hat{\theta}_{\mu_{t+1}}^{g_1,t+1}}{\hat{\theta}_{\mu_{t+1}}^{g_1,t}}\right)}\times\sqrt{\left(\frac{\hat{\theta}_{\mu_t}^{g_2,t}}{\hat{\theta}_{\mu_t}^{g_2,t+1}}\right)\times\left(\frac{\hat{\theta}_{\mu_{t+1}}^{g_2,t}}{\hat{\theta}_{\mu_{t+1}}^{g_2,t+1}}\right)}$$

$$\times\sqrt{\frac{\left(\frac{\hat{\theta}_{\mu_t}^{g_2,t}}{\hat{\theta}_{\mu_t}^{g_1,t}}\right)\times\left(\frac{\hat{\theta}_{\mu_t}^{g_2,t+1}}{\hat{\theta}_{\mu_t}^{g_1,t+1}}\right)}{\left(\frac{\hat{\theta}_{\mu_{t+1}}^{g_2,t}}{\hat{\theta}_{\mu_{+1}}^{g_1,t}}\right)\times\left(\frac{\hat{\theta}_{\mu_{t+1}}^{g_2,t+1}}{\hat{\theta}_{\mu_{t+1}}^{g_1,t+1}}\right)}}$$

$$CC\left(SE_{\mu_t}^{t}\left(\hat{\Psi}^{g_1,t},\hat{\Psi}^{g_2,t}\right),SE_{\mu_t}^{t+1}\left(\hat{\Psi}^{g_1,t+1},\hat{\Psi}^{g_2,t+1}\right)\right)=\Delta CC^{g_1}\times\Delta CC^{g_2}\times\Delta EC$$

The first component ΔCC^{g_1} evaluates shifts in g_1's capability set frontier over period *t* and *t+1*. If ΔCC^{g_1} is larger (smaller) than one, then the average distance between the origin and the intersection with the group capability set frontier of $\hat{\Psi}^{g_1,t+1}$ along the rays that go through μ_t and μ_{t+1} is larger (smaller) than the average distance between the origin and the intersection with the group capability set frontier of $\hat{\Psi}^{g_1,t}$ along the same rays. It can then be concluded that the capability set of g_1 has expanded (contracted) moving from period *t* to period *t+1*. If ΔCC^{g_2} is larger (smaller) than one, then the average distance between the origin and the intersection with the group capability set frontier of $\hat{\Psi}^{g_2,t+1}$ along the rays that go through μ_t and μ_{t+1} is smaller (larger) than the average distance between the origin and the intersection with the group capability set frontier of $\hat{\Psi}^{g_2,t}$ along the same rays. It can then be concluded that the capability set of g_2 has contracted (expanded) during period *t* and *t+1*. Note that for both components the effects of the ray that is used to evaluate sets is averaged out. The third component, ΔEC, neutralizes the effects of changes of the shifts in the frontiers to focus on the impact of changes in the ray that is used to evaluate the capability sets. If ΔEC is larger (smaller) than one, then this suggest that ray employed in period *t+1* (μ_{t+1}) is relatively less (more) favourable towards group g_2, as compared to the ray that was employed in period *t* (μ_t). Note that this decomposition can be easily extended to the bilateral evaluation framework in equation (7) of Chapter 4.

I believe this inter-temporal decomposition can lead to interesting applications and is particularly appropriate to monitor Member States' social inclusion performances. The illustration of this inter-temporal decomposition is left as an avenue for further research. In this regard, I would like to point out that the geometric mean quantity index framework in Chapter 3 and the capability set evaluation framework in Chapter 4 align well with the current way of analysing employment and social developments and levels in the Joint Employment Report, published yearly by the European Commission. The Joint Employment Report uses three tools to analyse individual social inclusion indicators (Council of the European Union, 2014):

> The *historical trend* gives for each Member State the change in the indicator in a certain year as compared with earlier periods in time; the synthetic change in social inclusion performances is given by the ΔOWN component in the decomposition in (10) in Chapter 3; the synthetic change in capability sets is captured by the ΔCC^{g_1} component in the abovementioned decomposition (with g_1 representing an EU Member State).
>
> The *snapshot of existing employment and social disparities* gives for each Member State the difference from the EU and the euro area average rates in the same year; the synthetic differences in social inclusion performance of a Member State relative to the EU level is captured by the geometric mean quantity index CI in (1) in Chapter 3; the synthetic differences in capability sets is captured by the capability set evaluation metrics $SE_{\mu}\left(\hat{\Psi}^{g_1}, \hat{\Psi}^{g_2}\right)$ (6) or $\overline{SE}\left(\hat{\Psi}^{g_1}, \hat{\Psi}^{g_2}\right)$ (7) in Chapter 4 (with g_1 representing an EU Member State and g_2 representing a pan-European group).
>
> *Dynamics of socio-economic convergence/divergence* gives the change in an indicator between two consecutive years in each Member State relative to the change at the EU and euro area levels; The synthetic change in social inclusion performances of a Member State relative to change at the EU level is captured by $\Delta OWN \times \Delta BP$ in the decomposition in (10) as in Chapter 3; The synthetic change in capability sets of a Member State relative to change at the EU level is captured by $\Delta CC^{g_1} \times \Delta CC^{g_2}$ in the abovementioned decomposition (with g_1 representing an EU Member State and g_2 representing a pan-European group).

The frameworks presented in Chapter 3 and Chapter 4 allow for the replication of these tools, with the key difference that a synthetic picture is provided.

