andos józsef zsigmond juhász

# ECONOMETRIC ANALYSES OF SUBJECTIVE WELFARE AND INCOME INEQUALITY

# Econometric Analysis of Subjective Welfare and Income Inequality

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Andos Juhász József Zsigmond
aus Budapest

Tübingen

2013

| | |
|---|---|
| Tag der mündlichen Prüfung: | 29.10.2013 |
| Dekan: | Professor Dr. rer. soc. Josef Schmid |
| 1. Gutachter: | Professor Dr. rer. pol. Martin Biewen |
| 2. Gutachter: | Professor Dr. rer. pol. Joachim Grammig |

# ECONOMETRIC ANALYSES OF SUBJECTIVE WELFARE AND INCOME INEQUALITY

ANDOS JÓZSEF ZSIGMOND JUHÁSZ

Dissertation
zur Erlangung des Doktorgrades (Dr. rer. pol.)

Lehrstuhl für Statistik, Ökonometrie und Quantitative Methoden
Fachbereich Sozial- und Wirtschaftswissenschaften
Eberhard Karls Universität Tübingen

April 2013

``To my beloved family, to Mina & Kelvin´´

*„[. . . ] on ne voit bien qu'avec le coeur.*
*L'essentiel est invisible pour les yeux.“*

— Antoine de Saint-Exupéry

# PREAMBLE

When I started my studies, it was my first semester and my first time to sit in the introductory analysis reading. We neatly called it "infinitesimal calculus" back then. While most other details have vanished with time from those days, there was one comment of our lecturer, Christopher Deninger, that I still clearly remember. I remember him walking down from the entrance door to the front of the lecture hall, stopping on the right side on the stairs that would lead him to the front to hold his introductory speech not at the blackboard ahead, but while standing with us, students, at the observing side of the hall. Leaning against the wall with his back and putting on a big smile, one of the first things he said took the excited, young audience by surprise, leading to initial silence and to vivid discussions, shortly after. In essence, what he gave us, was a friendly warning: if we went on to study mathematics, the process would irreversibly change our brain and we would never see the world as we saw it on that very day again. He warned us that this would be our last chance to turn around, to walk through the door and to keep our present vision intact. Well, back then, it felt like nothing more than a neat gag of a young and exceptionally able academic, but as time passed, I, myself, started to feel that I'd frequently question more and more in my everyday life, and with time, all the magic in the world just seemed to vanish to give place for some logical explanation. In my experience, this process is less desirable than most people may think. It takes more from you than it gives you. Albert Einstein said: "There are two ways to live: You can live as if nothing is a miracle; or, you can live as if everything is a miracle." While he captured the essence of the problem, one can hardly imagine what it takes to change one's view from the first to the second, which is the real challenge. The other way around, it is much less of an effort. The latter is like loosing something. That is always easier than gaining. At least the process, not the consequences.

There are many possible ways to look at my academic studies retrospectively. A very superficial way is, maybe, to describe it in turns of a physical movement over the west part of Germany, starting at the University of Münster and moving down over Mainz to the University of Tübingen. On the layer of academic alignment, the same process can be described as a movement from theoretical mathematics to applied one, to arrive at econometrics and applied economic research: a process to be accompanied by the one of self-discovery, meaning to find one's true, unreflected self-image in a sea of reflections from the outside world. The process of such a self-discovery is also a process of maturing. While there are several facets of that, like gathering professional expertise or learning self-discipline, which studying is at least shallowly all about, changing one's point of view of the world, on the other hand, may be seen as a higher aim in the studying process. My personal change in this regards can be probably sketched by a deal of juvenescent naivety to predominating scepticism and over-analytic world perception, to the idea of optimism and the revived belief in the existence of something behind the curtains.

It is my firm conviction that big plans, whether a doctoral thesis or some other path of even higher order, cannot succeed without people that surround us and support us on our way. In the first place, in my professional life, I own my deepest gratitude to Martin Biewen, my Ph.D. advisor, thanking him for everything in the past years, for his guidance, his support, his sustained friendliness and enthusiasm of every good kind, his unbreakable optimism and his confidence into my abilities, even in times when I had less of my own. I also thank Joachim Grammig, who's liveliness, friendliness and never-resting love to econometrics steadily lightens our spirits on the Third Floor to feel not only as colleagues, but also as a big, productive, econometric family. I also thank Markus Niedergesäss, Luis Huergo, Gideon Becker, Stefanie Seifert, Thomas Dimpfl, Stephan Jank, Franziska Peter, Kerstin Kehrle and Miriam Sperl for everything they did for me. Many thanks also to Frau Eiting and Frau Bürger, without whom I could not even imagine our Third Floor. Special thanks to Martin Rosemann, Bernhard Boockmann, Kai Schmid, Christian Arndt, Rolf Kleimann and all other IAW members for a productive and friendly working atmosphere in the joint Armutsbericht project during my doctorate.

My undivided and infinite love belongs to my wife, Mina, and my son, Kelvin. My work separated me from You far too often and for far too long, feeling like an infinite sum of eternities. You Two are the meaning of my life.

*Friedrichshafen, Bodensee, 28 April 2013*

# CONTENTS

I DISSERTATION INTRODUCTION 1

II A SATISFACTION-DRIVEN POVERTY INDICATOR - A BUSTLE AROUND THE POVERTY LINE 9

1.1 Introduction 11
1.2 Data and Variable Selection 13
1.3 Methodology 21
1.3.1 Motivation 21
1.3.2 The Grid Search Approach 23
1.3.3 The Nonlinear Least Squares Approach 25
1.3.4 The Fixed Effects Nonlinear Least Squares Approach 28
1.4 Empirical Results 29
1.4.1 SDPLs based on the SOEP 29
1.4.2 The Poverty Transition Region 37
1.5 Poverty Lines in European Countries 39
1.6 Conclusion 44
1.7 Appendix 46

III UNDERSTANDING RISING INCOME INEQUALITY IN GERMANY 47

2.1 Introduction 49
2.2 Possible sources of increasing inequality 51
2.3 Estimation of counterfactual income densities 58
2.4 Data and specification issues 63
2.5 Empirical results 67
2.5.1 Explaining changes between 1999/2000 and 2005/2006 67
2.5.2 Explaining changes between 1999/2000 and 2007/2008 79
2.6 Conclusion 82
2.7 Appendix 84
2.7.1 Details on the simulation of transfer changes 84

2.7.2 Additional Tables 87

IV THE INCOME DISTRIBUTION AND THE BUSINESS CYCLE IN GERMANY - A SEMIPARAMETRIC APPROACH 97
3.1 Introduction 99
3.2 Literature 101
3.2.1 The Theoretical Role of Macroeconomic Conditions 101
3.2.2 Empirical Approaches 103
3.3 Methodology 106
3.4 Data and Descriptive Analysis 110
3.5 Results 114
3.5.1 The Income Distribution and Macro-economic states 114
3.5.2 Counterfactual Analysis 121
3.6 Conclusion 126
3.7 Appendix 128
3.7.1 Subsampling and speed of convergence 128
3.7.2 Complementary Information 131

V DISSERTATION SUMMARY AND CONCLUSION 133

BIBLIOGRAPHY 137

LIST OF FIGURES 149

LIST OF TABLES 152

ACRONYMS 155

PUBLICATION NOTES 157

# Part I

# DISSERTATION INTRODUCTION

## Dissertation Introduction

There is a manifold of possible ways to analyze the state of the economy of a country and its development over time. Microeconomic approaches based on personal level data can generally help to describe the economic state from the point of view of the country's residents and economic agents. In a capitalist state, it is most natural to link residents and economic agents to economic mechanisms via a personal level productivity proxy, like monetary income. The overall structure of monetary income in a state can be represented by the size distribution of incomes and its analysis can reveal substantial economic mechanisms to understand the economic system in a state in general. This doctoral thesis is concerned with the analysis of the income distribution in Germany from three different perspectives.

### *A Satisfaction-Driven Poverty Indicator - A Bustle Around the Poverty Line*

Qualitative assessments of the income distribution frequently involve the classification of individual incomes into pre-defined social layers. While there are wide-spread and well-know definitions of such layers, like the one of the poor, the most widely used definitions are very ad-hoc in nature. The most prominent examples are the poverty line definitions, defining individuals as poor, when their incomes are below 60% of the median income (OECD) or if they receive less than 1.25 USD per day (World Bank), respectively. While there are certain admirable properties of these definitions, such as their easy interpretation, there is substantial need for a definition of poverty that is scientifically well-founded.

The first part of this doctoral thesis gives a new definition of poverty, based on a data-driven approach, derived from the subjective self-assessment of individuals regarding the valuation of their own income. In accordance with this, the derived 'Satisfaction-Driven Poverty Line' (SDPL) serves as a separator between poor and non-poor in the space of monetary incomes. While some approaches to define poverty lines based on individual self-assessment exist, e.g. the so called

Leyden Poverty Line of Goedhart et al. (1977) or the Subjective Poverty Line as in Kapteyn et al. (1988), they have undesirable properties. The most important limitation of these two approaches is the following. While the SDPL approach in this thesis only assumes that "people are capable and willing to give meaningful answers to questions about their [own] well-being" (Frey and Stutzer (2002)), the former two concepts implicitly make the assumption that people are also willing and able to give meaningful answers to questions about hypothetical life situations, that they are not living in themselves. The latter assumption may be seen as being unrealistic in real-life situations and being able to circumvent this problem is an important novelty.

To motivate the forthcoming poverty line definition, imagine two citizens with equal attributes and life circumstances. Imagine that both of them live in identical homes paying the same rent. Imagine now, one of the two had the exact amount of money to pay his rent and the other one had just one euro less. Unlike the first one, the second one is forced to leave his home, let us say, due to a very rigid contract, and is confronted with an uncertain future.

While this story sketches real life just in a very simplified manner, it indicates an everyday life situation people may be confronted with when they lack the money to pay for their basic needs. From the view of an economist, we witness two situations with just a marginally small difference in incomes, resulting in two completely different utilities and most probably in two strongly different life satisfactions. This means that the marginal utility of the above mentioned euro is unusually high in this situation. Starting from the hypothesis that comparable situations occur frequently in society, this part of the thesis is concerned with the question, whether such a phenomenon also exists on an aggregate level. If there is evidence for a significant jump in the utility of income just above a common basic-needs level, it can be interpreted as an income poverty line as it divides the population into two domains of utility measured by incomes. Following this motivation, the SDPL is defined as the best dichotomization of income to explain self-reported income-dissatisfaction.

In addition to finding an SDPL for Germany, our method also turns out to be able to test for another question of comparable importance. Known poverty line definitions in the space of monetary income make the implicit assumption that drawing a single line between the poor and the non-poor leads to a meaningful separation, saying that there exists a sharp separation of this kind. Alternatively, it might be that a poverty transition area definition would make more sense. The approach in this doctoral thesis is able to statically test the assumption that the poverty separator has actually more the characteristic of a line, which is also a considerable novelty.

Finally, the approach presents another novelty in that it can be used to establish and investigate cross-nationally valid poverty lines. Using data from the European Community Household Panel, further evidence is presented for satisfaction-based poverty lines across Europe and their cross-country differences are investigated. The results indicate that although the static 60%-definition is in line with the SDPL in Germany, the SDPLs of other European countries may be different. Nevertheless, the differences are shown to have a systematic character, depending on country specific inequality aversion and wealth level.

*Understanding Rising Income Inequality in Germany*

In the second part of this doctoral thesis, a quantitative analysis is conducted to investigate the factors behind the historic rise in income inequality and poverty in Germany around the turn of the millennium (1999-2006). During this time period, unemployment rose to record levels, part-time and marginal part-time work grew, and there was also evidence for a widening distribution of labor incomes. Apart from these changes, other factors may have contributed to the rise in income inequality, like the changes in the tax system (e.g. unequally reduced tax rates), changes in the transfer system (Hartz-reforms), changes in the household structure (e.g. a tendency to more single and elderly households), and changes in other socio-economic characteristics (like age or education). The question of which of these factors contributed to what extent to the observed inequality and poverty increase is of much interest to science and to the public debate.

There is some literature addressing specific aspects of the income distribution and its development in Germany. Most recently, Dustmann et al. (2009), Fuchs-Schündeln et al. (2010), Antonczyk et al. (2010) analyzed the effects derived from the changes in the labor market, Becker and Hauser (2006) and Arntz et al. (2007) the effects of the changed transfer system, while Peichl et al. (2012) analyzed the effects of the changed household structure.

As compared to existing research, the novelty of the approach in this thesis is to calculate the relative contribution of each of the factors to the change in the German income distribution in a unified framework. For this purpose, a counterfactual analysis is conducted, based on the semi-parametric kernel density reweighting method as proposed in DiNardo et al. (1996) and applied in Hyslop and Mare (2005) for New-Zealand and Daly and Valletta (2006) for the United States. As result, the thesis presents explicit ceteris paribus and sequential decompositions of the effects of the changed household structure, changed household characteristics, changed employment outcomes, changed labor market returns, changed transfer system and changed tax system, based on the pools of years 1999/2000 and 2005/2006. The results suggest that the most important factors behind the rise of income inequality are changes in employment outcomes, changes in labor market returns and changes in the tax system, while other factors play only a minor role.

*The Income Distribution and the Business Cycle in Germany - A Semi-parametric Approach*

The question of how the income distribution behaves given different macroeconomic conditions is important in order to disclose the relationship between micro- and macroeconomic processes. There is no generally applicable theory on how the distribution of incomes is influenced by changed macroeconomic conditions. The third part of this doctoral thesis adapts an empirical approach in order to investigate this relationship in Germany, based on data between 1996 and 2010. The method employed is a semi-parametric double-index model, pioneered by Ichimura and Lee (1991), without restrictions on the shape of the link function between indices of micro- and macro-level variables and individual incomes.

Until recently, prominent research in this area was conducted using aggregate summary measures of the income distribution. Blinder and Esaki (1978) and Jäntti (1994), for example, used distribution quantiles, while Thurow (1970), Salem and Mount (1974), or more recently Jäntti and Jenkins (2010) used distributional shape parameters to analyze the connection between macroeconomic states and the income distribution. While earlier work involved only simple system regressions, later ones applied cointegration approaches in order to avoid spurious results.

In this doctoral thesis, potential mechanisms between the macroeconomic variables GDP, inflation, government expenditure and unemployment on the one hand and the income distribution on the other hand are analyzed, following Farré and Vella (2008). In contrast to other approaches, this method allows for making use of the full available distributional information and individual level data, permitting to discover highly flexible macroeconomic effects on the income distribution. The results suggest that the influence of macroeconomic factors varies with individual characteristics and that the effects are in parts statistically significant, but are much less important for the shape of the income distribution than microeconomic factors.

# Part II

# A SATISFACTION-DRIVEN POVERTY INDICATOR - A BUSTLE AROUND THE POVERTY LINE

When it comes to poverty line definitions, the literature offers a wide variety of concepts. There is extensive work on the question of how to define a poverty line according to underlying axioms and philosophical or conceptional perspectives. In an empirical paper, Ravallion (2010) analyzes national poverty lines of 95 countries reviewing their strongly varying national concepts. One of his important observations is that while relative, income-based poverty line definitions are typically chosen in developed countries, developing countries mostly use consumption based and absolute measures.[1] The variety of national definitions mirrors different general concepts in poverty measurement such as the absolutistic 'basic needs concept' (related to physical survival) and the concept of 'relative deprivation' (related to social inclusion).[2] It turns out that empirical poverty measures can be seen as combinations of the two main concepts.[3] The classic example for an absolute poverty line definition is the international poverty line of 1.25 USD as proposed in Ravallion (2008), updating the previous value of the World Bank (1990) of 1 USD. A prominent example for a relative definition of poverty is the one used by the Statistical Office of the European Commission as 60% of the median income[4]. Such choices do not underrun a statistical optimization process in a rigid sense though.[5]

Moreover, a variety of approaches exists that are neither purely absolute nor relative. Prominent examples are called subjective poverty lines, see, e.g., Hagenaars and van Praag (1985).[6] These poverty lines depend on how individuals perceive poverty in society. Two well-known concepts are the Leyden Poverty Line of Goedhart et al. (1977) and the Subjective Poverty Line as in Kapteyn et al. (1988).

1 Focusing on consumption instead of monetary income helps to overcome the problem of measurability posed by agricultural production without explicit pricing.
2 See Atkinson and Bourguignon (2001).
3 This combination is also called 'weakly relative poverty', see Ravallion and Chen (2011).
4 Throughout this chapter we will refer to the equivalized nominal household income with the term 'income'.
5 Krämer (1994) investigates the source of this definition in more detail. He accounts the first appearance of such a definition to Fuchs (1967): "I propose that we define as poor any family whose income is less than one-half of the median family income.[...] no special claim is made for the precise figure of one-half."
6 Our poverty line definition belongs to this group.

The theoretic background of these approaches is appealing in many aspects. The suggestion that poverty line definitions should be based on answers to survey questions is central. Nevertheless, both concepts rely on strong assumptions. A basic conceptual difference between the two approaches above and our definition of the Satisfaction-Driven Poverty Line (SDPL) given in this chapter is the following one. While our SDPL approach is based on the hypothesis that "people are capable and willing to give meaningful answers to questions about their well-being"[7], meaning that direct questions on individual well-being and utility do make good proxies for their own 'true' utility, the former two concepts implicitly make a much stronger assumption, namely that people are capable and willing to give meaningful answers to questions about life situations that they are not necessarily living in themselves. Although 'meaningfulness' can mean less than statistical unbiasedness, our belief is that while people are experts of their own lifes[8], they are unlikely to be experts of other theoretical life situations.

To calculate the Satisfaction-Driven Poverty Line for Germany and twelve other European countries, we use a financial dissatisfaction indicator based on the variable 'satisfaction with household income' taken from the German Socio-Economic Panel (SOEP) and 'satisfaction with the household's financial situation' from the European Community Household Panel (ECHP). Controlling for a function of equivalized household income and other socio-demographic variables, we add an a-priori unspecified binary variable of income as an independent variable. Assuming, the true poverty line[9] has the unique property of best explaining income dissatisfaction, we identify the best choice of dichotomization in this sense and call it SDPL. Empirically, the poverty line is given as the maximizer of the goodness-of-fit of the underlying regression.
It is important to point out that this approach is not restricted to an absolute or relative definition of the poverty line.[10] A-priori, it may well be constant over time or a constant percentage point of the median or have any other behavior.

7 See Frey and Stutzer (2002).
8 This view is supported by recent research. For phychological justification see Diener et al. (1999) and Kahneman (1999) or Blanchflower and Oswald (2008) and Steptoe and Wardle (2005) for health related arguments. For a more general view see Ferrer-i Carbonell (2011).
9 The question of existence will also be investigated.
10 For why such choices may lead to too restrictive measures, see e.g. Ravallion (2010) p.17.

Furthermore, our approach does not rely on restricting assumptions concerning the explicit functional relationship between income and income utility. It is assumed though that reported income dissatisfaction is classified correctly, that the same classification translates to true (theoretical) monetary disutility and that the underlying disutility classification is best explained by true poverty. The reduction to monetary disutility arises naturally as our space of poverty classifications is itself restricted to income.

The aim of this chapter is to reveal a so far unexploited relationship between income satisfaction and income in order to construct a poverty line. Furthermore we show that this relationship contains a characterization of the widely-used definition of the poverty line by the Statistical Office of the European Commission at 60% of the median based on a statistically founded optimization criterion for Germany. Additionally, we show that our approach provides evidence for a sharp, discrete poverty line, rather than a fuzzy one[11]. For other European countries we show that the above relationship is not only a Germany-specific phenomenon, but it also exists in other European countries. There is evidence though that the country-specific optima deviate from the 60%-definition to some extent elsewhere. Nevertheless, results show that differences among countries in their estimated poverty lines can be explained quite well by the heterogeneity of macroeconomic characteristics.

The rest of this chapter is organized as follows. In Section 2 we introduce both data sets and motivate the choice of variables. Section 3 explains our methodologies in detail. Section 4 presents the empirical results using data from the SOEP for Germany and Section 5 the ECHP for international poverty lines. Section 6 concludes.

## 1.2 DATA AND VARIABLE SELECTION

The German Socio-Economic Panel (SOEP) is provided by the German Institute for Economic Research (DIW) in Berlin.[12] The SOEP is a representative yearly

11 An example for a class of such poverty lines can be found in Belhadj (2011).
12 For more details see Haisken-DeNew and Frick (2005) or Wagner et al. (2007).

panel study of private households in Germany. In 2009 almost 25,000 individuals living in about 11,000 households were interviewed.[13] The survey contains detailed information on a wide variety of personal and household level characteristics covering social, demographic, economic variables and variables of subjective well-being. The European Community Household Panel (ECHP) on the other hand is a representative panel survey provided by the Statistical Office of the European Commission, Eurostat. For a period of eight years, 1994-2001, households were interviewed to collect information on their income and living conditions on personal and household level on a yearly basis. The interviews cover a wide range of topics on living conditions such as income information, financial and housing situation, working life, social relations, health and biographical information.[14] In the first wave in 1994, a sample of around 60,500 households with around 130,000 adults were interviewed in twelve European Countries. Those countries were Germany, Denmark, the Netherlands, Belgium, Luxembourg, France, United Kingdom, Ireland, Italy, Greece, Spain and Portugal. Austria and Finland joined the project in 1995 and 1996 respectively and Sweden in the year 1997, based on the Swedish Living Conditions Survey. After 1996, German data was derived from the SOEP, data for Luxembourg from the Luxembourg Income Study and data for the United Kingdom from the British Household Panel Survey. Due to missing ECHP satisfaction variable we cannot make use of the data for Sweden. Also data for Germany, Luxembourg and the United Kingdom lack a satisfaction variable with ECHP-consistent definition when using national databases. Data for France cannot be used as household income is only available as gross income. With these limitations the full set of data with respondents of age 17+ with non-proxy interviews and available satisfaction and income data is used and we consider the following countries.

- Denmark, the Netherlands, Belgium, Ireland, Italy, Greece, Spain and Portugal (1994-2001),
- Austria (1995-2001), Finland (1996-2001),
- Germany, Luxembourg and the United Kingdom (1994-1996).

13 We did not use households with missing income information or incomplete age structure, but besides that we made no other restrictions.

14 For more on the ECHP, see http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/echp.

### *Variable selection*

In the field of happiness economics researchers frequently use direct questionnaire-based overall life satisfaction to proxy the ‘true’ overall welfare. At first glance it is hard to believe that happiness or unhappiness can be put in numbers in a proper and scientifically useable way by regular respondents. One may think that it is an impossible task to express the notion of happiness quantitatively, or one might think that - if it was possible conceptually - it is unlikely to work if non-experts are asked with only a couple of seconds to decide. However, research on the subject tells a quite different story. As summarized by Ferrer-i Carbonell (2011), there seems to be a strong statistical relationship between the proxy and true satisfaction. The connection can be revealed by using objective psychological measures of happiness, based e.g. on facial expressions or body language as investigated by Sandvik et al. (1993) and Kahneman (1999) or objectively measurable brain activity as reported in Urry et al. (2004). Furthermore, researchers found that there is evidence for the existence of a commonly shared context of happiness, so that the comparison of answers of different individuals is generally possible.[15] When putting this picture together, it seems to be that such ”... happiness measures are consistent, valid, and reliable. In sum, it appears that human happiness is a real phenomenon that we can measure.”[16]

*Income Satisfaction as dependent variable*

When it comes to subjective monetary poverty line definitions, personal utility of income or income-driven welfare is the center of interest. The limitation to this very aspect of welfare makes sense as the one-dimensional monetary approach is limited on its own, likely to be driven less substantially by information arising from income-complementary welfare aspects. While it is well-known that “money alone does not make happy”[17], shown by a weak - but significant - relationship between income and life satisfaction, we may expect much more in this regards when using income satisfaction instead. Indeed, the SOEP shows only a weak correlation of around 0.15 between income and life satisfaction, while the

15 See, for example, Van Praag (1991).
16 See Frank (2005).
17 See e.g. Ferrer-i Carbonell and Frijters (2004).

correlation is around 0.35 when using income satisfaction instead. On the other hand, it is important to see that while satisfaction with income has a rather specific formulation, responses are correlated with the ones about life as a whole to an extent of around 0.5. This gives rise to the assumption that income satisfaction is influenced by non-pecuniary life circumstances. Consequently, our particular interest lies in reported income satisfaction, the answer to the question "How satisfied are you today with your household income?" on a scale from 0 to 10. The scale in use here is called a bipolar 11-point-scale[18] with the numbers 0-2 explicitly tagged 'totally unhappy' and 9-10 tagged 'totally happy'.[19] Such a scale has advantages over other possible scales (see Abrams (1973) or the more recent Kroh (2005)). A typical distribution of the answers to this question is given in figure 1. The response rate of over 90% shows the willingness of the interviewees to answer to this question.[20] The fact that the scale is used in entirety without dominating corner solutions, suggests that the respondents are also willing to maximize the submitted information.[21]

As regards the ECHP, the data set offers satisfaction with the household's financial situation for the dependent variable. This is slightly different from the variable available from the SOEP. To obtain a comparable relationship to actual income it is more important to account for the overall financial situation.[22] We will go into detail on the additional controls used to handle this issue at the end of this subsection. Another important difference between the two satisfaction variables is that while the SOEP variable is given on an 11-point-scale, the ECHP variable is given on a 6-point scale. This may be a limitation leading to less detailed results.

18 For more on the origins of the scale see, e.g., Wagner (2007)).

19 As opposed to a bipolar scale a unipolar scale would run from 'totally not happy' to 'totally happy', that is from the total absence to the total presence of the same issue. This can be seen as restrictive if the lack of happiness is not necessarily interpreted as unhappiness.

20 For our analyzes we make use of the unrestricted sample including every respondent to the income satisfaction question with known household income.

21 The opposite would be the case if people just used 0 or 10 to communicate their being unhappy or being happy respectively, as we would not be able to differentiate between answers within the two groups.

22 This is less important when using panel data as slowly changing characteristics such as wealth can be netted out.

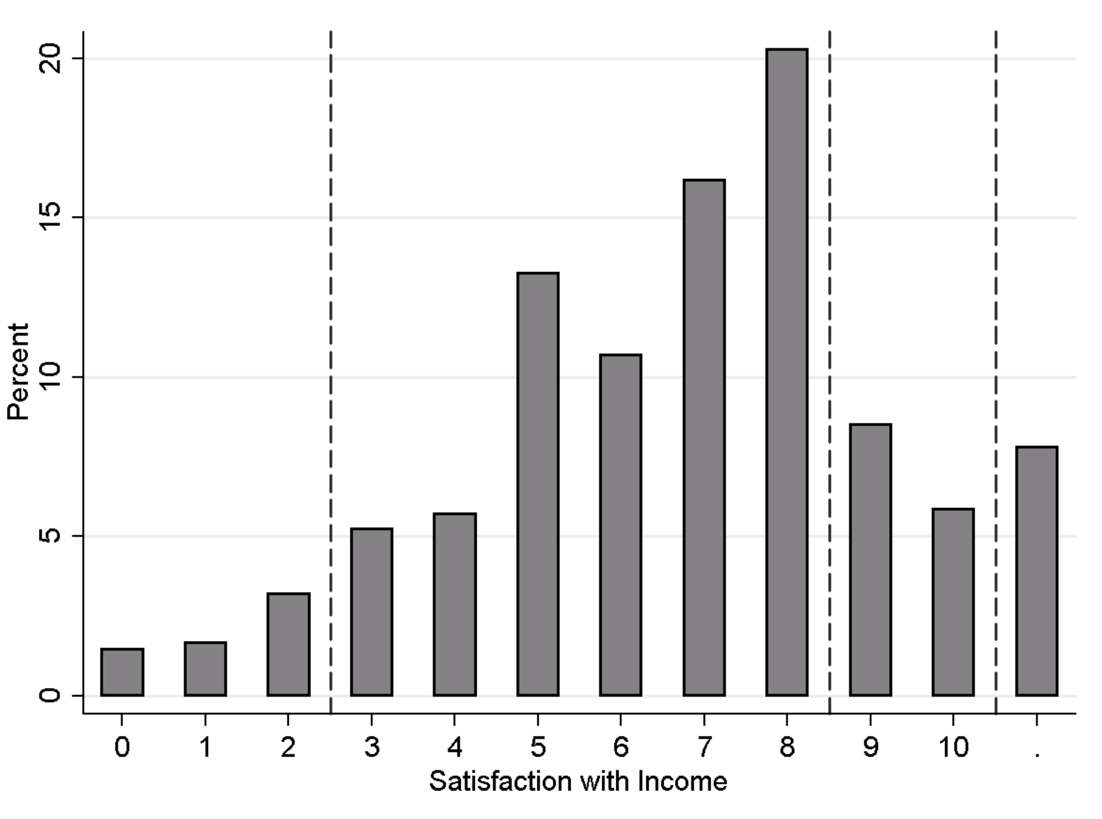



Figure 1: Distribution of answers to the SOEP income satisfaction question in 2009. Dashed lines separate the three sections outlined by the questionnaires formulation and the missing values. Own calculations.

*Income as independent variable*

A central point is to explain income-driven welfare based on household income. Using the household income instead of the individual income is a very common approach. Households are well-defined economic units that usually share their income sources and exhibit economy-of-scale effects. Both have to be accounted for when calculating the households' individual incomes. The most common approach is to apply equivalence scales[23] to account for the intra-household redistribution of income. Once we have accounted for this redistribution, a third

23 We use the commonly used scale, called new OECD scale, assigning the weight 1 to the household head, 0.5 for every other adult in the household and 0.3 for every child, but we also check for the robustness of this choice by using the Luxembourg scale with no qualitative difference in the results.

degree polynomial[24] of the resulting equivalized personal income is used to explain the income-driven satisfaction.

*Direct effects on income perception*

To uncover the direct relationship between income and income satisfaction, we control for other factors that may influence reported income satisfaction. Firstly, the overall financial situation is not captured by income, but may well have an important effect on answers, e.g. as high wealth makes a person less dependent on income.[25] We take two measures to control for this. As we do not have yearly data on financial wealth[26], we use house ownership as proxy. Furthermore, fixed effects approaches will net out such effects, assumed they are time constant. Secondly, we notice that relative income perception may be a part of the relationship between income and income satisfaction. To model such effects though, relevant social groups that the individual relates to and potentially compares himself with need to be identified.[27] Generally, it is not obvious which group a person relates to (colleagues, friends, neighbors, family members, etc.). However, we control for a very general way of relative income effects by including the unspecified poverty indicator in our regressions, which will also pick up relative income concerns as shown in section 5 (by the significance of the Gini coefficents in the country panel).

A third point that can be efficiently controlled for is the source of income. It may be important whether people work for their income on their own or whether they receive income from the state, e.g., in form of social welfare. Besides the direct monetary effects, strong non-monetary effects are expected. The latter effect may also be interpreted as undervaluation of the income, e.g., from social welfare. For more details concerning different effects of unemployment benefits see Winkelmann and Winkelmann (1998). Here we control for the number of months in the previous year the household received social welfare, unemployment benefits

24 Robustness tests were performed by using only a second degree polynomial and by using logarithmic income instead of a polynomial. This led to very similar results.
25 For more details see e.g. Headey and Wooden (2004).
26 The SOEP only contains detailed wealth related data for two selected years.
27 For more details see D'Ambrosio and Frick (2007), Clark and Senik (2010), Clark and Oswald (1996) or Luttmer (2005) among others.

and unemployment assistance. Furthermore we control for part-time employment, retirement status, unemployment and being out of labor force. A further and last point is income adaptation that generally plays a role for subjective income valuation.[28] In this regards though, we refer to Di Tella and MacCulloch (2008), who find that income adaptation is likely to be less pronounced for poverty-relevant subgroups. Nevertheless, we control for the number of month, the household received social welfare as mentioned above. This also controls for potential adaptation effects for the recipients. Furthermore, long term adaptation effects are netted when using fixed effects.

*Indirect effects on income perception*

Another set of controls has a more indirect impact on income perception. It is rather connected to complementary aspects of overall welfare as opposed to income-driven welfare. It is likely that impacts on aspects of life other than income may indirectly influence the perception of income, leading to an alteration of its perception. For example, people may not really refer to their income only if asked, but they may also be influenced and biased by other circumstances. Four subgroups that we see as relevant in this regards are attitude and personality, major life events, health and social network. Attitude and personality summarize the individual specific view of the world. Attitude typically cannot be measured directly, but assuming its mainly constant character it can be accounted for when panel data is available. As numerous other aspects of individual personality traits can be at least indirectly controlled for, we account for this group with the variables gender, age and education, where we use a dummy of high school degree and a dummy for apprenticeship or German Abitur as highest degree for education. Additionally, we control for being in education at the time of the interview.

Major life events are positive or negative shocks which may also bias the answers on income satisfaction. For example, if the beloved dog of the family dies just the day before the interview, the interviewee's answer to the question may be down-

28 For an early work on adaptation see Brickman et al. (1978) or for an ongoing discussion Ferrer-i-Carbonell and Van Praag (2009).

ward biased due to his bad mood.[29] As controls we use a dummy for a child born in the past six months, newly married in the last twelve months, divorced in the last six months, newly disabled in the last twelve months. Furthermore we control for deaths in the household in the last twelve months and children under fourteen that left the household in the last six months. In addition, we control for events in the household's near future. This clearly only makes sense for events that people can anticipate and that can be associated with psychological effects. Our choice here is to take divorce in the next twelve months and marriage in the next six months. Another category is health. Here we control for the number of days spent ill in the last year and the disability grade. The last category is social network. Controls belonging to this set relate to information about other individuals that the individual in question is aware of. While it is hard to characterize the influence of the social network in general, we may say that if a person is living with a social network that is in harmony with his personal attributes, he should feel happier than someone, who's social network is very different from the desired one. This can be people who desire more friends or a bigger family. To control for the social network, we control for marital status including being separated and children including age structure. We control for children between zero and three years, four to eleven years and twelve to seventeen years in the household with dummy variables. Controlling for other effects requires knowledge about the relevant social structure and the perception of it, both of which are likely to vary with the individual and are beyond the scope of our work. However, also here, fixed effects will net out such effects to a fair extent.

As already pointed out, there are some differences in the choice of controls when using the ECHP. The differences result from the different data sets on the one hand and from the fact that the financial situation has to be accounted for more explicitly on the other. To additionally control for the latter, we use the quartile membership of household capital income and rental income, debts and loan dummies and house ownership. Furthermore we control for unemployment, unemployment frequency since 1989, highest education dummies, disability in two stages of severity, new deaths, new births, marital status, employment, retire-

29 For more on the direct effects of such events on general life satisfaction see Mentzakis (2011).

ment, positive and negative subjective income shock since last year and EU and non-EU foreigner dummies. Finally, we control for the quartile position in the distribution of doctor visits as an objective control for health.

## 1.3 METHODOLOGY

### 1.3.1 *Motivation*

There is a wide range of econometric approaches used in recent literature about self-reported satisfaction. Much has changed due to the availability of new methods, advance in computation and due to new insights into the nature of the satisfaction variable. Typically, in the field of psychology simple cross sectional regression approaches were used to explain the role of different influence factors on self-reported satisfaction. Such approaches clearly rely on the cardinality assumption of the dependent variable.[30] If we rather want to assume ordinality only, nonlinear models like ordered logit models are the methods of choice. Such models have been widely used in economic literature, see e.g. Blanchflower and Oswald (2004). A drawback of cross-sectional approaches is that they cannot account for time constant individual heterogeneity. When assuming ordinality, the fixed effect logit model in Chamberlain (1980) can be used to overcome this problem. For an application see Winkelmann and Winkelmann (1998). As the latter model needs a binary dependent variable, a dichotomization of the satisfaction variable has to be used, generally leading to a loss of information.

According to Huppert and Whittington (2003), an additional issue arises if self-reported satisfaction is used. They state that the determinants of low satisfaction (dissatisfaction) and high satisfaction are different. For their analysis they use two distinct satisfaction scales: one explicitly describing well-being, the other one describing the presence and absence of symptoms for mental and physical health related problems. As standard ordered logit based models are structurally limited through their single crossing property[31], they are likely to be too restrictive to capture heterogenous impacts according to the above findings properly.

30 For examples see e.g. Diener et al. (1999).
31 See Maddala (1983) for more details.

Based on the cross-sectional generalized ordered probit model in Terza (1985), Boes and Winkelmann (2010) introduce a panel based extension with correlated random effects, assuming ordinality, while accounting for scale heterogeneity and unobserved individual heterogeneity.

An aim of our study is to reveal a relationship between income satisfaction and income, implicitly defining an income poverty line. We start by defining the poverty line in a strict binary sense based on a dummy variable of income. Starting from a relationship given by a third degree polynomial in income and controls mentioned in the last section, we introduce an additional explanatory variable in form of the above mentioned dummy without defining its exact position. We have already motivated the use of satisfaction with income instead of general life satisfaction for the estimation of the satisfaction-driven poverty line (SDPL) in the last section. As we are only interested in the relationship between poverty and clear income dissatisfaction, we use a dichotomization of income satisfaction instead of income satisfaction itself. The resulting binary variable has the value 0 for all people that are dissatisfied with their incomes and 1 for all others. As the SOEP's income satisfaction question explicitly labels the answers 0, 1 and 2 as 'totally unhappy', we see our choice to take the cutoff point 2 as well-founded.[32] In the case of the financial satisfaction variable of the ECHP, choice is much more limited. Here only the value of 1 is explicitly labeled as 'not satisfied at all', making other choices speculative. We therefore chose 1 as our cutoff point. Using the binary dependent variable defined above and a simple regression yield a linear probability model in which the marginal effect of the dummy variable for income has the straightforward interpretation of a change in probability to not being totally unhappy with income. Such a jump in probability can then be interpreted as the location of leaving poverty as argued in the first section. We define the best dummy variable to explain the relationship between income and income satisfaction as the satisfaction- driven poverty indicator.

32 Additionally, it turns out that taking 1 or 3 do not make a qualitative difference.

The dichotomization of income satisfaction has the additional advantage of circumventing the scale heterogeneity problem mentioned above.[33] Otherwise we would need to deal with generally different impacts of income on different levels of satisfaction to avoid mixture-driven outcomes. In addition to this, there are some computational reasons for us to prefer models based on the cardinality assumption. To perform a detailed grid search on the dummy variable, we need to repeat the model estimations a large number of times. While maximum likelihood methods generally take much longer time to calculate, a number of them also suffer from numerical drawbacks due to local extrema and saddle points with suboptimal solutions that are hard to identify. This problem is especially pronounced with high numbers of repetitions leading potentially to suboptimal grid maxima. Fortunately, empirical research frequently comes to the conclusion that while the ordinality assumption is the right choice in theory, there are no practical differences to the results when compared to cardinality-based methods. See Boes and Winkelmann (2010) for such a conclusion. Following these arguments we use the cardinality assumption in our research leading to linear models when performing grid search.[34] As controlling for individual heterogeneity is also empirically important in general[35], we extend our analysis to the use of fixed effects panel models. Additionally, we introduce a more flexible approach based on nonlinear least squares both in the cross-sectional and the fixed effects panel setting.

### 1.3.2 *The Grid Search Approach*

The idea of the following approaches is to estimate the poverty line as the income value that maximizes the goodness-of-fit of a model explaining income satisfaction by the variables defined above. Assume the following underlying relationship

$$s_{it} = \beta_0 + \mathbf{x}_{it}'\beta + g(\pi, \iota_{it})\gamma(\pi) + u_{it}, \quad i \in \{1, \dots, N\},\ t \in \{1, \dots, T\}, \tag{1.1}$$

33 See Boes and Winkelmann (2010).

34 Nevertheless, for robustness checks we use a cross sectional logit model for comparison, yielding very similar results.