Inequalities should be addressed in a comprehensive way

The Stiglitz-Sen-Fitoussi (2009, p. 59) stresses that "inequalities in human conditions are integral to any assessment of quality of life across countries and the way that it is developing over time". The authors further argue that "social progress depend not only on the average conditions in each country but also on the inequalities in people's conditions. *Inequality in each of the dimensions* of quality of life is significant in itself, and this underscores the importance of avoiding the presumption that any single dimension will always encompass all the others. At the same time, because of the links between the dimensions of quality of life, various inequalities may also strengthen each other (Stiglitz et al., 2009, p. 217). Finally, the authors *recommend* to account "for the diversity of experience to fill the gap between country-wide estimates and people's feelings about their own conditions" (Stiglitz et al., 2009, p. 217). *Inequalities in quality of life* should be assessed *across people, socio-economic groups, gender and generations*, with special attention to inequalities that have arisen more recently, such as those linked to immigration" (Stiglitz et al., 2009, p. 15) .

The methodological frameworks presented in Chapter 3 and Chapter 4 *contribute* to the implementation of this recommendation in that inequality issues can be taken into account. Both methodologies were empirically illustrated by providing a comprehensive picture of average social living conditions in European countries. However, both approaches also allow for a more meso-level assessment of well-being, although not illustrated, by benchmarking performances of certain subgroups versus the country average (Chapter 3[83]) or by comparing the capability set of a certain subgroup versus a country-specific capability set that includes all possible subgroups (Chapter 4[84]). The inter-temporal decomposition presented in Chapter 3 and in the previous Section "The development of single summary measures of quality-of-life" can be further employed to account for changes in disparities in social living conditions between groups. Next, a key motivation behind the geometric aggregation function presented in Chapter 3 was to penalize *inequalities in social outcomes across dimensions*. The geometric aggregation function entails that poor performances on one sub-indicator cannot be fully compensated by sufficiently high values on other sub-indicators. It was furthermore argued in Chapter 3 that by employing both optimistic and pessimistic weighting schemes, one obtains an idea of the "degree of unbalance in social performances". The Bortkiewicz decomposition allows to account for such inequalities across

[83] Several EU social indicators can be broken down by subgroup performances.
[84] Or, alternatively, by a subgroup that is considered the best or worst off.

dimensions in terms of three factors: i.e., (1) the effectiveness of the optimistic and pessimistic weights in, respectively, maximizing and minimizing the geometric index, (2) the dissimilarity of the optimistic and pessimistic weights and their divergences from an equal weighting scheme and (3) the divergence of performances on the sub-indicators.

There are several *open research paths* to further account for inequality issues within both frameworks. A limitation of the Bortkiewicz-decomposition in Chapter 3 is that the components are weighted in terms of optimistic (i.e. $\frac{CI_i^+}{CI_i^-} = e^{R_i^{\omega_{r,i}^+}\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right)\right)\times CV_i^{\omega_{r,i}^+}\left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right)\times\sigma_i^{\omega_{r,i}^+}\left(\frac{y_{r,i}}{y_{r,B}}\right)}$) or pessimistic (i.e. $\frac{CI_i^+}{CI_i^-} = e^{R_i^{\omega_{r,i}^-}\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right)\times CV_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\times\sigma_i^{\omega_{r,i}^-}\left(\frac{y_{r,i}}{y_{r,B}}\right)}$) weights. One way to integrate both perspective would be to take a geometric average of the optimistically and pessimistically weighted decompositions, leading to a six-way decomposition:

$$\frac{CI_i^+}{CI_i^-} = e^{0.5\times\left(R_i^{\omega_{r,i}^-}\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right)\times CV_i^{\omega_{r,i}^-}\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\times\sigma_i^{\omega_{r,i}^-}\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)+0.5\times\left(R_i^{\omega_{r,i}^+}\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right)\right)\times CV_i^{\omega_{r,i}^+}\left(\frac{\omega_{r,i}^-}{\omega_{r,i}^+}\right)\times\sigma_i^{\omega_{r,i}^+}\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)}$$

An alternative approach without weighted components can be achieved by multiplying $\frac{CI_i^+}{CI_i^-}$ with $\frac{CI_i^=}{CI_i^=}$, where $CI_i^=$ represents a geometric mean composite indicator under equal weighting ($\omega_{r,i}^= = \frac{1}{r}$):

$$\frac{\frac{CI_i^+}{CI_i^=}}{\frac{CI_i^-}{CI_i^=}} = \frac{e^{R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(r\omega_{r,i}^+\right)\right)\times\sigma_i\left(r\omega_{r,i}^+\right)\times\sigma_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)}}{e^{R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(r\omega_{r,i}^-\right)\right)\times\sigma_i\left(r\omega_{r,i}^-\right)\times\sigma_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)}}$$

Using the fact that $-R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right), r\omega_{r,i}^-\right)$ equals $R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right), -r\omega_{r,i}^-\right)$, this equation can be further simplified to:

$$\frac{CI_i^+}{CI_i^-} = e^{\left(R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(r\omega_{r,i}^+\right)\right)\times\sigma_i\left(r\omega_{r,i}^+\right)+R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(-r\omega_{r,i}^-\right)\right)\times\sigma_i\left(r\omega_{r,i}^-\right)\right)\times\sigma_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)}$$

This alternative decomposition shows that the "degree of unbalance" $\frac{CI_i^+}{CI_i^-}$ depends on five factors:

(i) The first factor, $R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right), r\omega_{r,i}^+\right)$, gives the correlation between the optimistic weights $\omega_{r,i}^+$ and the sub-indicators $\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$. A positive correlation implies that strong performances on the sub-indicators (i.e. high $\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$ values) are correlated with high $\omega_{r,i}^+$ -values. The higher the correlation, the larger CI_i^+ and the larger the potential disparities between CI_i^+ and CI_i^-.

(ii) The second factor, $\sigma_i\left(r\omega_{r,i}^+\right)$, captures the standard error or the degree of concentration of the optimistic Benefit-of-the-Doubt-based weights $\omega_{r,i}^+$. The more concentrated the optimistic weights are, the higher the weight that will be attached to indicators of comparative strength and, hence, the higher the CI_i^+-value (provided that $R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right), r\omega_{r,i}^+\right)$ is positive). This is because, as explained in the Chapter 3, the multiplicative aggregator function penalizes inequality among the sub-indicators.

(iii) The third factor, $R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),-r\omega_{r,i}^{-}\right)$, gives the correlation between the pessimistic weights $-\omega_{r,i}^{-}$ and the normalized sub-indicators $\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$. A positive correlation implies that strong performances on the sub-indicators (i.e. high $\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$ values) are correlated with low $\omega_{r,i}^{-}$-values (or with high $-\omega_{r,i}^{-}$-values). Thus, the higher the correlation, the smaller CI_i^{-} and the larger the disparities between CI_i^{+} and CI_i^{-}.

(iv) The fourth factor, $\sigma_i\left(r\omega_{r,i}^{-}\right)$, captures the standard error of the pessimistic Benefit-of-the-Doubt-based weights. The more concentrated the pessimistic weights are, the lower the weight that will be attached to indicators of comparative weakness and, hence, the lower the CI_i^{-}-value (provided that $R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),-r\omega_{r,i}^{-}\right)$ is positive).

(v) The final factor, $\sigma_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)\right)$, represents the standard deviation of $\ln\left(\frac{y_{r,i}}{y_{r,B}}\right)$. This factor captures the intensity of the differences in the evaluated performances on the set of sub-indicators vis-à-vis the baseline sub-indicator performance values. If an entity shows a strong variation in its performances on the set of sub-indicators, the aggregation of these performances are more likely to diverge under different weighting schemes.

The main differences with this alternative version of the Bortkiewicz-decomposition is that it allows the effectiveness of the optimistic and pessimistic weighting models in, respectively, maximizing and minimizing the geometric mean composite indicator to be separately assessed (e.g. $R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\omega_{r,i}^{+}\right)$ and $R_i\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),-\omega_{r,i}^{-}\right)$). The decomposition presented in Chapter 3 jointly assesses the effectiveness of both weighting models (e.g. $R_i^{\omega_{r,i}^{+}}\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(\frac{\omega_{r,i}^{-}}{\omega_{r,i}^{+}}\right)\right)$ or

$R_i^{\omega_{r,i}^-}\left(\ln\left(\frac{y_{r,i}}{y_{r,B}}\right),\left(\frac{\omega_{r,i}^+}{\omega_{r,i}^-}\right)\right)$). In addition, the alternative decomposition is less arbitrary in that the components do not depend on whether an optimistic or pessimistic weighting perspective is maintained.

I see two major paths for further research in the integration of the measurement of inequalities within the capability set estimation and evaluation framework presented in Chapter 4. First, one could further account for inequalities in capabilities within groups by moving from a robust order-*m* frontier estimator to the so-called robust *order-α frontier estimator*. In the order-*m* model, one first sets an *m* as a trimming parameter which allows to tune the percentage of individuals that will lie above the (robust) group capability set frontier. The order-*α* type of frontier estimator goes the other way round: in the order-*α* model, the (robust) group capability set frontier is determined by first fixing a probability 1-*α* of observing achieved functioning bundles that lie above the order-*α* frontier (Daraio & Simar, 2007, p. 72). These *α*-values can be distributionally interpreted: an *α* set to 0.5 gives a "median" group capability set frontier where half of the achieved functioning bundles lie above this frontier, an *α* set to 0.8 equal an "upper class" group capability set frontier where 20 per cent of the achieved functioning bundles lie above the frontier, etc.. I believe that several interesting (multidimensional) distributional metrics can be constructed by applying the order-*α* model. One suggestion is to use the order-*α* model to estimate distributional group capability set frontiers under varying *α* values. In a next step, the framework presented in Chapter 4 could be used to estimate the distances between several distributional order-*α* frontiers within groups. This would say something about the inequalities in capabilities within groups, which can be, in turn, compared across groups. One could also use the order-*α* model to make between-group evaluations of capability sets that have a more consistent distributional interpretation. It is common practice in the non-parametric partial frontier literature to estimate and compare sets that are based on the same size of *m* (Rogge et al., 2013). Even though it was shown in the last chapter that the percentage of functioning bundles that lie outside of the estimated group capability set frontier is rather similar for the majority of the countries for the same size of *m*, I believe a more consistent evaluation of groups' capability sets can be ensured by employing the order-*α* model. That is, the order-*α* model seems more appropriate in that the estimated order-*α* partial group capability sets have exactly the same distributional interpretation for the same *α*-values. That being said, the locally convex order-*m* estimator was employed, as the algorithm behind non-convex order-*α* frontier estimator is computationally complex and expensive (Ferreira & Marques, 2017). When a simplified algorithm would be available, the obvious way to proceed would be to employ the convex order-α frontier estimator.

Second, given that a deep concern for the worst-off is a key motivation behind the capability approach (Decancq et al., 2016, p. 3), one could turn to the so-called "pessimistic" frontier estimation model to estimate a "*lack-of-capability-frontier*" of the group capability set (Zhou et al., 2007). The lack-of-capability-frontier represents the worst possible functioning achievements within the group capability set and could help identify the (minimal) capabilities that the worst-off are achieving. One could then use the framework presented in Chapter 4 to estimate the gap between a group's capability set frontier and a group's "lack-of-capability-frontier". This gap would measure the difference in the functionings that are maximally achievable in the group and the functionings that are minimally achieved in the group, - again - touching upon the notion of within-group inequality. The pessimistic frontier estimation model can also be used to compare the lack-of-capability-frontier across groups, allowing to quantify between-group differences in minimally realized capabilities.

References

Aaberge, R., Langørgen, A. & Lindgren, P. (2017). The distributional impact of public services in European countries. In A.B. Atkinson, A.C. Guio & E. Marlier (Eds.), *Monitoring social inclusion in Europe* (pp. 159-199). Luxembourg: Publications Office of the European Union.

Addabbo, T., Di Tommaso, M. L., & Maccagnan, A. (2014). Gender differences in Italian children's capabilities. *Feminist Economics*, *20*(2), 90-121.