35 See, for example, Ferrer-i Carbonell and Frijters (2004).

with $i$ as cross sectional and $t$ as time index respectively, $\beta_0$ as intercept, $\beta := (\beta_1, \beta_2, \ldots, \beta_K)'$ as $K \times 1$ real vector of coefficients, $\mathbf{x}_{it} := (x_{it1}, x_{it2}, \ldots, x_{itK})'$ as independent variables, $s_{it}$ as the dependent variable income satisfaction, $\iota_{it}$ as the income of person $i$ at time $t$ and $u_{it}$ as an additive error term for all $i$ and $t$.[36] Additionally, let $g(\pi, a) := \mathbb{1}_{\{a>\pi\}}(a)$ be a dichotomous function and $\gamma(\pi)$ a real valued coefficient that generally depends on the value of $\pi$, where $\pi$ is the point of income dichotomization.
In the special case of $T = 1$ we have

$$s_i = \beta_0 + \mathbf{x}_i'\beta + g(\pi, \iota_i)\gamma(\pi) + u_i, \tag{1.2}$$

which can be estimated by ordinary least squares given the usual regularity conditions for any fixed, positive $\pi$ yielding the usual

$$R^2(\pi) = \max_{\beta,\gamma(\pi)} \sum_{i\in N} (\mathbf{x}_i'\beta + g(\pi, \iota_i)\gamma(\pi) - \bar{s})^2 \cdot \left(\sum_{i\in N} (s_i - \bar{s})^2\right)^{-1} \tag{1.3}$$

as solution of the maximization problem with $R^2(\pi)$ as the standard goodness-of-fit measure and $\bar{s}$ as the arithmetic mean of the $s_i$.
It follows that for a finite grid $G \subset \mathbb{R}$, $\#G < \infty$, we can extend our maximization problem and get

$$R_G^2 := \max_{\pi\in G} R^2(\pi) = \max_{\beta,\gamma,\pi} \sum_{i\in N} (\mathbf{x}_i'\beta + g(\pi, \iota_i)\gamma(\pi) - \bar{s})^2 \cdot \left(\sum_{i\in N} (s_i - \bar{s})^2\right)^{-1} \tag{1.4}$$

yielding the SDPL

$$\hat{\pi} := \text{argmax}_{\pi\in G} R^2(\pi), \tag{1.5}$$

i.e. the poverty line that makes the relationship between income satisfaction and poverty status as strong as possible.

Consider now another configuration that makes use of a possible panel structure of the data. Given $T > 1$ and putting $u_{it} := \nu_i + \epsilon_{it}$ as the sum of individual

36 We suppress the $i$ in $T_i$ allowing for unbalanced panels for notational simplicity.

unobservable heterogeneity $\nu_i$ and an idiosyncratic error term $\epsilon_{it}$ in (1.1) we obtain

$$s_{it} = \beta_0 + \mathbf{x}'_{it}\beta + g(\pi, \iota_{it})\gamma(\pi) + \nu_i + \epsilon_{it}, \quad i \in \{1, \dots, N\}, \, t \in \{1, \dots, T\}. \quad (1.6)$$

As the above strategy remains basically unchanged for either of the usual panel data models using least squares estimation, we keep the exposition brief. It is important to point out that $g(\pi, \iota_{it})$ is typically time variant, so time demeaning or usual differentiating strategies do not endanger the identifiability of $\hat{\pi}$. Furthermore, the goodness-of-fit measure we use is the $R^2$ based on the time demeaned least squares regression, also called 'within' $R^2$.[37]

While the above approach is conceptually simple, it has two main drawbacks. The first - theoretical - one is that the solution generally depends on the choice of G and it is unlikely to equal the global solution on $\mathbb{R}$. The second - practical - one is that it may require a fair amount of computational time to calculate #G regressions, especially if panel data models are involved in the calculation. Both of these problems can be solved as shown in the next subsection.

### 1.3.3 *The Nonlinear Least Squares Approach*

To put the grid search in a more familiar econometric setting consider the following cross sectional relationship

$$s_i = f(\tilde{\mathbf{x}}_i, \tilde{\beta}) + u_i \quad (1.7)$$

with f as a not necessarily linear function of the independent variables $\tilde{\mathbf{x}}_i$, a vector of coefficients $\tilde{\beta}$ and an additive error term $u_i$. In our case let

$$f(\tilde{\mathbf{x}}_i, \tilde{\beta}) = \beta_0 + \mathbf{x}'_{it}\beta + \mathbb{1}_{\{a>\pi\}}(a) \cdot \gamma(\pi) \quad (1.8)$$

To estimate the coefficient vector $\tilde{\beta} = (\beta, \gamma, \pi)$, a nonlinear least squares (NLS) approach can be carried out given sufficient regularity conditions and conditions

37 Also compare equation (1.14) below.

of identification as described in Wooldridge (2002), Chapter 12. While most of the conditions are rather unproblematic for empirical application, in our special case $f$ violates the necessary condition of being continuous on the parameter space. This is why (1.7) given (1.8) cannot be estimated directly.

To overcome this problem we choose a function that is very similar to the indicator function but is sufficiently smooth for a consistent NLS estimation. It can be seen easily that the cumulative distribution function (CDF) of the logistic distribution

$$\Lambda(x, \alpha, \beta) := \frac{1}{1 + e^{-(x-\alpha)/\beta}} \tag{1.9}$$

is a smooth function with location parameter $\alpha$ and scale parameter $\beta$. Furthermore it can be shown that for all $\alpha \in \mathbb{R}$

$$\Lambda(x, \alpha, \beta) \underset{\beta \searrow 0}{\longrightarrow} \mathbf{1}_{\{x>\alpha\}}(x)$$

pointwise, making the choice especially attractive. Nevertheless, there are other choices that approximate the indicator function even better. Note that $\Lambda(x, \alpha, \beta) \neq \mathbf{1}_{\{x>\alpha\}}(x) \forall x \in \mathbb{R}$ given any $\beta > 0$ as the CDF of the logistic distribution never equals zero or one for finite $x$. This problem is not present if one takes

$$\Pi(x, \alpha, \beta) := \begin{cases} 0 & \text{if } x < \alpha \\ 10 \cdot (\frac{x-\alpha}{\beta})^3 - 15 \cdot (\frac{x-\alpha}{\beta})^4 + 6 \cdot (\frac{x-\alpha}{\beta})^5 & \text{if } \alpha \leqslant x \leqslant \alpha + \beta \\ 1 & \text{if } x > \alpha + \beta, \end{cases} \tag{1.10}$$

constructed to be sufficiently smooth in $\alpha$ and $\alpha + \beta$ and again

$$\Pi(x, \alpha, \beta) \underset{\beta \searrow 0}{\longrightarrow} \mathbf{1}_{\{x>\alpha\}}(x).$$

The advantage of this function is that it equals the indicator function on the complement of its middle section for all choices of $\beta > 0$. Besides that, in the next section we can use both of the above approaches for robustness checks.

For a clearer comparison of $\Lambda$, $\Pi$ and the indicator function see figure 2 for an illustrative example.

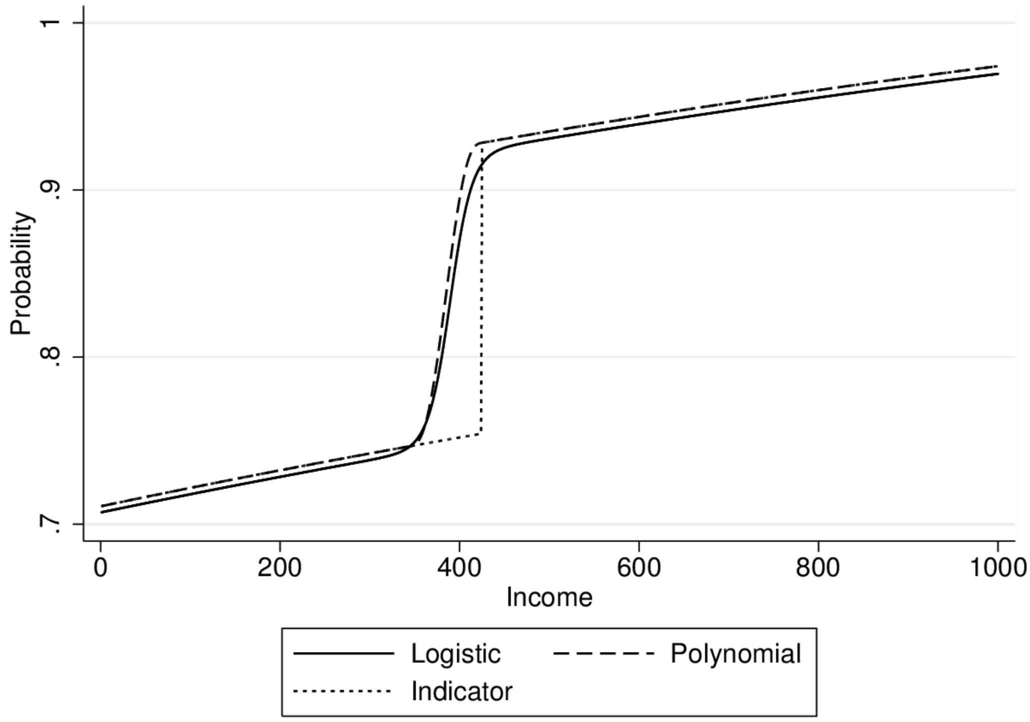



Figure 2: Illustrative example of the indicator smoothing concept applied to the relationship between income satisfaction and income based on real SOEP data of the year 1985. Here the jump function is combined with a third degree polynomial. Own calculations.

When inserting either (1.9) or (1.10) into (1.8) the model parameters can be estimated by nonlinear least squares. Besides that, we also gain a more flexible model with the additional parameter $\beta$. While in the case of $\Pi$ we have a direct interpretation of $\beta$ as the length of the transition region between $0$ and $1$, a similar interpretation applies in the case of $\Lambda$ if we define the length of the transition region as the one between $\Lambda^{-1}(\epsilon)$ and $\Lambda^{-1}(1-\epsilon)$, with $\epsilon = 0.05$ for example, as the values zero and one are never achieved. This yields for $\alpha = 0$

$$\Lambda^{-1}(0.95, 0, \beta) - \Lambda^{-1}(0.05, 0, \beta) = 2 \cdot \Lambda^{-1}(0.05, 0, \beta) \approx 5.889 \cdot \beta, \tag{1.11}$$

without loss of generality. If we take a look at figure 2 we may recall the 3 elements of the region of interest explicitly. The height of the jump that we may also interpret as the marginal probability effect of the state change of being poor by SDPL definition to not being poor by SDPL definition is represented by the vertical section of the dotted line and is of the magnitude of around 17%. It means that given a ceteris paribus state change to being not poor comes along with a 17% higher probability, namely around 92% probability, of not being dissatisfied with income. The second element is the jump distance or the length of the poverty transition region (PTR). While it is zero in the indicator case, it is around 80 Euros in the logistic and polynomial cases, meaning that in this case changing one's state from poor to not poor is not a discrete phenomenon. The third element is the location of the jump. In the logistic and polynomial cases it may be defined as the end of the jump for example and in the indicator case it is more clearly the point of dichotomization. The number here is somewhat around 400 Euros, the estimate for the SDPL for Germany in 1985.

### 1.3.4 *The Fixed Effects Nonlinear Least Squares Approach*

Given

$$s_{it} = f(\tilde{x}_{it}, \tilde{\beta}) + \nu_i + \epsilon_{it} \tag{1.12}$$

time demeaning leads to

$$\ddot{s}_{it} = \ddot{f}(\tilde{x}_{it}, \tilde{\beta}) + \ddot{\epsilon}_{it} \tag{1.13}$$

with $\ddot{s}_{it} := s_{it} - 1/J \cdot \sum_{j=1}^{J} s_{it-j+1}$, $\ddot{f}_{it} := f_{it} - 1/J \cdot \sum_{j=1}^{J} f_{it-j+1}$ and $\ddot{\epsilon}_{it} := \epsilon_{it} - 1/J \cdot \sum_{j=1}^{J} \epsilon_{it-j+1}$ respectively for time horizon J.

While there is no simplified expression for $\ddot{f}$ in general, estimation is still possible if f follows necessary regularity conditions emerging directly from the cross sectional NLS case. Minimization of the empirical counterpart of $\ddot{\epsilon}_{it}$ is equivalent to minimization of the population variable $E(\epsilon_{it}^2)$ in (12) because

$$E\left(\ddot{\epsilon}_{it}^2\right) = E\left(\epsilon_{it}^2\right) \cdot (1 - 1/J) \tag{1.14}$$

for all fixed J.[38]

We call this approach fixed effects nonlinear least squares approach (FENLS) throughout the rest of this chapter. Using this model is particularly desirable as it both accounts for individual heterogeneity and posses a higher degree of freedom to adjust to the problem at hand, allowing to test for additional hypotheses.

## 1.4 EMPIRICAL RESULTS

### 1.4.1 *SDPLs based on the SOEP*

After applying the grid search method as described in Section 3 to the SOEP, we obtain results as shown in the Figures 3a-d and 4a-d.[39] Each plot can be interpreted as follows. The value zero is the baseline $R^2$ of the underlying regression, without the poverty dummy. The additional gain in $R^2$ is given by the solid line with the poverty line set for the value as depicted on the x-axis. The highest gain is normalized to one. The grid applied to the search is between 150 and 4000 Euros in nominal value in 5 Euro steps. The first four figures show grid search results based on the linear probability model for the years 1985, 1995, 2005 and 2009. The last four figures show grid search results based on the fixed effects model of three consecutive years. Results here are labeled with the middle year as 1986, 1996, 2006 and 2008. The dotted vertical lines show 60% of the median. Despite the time span of 25 years, results are very robust. Dichotomizations at around 60% of the median deliver the highest explanatory power of the variable with pronounced unimodal shapes from 1994.[40] Both cross sectional and panel approaches yield very similar results.

---

38 See Wooldridge (2002), equation (10.51).
39 The results shown are also qualitatively representative for the years not depicted.
40 Two consecutive years shortly after the German unification are generally the only bimodal exceptions.

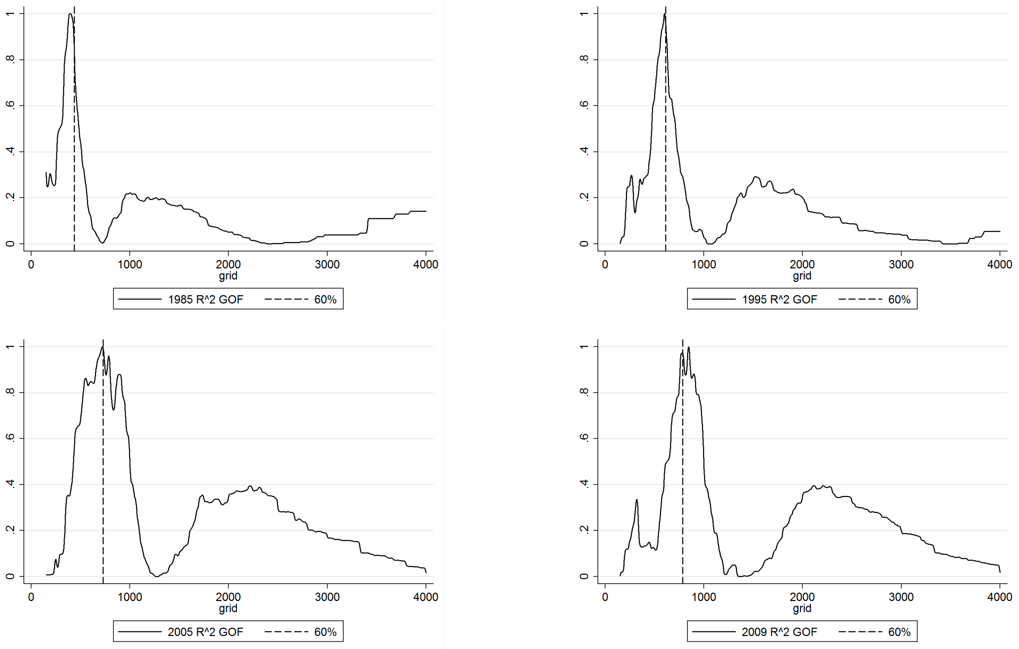



Figure 3: Grid search results in 5 Euro steps using the linear probability model. Baseline is set as the $R^2$ of regressions without dummy. The highest value is normalized to 1. Top left (a) result for 1985, top right (b) for 1995, bottom left (c) for 2005 and bottom right (d) for 2009. Own calculations.

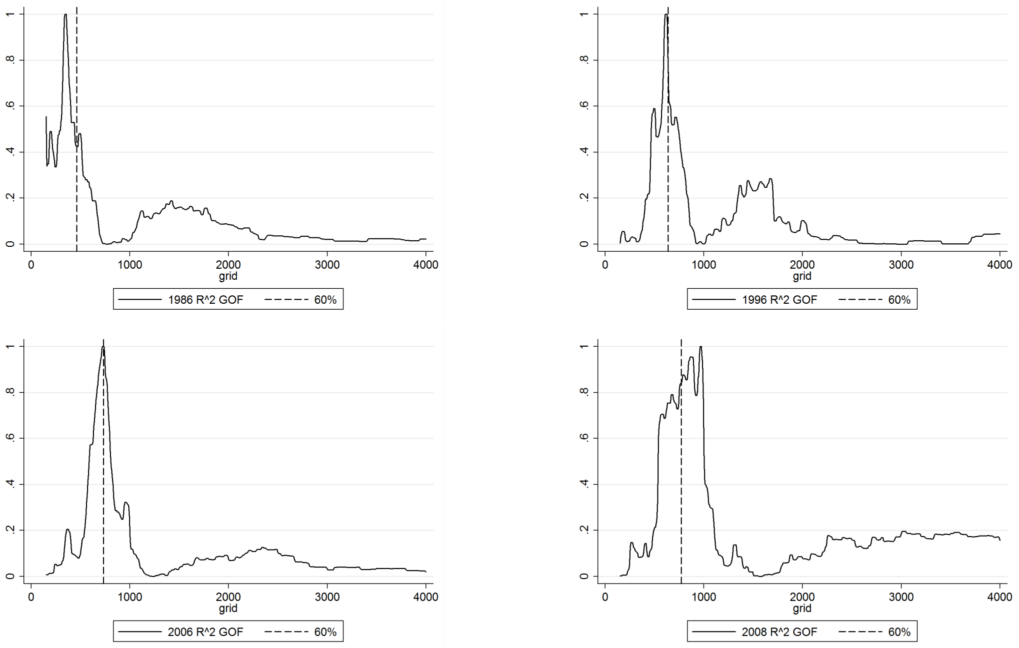



Figure 4: Grid search results in 5 Euro steps using the 3-years fixed effects panel model. Baseline is set as the $R^2$ of regressions without dummy. The highest value is normalized to 1. Top left (a) result for 1985-1987, top right (b) for 1995-1997, bottom left (c) for 2005-2007 and bottom right (d) for 2007-2009. Own calculations.

| dep. var. : income satisfaction | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| income | .0002088*** | .0001552*** | .0002376*** | .0001902*** |
| $income^2$ | -6.41e-08*** | -3.59e-08*** | -5.69e-08*** | -4.29e-08*** |
| $income^3$ | 5.69e-12*** | 2.00e-12*** | 3.24e-12*** | 2.25e-12*** |
| **incdum** | .0902269*** | .0799031*** | | .0873776*** |
| married | .0475478*** | -.0005017 | 6.54e-06 | |
| separated | -.0890270** | -.0265090 | -.0270401 | |
| sex | .0104854* | | | |
| age | -.0046393*** | | | |
| $age^2$ | .0000521*** | -.0000744 | -.0000708 | |
| e | -.0179378** | | | |
| child 0-3 | .0163940 | -.0053448 | -.0011421 | |
| child 4-11 | .0080097 | .0365091** | .0381299** | |
| child 12-17 | .0226381*** | .0288710*** | .0289072*** | |
| jobless X jobless other | | | | |
| 0 1 | -.0041254 | -.0024654 | -.0025651 | |
| 1 0 | -.1467259*** | -.1056438*** | -.1100456*** | |
| 1 1 | -.1323874*** | -.1583044*** | -.1674143*** | |
| in education | -.0096857 | -.0317808** | -.0308935** | |
| retired | -.0002229* | -.0232619 | -.0199658 | |
| part-time | .0032601 | -.0316423* | -.0313248* | |
| out of labor force | -.0238282 | -.0781198*** | -.0787139*** | |
| owner | .0213148*** | | | |
| months unempl. benefits | -.0017560 | .0007395 | .0008081 | |
| months unempl. assistance | -.0058263 | -.0078976** | -.0080673*** | |
| months social assistance | -.0030641 | .0015851 | .0010559 | |
| university | -.0232552** | | | |
| abitur/voc. training | -.0096111 | | | |
| disability degree | -.0002323* | .0000624 | .0000258 | |
| ill | -.0003014* | 1.25e-08 | -4.74e-06 | |
| newborn | -.0943048* | | | |
| newly married | -.0582913** | | | |
| newly disabled | -.0338399 | | | |
| new death | .0840644*** | | | |
| newly divorced | -.2116805** | | | |
| new child u14 left | -.0373934 | | | |
| soon divorced | .0673761 | | | |
| soon married | .0676058*** | | | |
| _cons | .7468746*** | .9253191*** | .9192893*** | .6858231*** |
| Observations | 13151 | 35771 | 35771 | 35771 |
| $R^2$ | 0.1316 | 0.0390 | 0.0341 | 0.0247 |
| Location of Maximum | 675 | 675 | - | 675 |

* p<0.1 ** p<0.05 *** p<0.01

Table 1: Year 1998, income satisfaction dummy as regressand. (1) LPM with the best income dummy fit, (2) 3-years FE panel model with the best income dummy fit, (3) as (2) without dummy and (4) as (2) without additional controls. Robust standard errors. Own calculations.

| dep. var. : income satisfaction | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| income | .0000527*** | .0000437*** | .0000795*** | .0000530*** |
| $income^2$ | -3.55e-09*** | -2.61e-09*** | -4.70e-09*** | -3.09e-09*** |
| $income^3$ | 2.37e-12*** | 3.23e-14*** | 5.79e-14*** | 3.78e-14*** |
| **incdum** | .1337918*** | .0900829*** | | .1010768*** |
| married | .0219009*** | .0271412 | .0139223 | |
| separated | -.0000279 | .0329272 | .0369348 | |
| sex | .0115466* | | | |
| age | -.0072594*** | | | |
| $age^2$ | .0000676*** | -.0000358 | -.0000437 | |
| e | -.0061185 | | | |
| child 0-3 | .0391889*** | .0484145*** | .0519664** | |
| child 4-11 | .0095827 | .0223486 | .0239098 | |
| child 12-17 | .0090761 | .0059450 | .0074916 | |
| jobless X jobless other | | | | |
| 0 1 | -.0105735 | -.0196849 | -.0230573 | |
| 1 0 | -.1265161*** | -.1656058*** | -.1752323*** | |
| 1 1 | -.0376719 | -.1406440*** | -.1559932*** | |
| in education | -.0122396 | -.0179335 | -.0212733 | |
| retired | .0211099 | -.0202260 | -.0223186 | |
| part-time | -.0081407 | -.0261201 | -.0299941* | |
| out of labor force | .0065805 | -.1089817*** | -.1130105*** | |
| owner | .0135794** | | | |
| months unempl. benefits | -.0053308 | .0025754 | .0024671 | |
| months unempl. assistance | -.0083573*** | .0017857 | .0020511 | |
| months social assistance | -.0060898 | .0014436 | .0009512 | |
| university | -.0042473 | | | |
| abitur/voc. training | -.0009225 | | | |
| disability degree | -.0004776*** | .0005772* | .0005917* | |
| ill | -.0004839** | -.0003344** | -.0003449** | |
| newborn | -.0608937 | | | |
| newly married | -.0013687 | | | |
| newly disabled | .0027484 | | | |
| new death | .0965823*** | | | |
| newly divorced | -.0020439 | | | |
| new child u14 left | .0501440** | | | |
| soon divorced | -.1353558 | | | |
| soon married | -.0778485 | | | |
| _cons | .8866412*** | .8850434*** | .9270529*** | .7641366*** |
| Observations | 17882 | 53704 | 53704 | 53704 |
| $R^2$ | 0.1553 | 0.0395 | 0.0313 | 0.0213 |
| Location of Maximum | 875 | 965 | - | 965 |

* p<0.1 ** p<0.05 *** p<0.01

Table 2: Year 2008, income satisfaction dummy as regressand. (1) LPM with the best income dummy fit, (2) 3-years FE panel model with the best income dummy fit, (3) as (2) without dummy and (4) as (2) without additional controls. Robust standard errors. Own calculations.

Table 2 shows the regression results at the maxima based on data between 2007 and 2009. Additionally, we present results based on data of one decade earlier for comparison in table 1. Columns marked by (1) show the results based on the linear probability model, (2) based on the fixed effects model with full set of controls, (3) shows fixed effects without the income dummy variable and (4) shows results for the case, where no additional controls are used besides income.

While for (1) the full set of controls is used as described in Section 2, we drop some of them for the fixed effects models. We cannot make use of the time invariant regressors gender and East German origin, and age is dropped as year dummies are employed in the fixed effects regressions. Furthermore we excluded variables with very weak time variation such as house ownership, university degree and Abitur/vocational training. Additionally, we dropped the controls for major life events as their impacts were already erratic in the cross sectional case, and as they were highly insignificant in the fixed effects model.

Despite the differences in the models, the coefficients of the third degree polynomial of income have the same magnitude. In the relevant income segment the marginal effect of income is positive and falling in income, in line with the general assumption that the marginal utility of income is a convex function. The additional dummy variable of income is always positive and highly significant ranging between 8% and 9% for 1998 and between 9% and 13.4% for 2008, always smaller in the fixed effects model. When plotting the probability effects arising from the income variables only, figure 5 shows strong differences between the dummy-included and dummy-excluded models. Exclusion leads to local underestimation of the effect of income of up to 5% around the jump. When comparing (2) and (3) we see for both years that the inclusion of the dummy variable is not only reasonable in terms of the high significance of the variable, but also raises the fit of the regressions from 0.0313 to 0.0395 in 1998 and from 0.0341 to 0.0390 in 2008 by a considerable amount. When we take a look at the constants of the regressions, they range between 68.5% and 92.5% for 1998 and between 76% and 93% for 2008. Naturally, these baseline probabilities of not being dissatisfied with income cannot be compared directly as they are valid for different reference individuals. For example, the effect of age ranges between 0 and -10%

with a minimum at around 42 years of age in 1998 and around 50 for 2008 in the linear probability model, while age is only included squared in the fixed effects regressions and is never significant.

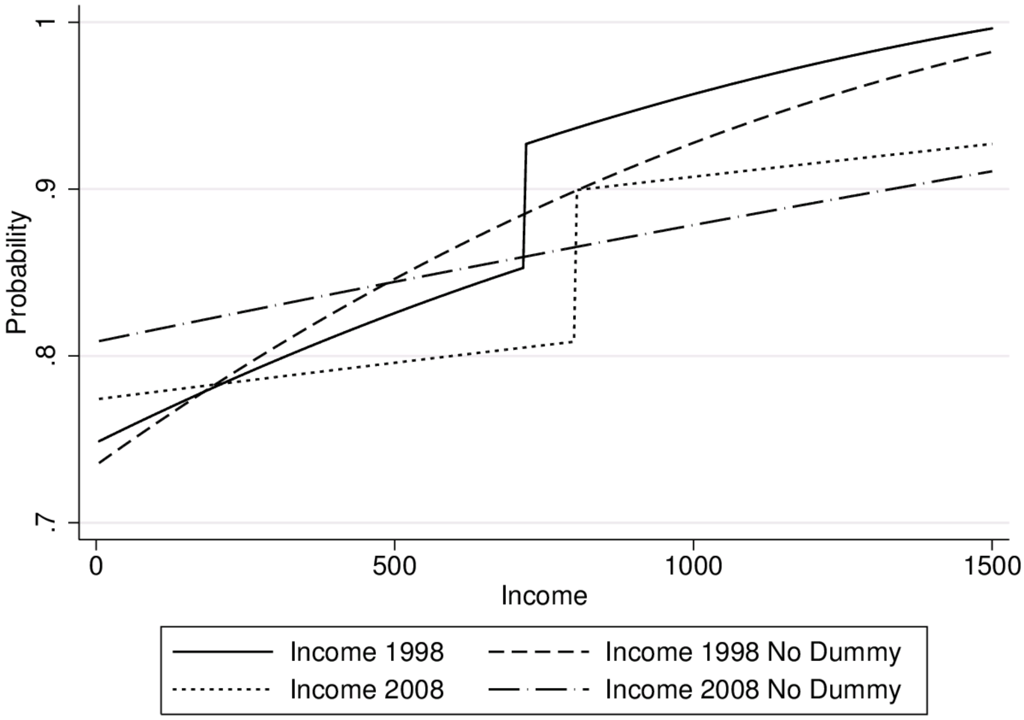



Figure 5: Probability plots for not being dissatisfied with income based on fixed effects estimation with and without the income dummy variable. The plots are valid for a reference person with zero valued control variables as in table 2, models (2) and (3). Own calculations.

While the coefficient of married is high and significant with 4.75% and 2.2% respectively, its effect vanishes when using the fixed effects model. The effect of separated varies strongly, but it is below zero as to be expected if significant in 1998. In the FE cases estimates of the latter two variables may be weak due to a weak time variation in the panel. The gender coefficient suggests a somewhat higher probability of not being dissatisfied with income for females, but the evidence is not strong. The East German dummy is significant in 1998, but not any more in 2008, indicating a vanishing heterogeneity in the German population. Both signs are negative as expected.

The presence of children has generally a positive effect, especially when the results are significant. The interactions between jobless and other persons jobless in the family have negative signs and the highest impacts besides income itself. If another person in the household is jobless, there is some weak evidence for a negative effect of up to around 2.3% in 2008. This effect may be a psychological impact of unemployment on the interviewee, as income itself is already accounted for. If the interviewee himself is without a job and no one else is unemployed in the household, the probability effect is highly significant and may well exceed 16% depending on year and model. As income is accounted for, this effect might be a strong psychological effect of unemployment. If we compare the differences between the model in (2) and (4) with respect to the coefficient of the income dummy, we see an overestimation of the dummies' probability effect. This turns out to mainly coincide with the absence of the unemployment dummy. Putting things the other way around, we see an overestimation of the effects of unemployment, when omitting the income dummy in (3) as compared to (2). Interestingly, if another person in the household is additionally without a job, there is some evidence that this effect is mitigated in 2008. Nevertheless, in 1998 the FE models tell the opposite story. The former effect may be explained with an income comparison in the family, mitigating relative deprivation, or it might be a residual effect arising from the fact that at least two persons in the household are present to share the burden of unemployment.

The four occupation dummies are to some extent complementary and may be interpreted in relation to fulltime employment. Participation in education has a weakly negative effect and may be explained with some unexplained correlation with age and lack of employment. The effect of being retired is mostly negative and mostly insignificant. Finally, part-time employment instead of full-time employment has a negative sign and being out of the labor force is mostly highly significant and negative. The linear probability models reveal a positive and strongly significant effect of around 1.4% to 2% for house ownership. It is not only a proxy for financial situation but also captures the effect of not paying rent, leaving a higher amount of household income for consumption.

The following three variables capture the degree of dependency of social transfers. The number of months receiving unemployment benefits is not significant, probably as we already controlled not only for income, but also for unemployment status. The number of months receiving unemployment assistance captures the presence of long term unemployment beyond the unemployment dummy. The effect is negative as expected and may be well over 10% if received at least 12 months. Social assistance is typically not unemployment related and has no significant effect in our models.

University degree and Abitur or vocational training have a weakly negative influence which may mirror higher expectations on income levels of the subgroups or potentially higher relative deprivation. Because of limited time variation, they were left out from the fixed effects estimations and they do not show significance in the cross section with one exception. Degree of disability and days spent ill last year are wealth proxies. While the degree of disability has a negative and significant effect in the cross sections, it partly vanishes in the panel estimations. Illness has an additional negative effect, especially significant in 2008. In model (2) of 2008 the effect of 8 weeks of illness is -1.8%. The last eight controls only appear in the cross sectional estimations. While they are mostly not significant, their values are also strongly erratic with changing signs. The fact that death is significantly positive in both years is merely a coincidence and typically does not apply to the other years analyzed. Because of this, these controls were not used in the fixed effects models. Year dummies were also included, but they were never significant and are not listed.

Figures 6a-b show the complete estimated SDPL time series. The left figure shows that all models yield similar time series with a very similar time trend compared to the time series of the official poverty line. The right figure shows the SDPL in terms of median income percentages. The values fluctuate nicely around the definition of the European Commission, meaning that this poverty line explains income dissatisfaction best among all potential poverty line definitions. If we take a look at figure 7a, we see that the fixed effects estimation yields lower marginal probability effects at the poverty line of around 10% for FE and 11% for FENLS, while the cross-section delivers 13.5% for LPM and 14.5% for NLS. It may be assumed here that cross sectional models overestimate the effects by not accounting for individual heterogeneity.

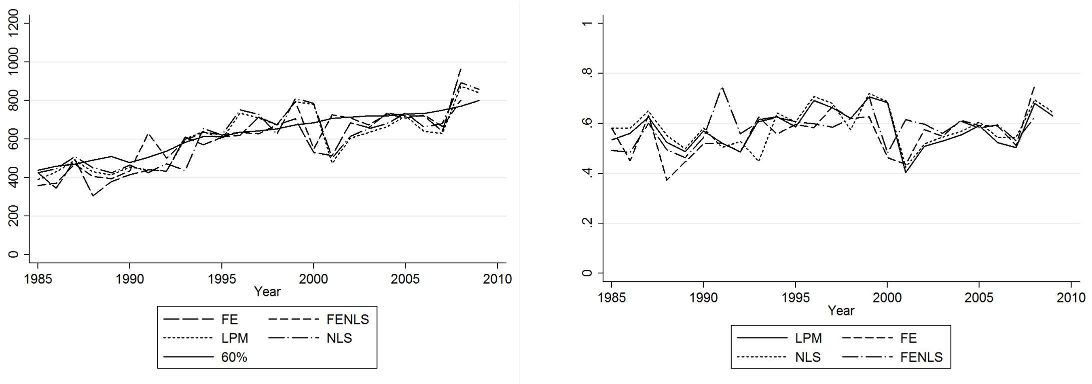



Figure 6: Time series: SDPLs based on LPM, FE, NLS and FENLS estimations and the 60% poverty line. Left (a) in Euros, right (b): as median percentage. Own calculations.

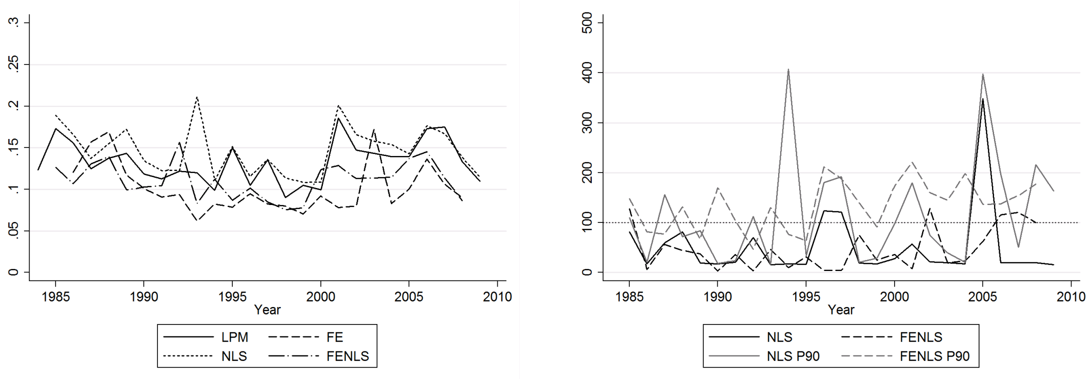



Figure 7: Left (a): Time series of the (marginal) probability effect of the PTR transition based on the LPM, FE, NLS and FENLS estimations. A single outlier in 1989 was set to the bootstrap mean for the FENLS. Right (b): Time series of the length of the PTR based on NLS and FENLS estimations and their one sided percentile based bootstrap 90% confidence intervals. Outliers in 1987, 1990, 2001 for the NLS were set to their bootstrap means. Own calculations.

### 1.4.2 *The Poverty Transition Region*

The nonlinear least squares approach is a generalization of the grid search method and it is not restricted to a search grid that almost surely misses global maxima on its uncountable superset $\mathbb{R}$. While this problem should only be minor in practice, the nonlinear least squares method also adds a straightforward way to statistically test for the maximum's location. Furthermore, it adds a degree of freedom to the characterization of poverty as it includes a jump length dimension of the local smoothing function. This dimension can be used to check whether the state change to poverty is discrete as assumed by the dummy grid search construction or whether there is a poverty transition region.

Both the NLS and FENLS procedures were carried out in two independent settings.[41] The first one involved the estimation of the location by fixing the length of the PTR to a small, but feasible value.[42] The purpose here was to underline the linear findings of the grid search, where PTR lengths of zero were assumed, and to obtain standard errors for the locations. Location results were already shown in the last subsection. The standard errors obtained are between 2 and 5 Euros in the NLS case and 2 and 8 Euros in the FENLS case with exception of 1990, where it is as big as 80 Euros. These standard errors underline that the estimation of the SDPL locations are rather precise.

The second setting starts with pre-estimated SDPLs based on the LPM and FE procedures respectively. The aim here is to assess the local behavior of the transition. To calculate a confidence intervall for the PTR's length, a two stage bootstrap is performed involving both the pre-estimation of the location and the nonlinear procedure. The results for the length of the PTR are given in figure 7b. The estimates based on the original sample are reported along with the 90th centiles of the empirical distributions of the estimators derived from the bootstrap procedure. The estimates are almost always below 100 Euros and under 50 Euros in over two third of the cases. The results cannot be statistically distinguished from any low value, so there is no evidence in this setting against the assumption that the PTR is vanishingly short and that the search for a non-fuzzy poverty line is justified.

41 This strategy avoids an arbitrary restriction for the smooth function to be local in the NLS context when the PTR's length and location are estimated simultaneously. Generally, a smooth function stretched on the full income spectrum yields a better fit than its local counterpart and may therefore be favored by the iteration procedure. Nevertheless, a better global fit does not indicate the non-existence of a meaningful local smooth, it only prevents the single step procedure to find it.

42 The value was chosen to be 20 Euros. As a robustness check, 7 different starting points were used and in cases of different estimation outcomes the one with the smallest RSS was chosen.

While results based on German data suggest that the SDPL is temporally stable and consistent with a constant percentage point of the median income, it is interesting to investigate whether this result applies more generally in a cross-country comparison. To take a closer look, we use the ECHP. In contrast to the SOEP, the ECHP contains a question about the satisfaction with the financial situation in general instead of the more specific income satisfaction. As this makes a more rigid control for the household's financial situation necessary, we apply the FE panel approach only, netting out the influence of wealth, given its assumed time constancy. Also, satisfaction is only measured on a 6-point scale here as compared to the 11-point scale of the SOEP. Additionally, there are fewer observations at country level. Nevertheless, the results can be useful to draw an international map on approximate SDPL estimates.
In figure 8 we present the estimated country specific poverty lines, based on country-wise fixed effects estimation. The estimated maxima are plotted along with 95% bootstrap confidence intervals.[43] In addition, results based on the British Household Panel Survey (BHPS) 1994-2001 and on the SOEP 1994-1996 are shown for comparison.