Alkire, S. (2008). Using the capability approach: Prospective and evaluative analyses. In F. Comin, M. Qizilbash & S. Alkire (Eds.), *The capability approach: Concepts, measures and applications* (pp. 26-49). Cambridge, England: Cambridge University Press.

Amado, C. A., São José, J. M., & Santos, S. P. (2016). Measuring active ageing: a data envelopment analysis approach. *European Journal of Operational Research, 255*(1), 207-223.

Anand, P., Hunter, G., Carter, I., Dowding, K., Guala, F., & Van Hees, M. (2009). The development of capability indicators. *Journal of Human Development and Capabilities, 10*(1), 125-152.

Aristei, D., & Bracalente, B. (2014). Measuring Multidimensional Inequality: Methods and Issues in Empirical Analysis. In F. Crescenzi & S. Mignani (Eds.), *Statistical Methods and Applications from a Historical Perspective* (pp. 203-215). Cham: Springer.

Armknecht, P., & Silver, M. (2014). Post-Laspeyres: The Case for a New Formula for Compiling Consumer Price Indexes. *Review of Income and Wealth*, *60*(2), 225-244.

Athanassoglou, S. (2016). Revisiting worst-case DEA for composite indicators. *Social Indicators Research*, *128*(3), 1259-1272.

Atkinson, T., Cantillon, B., Marlier, E. & Nolan, B. (2002). *Social Indicators: The EU and Social Inclusion*. Oxford: Oxford University press.

Bäckman, O., & Ferrarini, T. (2010). Combating child poverty? A multilevel assessment of family policy institutions and child poverty in 21 old and new welfare states. *Journal of Social Policy, 39*(2), 275-296.

Bădin, L., Daraio, C., & Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. *European journal of operational research, 201*(2), 633-640.

Bădin, L., Daraio, C., & Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research, 223*(3), 818-833.

Ballon, P., & Krishnakumar, J. (2008). A Model-Based Multidimensional Capability Deprivation Index. *The International Association for Research in Income and Wealth*. Internet source, accessed on 10 February, 2019. http://iariw.org/papers/2008/ballon.pdf

Barberà, S., Bossert, W., & Pattanaik, P. K. (2004). Ranking sets of objects. In S. Barbera, P.J. Hammond & C. Seidl (Eds.), *Handbook of utility theory*, (Vol. 2, pp. 893-977). Dordrecht: Kluwer.

Bárcena-Martín, E., Blanco-Arana, M. C., & Perez-Moreno, S. (2017a). *Dynamics of child poverty in the European countries* (ECINEQ, Society for the Study of Economic Inequality Working Paper Series No. 437).

Bárcena-Martín, E., Blanco-Arana, M. C., & Perez-Moreno, S. (2018). Social transfers and child poverty in European Countries: pro-poor targeting or pro-child targeting?. *Journal of Social Policy*, *47*(4), 739-758.

Bárcena-Martín, E., Blázquez, M., Budría, S., & Moro-Egido, A. I. (2017b). Child deprivation and social benefits: Europe in cross-national perspective. *Socio-Economic Review*, *15*(4), 717-744.

Bárcena-Martín, E., Lacomba, B., Moro-Egido, A. I., & Pérez-Moreno, S. (2014). Country differences in material deprivation in Europe. *Review of Income and Wealth*, *60*(4), 802-820.

Berthoud, R., & Bryan, M. (2011). Income, deprivation and poverty: a longitudinal analysis. *Journal of Social Policy*, *40*(1), 135-156.

Bhattacharya, J., & Banerjee, S. (2012). Women empowerment as multidimensional capability enhancement: An application of structural equation modeling. *Poverty & Public Policy*, *4*(3), 79-98.

Blancard, S., & Hoarau, J. F. (2013). A new sustainable human development indicator for small island developing states: A reappraisal from data envelopment analysis. *Economic Modelling*, *30*, 623-635.

Blancas, F. J., Contreras, I., & Ramírez-Hurtado, J. M. (2012). Constructing a composite indicator with multiplicative aggregation under the objective of ranking alternatives. *Journal of the Operational Research Society*, *64*(5), 668-678.

Boarini, R., & d'Ercole, M. M. (2006). *Measures of material deprivation in OECD countries*. (OECD Social, Employment and Migration Working Paper series No. 37). Paris: OECD publishing.

Bortkiewicz, L. V. (1923). Zweck und struktur einer preisindexzahl. *Nordisk Statistisk Tidsskrift*, *2*, 369-408.

Bosco, B., & Ambra P (2016). *Government Effectiveness, Middle Class and Poverty in the EU: A Dynamic Multilevel Analysis* (Management and Statistics Working Paper No. 344). Milan: University of Milan Bicocca Department of Economics.

Bradshaw, J., & Finch, N. (2003). Overlaps in dimensions of poverty. *Journal of social policy*, *32*(4), 513-525.

Brady, D. (2005). The welfare state and relative poverty in rich western democracies, 1967–1997. *Social Forces*, *83*(4), 1329-1364.

Brady, D., Fullerton, A. S., & Cross, J. M. (2009). Putting poverty in political context: a multi-level analysis of adult poverty across 18 affluent democracies. *Social Forces*, *88*(1), 271-299.

Brady, D., Giesselmann, M., Kohler, U., & Radenacker, A. (2018). How to measure and proxy permanent income: evidence from Germany and the US. *The Journal of Economic Inequality*, 1-25.

Brady, D., & Kall, D. (2008). Nearly universal, but somewhat distinct: The feminization of poverty in affluent Western democracies, 1969–2000. *Social Science Research*, *37*(3), 976-1007.

Brooks-Gunn, J. & G. J. Duncan (1997). The effects of poverty on children. *The future of children*, *7*(2), 55-71

Browne, W. J. (2015). MCMC estimation in MLwiN. *Centre for Multilevel Modelling, University of Bristol*.

Bryan, M. L., & Jenkins, S. P. (2015). Multilevel modelling of country effects: A cautionary tale. *European Sociological Review*, *32*(1), 3-22.