The country specific SDPLs can be found in a range of 32.27% for Portugal and 64.71% in Finland. In addition to Portugal the Mediterranean countries Italy (40.02%), Greece (44.66%) and Spain (50.70%) have rather low poverty lines. On the other hand, Denmark with its 44.85% has a significantly lower poverty line than Finland. The Benelux countries Luxembourg (56.06%), Belgium (48.37%) and the Netherlands (50.00%) are relatively homogenous. Furthermore, Austria with its 55.48% belongs to the middle field. The estimate for Germany of around 52.30% is lower than the estimate based on the SOEP at around 60.83%, but the former one is less exact and the two confidence intervals have a nonempty cross section, so the difference is not significant. As regards the data from the United Kingdom, the two results are only slightly different with 62.56% based on the ECHP and 67.58% based on the BHPS, with the difference being statistically in-

43 See Appendix A for how these bootstrap confidence intervals were computed.

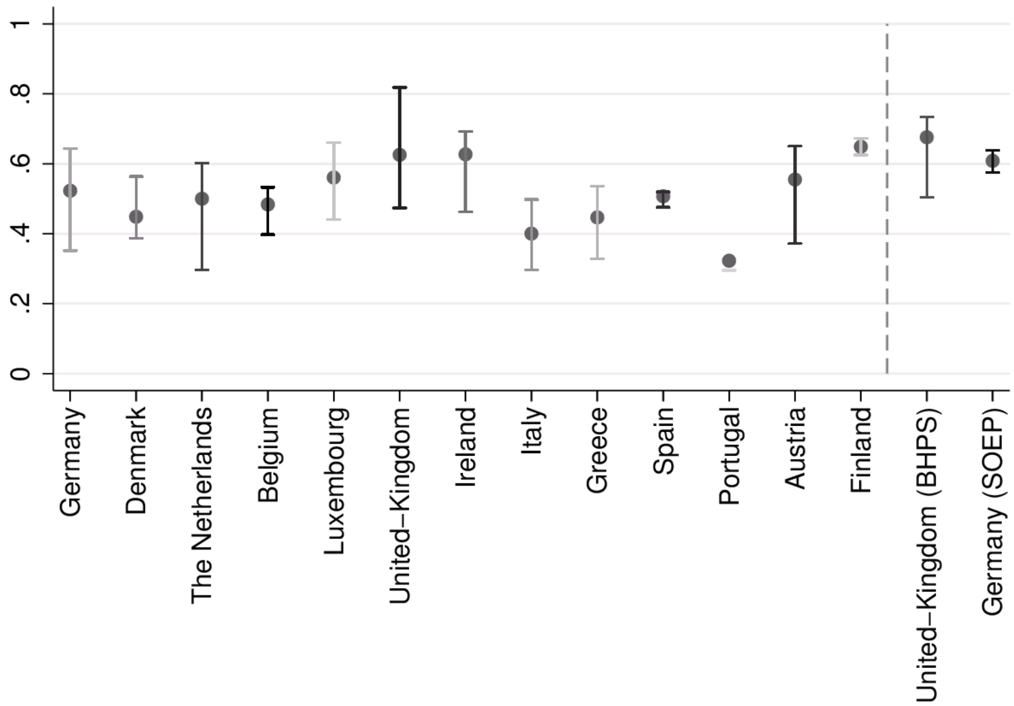



Figure 8: Fixed Effects SDPL poverty lines based on the complete ECHP panel information. 95% confidence intervals are provided based on bootstrap estimation. Results based on the BHPS 1994-2001 and on the SOEP 1994-1996 are provided for comparison. Own calculations.

significant. Finally, Ireland with its 62.71% is very similar to the UK.

In general, figure 8 provides a heterogenous picture of poverty lines across Europe and may be interpreted as evidence against the hypothesis of a constant relative poverty line across European countries. In particular, there is indication that the SDPLs of Denmark, Belgium, Italy, Greece, Spain and Portugal may be lower than the 60% definition of Eurostat. This is not necessarily the view we would share a-priori. From a global point of view, all these countries belong to the group of developed countries and their additional spatial proximity may also suggest that their subjective understanding of poverty does not differ that much.[44] If heterogeneity exists though as indicated, it is interesting to investigate

44 Also note that in many cases, the differences across countries are not statistically significant.

its nature, e.g. whether it is systematic in a sense that it can be explained by the diversity among these countries in a macroeconomic, social and cultural manner.

The differences among the countries are explored in more detail using a country panel based on 3-year subperiods. The dependent variable is taken to be the SDPL as a percentage point of the median for each country and year, while the independent variables are chosen to be macroeconomic characteristics.[45] As we are also particularly interested in country level characteristics that do not vary or only slightly vary over time, we see fixed effects estimation as too restrictive and use pooled OLS (POLS) with regional indicators instead. As macroeconomic characteristics we take the Gini coefficient of disposable income, GDP per capita as purchasing power parity in Euros, employment as a share of the employed in the employable population by the ILO definition, grade of urbanization as a share of the population living in cities, life expectancy at birth and the age dependency ratio of the old[46]. Furthermore, we control for regions of Europe including Mediterranean, Benelux and Scandinavian countries and their complement as the reference group. All the macroeconomic variables are centered at the overall sample mean excluding GDP per capita, which is expressed as the deviation from the overall sample mean in percentage points.

The first regression (1) in table 3 is a regular POLS with macroeconomic variables and regional indicators. In (2) the regional indicators are dropped to increase the degrees of freedom of the data. Regressions (3) and (4) are counterparts of (1) and (2) calculated with multiple imputation (MI) to fill in missing values of the Gini coefficient.[47] To control for time effects we assume a linear time trend.[48] As the number of observations is rather low, we performed stability checks by

45 These data were obtained from OECD StatExtract, Belgostat, the Central Statistics Office Ireland, Statistics Finland and Datastream.

46 That is the population size over 65 divided by the population size between 15 and 65 years of age.

47 About one quarter of the Gini coefficients for the countries and points of time in use are missing. We contacted Eurostat, but there is no further data available at this time. The MI procedure used is the STATA implemented predictive mean matching. Robustness checks showed that other imputation methods lead to very similar results.

48 Time dummies indicate that linearity is a good approximation and more importantly we save degrees of freedom for the estimations.

| dep. var.: SDPL | (1) coeff. | (1) P-value | (2) coeff. | (2) P-value | (3) coeff. | (3) P-value | (4) coeff. | (4) P-value |
|---|---|---|---|---|---|---|---|---|
| Gini | -1.548* | 0.059 | -1.287** | 0.017 | -1.060 | 0.171 | -1.356** | 0.011 |
| GDP per capita | -0.186** | 0.048 | -0.245** | 0.016 | -0.317*** | 0.005 | -0.297*** | 0.004 |
| Employment | -0.280 | 0.470 | -1.224*** | 0.002 | -1.089*** | 0.006 | -1.316*** | 0.001 |
| Urban | -0.335 | 0.189 | 0.292* | 0.074 | -0.021 | 0.913 | -0.222 | 0.142 |
| Life Expectancy | -0.647 | 0.713 | -2.454 | 0.156 | -2.658 | 0.179 | -2.999* | 0.093 |
| Age Dependency Ratio | 2.441 | 0.129 | -2.245*** | 0.004 | -0.980 | 0.291 | -2.209*** | 0.005 |
| Mediterranean | -0.857 | 0.884 | | | -7.064 | 0.311 | | |
| Benelux | 22.027** | 0.016 | | | 6.030 | 0.181 | | |
| Scandinavian | 7.071* | 0.051 | | | -0.089 | 0.988 | | |
| Year | -1.443 | 0.262 | 0.642 | 0.512 | 0.249 | 0.781 | 0.633 | 0.496 |
| cons | 46.428*** | 0.000 | 48.317*** | 0.000 | 48.967*** | 0.000 | 46.958*** | 0.000 |
| Number of Observations | 65 | | 65 | | 86 | | 86 | |
| $R^2$ | 0.408 | | 0.351 | | | | | |

* p<0.1 ** p<0.05 *** p<0.01

Table 3: POLS country panel estimations with the SDPL as poverty line for 13 ECHP countries. Models (3) and (4) make use of multiple imputation. Standard errors are clustered. Own calculations.

leaving out single countries from the regression to see whether the results are strongly influenced by the presence of a single country. Outcomes showed that this is not the case.

Taking a look at regression (1) and (2) we see that the coefficient of the Gini is negative and significant. An increase of the second decimal digit of the Gini by one goes along with a decrease of the SDPL by 1.6% in (1) and 1.3% points in (2) respectively. The fact that the regression coefficient of the Gini is negative, is a sign for inequality aversion. Countries with higher inequality typically harbor a higher share of the population with rather low income. Such countries have a tendency for a lower SDPL. The other way around, if a country has very low inequality, meaning that a high share of the population can be found around the median, the SDPL is typically a higher percentage of the median with a lower distance to the mid section as the population is more sparse in lower regions with a higher relative deprivation for those at the very bottom of the distribution.

The coefficient of per capita GDP is also negative and highly significant in all settings. Wealthier countries - proxied by GDP per capita - seem to go along ceteris paribus with a lower SDPL in relation to the median. This is in line with basic needs considerations, as it is easier for people to make ends meet in a wealthier society in terms of a given median income percentage, which is higher in absolute terms than in less wealthy societies.

These findings indicate that the SDPL concept has desirable cross country properties, as it can be expressed as a mixture of the absolutistic 'basic needs concept' and the concept of 'relative deprivation'. To put it in the words of Ravallion and Chen (2011): "While we can agree that people care about their relative position in society (at least above some level of living), it is very hard to accept that they do not also care about their absolute levels of living (at least for all except very rich societies). More plausibly, utility is derived from both absolute income and relative income."

Nevertheless, there are other factors to look at to describe a wider range of heterogeneity among countries. For example, employment has a negative coefficient in all four models, being highly significant in (2) to (4).[49] Higher employment is typically positively correlated with a higher wealth and the result may be interpreted similarly to the case of the GDP. Urbanization on the other hand exploits a positive and significant relationship with the SDPL in (2), else it is insignificant. This may weekly indicate that people living together in spatially condensed areas are more competitive and interpret poverty more in the relative sense in the direction of the median. The coefficient of life expectancy can be interpreted as a health indicator on societal level. It is always negative and it is significant in setting (4). This gives rise to the assumption that a healthier society is less concerned with income related issues of overall life satisfaction, being already satisfied with less income than in societies with lower health indicators. The age dependency ratio is an indicator for the countries' age structures. When its coefficient is significant, its sign is negative. As the ratio is high if the share of pensioners is high, a negative sign may indicate that pensioners tend

49 The tendency towards insignificance is generally higher in (1), as due to the higher number of regressors model (1) has less power to reject the null.

to think more in absolute terms about their income and that they are satisfied with a lower amount. Also the thinking, the elderly may be less 'competitive', is similarly in line with this finding. As regards general time trends, the estimates are highly insignificant in all cases. This is an indication for the SDPL being time stable between 1994 and 2001 in the region in scope.

Along with the variables just mentioned, we use controls for Mediterranean, Benelux and Scandinavian countries to capture further unspecified time constant effects. Results show that in (1) the coefficients of Benelux and Scandinavian are significant and positive, indicating higher SDPLs than in central European countries. After MI though, (3) shows no significance any more. Why (3) is more in line with the actual findings may be seen when looking at the coefficient of the Scandinavian countries. As we saw before, Denmark and Finland had very different SDPLs, a rather low and a rather high one, so it is more intuitive that the coefficient is insignificant. Thus, (3) indicates that with the amount of data we have at hand, differences are well-explained with the set of macroeconomic variables chosen above and that regional fixed effects have no additional explanatory power.

## 1.6 CONCLUSION

This chapter defines the satisfaction-driven poverty line as to maximize the relationship between poverty status and income dissatisfaction, based on the simple assumption that the true poverty status has the unique property of explaining income dissatisfaction best in any proper econometric setting. Application to data from the German Socio-Economic Panel yields a time stable poverty line similar to the definition provided by the Statistical Office of the European Commission as 60% of the median income. Our results characterize this definition of poverty as the best dichotomization to explain the relationship between income and income dissatisfaction measured by the subjective psychological variable of income satisfaction.

The results suggest that our approach is generally applicable and that estimation of the SDPL may serve as a unified strategy for poverty measurement across countries. Using data from the European Community Household Panel, additional evidence for satisfaction-based poverty lines and their cross-country differences were presented. The results further indicate that although the static 60%-definition is in line with the SDPL in Germany, the SDPLs of other European countries may be different. Analysis of the differences among European SDPLs suggests on the one hand that national poverty lines can be explained by the income inequality of the country in form of inequality aversion, resulting in poverty lines that are higher percentages of the median, when inequality is lower. On the other hand, the percentage points fall with higher wealth as basic needs are already satisfied with an income at a lower percentage of the median, if a country is wealthier.

In addition to this, the proposed approach also allows one to test whether the poverty line exists as a discrete phenomenon, or whether a fuzzy poverty line is more appropriate. Our analysis for Germany suggests that the state change to poverty is rather discrete, so that poverty line definitions rather than definitions of poverty transition regions are more appropriate.

When the bootstrap is applied to the SDPL based on grid search, the problem emerges that in some resamples the secondary maximum - also observable in Figures 3 and 4 - is coincidentally larger than the maximum of interest (the one in the poverty region). In order to adress this problem, we use a two-step strategy. In a first step we resample the data to obtain a bootstrap distribution of the maxima of the FE estimates' $R^2$-s. In the second step we estimate a finite mixture model given the bootstrap based maxima. In this way, we can control for observation disturbances caused by secondary and sometimes tertiary maxima that are inferior in the original sample, but deliver global maxima in some resamples and are thus handled as observation disturbances that prevent us from observing the position of the primary maximum. This situation suggests a latent class setting. A finite mixture distribution is then given as follows:

$$f_{mix}(x) = \sum_{d=1}^{D} \pi_d \cdot \phi(x|\mu_d, \sigma_d), \quad \phi(.|\mu, \sigma) \sim \mathbb{N}(\mu, \sigma).$$

In almost all cases $D = 2$ and $D$ is never greater than 3. We estimate via the STATA addon fmm[50] using maximum likelihood. As confidence intervals we report the 95% confidence intervals of the normal distribution the original maximum $m$ belongs to with highest a-posteriori probability $\max_{i \in \{1,...,D\}} \tau_i(m)$, given[51]

$$\tau_i(m|\pi_1 \ldots \pi_D, \mu_1 \ldots \mu_D, \sigma_1, \ldots, \sigma_D) = \pi_i \phi(m|, \mu_i, \sigma_i) / \sum_{d=1}^{D} \pi_d \phi(m|\mu_d, \sigma_d).$$

50 The addon was programmed by Partha Deb, Hunter College and the Graduate Center, City University of New York, USA partha.deb@hunter.cuny.edu.
51 For more on finite mixture models see McLachlan and Peel (2000).

Part III

# UNDERSTANDING RISING INCOME INEQUALITY IN GERMANY

There has been a clear trend of increasing income inequality in industrialized countries over the past three decades, although with differences in the timing and intensities across countries (see OECD (2008), OECD (2011)). This trend was first observed in Anglo-Saxon countries such as the United States, where pronounced changes in the wage and earnings distribution in the 1980s and 1990s sparked a large body of literature examining the possible causes of increasing inequalities in labour market returns (see, e.g., Bound and Johnson (1992), Levy and Murnane (1992), Murphy and Welch (1992), Juhn et al. (1993), and DiNardo et al. (1996)). Wage incomes, which have been the focus of many previous studies, are only one component in the distribution of *overall incomes*. The distribution of overall incomes seems particularly policy-relevant as the distribution of personal financial possibilities is closely linked to personal economic well-being. An analysis of the distribution of overall incomes requires a comprehensive view of the income distribution including its different economic, social and institutional determinants such as demographic aspects, employment outcomes, remuneration of market activities, taxes and government transfers (such a comprehensive view of the income distribution has been adopted in a recent study by Checchi and García-Peñalosa (2010)).

Germany's distribution of overall incomes is particularly interesting as it remained quite stable until the end of the 1990s (see Steiner and Wagner (1996), Biewen (2000), Prasad (2004)), but witnessed a sharp increase in inequality and poverty after 1999/2000 . At the same time, a number of factors that are likely candidates for explaining changes in the income distribution underwent substantial changes. For example, there was a steep increase in unemployment and an increase in part-time and marginal part-time work. Moreover, wage inequality grew in a pronounced way from the end of the 1990s onwards. There is a consensus that the pronounced changes in the structure of wages that were observed in other countries reached Germany with considerable delay, although the changes were less drastic than those observed in countries such as the United States (see Kohn (2006), Gernandt and Pfeiffer (2007), Dustmann et al. (2009), Fuchs-Schündeln et al. (2010), Antonczyk et al. (2010)). Changes in employment structures and rising wage dispersion are not the only factors that may have been responsible for the increasing inequality in Germany. Other factors include demographic changes, changes in living arrangements, changes in characteristics such as age or educational qualifications, and changes in the tax- and transfer system (see OECD (2008), OECD (2011)).

While a large number of studies has focussed on such individual factors, surprisingly little is known about the relative importance of the different factors for the observed changes in the overall distribution. Although possible reasons for changes in the distribution have been well-documented for many countries (see OECD (2008), OECD (2011) and the references therein), it remains unclear which of the many possible candidates are the main drivers of distributional change. This is all the more surprising as knowledge about which factors are most important is highly policy-relevant. For example, it is relevant to know whether rising income inequality in Germany is more the result of a widening wage distribution, or the result of rising unemployment or changes in employment structures.

In this chapter, we provide a detailed examination of the main reasons for rising income inequality in Germany in a unified framework. Building on previous work by Hyslop and Mare (2005), for New Zealand, and Daly and Valletta (2006), for the United States, we use the semi-parametric kernel density reweighting method originally developed by DiNardo et al. (1996), in order to shed light on the factors behind the increase in inequality and poverty between 1999/2000 and 2005/2006. We consider in particular i) changes in the distribution of household types, ii) changes in the distribution of household characteristics such as age or educational qualifications, iii) changes in employment outcomes conditional on such characteristics, iv) changes in labour market returns, v) changes in the transfer system, and vi) changes in the tax system. Our results complement previous studies on the German income distribution, which have documented some of the developments considered here, but which did not attempt to quantify their relative importance for the overall development of the distribution (see, e.g., Hauser and Becker (2002), Federal Government of Germany (2008), German Council of Economic Experts (2009), and Frick and Grabka (2010)). An exception is the study by Peichl et al. (2012), who provide an explicit estimate of how much changes in household structures contributed to changes in the German income distribution (they do not consider other factors, however).

The remainder of this chapter is structured as follows. Section 2 provides an informal discussion of the development of the German income distribution and its possible determinants. In section 3, we describe our methodological setup. Section 4 discusses some data and specification issues. In section 5, we present our empirical analysis. Section 6 concludes.