Burchardt, T., & Vizard, P. (2007). *Definition of equality and framework for measurement: Final Recommendations of the Equalities Review Steering Group on Measurement* (LSE STICERD Research Paper No. 120). London: London School of Economics.

Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, *106*(1), 1-25.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429-444.

Charpentier, A., & Mussard, S. (2011). Income inequality games. *The Journal of Economic Inequality, 9*(4), 529-554.

Chen, Y., Cook, W. D., Du, J., Hu, H., & Zhu, J. (2017). Bounded and discrete data and Likert scales in data envelopment analysis: application to regional energy efficiency in China. *Annals of Operations Research, 255*(1-2), 347-366.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social indicators research, 82*(1), 111-145.

Cherchye, L., Moesen, W., & Van Puyenbroeck, T. (2004). Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. *Journal of Common Market Studies, 42*(5), 919-955.

Chzhen, Y. (2014). *Child poverty and material deprivation in the European Union during the Great Recession* (Innocenti Working Paper No. 2014-06). Florence: UNICEF Office of Research.

Chzhen, Y., & Bradshaw, J. (2012). Lone parents, poverty and policy in the European Union. *Journal of European Social Policy, 22*(5), 487-506.

Copeland, P., & Daly, M. (2012). Varieties of poverty reduction: Inserting the poverty and social exclusion target into Europe 2020. *Journal of European Social Policy, 22*(3), 273-287.

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

Council of the European Communities (1975). *Council Decision of 22 July 1975 concerning a programme of pilot schemes and studies to combat poverty*. No L 199/34, 34-35.

Council of the European Union (2014). *Joint Employment Report: Adoption*. 7476/14. Brussels, 12 March.

Coveney, M., Garcia-Gomez, P., van Doorslaer, E., & Van Ourti, T. (2018). *Every crisis has a silver lining? Unravelling the pro-cyclical pattern of health inequalities by income* (Tinbergen Institute Discussion Paper No. 2018-066/V).

Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of productivity analysis, 24*(1), 93-121.

Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. New York: Springer.

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

Decancq, K., & Lugo, M. A. (2009). *Measuring inequality of well-being with a correlation-sensitive multidimensional Gini index* (University of Oxford Department of Economics Discussion Paper Series No. 459). Oxford: University of Oxford.

Decancq, K., Schokkaert, E., & Zuluaga, B. (2016). *Implementing the capability approach with respect for individual valuations: an illustration with Colombian data* (Discussion Paper Series No.16.09). Leuven: KU Leuven Department of Economics.

De Graaf-Zijl, M., & Nolan, B. (2011). Household joblessness and its impact on poverty and deprivation in Europe. *Journal of European Social Policy, 21*(5), 413-431.

de Neubourg, C., Bradshaw, J., Chzhen, Y., Main, G., Martorano, B. & Menchini, L. (2012). *Child deprivation, multidimensional poverty and monetary poverty in Europe* (Innocenti Working Paper No. 2012-02). Florence: UNICEF Office of Research.

Deutsch, J., & Silber, J. (2006). The “fuzzy set” approach to multidimensional poverty analysis: using the shapley decomposition to analyze the determinants of poverty in Israel. In A. A., Lemmi & G. Betti (Eds.) *Fuzzy Set Approach to Multidimensional Poverty Measurement* (Vol. 3, pp. 115-174). New York: Springer Science & Business Media.

Dewilde, C. (2008). Individual and institutional determinants of multidimensional poverty: A European comparison. *Social Indicators Research*, *86*(2), 233-256.

Diris, R., & Vandenbroucke, F. (2016). *How does early deprivation relate to later-life outcomes? A longitudinal analysis* (Discussion Paper Series No. 16.27). Leuven: KU Leuven Department of Economics.

Diris, R., Vandenbroucke, F., & Verbist, G. (2017). The impact of pensions, transfers and taxes on child poverty in Europe: the role of size, pro-poorness and child orientation. *Socio-Economic Review, 15*(4), 745-775.

Domínguez-Serrano, M., & Blancas, F. J. (2011). A gender wellbeing composite indicator: The best-worst global evaluation approach. *Social Indicators Research*, *102*(3), 477-496.

Douglas, M. & Ney, S. (1998). *Missing persons: A critique of the Social Sciences*. Berkeley, CA: University of California Press.

Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: a social choice approach. *Journal of Environmental Economics and Management*, *47*(2), 270-283.

European Commission (1992). *Toward a Europe of Solidarity: Intensifying the Fight against Social Exclusion, Fostering Integration*. COM (92) 542, Brussels.

European Commission (2011). *The social dimension of the Europe 2020 Strategy: A report of the Social protection committee*. Luxembourg: Publications Office of the European Union.

European Commission (2013a), *Investing in Children: breaking the cycle of disadvantage.* Commission Recommendation 2013/112/EU, Brussels.

European Commission (2013b). *Strengthening the social dimension of the Economic and Monetary Union*. Communication from the Commission to the European Parliament and the Council. Brussels, 2.10.2013.

European Commission (2015). *Report of the social protection committee. Portfolio of EU Social Indicators for the Monitoring of Progress Towards the EU Objectives for Social Protection and Social Inclusion*. Luxembourg: Publications Office of the European Union.

Farina, F., Peluso, E., & Savaglio, E. (2004). Ranking opportunity sets in the space of functionings. *The Journal of Economic Inequality, 2*(2), 105-116.

Ferreira, D. C., & Marques, R. C. (2017). A step forward on order-α robust nonparametric method: Inclusion of weight restrictions, convexity and non-variable returns to scale. *Operational Research*, 1-36.

Fleurbaey, M., Blanchet, D. (2013). *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford: Oxford University Press.

Fleurbaey, M., Decancq, K., & Schokkaert, E. (2009). *What good is happiness?* (Core discussion paper No. 2009/17).

Foster, J. E. (2011). Freedom, opportunity, and well-being. In K.J. Arrow, A.K. Sen & K. Suzumura (Eds.), *Handbook of social choice and welfare* (Vol. 2, pp. 687-728). Amsterdam: North-Holland.

Fusco, A. Guio, A-C & Marlier, E. (2011). Characterising the income poor and the materially deprived in European countries. In A. Atkinson & E. Marlier (Eds.), *Income and living conditions in Europe* (pp. 132-153). Luxembourg: Publications Office of the European Union.

Gaertner, W. (2012). Evaluating sets of objects in characteristics space. *Social Choice and Welfare, 39*(2-3), 303-321.

Gaertner, W., & Xu, Y. (2006). Capability sets as the basis of a new measure of human development. *Journal of Human Development*, *7*(3), 311-321.

Gaertner, W., & Xu, Y. (2008). A new class of measures of the standard of living based on functionings. *Economic Theory, 35*(2), 201-215.

Gaertner, W., & Xu, Y. (2011). Reference-dependent rankings of sets in characteristics space. *Social Choice and Welfare*, *37*(4), 717.

Giambona, F., & Vassallo, E. (2014). Composite indicator of social inclusion for European countries. *Social Indicators Research, 116*(1), 269-293.

Gordon, D., Adelman, A., Ashworth, K., Bradshaw, J., Levitas, R., Middleton, S., Pantazis, C., Patsios, D., Payne, S., Townsend, P. & Williams, J. (2000). *Poverty and social exclusion in Britain*. York: Joseph Rowntree Foundation.

Guio, A.-C. (2009). *What can be learned from deprivation indicators in Europe?*. Luxembourg: Publications office of the European Union.

Guio, A. C., Fusco, A., & Marlier, E. (2010). *Risk factors of income poverty and material deprivation in Belgium and Regions* (Discussion Papers IWEPS No. 1005).

Guio, A-C., Gordon, D. & Marlier, E. (2012). *Measuring Material Deprivation in the EU. Indicators for the whole Population and Child-Specific Indicators*. Luxembourg: Publications office of the European Union.

Guio, A-C., Gordon, D., Najera H. & Pomati M. (2017*). Revising the EU material deprivation variables (analysis of the final 2014 EU-SILC data)- Whole population MD indicator*. Luxembourg: Publications office of the European Union.

Guio, A. C., Gordon, D., Marlier, E., Najera, H., & Pomati, M. (2018). Towards an EU measure of child deprivation. *Child indicators research, 11*(3), 835-860.

Hatefi, S. M., & Torabi, S. A. (2010). A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics*, *70*(1), 114-120.

Israel S. & D. Spannagel (2013). *Material deprivation: an analysis of cross-country differences and European convergence*. Brussels: European Commission.

Kenworthy, L. (2011). Public Services Are an Important Antipoverty Tool. In L. Kenworthy (Ed.). *Progress for the Poor*. Oxford: Oxford University Press.

Kenworthy, L., Epstein, J. & Duerr, D. (2011). General social policy reduces material deprivation. In L. Kenworthy (Ed.). *Progress for the Poor*. Oxford: Oxford University Press.

Kim, K. S., Lee, Y., & Lee, Y. J. (2010). A multilevel analysis of factors related to poverty in welfare states. *Social indicators research*, *99*(3), 391-404.

Korpi, W., & Palme, J. (1998). The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and poverty in the Western countries. *American sociological review, 63*(5), 661-687.

Krishnakumar, J. (2007). Going beyond functionings to capabilities: An econometric model to explain and estimate capabilities. *Journal of Human Development, 8*(1), 39-63.

Krishnakumar, J., & Ballon, P. (2008). Estimating basic capabilities: A structural equation model applied to Bolivia. *World Development, 36*(6), 992-1010.

Krishnakumar, J., & Chávez-Juárez, F. (2016). Estimating capabilities with structural equation models: How well are we doing in a 'real'world?. *Social Indicators Research, 129*(2), 717-737.

Kuklys, W. (2005). *Amartya Sen's Capability Approach: Theoretical Insights and Empirical Applications*. Berlin: Springer.

Kynch, J., & Sen, A. (1983). Indian women: well-being and survival. *Cambridge Journal of Economics, 7*(3/4), 363-380.

Lohmann, H. (2009). Welfare states, labour market institutions and the working poor: A comparative analysis of 20 European countries. *European Sociological Review*, *25*(4), 489-504.

Lokshin, M., & Ravallion, M. (2008). Testing for an economic gradient in health status using subjective data. *Health Economics, 17*(11), 1237-1259.

Lollivier, S., & Verger, D. (1998). Pauvreté d'existence, monétaire ou subjective sont distinctes. *Économie et statistique, 308*(1), 113-142.

Madden, D. (2011). Health and income poverty in Ireland, 2003–2006. *The Journal of Economic Inequality*, *9*(1), 23-33.

Marlier, E., Atkinson, A.B., Cantillon, B. & Nolan, B. (2007). *The EU and social inclusion: facing the challenges*. Bodmin: MPG books.

Marx, I., Salanauskaite, L., & Verbist, G. (2013). *The paradox of redistribution revisited: and that it may rest in peace?* (GINI Discussion Paper No. 82).

McKay, S. (2004). Poverty or preference: what do 'consensual deprivation indicators' really mean?. *Fiscal studies, 25*(2), 201-223.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of mathematical sociology*, *4*(1), 103-120.

McKnight, A. (2013). *Material deprivation among children – how parents evaluate 'need' in relation to available income* (CASE Annual Report 2012). London: Centre for Analysis of Social Exclusion, London.

Melyn, W., & Moesen, W. (1991). *Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available* (Public economics research papers No. 17). Leuven: CES KU Leuven.

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM?. *Survey research methods, 3*(1), 45-58.

Micklewright, J. (2002). *Social exclusion and children: A European view for US debate* (Innocenti Working Papers No. 90).

Moller, S., Huber, E., Stephens, J. D., Bradley, D., & Nielsen, F. (2003). Determinants of relative poverty in advanced capitalist democracies. *American Sociological Review*, *68*(1), 22-51.