## 2.2 POSSIBLE SOURCES OF INCREASING INEQUALITY

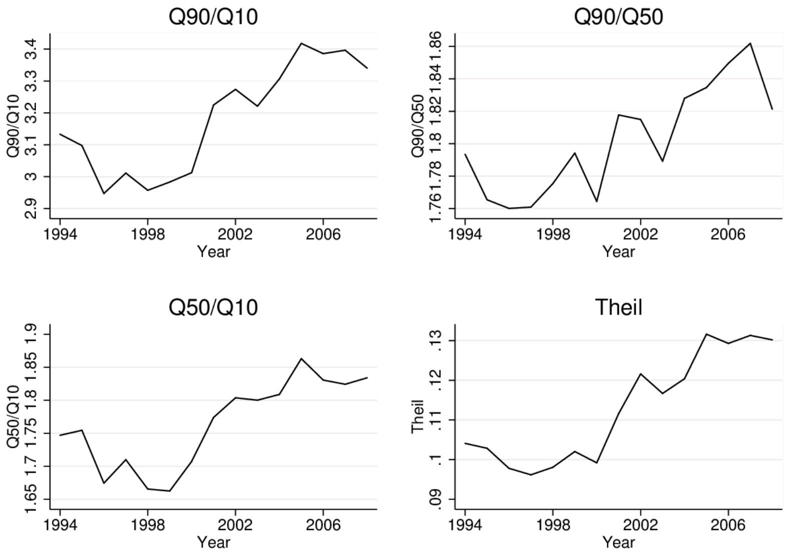
As mentioned in the introduction, there was a significant increase in inequality and poverty in Germany between 1999/2000 and 2005/06. The aim of this section is to embed the period of interest 1999/2000 to 2005/2006 into the more comprehensive period 1994 to 2008 in order to prove that 1999/2000-2005/2006 is the period in which 'there is something to explain', and to give an informal discussion of possible factors behind the increase. As figures 9 and 10 show, the inequality increase between 1999/2000 and 2005/2006 was considerable.[52] In the following, we provide an informal discussion of possible factors behind these changes in the distribution.

*Changes in employment and unemployment*

As figure 11 shows, the period 1999 to 2005 was one of steep unemployment growth. At the peak in 2005, there were almost 5 million registered unemployed in Germany. Figure 11 also shows that overall employment stagnated during the period 1999 to 2005. After 2006, employment started to grow significantly, while unemployment fell back to a level comparable to that in 1999. The fact

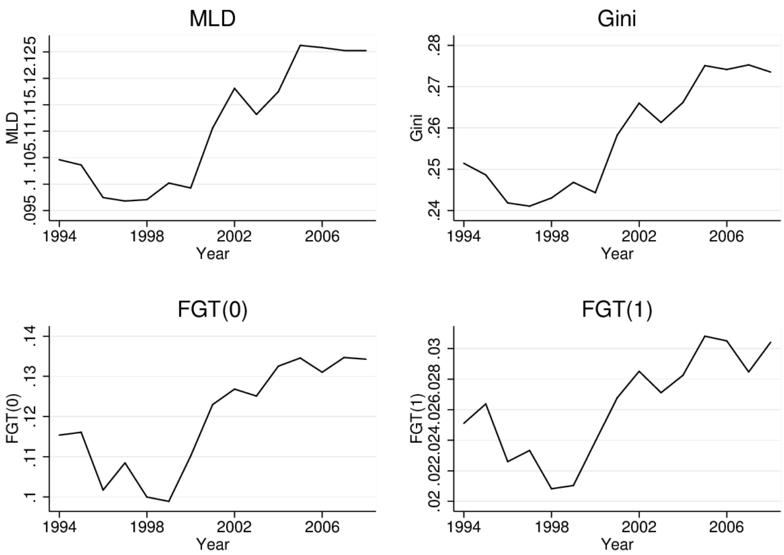
52 Inequality stayed relatively constant before 1999 and after 2005 (see Biewen (2000), and Frick and Grabka (2010)). Our income concept is yearly equivalized post-government personal income, which is calculated as the sum of income from all sources in a given household (including government transfers), net of taxes and social security contributions. The resulting value is then divided by an equivalence scale and distributed equally among household members. More details on the definition of our variables are given below.

Figure 9: Trends in inequality and poverty 1994-2008 (a)




Source: SOEP. Inequality in yearly equivalized post-government personal income.

Figure 10: Trends in inequality and poverty 1994-2008 (b)




Source: SOEP. Inequality in yearly equivalized post-government personal income.

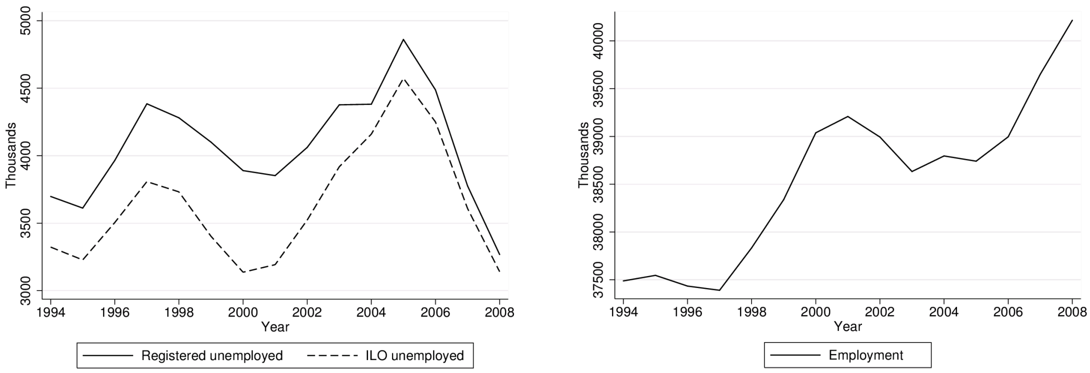
that unemployment fell again after 2005 while inequality and poverty remained at their high levels suggests that the rise in unemployment is unlikely to be the only reason for the inequality increase between 1999 and 2005.

In addition to changes in unemployment, there were other changes in employment that may have influenced the distribution of incomes. Figure 12 displays the evolution of the share of individuals living in households with different employment outcomes. On the one hand, the figure reflects the development of unemployment as, for example, the share of individuals living in households with no employment continuously increased between 1999 and 2005, but fell after 2005. Similarly, the share of individuals living in households with at least two full-time workers fell between 1999 and 2005, and increased again after 2005. On the other hand, the share of individuals in households with exactly one full-time worker kept decreasing independently of the development of unemployment, while the share of individuals in households with at least one part-time worker – including marginal part-time – steadily increased. The growth of this kind of households even accelerated after 2005.

*Changes in labour market returns*

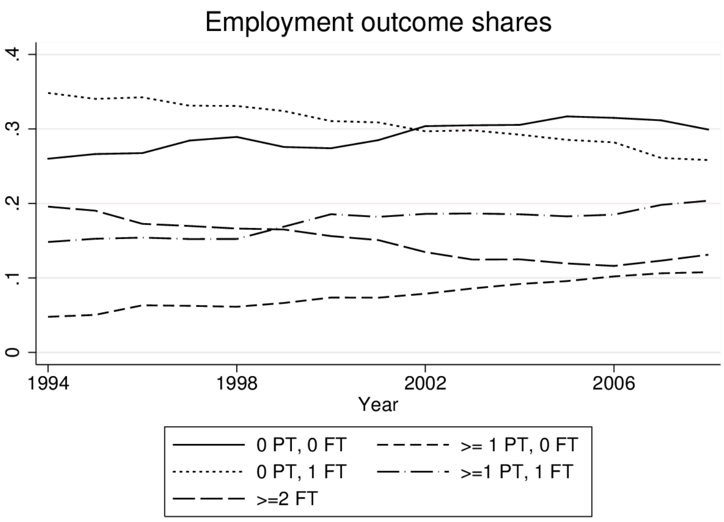
A second possible source of increasing income inequality is increasing inequality in labour market returns. This has been the focus of many previous studies. The common perception is that the effects of skill-biased technological progress, which is seen as the main cause for the widening wage distribution in Anglo-Saxon countries since the 1980s (Bound and Johnson (1992), Levy and Murnane (1992), Murphy and Welch (1992), Gosling et al. (2000)) reached the German labour market with a delay. In Germany, wage inequality started to grow in a clear way from the mid-1990s onwards (Kohn (2006), Gernandt and Pfeiffer (2007), Dustmann et al. (2009), Fuchs-Schündeln et al. (2010), Antonczyk et al. (2010)). The evidence suggests that wage inequality increased both between and within skill groups, and that increases at the top are well explained by skill-biased technical progress, while increases in the lower tail of the distribution are better explained by additional factors such as deunionization and supply side effects (Dustmann et al. (2009), Antonczyk et al. (2010)).

Figure 11: Trends in employment and unemployment 1994-2008




Source: German Federal Employment Office

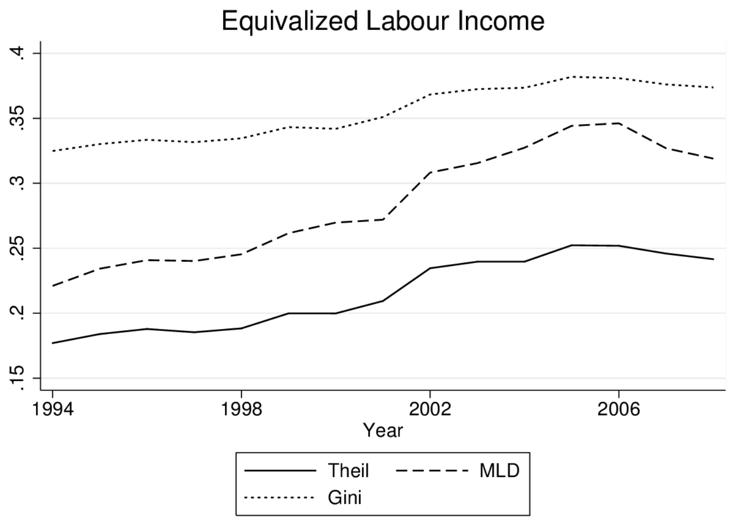
Figure 12: Share of individuals living in households with different employment outcomes 1994-2008




Source: SOEP. FT = full-time, PT = part-time or marginal employment.

Figure 13 shows that growing inequality in labour market returns also translated into growing inequality of labour incomes at the household level. The figure displays the evolution of inequality in equivalized household labour income, i.e. in household labour income divided by an equivalent scale and equally distributed among household members. There is a clear trend of growing inequality, especially between 1999 and 2005. This trend ended in 2006, after which inequality in equivalized household labour income slightly fell.

Figure 13: Inequality in equivalized labour income 1994-2008




Source: SOEP. See text for the definition of income variables.

*Changes in the transfer system*

In a highly developed welfare state like Germany, personal disposable incomes in Germany are to a large extent influenced by government transfers, especially at the bottom of the distribution. Changes in the transfer system may therefore directly affect the income distribution. In fact, a major set of labour market reforms, the so-called Hartz-reforms, was enacted in 2005. One of the key elements of the Hartz-reforms was the introduction of the so-called unemployment benefit II which replaced both the former means-tested unemployment assistance for the unemployed and social assistance payments for all other individuals who are (at least in principle) able to take part in the labour market.

The introduction of the unemployment benefit II for former recipients of unemployment assistance had a potentially substantial impact on the income distribution as the old unemployment assistance depended on the former income of the unemployed, while the new unemployment benefit II only provides a basic income independent of any former income.[53] Apart from these income reducing features of the unemployment benefit II, some population subgroups also benefited from its introduction. This was especially true of former recipients of social assistance who benefited from the slightly higher level of the unemployment benefit II, and individuals who, intentionally or unintentionally, failed to claim social assistance under the old system. In fact, contrary to the perception of the Hartz-reforms as being antisocial, the introduction of the unemployment benefit II led to a major increase in government spending.

Another potentially inequality increasing feature of the Hartz-reforms was the reduction of the age-dependent maximum entitlement period for the unemployment benefit I from up to 36 months to generally 12 months (18 months for individuals aged over 55 years). As unemployment benefit I also depends on the former income of the unemployed, this typical leads to a substantial drop in income unless the person in question succeeds in finding a job.

*Changes in the tax system*

As in many other countries, the German tax schedule experienced several changes between 1999 and 2008. The main changes are summarized in table 4. Tax rates were generally reduced, but reductions were somewhat higher at the top of the distribution. In 2007, the so-called 'rich tax' took back some of the reductions for

53 However, for former recipients of the ordinary unemployment benefit (i.e. unemployment benefit I), the drop in income to the basic level is cushioned over a period of two years, during which part of the difference between the higher unemployment benefit I and the basic income level is covered by extra payments. For more details on transfer changes in the course of the Hartz-reforms and an analysis of their distributional impacts, see Hauser and Becker (2002), and Arntz et al. (2007).

Table 4: Changes in the German tax schedule

| Year | Basic Allowance | Min. Marginal Tax Rate | End of Progression Zone | Max. Marginal Tax Rate |
|---|---|---|---|---|
| 1999 | 6,681 EUR | 23.9% | 61,376 EUR | 53% |
| 2000 | 6,902 EUR | 22.9% | 58,643 EUR | 51% |
| 2001 | 7,206 EUR | 19.9% | 54,998 EUR | 48.5% |
| 2002/2003 | 7,235 EUR | 19.9% | 55,008 EUR | 48.8% |
| 2004 | 7,664 EUR | 16.0% | 52,152 EUR | 45% |
| 2005/2006 | 7,664 EUR | 15.0% | 52,152 EUR | 42% |
| 2007/2008 | 7,664 EUR | 15.0% | 52,152 EUR | 42% (45%) |

From 2007 onwards, the marginal tax rate for taxable incomes over 250,000 Euros was 45%.
Source: German Federal Ministry of Finance.

tax payers in the upper part of the distribution. Given that some of the changes were considerable, it seems likely that these changes had some impact on the final distribution of disposable income.

*Changes in household structures*

There are clear trends in the way household structures change in industrialized countries (see OECD (2008), ch. 2.). In particular, there is a trend towards smaller households and towards untypical household forms such as single parents. The effect of the secular decline of household size on the income distribution in Germany is studied in Peichl et al. (2012). Not explicitly considering other influences on the income distribution, they find that the effect of declining household sizes is moderate, even over a period of 20 years. Nevertheless, it seems necessary to account for such changes when studying the effect of other factors such as employment or labour market returns.

*Changes in other household characteristics*

There are, apart from the household form, a number of other characteristics whose change over time may potentially influence the income distribution. These are in particular changes in the age structure of the population (increasing share of the elderly, and the decreasing shares of children and young persons), changes in educational qualifications (secular skill-upgrading), and other changes in the composition of the population, e.g. due to immigration. Again, it seems necessary to account for such changes when studying the effect of other factors such as employment or labour market returns.

*Other changes*

We will capture distributional changes induced by factors other than the ones listed above in the 'residual' of our decomposition analysis. It turns out that the unexplained 'residual' of our analysis is limited so that the factors listed above successfully account for most of the observed distributional changes. One residual factor that is worth mentioning is inequality coming from households' capital incomes. There is evidence that inequality in household wealth increased over the period 1999 to 2005 (see German Council of Economic Experts (2009), and Frick and Grabka (2009)), implying that capital incomes also grew more unequal. It turns out that inequality in equivalized capital incomes displays a similar pattern as inequality in equivalized labour incomes, i.e. an increase between 1999 and 2005, and a slight decrease after 2005 (results are available on request). There is also evidence that capital incomes increased their share in overall equivalized income from about 5 percent in 1994 to about 8 percent in 2007. However, given that the share of capital income in overall income is so small, and given the relative small changes over time, we expect only very moderate influences of capital income.

## 2.3 ESTIMATION OF COUNTERFACTUAL INCOME DENSITIES

Following DiNardo et al. (1996), and Hyslop and Mare (2005), we use a semiparametric decomposition technique to analyze the development of the distribution of equivalized net incomes over the period 1999 to 2008. For sample size reasons and in order to make our results less dependent on individual years, we pool in our analysis two adjacent years.[54] Our main interest lies in the analysis of the change between 1999/2000 ('period 0') and 2005/2006 ('period 1') as this marks the period over which the distribution experienced a major inequality increase. In order check the robustness of our results and in order to gain further insights, we also analyze the change between 1999/2000 and the more recent period 2007/2008, which also marks the end of our data.

The basic idea of DiNardo et al.'s decomposition method is that of a shift-share analysis, in which observations are reweighted according to whether they are over- or underrepresented in a counterfactual situation. These are combined with simulated changes of individual income components. Counterfactual situations are obtained by holding some aspects of interest fixed at the period 0 level, while changing others to the period 1 level. The method has its limitations in that it cannot account for interactions between the different factors in the form of behavioural reactions or general equilibrium effects. Despite these limitations, it is generally believed that counterfactual reweighting and simulation exercises convey important information about the main drivers of distributional changes. However, for the reasons mentioned, one must be cautious when interpreting the results in a strict causal way (for a general discussion of the limits of decomposition methods, see Fortin et al. (2010)).

*Stage 1: Changes in the distribution of household types*

As a first stage, we consider the effect of shifts in the composition of the population with respect to a number of household types (we will distinguish between the six household types, see below). The counterfactual income distribution in

54 Note that we pool observations not incomes. The pooling is necessary in order to increase the precision of the results which would otherwise be too imprecise to draw valid conclusions. Pooling in order to increase sampling precision is very common, see e.g. Hyslop and Mare (2005), or Blundell et al. (2007). Pooling observations for two years also has the additional advantage of making the analysis less dependent on the specific choice of year.

which everything is as in period 0, but the distribution of household types is shifted to that of period 1 is given by

$$f_0(y|t_h = 1) = \sum_{j=1}^{6} w_{1j} f_{0j}(y), \tag{2.1}$$

where $y$ denotes net equivalized personal income, $w_{1j}$ is the population share of household type j in period 1, and $f_{0j}(y)$ the income distribution of individuals from household type j in period 0. Analogously, $f_0(y|t_h = 0)$ would be the factual income distribution in period 0, where $w_{1j}$ is replaced by the factual population shares $w_{0j}$.

*Stages 2 and 3: Changes in household characteristics and employment outcomes*

The second and third stages of our decompositions account for changes in the distribution of household characteristics $x$ (e.g. the age and educational composition of the household, see below for more details) and changes in household employment outcomes $e$ conditional on these characteristics $x$. For example, the counterfactual income density for individuals living in household type j in which everything is as in period 0 but the distribution of household characteristics $x$ and the distribution of household employment outcomes $e$ conditional on these characteristics are as in period 1, is given by

$$f_{0j}(y|t_x = 1, t_e = 1) = \int_e \int_x f_{0j}(y|x, e)\, dF_{1j}(e|x)\, dF_{1j}(x) \tag{2.2}$$

$$= \int_e \int_x f_{0j}(y|x, e) \left[\frac{dF_{1j}(e|x)}{dF_{0j}(e|x)}\right] dF_{0j}(e|x) \left[\frac{dF_{1j}(x)}{dF_{0j}(x)}\right] dF_{0j}(x) \tag{2.3}$$

$$= \int_e \int_x \Psi_{e|x,j} \cdot \Psi_{x|j} \cdot f_{0j}(y|x, e)\, dF_{0j}(e|x)\, dF_{0j}(x). \tag{2.4}$$

This means the counterfactual distribution $f_{0j}(y|t_x = 1, t_e = 1)$ is just a reweighted version of the factual distribution $f_{j0}(y)$ with reweighting factors $\Psi_{e|x,j}$ and $\Psi_{x|j}$. The factual distribution $f_{j0}(y)$ can be obtained by setting $\Psi_{e|x,j} = \Psi_{x|j} = 1$. Analogously, $f_{0j}(y|t_x = 1, t_e = 0)$ with $\Psi_{e|x,j} = 1$ is the counterfactual distribution where only the distribution of characteristics $x$ is shifted to that of period 1 (while the conditional employment and everything else is held fixed at its period 0 level). Finally, $f_{0j}(y|t_x = 0, t_e = 1)$ with $\Psi_{x|j} = 1$ would be the distribution where only conditional employment outcomes are changed to the period 1 level, but everything else is held fixed at the period 0 level.

The reweighting factors $\Psi_{e|x,j}, \Psi_{x|j}$ can be rewritten as

$$\Psi_{x,j} = \frac{P_j(x|t=1)}{P_j(x|t=0)} = \frac{P_j(t=1|x)}{P_j(t=0|x)} \cdot \frac{P_j(t=0)}{P_j(t=1)}, \tag{2.5}$$

$$\Psi_{e|x,j} = \frac{dF_{1j}(e|x)}{dF_{0j}(e|x)} = \frac{P_{1j}(e|x)}{P_{0j}(e|x)}. \tag{2.6}$$

Following Hyslop and Mare (2005), we define household employment outcomes $e$ as an ordinal variable (see below), so that reweighting factor $\Psi_{e|x,j}$ can be estimated using predictions from ordinal logit models $P_{1j}(e|x)$ and $P_{0j}(e|x)$. Analogously, reweighting factor $\Psi_{x|j}$ can be estimated using predictions from logit models $P_j(t = 1|x), P_j(t = 0|x)$ and the ratio of observational mass in period 0 and period 1.

*Stages 4,5 and 6: Changes in labour market returns, transfers, and taxes*

In stages 4 to 6 of our decomposition, we consider changes in labour market returns to household characteristics and employment outcomes as summarized in a vector $z$ (stage 4) as well as selected changes in the transfer system (stage 5) and the tax schedule (stage 6). The vector of characteristics $z$ is understood to include household characteristics $x$, household employment outcomes $e$, and suitable interactions between $x$ and $e$. Let $\widehat{\Delta y}_{lab} = z_0{}'\hat{\beta}_{1j} - z_0{}'\hat{\beta}_{0j}$ be the expected change in household labour income due to changes $\Delta\hat{\beta}_j = \hat{\beta}_{1j} - \hat{\beta}_{0j}$ in returns to $z$. The counterfactual income $y_0^{cf}$ in period 0 that accounts for the expected change in household labour income due to changes in labour market returns, changes in the transfer system and changes in the tax schedule is then given by

$$y_0^{cf} = y_{gross,0} + \widehat{\Delta y}_{lab} + y_{transf,1} - y_{sscontr,0} - tax_1(y_{gross,0} + \widehat{\Delta y}_{lab}), \tag{2.7}$$

where $y_{gross,0}$ denotes period 0 market incomes from all sources, $y_{transf,1}$ are government transfers that possibly include counterfactual changes, $y_{sscontr,0}$ are factual period 0 household social security contributions, and $tax_1(\cdot)$ is the counterfactual tax schedule of period 1. If changes in labour market returns are not desired in the counterfactual situation, then $\widehat{\Delta} y_{lab}$ is set to zero. Analogously, if changes in the transfer system are not of interest, $y_{transf,1}$ is replaced by factual transfers $y_{transf,0}$. Finally, if changes in the tax schedule are not considered, the counterfactual tax schedule $tax_1(\cdot)$ is replaced by its factual counterpart $tax_0(\cdot)$.

Using (2.7), we only predict *changes* due to counterfactual variations. Our reference point is always factual household net income $y_{net,0} = y_{gross,0} + y_{transf,0} - y_{sscontr,0} - \; tax_0$. In this way we preserve as much as possible of the information on incomes and their heterogeneity as given in the sample. In short-hand notation, we express the changes to income in period 0 due to counterfactual variations as

$$y_0^{cf} = y_{net,0} + \widehat{\Delta} y_{lab} + \widehat{\Delta} tr - \widehat{\Delta} t,, \tag{2.8}$$

where $\widehat{\Delta} y_{lab}, \widehat{\Delta} tr, \widehat{\Delta} t$ represent the shifts due to counterfactual changes in labour market returns, the transfer system, and the tax schedule.

*Counterfactual densities incorporating stages 1 to 6*

Combining equations (2.1) through (2.8) one can define counterfactual income densities that combine any desired set of counterfactual variations. For example, the overall income distribution in period 0 that results if household structures, employment outcomes, labour market returns, and government transfers are fixed at their period 0 levels but the distribution of characteristics $x$ and the tax schedule are counterfactually set to their period 1 levels, is given by

$$f_0(y|t_h = 0, t_x = 1, t_e = 0, t_r = 0, t_{tr} = 0, t_t = 1) = f_0(y|0, 1, 0, 0, 0, 1). \tag{2.9}$$

Following DiNardo et al. (1996), counterfactual densities $f_0(y|t_h, t_x, t_e, t_r, t_{tr}, t_t)$ can be estimated as

$$\hat{f}(y|t_h, t_x, t_e, t_r, t_{tr}, t_t) =$$
$$= \sum_{j=1}^{6} \sum_{i=1}^{n_j} \theta_i \Psi_j \Psi^i_{x|j} \Psi^i_{e|x,j} K \left( \frac{y - (y_{net,0,i} + \widehat{\Delta} y_{lab,i} + \widehat{\Delta} tr_i - \widehat{\Delta} t_i)}{h} \right) \frac{1}{h}, \tag{2.10}$$

where $\theta_i$ denotes the sample weight of individual $i$, $n_j$ is the number of individuals in household type $j$, $K(\cdot)$ a kernel function, $h$ a bandwidth, and $\Psi_j = w_{1j}/w_{0j}$. If a particular counterfactual variation is not desired, the corresponding weighting factor $\Psi_j, \Psi^i_{x|j}, \Psi^i_{e|x,j}$ is set to 1, or the corresponding shift factors $\widehat{\Delta} y_{lab,i}, \widehat{\Delta} tr_i, \widehat{\Delta} t_i$ are equal to zero, respectively.

*Estimation of inequality and poverty indices*

Given an estimated income density $\hat{f}(y)$, we use numerical integration methods to calculate the inequality and poverty indices shown in table 5 (for the definition and properties of these indices, see Cowell (2000)).

*Statistical inference*

We compute bootstrap standard errors for all our decomposition results. These bootstrap standard errors correctly take into account the serial correlation of individual observations over time as well as their clustering in households. This is achieved by resampling from the universe of longitudinal household observations (see Biewen (2002) for a discussion of these issues).

## 2.4 DATA AND SPECIFICATION ISSUES

We base our analysis on data from the German Socio-Economic Panel (SOEP). Our data refers to individuals (including children). We use all available SOEP subsamples and all our calculations are weighted with the appropriate sample weights. Our main income variable is real annual equivalized personal net in-

Table 5: Inequality and poverty indices

| Index | Abbr. | Estimator |
|---|---|---|
| Quantile ratio 90/10 | Q90/Q10 | $\widehat{q90q10}(\hat{f}) = \hat{q}_{90}/\hat{q}_{10}$ |
| Quantile ratio 90/50 | Q90/Q50 | $\widehat{q90q50}(\hat{f}) = \hat{q}_{90}/\hat{q}_{50}$ |
| Quantile ratio 50/10 | Q50/Q10 | $\widehat{q50q10}(\hat{f}) = \hat{q}_{50}/\hat{q}_{10}$ |
| Theil's measure | Theil | $\widehat{\text{theil}}(\hat{f}) = \int \frac{y}{\mu(\hat{f})} \log\left(\frac{y}{\mu(\hat{f})}\right) \hat{f}(y)\, dy$ |
| Mean log deviation | MLD | $\widehat{\text{mld}}(\hat{f}) = -\int \log\left(\frac{y}{\mu(\hat{f})}\right) \hat{f}(y)\, dy$ |
| Gini coefficient | Gini | $\widehat{\text{gini}}(\hat{f}) = \int y(2\hat{F}(y) - 1)\hat{f}(y)\, dy$ |
| Foster et al. | FGT($\alpha$) | $\widehat{\text{FGT}}(\hat{f}, \alpha) = \int_{\{y<p(\hat{f})\}} \left(\frac{p(\hat{f})-y}{p(\hat{f})}\right)^{\alpha} \hat{f}(y) dy$ |

Note: $\text{FGT}(0) =$ poverty headcount, $\text{FGT}(1) =$ poverty gap measure, $p(\hat{f})$ = poverty line, $\alpha \geqslant 0$

come which is calculated from annual net household income. Annual net household income is given by gross income plus transfers minus social security contributions and taxes. Our data set contains information on each of these components of net income. Taxes were calculated by the data provider, the DIW Berlin, using the official rules. For more details on the definition of the different variables, see Grabka (2009). Our definitions are in general similar to the ones used in the official 'Report on Poverty and Richness' (Federal Government of Germany (2008)). There are two important differences, however. We do not consider imputed rental values and modifications to gross income due to differential treatment of population subgroups with respect to social security contributions.

In order to compute the individual income of the members of a given household, household net income is divided by the sum of equivalence weights defined by the OECD equivalence scale (the household head receives a weight of 1, additional household members over 14 years receive a weight of 0.5, household aged 14 years or less receive a weight of 0.3). In a robustness analysis, we consider two alternative equivalence scales to see whether our results depend on this particular choice (see below). Following recommendations and practice of the Statistical Office of the European Commission, we set the poverty line to 60 percent of the median of equivalized personal incomes in a given year.

As indicated above, we define six different household types: i) single pensioner households (65 years or older), ii) multiple pensioner households (at least one household member is 65 years or older and no household member is under 55), iii) single adults without children, iv) multiple adults without children, v) single adults with children, and vi) multiple adults with children. Our motivation for this classification is that it combines information on both the principal age composition and the structure of households, thus identifying population subgroups that share a similar economic and social position.

As further household characteristics $x$ we consider the number of adults in the household, the fraction of female adults in the household, the fraction of adult household members with different educational qualifications (university degree, high school and/or vocational training, no such degree or qualification), the fraction of adult household members with non-German nationality, the fraction of adult household members with disabilities, the fraction of married adults in the household, the fraction of household members in different age groups (0-3 years, 4-11 years, 12-17 years, 18-30 years, 31-50 years, 51-64 years, 65 or older), and a dummy indicating whether the household resided in East Germany (see table 11 in the appendix for details).

Employment outcomes $e$ are defined in an ordinal way: i) no part-time or full-time workers in the household, ii) no full-time workers but at least one part-time worker, iii) one full-time worker but no part-time workers iv) one full-time worker and at least one part-time worker, v) at least two full-time workers. The category 'part-time work' also includes marginal employment ('geringfügige Beschäftigung'). Category iii) also includes the case where one individual holds a part-time or marginal part-time job in addition to a full-time job. The evolution of the share of individuals living in households with each of the six possible outcomes is given in figure 12. We estimate the probability for each household employment outcome $e$ conditional on household characteristics $x$ using ordinal logit models. All estimations are carried out for each household type separately (see table 12 in the appendix for details).

In order to estimate labour market returns, we regress log household labour income on household characteristics, employment outcome categories, and a full set of interactions. We drop regressors that turn out insignificant. Again, all regressions are carried out separately for each household type (see table 13 in the appendix for details).

In order to analyze the effects the key changes in the transfer system, we simulate unemployment benefits II for former recipients of unemployment assistance and for former recipients of social assistance who are able to take part in the labour market. We also simulate the reduction of the maximum entitlement period for recipients of unemployment benefit I who will fall back to unemployment benefit II payments after the end of unemployment benefit I. We do not consider behavourial reactions to these changes in the transfer system. In this way, we will probably overestimate the distributional effects of these changes as individuals will typically try to take actions in order to make up for income losses incurred (in a related analysis however, Arntz et al. (2007), suggest that taking into account behavioural reactions makes little difference). In the sensitivity analysis involving the years 2007/2008, we in addition take into account the introduction of a new parental leave benefit which was introduced in 2007. More details on the simulation of transfer changes are given in the appendix.

The tax schedule is estimated using a flexible polynomial of third degree in household gross income along with suitable interactions with variables such as marital status or children (i.e. we regress the household tax variable as given in the data on a polynomial in household taxable income and interactions with other characteristics, results are available on request.) The method of calculating taxes using regressions is also used in Frenette et al. (2007), and can be seen as a parsimonious variant of micro-simulation.[55] The regressions are only carried out for nonzero tax values. A household pays no tax if its gross income is below the sum of personal tax exemptions. When calculating counterfactual tax values, we first check whether this is the case. We consider different kinds of personal

55 For more advanced uses of micro-simulation methods for studying the effect of taxes on the income distribution, see e.g. Bargain and Callan (2010), Bargain (2012a) or Bargain (2012b). These articles also discuss the possibility of decomposing the effect of tax changes into effects due to 'fiscal drag', and other effects. This is an interesting question which we defer to future research.

tax exemptions including the rules for the (changing) taxable share of old age pensions as well as standard deductions for labour income, capital income, and insurance contributions. We calculate positive tax values using the estimated tax schedule only if household gross income exceeds the sum of personal tax exemptions and impute a value of zero otherwise. We note that, due to their complexity and due to the nature of our tax variable (which is at the household level) we are not able to replicate all the details of household taxation in Germany. Our regressions fit the tax values given in the data extremely well, making us confident that we use the correct tax schedule.The fit of our tax predictions (including zero predictions) as measured by the squared correlation between predictions and actual values is usually around 95 percent with the exception of the smaller group of pensioners for which we obtain a fit of around 70 percent.

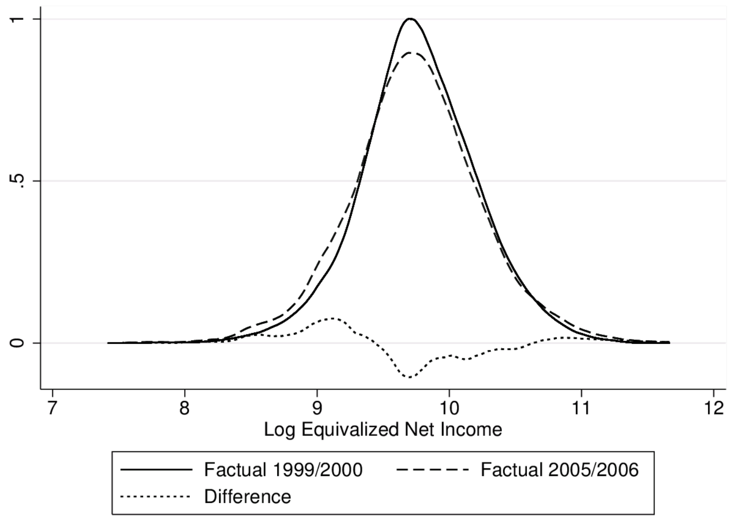
Finally, note that our analysis refers to inequality in net income between individuals, not households. All data are individual data but individuals are attributed the characteristics and the (equivalized) incomes of the households they live in. Incomes are expressed in year 2000 Euros (except for tax calculations which require nominal incomes). In order to minimize the outlier sensitivity of our regressions we exclude the bottom and top .5 percent of observations in our calculations. For expositional reasons we consider log equivalized incomes, which we appropriately transform back when calculating the inequality and poverty indices in table 5.

## 2.5 EMPIRICAL RESULTS

### 2.5.1 *Explaining changes between 1999/2000 and 2005/2006*

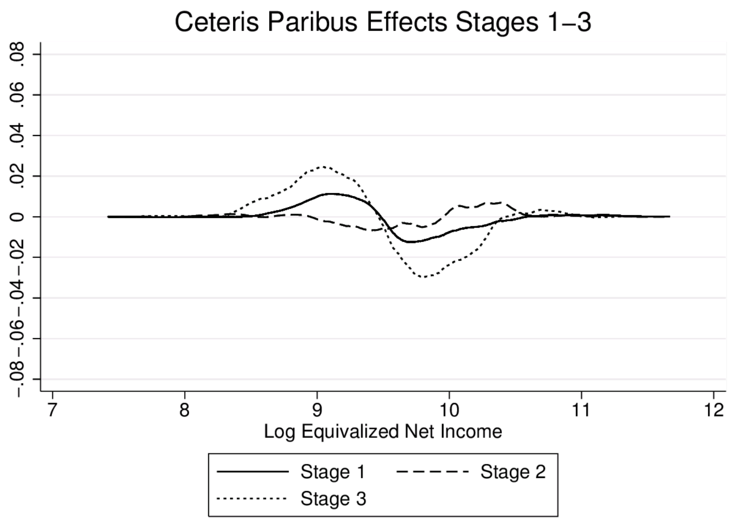
Figure 14 shows how the overall shape of the (log) income distribution changed from 1999/2000 ('period 0') to 2005/2006 ('period 1'). The picture that emerges is one of increasing spread, i.e. the distribution in 2005/2006 has a lower peak and fatter tails than the one in 1999/2000. However, the widening of the distribution is not symmetric. The changes seem particularly pronounced in the lower tail of the distribution, implying that low incomes were particularly affected by increasing inequality.

Figure 14: Overall change in density from 1999/2000 to 2005/2006




Source: SOEP, own calculations

Figure 15: Density change if only one factor is changed




Source: SOEP, own calculations. The graph shows the difference between the factual log-income density 1999/2000 and the counterfactual density that results if only one factor is changed to its 2005/2006 level. Stage 1 = density change if only the distribution of household types is changed, Stage 2 = density change if only the distribution of household characteristics is changed, Stage 3 = density change if only conditional employment outcomes are changed.

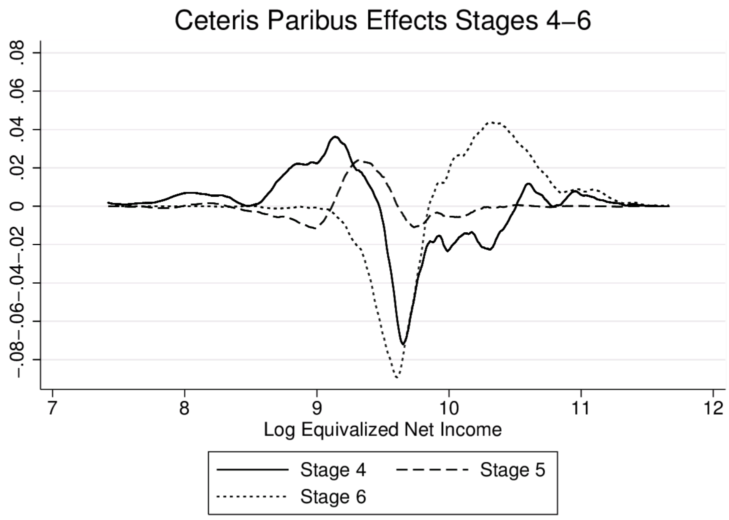
*'Ceteris paribus' effects of individual factors*

We now consider 'ceteris paribus' effects of the different factors, i.e. we change only one factor at a time to its period 1 level, but hold everything else fixed at the level of the base period 0. We believe that such an exercise comes closest to what one has in mind when asking about the 'effect' of a particular factor on the overall change. For example, the bold line in figure 15 shows the difference between the factual distribution in 1999/2000 and the income distribution that would prevail if the distribution of household types was changed to that of period 1, but everything else was held fixed at its period 0 level. The figure suggests that changes in household structures alone did not contribute much to the overall change in the distribution between 1999/2000 to 2005/2006. The overall density change is of the order of around 0.1 density points (see figure 14), while the differences due to changes in household types do not exceed 0.01 density points. The fact that changes in the distribution of household types do not contribute much over a period of five to six years is not surprising given that demographic change is slow.

In a similar way, the difference in figure 15 referring to 'stage 2' confirms that changing only the distribution of other household characteristics to its period 1 level had an even smaller effect on the distribution. By contrast, the dotted line shown in figure 15 relating to 'stage 3' demonstrates that a ceteris paribus change in conditional employment outcomes leads to a noticeable redistribution of mass from the middle to the bottom of the distribution. This suggests that changes in unemployment and part-time employment affected in particular individuals in the middle and lower part of the distribution. High income households (i.e. over 10.5 log-income points) appeared to be largely unaffected by such changes.