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, *26*(1), 67-82.

Muellbauer, J. (1987). Professor Sen on the standard of living. In A. Sen (Ed.), *The standard of living* (pp. 39-58). Cambridge: Cambridge University Press.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). *Handbook on constructing composite indicators: Methodology and users guide. Paris/Ispra*. OECD Publication.

Nelson, K. (2012). Counteracting material deprivation: The role of social assistance in Europe. *Journal of European Social Policy*, *22*(2), 148-163.

Nolan, B. & Whelan, C. T. (2011). *The EU 2020 poverty target* (GINI Discussion Paper No. 19). Amsterdam: AIAS.

Notten, G & Guio, A.C. (2018). By how much do social transfers reduce income poverty and material deprivation in Europe?. Paper presented at Net-SILC3 International Conference Comparative EU Statistics on Income and Living Conditions, Athens, April 19 2018.

Nussbaum, M. C. (2000). *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

Nussbaum, M. C., & Sen, A. (1993). *The quality of life*. Oxford: Oxford University Press.

OECD (2008). *Growing Unequal? Income Distribution and Poverty in OECD Countries*. Paris: OECD Publishing.

OECD (2011). *Divided We Stand. Why Inequality Keeps Rising*. Paris: OECD Publishing.

Papadopoulos, F., & Tsakloglou, P. (2008). Social exclusion in the EU: A capability-based approach. In F. Comim, M. Qizilbash, & S. Alkire (Eds.), *The Capability Approach: Concepts, Measures and Applications* (pp. 242-267). Cambridge: Cambridge University Press.

Pattanaik, P. K., & Xu, Y. (1990). On ranking opportunity sets in terms of freedom of choice. *Recherches Économiques de Louvain/Louvain Economic Review*, *56*(3-4), 383-390.

Perry, B. (2002). The mismatch between income measures and direct outcome measures of poverty. *Social Policy Journal of New Zealand*, *19*, 101-127.

Raftery, A. E. & Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In J. M. Bernado, J. O. Berfer, A. P. Dawid & A. FM. Smith (Eds.), *Bayesian Statistics 4* (pp. 765-76). Oxford: Oxford University Press.

Ravallion, M. (2012). Troubling tradeoffs in the human development index. *Journal of Development Economics, 99*(2), 201-209.

Reinstadler, A. & Ray, J-C. (2010). *Macro Determinants of Individual Income Poverty in 93 Regions of Europe* (LISER Working Paper Series No. 2010-13).

Ridge, T. (2002). *Childhood Poverty and Social Exclusion: From a Child's Perspective*. Bristol: The Policy Press, Bristol.

Ridge, T. (2011). The everyday costs of poverty in childhood: A review of qualitative research exploring the lives and experiences of low-income children in the UK. *Children & society, 25*(1), 73-84.

Robeyns, I. (2003). Sen's capability approach and gender inequality: selecting relevant capabilities. *Feminist economics, 9*(2-3), 61-92.

Robeyns, I. (2006). Measuring Gender Inequality in Functionings and Capabilities: Findings from the British Household Panel Survey. In P. Bharti & M. Pal (Eds.), *Gender Disparity: Manifestations, Causes and Implication* (pp. 236-277). New Delhi: Anmol

Robeyns, I. (2018). *Wellbeing, freedom and social justice: The capability approach re-examined*. Open Book Publishers.

Roche, J. M. (2009). *Capability and Group Inequalities: Revealing the latent structure* (OPHI, Research in Progress series). Oxford: Department of International Development (QEH), University of Oxford.

Rogge, N. (2012). Undesirable specialization in the construction of composite policy indicators: The Environmental Performance Index. *Ecological indicators, 23*, 143-154.

Rogge N. (2017). Measuring the impact of the economic crisis on the level of change in EU social inclusion: period 2005-2012. *Journal of Productivity Analysis, 47*(2), 103-116.

Rogge, N., & Konttinen, E. (2018). Social Inclusion in the EU Since the Enlargement: Progress or Regress?. *Social Indicators Research*, *135*(2), 563-584.

Rogge, N., & Self, R. (2018). Measuring regional social inclusion performances in the EU: Looking for unity in diversity. *Journal of European Social Policy*.

Rogge, N., & Van Nijverseel, I. (2018). Quality of Life in the European Union: A Multidimensional Analysis. *Social Indicators Research*, 1-25.

Rogge, N., Van Reeth, D., & Van Puyenbroeck, T. (2013). Performance evaluation of Tour de France cycling teams using data envelopment analysis. *International Journal of Sport Finance, 8*(3), 236-257.

Saltkjel, T., & Malmberg-Heimonen, I. (2017). Welfare Generosity in Europe: A Multi-level Study of Material Deprivation and Income Poverty among Disadvantaged Groups. *Social Policy & Administration*, *51*(7), 1287-1310.

Schokkaert, E. (2007). Capabilities and satisfaction with life. *Journal of Human Development*, *8*(3), 415-430.

Sen, A. K. (1985) *Commodities and Capabilities*. Amsterdam: North Holland.

Sen, A. K. (1987) *The Standard of Living*. Cambridge: Cambridge University Press.

Sen, A. (1988). Freedom of choice: concept and content. *European Economic Review, 32*(2-3), 269-294.

Sen, A. K. (1993). Capability and Well-Being. In M. Nussbaum & A. Sen (Eds.), *The Quality of Life* (pp. 30-53). Oxford: Oxford University Press.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, *2*(28), 307-317.

Shwartz, M., Burgess, J. F., & Zhu, J. (2016). A DEA based composite measure of quality and its associated data uncertainty interval for health care provider profiling and pay-for-performance. *European Journal of Operational Research*, *253*(2), 489-502.

Simar, L., & Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: recent developments and perspectives. In H. Fried, C.A.K. Lovell & S. Schmidt (Eds.), *The measurement of productive efficiency and productivity growth* (pp. 421-521). New York: Oxford University Press.

Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic & Advanced Modeling, second edition*. London: Sage Publishers.

Social Protection Committee (2015). *Portfolio of EU social indicators for the monitoring of progress towards the EU objectives for social protection and social inclusion*. Luxembourg: Publications Office of the European Union.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583-639.

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, *57*(3), 748-761.

Stewart, F. (2005). Groups and capabilities. *Journal of human development, 6*(2), 185-204.

Stewart, F. (Ed.) (2008). *Horizontal inequalities and conflict: understanding group violence in multiethnic societies.* Basingstoke: Palgrave Macmillan.

Stiglitz, J., Sen, A., & Fitoussi, J. P. (2009). *The measurement of economic performance and social progress revisited. Reflections and overview*. Paris: Commission on the Measurement of Economic Performance and Social Progress.

Tárki (2011). *Child well-being in the European Union – Better monitoring instruments for better policies*. Report prepared for the State Secretariat for Social Inclusion of the Ministry of Public Administration and Justice. Budapest.

Tofallis, C. (2014). On constructing a composite indicator with multiplicative aggregation and the avoidance of zero weights in DEA. *Journal of the Operational Research Society, 65*(5), 791-792.

Townsend, P. (1979). *Poverty in the United Kingdom*. Hardmonsworth: Penguin Books.

United Nations Development Program (1995). *Human Development Report*. New York: UNDP.

United Nations Development Program (2004). *Human Development Report*. New York: UNDP.

Van Doorslaer, E., & Jones, A. M. (2003). Inequalities in self-reported health: validation of a new approach to measurement. *Journal of health economics, 22*(1), 61-87.

Van Puyenbroeck, T., & Rogge, N. (2017). Geometric mean quantity index numbers with Benefit-of-the-Doubt weights. *European Journal of Operational Research, 256*(3), 1004-1014.

Veall, M. R., & Zimmermann, K. F. (1996). Pseudo-R2 Measures For Some Common Limited Dependent Variable Models. *Journal of Economic surveys, 10*(3), 241-259.

Verbist, G., & Matsaganis, M. (2014). The Redistributive Capacity of Services in the European Union. In B. Cantillon & F. Vandenbroucke (Eds.) (2014). *Reconciling work and poverty reduction: how successful are European welfare states?*. Oxford: Oxford University Press.

Verschelde, M., & Rogge, N. (2012). An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: Evidence from a conditional Data Envelopment Analysis approach. *European Journal of Operational Research*, *223*(1), 214-225.

Visser, M., Gesthuizen, M., & Scheepers, P. (2014). The Impact of Macro-Economic Circumstances and Social Protection Expenditure on Economic Deprivation in 25 European Countries, 2007–2011. *Social indicators research*, *115*(3), 1179-1203.

Whelan, C. T., Layte, R., & Maitre, B. (2002). Multiple deprivation and persistent poverty in the European Union. *Journal of European Social Policy*, *12*(2), 91-105.

Whelan, C. T., Layte, R., & Maitre, B. (2003). Persistent income poverty and deprivation in the European Union: an analysis of the first three waves of the European Community Household Panel. *Journal of Social Policy, 32*(01), 1-18.

Whelan, C. T., Layte, R., & Maître, B. (2004). Understanding the mismatch between income poverty and deprivation: a dynamic comparative analysis. *European Sociological Review*, 287-302.

Whelan, C. T., Layte, R., Maître, B., & Nolan, B. (2001). Income, deprivation, and economic strain. An analysis of the European community household panel. *European sociological review*, *17*(4), 357-372.

Whelan, C. T., & Maître, B. (2006). Comparing poverty and deprivation dynamics: Issues of reliability and validity. *The Journal of Economic Inequality*, *4*(3), 303-323.

Whelan, C. T., & Maître, B. (2007). Income, deprivation and economic stress in the enlarged European Union. *Social Indicators Research*, *83*(2), 309-329.

Whelan, C. T., & Maître, B. (2012). Understanding material deprivation: A comparative European analysis. *Research in Social Stratification and Mobility*, *30*(4), 489-503.

Whelan, C. T., & Maître, B. (2013). Material deprivation, economic stress, and reference groups in Europe: An analysis of EU-SILC 2009. *European Sociological Review*, *29*(6), 1162-1174.

Wiepking, P., & Maas, I. (2005). Gender differences in poverty: A cross-national study. *European Sociological Review*, *21*(3), 187-200.

Wong, Y.-H. B., & Beasley, J.E. (1990). Restricting Weight Flexibility in Data Envelopment Analysis. *The Journal of the Operational Research Society, 41*(9), 829-835.

Xara, H. X., & Schokkaert, E. (2017). Putting measures of individual well-being to use for ex-ante policy evaluation. *The Journal of Economic Inequality*, *15*(4), 421-440.

Xu, Y. (2002). Functioning, capability and the standard of living–an axiomatic approach. *Economic theory*, *20*(2), 387-399.

Young, H. P. (1985). Monotonic solutions of cooperative games. *International Journal of Game Theory*, *14*(2), 65-72.

Zaidi, A., & Burchardt, T. (2005). Comparing incomes when needs differ: equivalization for the extra costs of disability in the UK. *Review of Income and Wealth, 51*(1), 89-114.

Zhou, P., Ang, B. W., & Poh, K. L. (2007). A mathematical programming approach to constructing composite indicators. *Ecological Economics*, *62*(2), 291-297.

Zhou, P., Ang, B. W., & Zhou, D. Q. (2010). Weighting and aggregation in composite indicator construction: a multiplicative optimization approach. *Social Indicators Research, 96*(1), 169-181.

Zhu, J. (2004). A buyer–seller game model for selection and negotiation of purchasing bids: extensions and new models. *European Journal of Operational Research*, *154*(1), 150-156.

List of Doctoral Dissertations

Doctoral dissertations from the Faculty of Business and Economics,

See: http://www.kuleuven.ac.be/doctoraatsverdediging/archief.htm