The bold line in figure 16 shows the considerable effects of a ceteris paribus change in labour market returns on the distribution of incomes. The changes mainly affect the middle and the bottom of the distribution, but in contrast to the case of changing employment outcomes, the very top of the distribution is also slightly affected. The ceteris paribus effect of the transfer changes due to the Hartz-reforms are given by the dashed line in figure 16. The impact of these

Figure 16: Density change if only one factor is changed




Source: SOEP, own calculations. The graph shows the difference between the factual log-income density 1999/2000 and the counterfactual density that results if only one factor is changed to its 2005/2006 level. Stage 4 = density change if only labour market returns are changed, Stage 5 = density change if only the transfer system is changed, Stage 6 = density change if only the tax schedule is changed.

changes on the overall distribution seems limited, but as expected, there is a shift from the middle and the very bottom of the distribution to the area between the middle and the bottom. This is consistent with the view that the Hartz-reforms hit middle income earners by replacing income dependent unemployment assistance by the basic income (i.e. unemployment benefit II), and by cutting the maximum entitlement period for unemployment benefit I. On the other hand, households with very low incomes benefited from the introduction of unemployment benefit II as its level was slightly higher than that of the former social assistance, and more households were eligible. Finally, the dotted line in figure 16 presents the ceteris paribus effect of changes in the tax schedule. These led to a considerable shift of the distribution to the right, but much more so for middle and especially for high incomes.[56] This suggests that middle and high incomes

56 Note that, due to the log-transformation, the graphs tend to overstate the importance of changes at the very top of the distribution. In a graph displaying densities of unlogged incomes, differences at the top of the distribution would be spread over fairly wide income intervals.

benefited overproportionally from reduced tax rates, while the density in the very low end of the distribution remained close to constant as these households usually do not pay any tax at all.

Our ultimate goal is to measure what percentage of the inequality increase between 1999/2000 to 2005/2006 can be accounted for by the different factors. Table 6 therefore summarizes what percentage of the overall increase as measured by various inequality and poverty indices can be explained by changing one factor at a time. The numbers largely confirm the findings from the graphical analysis. Only a relatively small percentage of the overall inequality increase can be explained by ceteris paribus changes in the distribution of household types (around 9 percent on average) or by changes in socio-economic attributes (around 3 percent on average). Ceteris paribus changes in conditional employment outcomes explain on average 29 percent of the total increase (column 3), which is a substantial contribution. Isolated changes in labour market returns account for an even larger share of the overall increase, namely on average 47 percent (column 4).

At -6 percent on average (column 5), changes in the transfer system have a slightly negative net effect on inequality (which is generally not significantly different from zero). At first glance, the finding that the effect of these reforms was inequality reducing rather than inequality increasing may seem surprising. However, as explained above, many individuals at the very bottom of the distribution actually benefitted from the reforms (former recipients of social assistance and individuals who were not eligible before the reforms). Moreover, there was a shift of mass from the middle of the distribution to the area below the middle. Together, this resulted in an equivalization in the lower half of the distribution (see dashed line in figure 16). Finally, it is well known that the reforms generally led to a major increase in government spending which exclusively accrued to the lower end of the distribution.

Table 6:

Ceteris paribus effects 1999/2000 – 2005/2006

| | Percentage of the overall inequality increase explained by ceteris paribus change of | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Household structure** | | **Household characteristics** | | **Employment outcomes** | | **Labour Market returns** | | **Transfer system** | | **Tax system** | |
| | **(1)** | | **(2)** | | **(3)** | | **(4)** | | **(5)** | | **(6)** | |
| **Q90/Q10** | 11.09 | (2.95) | 3.45 | (2.72) | 34.08 | (7.69) | 48.55 | (10.96) | -6.85 | (5.38) | 35.52 | (5.96) |
| **Q90/Q50** | 12.02 | (5.38) | 4.00 | (5.75) | 36.25 | (16.03) | 41.33 | (23.89) | 7.01 | (11.01) | 38.28 | (17.41) |
| **Q50/Q10** | 10.99 | (3.75) | 3.29 | (3.27) | 33.82 | (9.94) | 53.62 | (13.34) | -14.69 | (7.29) | 34.94 | (6.45) |
| **Theil** | 9.07 | (2.00) | 2.66 | (1.99) | 20.07 | (4.02) | 38.39 | (8.25) | -1.20 | (4.91) | 24.85 | (4.07) |
| **MLD** | 6.39 | (2.11) | 0.77 | (2.25) | 21.14 | (4.72) | 41.47 | (9.03) | -4.39 | (5.99) | 24.75 | (4.42) |
| **Gini** | 7.78 | (2.33) | 0.38 | (2.42) | 22.63 | (4.90) | 36.37 | (9.07) | -2.67 | (5.86) | 27.11 | (4.92) |
| **FGT(0)** | 11.48 | (3.65) | 3.33 | (3.16) | 32.75 | (8.73) | 53.33 | (12.52) | -13.71 | (7.18) | 37.58 | (6.62) |
| **FGT(1)** | 7.25 | (3.27) | 5.84 | (3.84) | 33.00 | (10.23) | 62.69 | (15.47) | -19.31 | (10.29) | 31.65 | (6.19) |

Source: SOEP, own calculations. The numbers in parentheses are bootstrap standard errors taking into account the longitudinal sample design, stratification, and the clustering of individuals in households.

Taken together, it is not implausible that the overall effect of the reform was in fact inequality reducing rather than inequality increasing.[57] As to the last factor considered in our analysis, ceteris paribus changes in the tax system account for an average 31 percent of the total inequality increase between 1999/2000 and 2005/2006 (column 6). This is also a considerable effect.

The general conclusion is that changes in employment outcomes, changes in the tax schedule and especially changes in labour market returns explain a major share of the the overall inequality increase while changes in household structures, household characteristics, and changes in the transfer system play a smaller role.

*Sequential decomposition of the increase in inequality and poverty*

A drawback of the ceteris paribus analysis presented so far is that the percentages contributed by each factor do not add up to the complete change and that nothing can be said about the importance of residual factors. We therefore proceed in the usual fashion (DiNardo et al. (1996), and Hyslop and Mare (2005)), and decompose the inequality increase 1999/2000 to 2005/2006 into a sequence of incremental changes that result when changes of the individual factors are accumulated. Forcing the sum of contributions to add up to 100 percent comes at the cost that the results may be path-dependent, i.e. they may depend on the particular order in which the different factors are sequentially changed, something which we address in a sensitivity analysis.[58] Given the shortcomings of both the ceteris paribus and the sequential decomposition analysis, our final assessment of the importance of the different factors will critically have to take into account all the available evidence from the different approaches.

57 Using the same data, the related study by Arntz et al. (2007), obtains exactly the same result. We have to add one qualification to our and Arntz et al.'s result, however. As mentioned, the introduction of unemployment benefit II was accompanied by a slight increase in benefit levels for former recipients of social assistance. However, some of this increase was only meant to compensate for one-time payments that were a part of the old social assistance system but absent in the new unemployment benefits II. To the extent that former social assistance recipients did not report these one-time payments as income in our data, these individuals gained more in our analysis than they did in reality, overstating possible inequality reducing effects of the reforms.

58 Biewen (2001), illustrates the problems of possible path dependencies in sequential decompositions such as the one considered here.

Using the idea of a sequential decomposition, the change in inequality between 1999/2000 and 2005/2006 can be decomposed as

$$I\left(\hat{f}_1(y)\right) - I\left(\hat{f}_0(y)\right) = \left[I\left(\hat{f}_0(y|1,0,0,0,0,0)\right) - I\left(\hat{f}_0(y|0,0,0,0,0,0)\right)\right] \tag{2.11}$$

$$+ \quad \left[I\left(\hat{f}_0(y|1,1,0,0,0,0)\right) - I\left(\hat{f}_0(y|1,0,0,0,0,0)\right)\right] \tag{2.12}$$

$$+ \quad \left[I\left(\hat{f}_0(y|1,1,1,0,0,0)\right) - I\left(\hat{f}_0(y|1,1,0,0,0,0)\right)\right] \tag{2.13}$$

$$+ \quad \left[I\left(\hat{f}_0(y|1,1,1,1,0,0)\right) - I\left(\hat{f}_0(y|1,1,1,0,0,0)\right)\right] \tag{2.14}$$

$$+ \quad \left[I\left(\hat{f}_0(y|1,1,1,1,1,0)\right) - I\left(\hat{f}_0(y|1,1,1,1,0,0)\right)\right] \tag{2.15}$$

$$+ \quad \left[I\left(\hat{f}_0(y|1,1,1,1,1,1)\right) - I\left(\hat{f}_0(y|1,1,1,1,1,0)\right)\right] \tag{2.16}$$

$$+ \quad \left[I\left(\hat{f}_1(y)\right) - I\left(\hat{f}_0(y|1,1,1,1,1,1)\right)\right] \tag{2.17}$$

where $I(\cdot)$ is one of the inequality or poverty indices in table 5. The overall inequality change $I\left(\hat{f}_1(y)\right) - I\left(\hat{f}_0(y)\right)$ is split up into parts contributed by changes in the household structure (2.11), changes in household characteristics (2.12), changes in conditional employment outcomes (2.13), changes in labour market returns (2.14), changes in the transfer system (2.15), changes in the tax schedule (2.16), and an unexplained residual (2.17).

Table 7 shows the contributions of each of the factors as a percentage of the overall inequality increase. For example, around 7.78 percent of the increase of the Gini coefficient from 1999/2000 to 2005/2006 are attributable to changes in household structures. The results largely reproduce the findings from the ceteris paribus analysis with the exception of the last two stages (transfer and tax changes). The effect of transfer changes are larger in magnitude, while that of the changes in the tax system are smaller when compared to the ceteris paribus analysis. The explanation is that in the sequential decomposition, changes in the transfer system are applied *after* employment outcomes (stage 3) and labour incomes (stage 4) are set to the situation of 2005/2006. As there are more low labour incomes and more unemployment in this situation, this increases the scope for effects of changes in the transfer system. Similarly, the effects of changes in the tax system may become smaller because low incomes are to a large extent exempt from taxes.

The unexplained residuals shown in the last column of table 7 suggest that the six factors taken together generally explain around 80 percent of the overall inequality increase. Exceptions are the Q90/Q50 ratio for which the residual is negative, and the Theil index for which it amounts to around 27 percent. In the case of the Q90/Q50 ratio, this may be explained by the generally high variability of the results for this index which is also reflected in large standard errors. It is unclear, why the standard errors for the Q90/Q50 ratio are larger than for the other indices, but this means that the results for this index are less reliable than for the other indices. The otherwise relatively small size of the residual suggests that most of the inequality increase is successfully accounted for by the factors considered above. The remaining residuals reflect the influence of rising inequality in capital incomes (see above) as well as all other unmodeled influences.

*Sensitivity analysis*

In order to check how sensitive the sequential decomposition is with respect to the decomposition order, we calculated the decomposition for all possible orders. As a simplification, we treated the changes in the tax- and transfer system as one stage, i.e. we considered orders in which the tax- and transfer system were changed at the same time. This results in $5! = 120$ different orderings. The results of this exercise are shown in table 8. Although the averaged results convey a picture that is qualitatively similar to what we presented above, it turns out that there is quite a lot variability in the contributions of the different factors, depending on the decomposition order chosen. This calls into question the common practice of carrying out such sequential decompositions and makes it necessary to better justify why a particular decomposition order is chosen.

Table 7:

Decomposition 1999/2000 – 2005/2006

| | Percentage of the overall inequality increase explained by sequential change of | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Household structure (1)** | | **Household characteristics (2)** | | **Employment outcomes (3)** | | **Labour Market returns (4)** | | **Transfer system (5)** | | **Tax system (6)** | | **Rest** |
| **Q90/Q10** | 11.09 | (2.95) | 3.85 | (2.95) | 31.43 | (7.67) | 54.03 | (12.62) | -24.42 | (8.73) | 18.64 | (6.40) | 5.40 |
| **Q90/Q50** | 12.02 | (5.38) | 3.02 | (5.85) | 34.44 | (15.57) | 39.09 | (24.20) | 10.39 | (14.77) | 29.33 | (18.53) | -28.29 |
| **Q50/Q10** | 10.99 | (3.75) | 4.42 | (3.65) | 30.25 | (9.70) | 61.43 | (16.10) | -43.18 | (12.75) | 12.12 | (5.75) | 23.97 |
| **Theil** | 9.07 | (2.00) | 2.03 | (2.09) | 17.56 | (3.70) | 41.27 | (8.62) | -5.55 | (6.48) | 8.44 | (7.82) | 27.18 |
| **MLD** | 6.39 | (2.11) | 3.11 | (2.29) | 21.00 | (4.45) | 49.45 | (10.16) | -11.58 | (9.00) | 17.39 | (9.54) | 14.24 |
| **Gini** | 7.78 | (2.33) | 2.74 | (2.43) | 22.38 | (4.66) | 40.45 | (9.51) | -7.03 | (8.73) | 17.12 | (9.82) | 16.56 |
| **FGT(0)** | 11.48 | (3.65) | 3.78 | (3.42) | 28.56 | (8.38) | 57.82 | (13.92) | -32.05 | (10.72) | 14.74 | (5.36) | 15.67 |
| **FGT(1)** | 7.25 | (3.27) | 6.85 | (4.22) | 28.36 | (9.21) | 78.62 | (19.38) | -46.69 | (14.14) | 15.64 | (6.17) | 9.98 |

Source: SOEP, own calculations. The numbers in parentheses are bootstrap standard errors taking into account the longitudinal sample design, stratification, and the clustering of individuals in households.

Table 8: Results from 120 possible sequential decompositions for the decomposition 1999/2000 – 2005/2006

| | Marginal percentage change attributable to | | | | | |
|---|---|---|---|---|---|---|
| | **Household structure (1)** | **Household charact. (2)** | **Employment outcomes (3)** | **Lab. market returns (4)** | **Transfer system (5)** | **Tax system (6)** |
| | | | **Q90/Q10 (Total Change = .379)** | | | |
| **Benchmark** | 11.09 | 3.85 | 31.43 | 54.03 | -24.42 | 18.64 |
| **Mean** | 8.99 | 3.94 | 30.94 | 39.38 | -15.50 | 26.83 |
| **Std.dev.** | 1.69 | 1.26 | 4.69 | 11.28 | 7.09 | 8.54 |
| **Min** | 6.47 | 1.07 | 24.19 | 24.09 | -24.42 | 17.22 |
| **Max** | 11.48 | 5.94 | 39.92 | 57.14 | -6.85 | 37.01 |
| | | | **Q90/Q50 (Total Change = .076)** | | | |
| **Benchmark** | 12.02 | 3.01 | 34.44 | 39.08 | 10.39 | 29.33 |
| **Mean** | 11.77 | 2.17 | 36.67 | 37.06 | 8.18 | 32.41 |
| **Std.dev.** | 1.54 | 1.18 | 1.22 | 3.51 | 2.35 | 4.57 |
| **Min** | 8.27 | -.01 | 34.28 | 32.45 | 4.00 | 26.07 |
| **Max** | 14.51 | 4.00 | 39.17 | 41.96 | 11.42 | 41.24 |
| | | | **Q50/Q10 (Total Change = .133)** | | | |
| **Benchmark** | 10.99 | 4.42 | 30.25 | 61.43 | -43.18 | 12.12 |
| **Mean** | 7.46 | 4.90 | 27.69 | 40.61 | -28.53 | 23.88 |
| **Std.dev.** | 2.91 | 2.10 | 7.84 | 16.16 | 11.64 | 11.48 |
| **Min** | 3.45 | .56 | 16.37 | 18.59 | -43.36 | 10.91 |
| **Max** | 12.12 | 8.05 | 41.08 | 65.62 | -13.64 | 36.24 |
| | | | **Theil (Total Change = .027)** | | | |
| **Benchmark** | 9.07 | 2.03 | 17.56 | 41.27 | -5.55 | 8.44 |
| **Mean** | 7.38 | 2.21 | 19.36 | 32.08 | -3.73 | 15.48 |
| **Std.dev.** | 1.10 | .79 | 1.77 | 7.21 | 1.75 | 7.14 |
| **Min** | 5.60 | .50 | 16.51 | 22.19 | -5.55 | 7.85 |
| **Max** | 9.10 | 3.62 | 23.31 | 43.16 | -1.20 | 23.71 |
| | | | **MLD (Total Change = .024)** | | | |
| **Benchmark** | 6.39 | 3.11 | 21.00 | 49.45 | -11.58 | 17.39 |
| **Mean** | 6.94 | 2.60 | 22.47 | 36.91 | -7.54 | 24.35 |
| **Std.dev.** | .69 | 1.28 | 2.78 | 8.35 | 3.17 | 6.88 |
| **Min** | 5.96 | .76 | 18.33 | 25.06 | -11.58 | 17.24 |
| **Max** | 8.73 | 4.91 | 29.19 | 51.00 | -3.04 | 32.38 |
| | | | **Gini (Total Change = .026)** | | | |
| **Benchmark** | 7.78 | 2.74 | 22.38 | 40.45 | -7.03 | 17.12 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Mean** | 7.90 | 1.70 | 23.23 | 29.19 | -4.53 | 25.93 |
| **Std.dev.** | .86 | .90 | 2.04 | 9.26 | 2.04 | 8.53 |
| **Min** | 6.59 | .38 | 20.12 | 17.79 | -7.09 | 16.92 |
| **Max** | 10.13 | 3.32 | 28.45 | 42.50 | -1.26 | 35.55 |
| | | | **FGT(0) (Total Change = .028)** | | | |
| **Benchmark** | 11.48 | 3.78 | 28.56 | 57.82 | -32.05 | 14.74 |
| **Mean** | 7.16 | 4.36 | 27.73 | 39.35 | -22.87 | 28.59 |
| **Std.dev.** | 3.43 | 1.97 | 5.67 | 15.71 | 7.65 | 13.28 |
| **Min** | 2.44 | .26 | 19.18 | 18.10 | -32.91 | 13.88 |
| **Max** | 11.92 | 7.07 | 37.48 | 61.85 | -13.00 | 43.89 |
| | | | **FGT(1) (Total Change = .008)** | | | |
| **Benchmark** | 7.25 | 6.85 | 28.36 | 78.62 | -46.69 | 15.64 |
| **Mean** | 5.71 | 6.98 | 29.27 | 58.40 | -32.79 | 22.43 |
| **Std.dev.** | 1.30 | 2.14 | 8.02 | 11.59 | 11.17 | 6.69 |
| **Min** | 3.94 | 2.87 | 14.97 | 38.38 | -47.01 | 14.34 |
| **Max** | 8.25 | 11.79 | 45.16 | 79.96 | -18.99 | 29.84 |

Source: SOEP, own calculations. We consider only sequences in which the tax- and transfer system are changed at the same time.

We give the following reasons why the order decribed in (2.11) to (2.16) is more plausible than other orders. First, in decomposition (2.11) to (2.16) factors are basically changed in the order of their 'pre-determinedness', i.e. household type and household characteristics are chosen before employment outcomes, labour incomes are the result of household characteristics and employment outcomes, and taxes and transfers are both the result of labour incomes and household characteristics. Second, the order used in (2.11) to (2.16) essentially reproduces the contributions that result from the ceteris paribus analysis, which is appealing on a-priori grounds, but which has the disadvantage of non-additivity. We go one step further and claim that the ceteris paribus analysis is more informative with respect to the 'effects' of the different factors as they directly answer the question of what happens if only one factor is changed in isolation.

We have also carried out further sensitivity checks, in particular we varied the bandwidth used in our density estimations and the equivalence scale used to make incomes comparable across household types. A combination of graphical inspection and Silverman's rule of thumb led us to use a fixed bandwidth of 0.175 throughout our analysis (Hyslop and Mare (2005) use a similar fixed bandwidth). Our numerical results change only little if we vary the bandwidth between 0.1 and 0.3, and qualitative results remain unchanged. The same applies if we use two alternative equivalence scales (we used the so-called Luxembourg scale which deflates household incomes by the square root of household size, and another scale which assigns a weight of 1 to the household head, and weights of 0.7 and 0.5 to additional household members over 14 years, and up to 14 years, respectively).

### 2.5.2 *Explaining changes between 1999/2000 and 2007/2008*

As another sensitivity test, we also carry out our analysis for the period 1999/2000 to 2007/2008. In view of the considerable changes in employment between 2005/2006 and 2007/2008 (figures 11 and 12), and the reversing trend in labour income inequality after 2005 (figure 13), this will allow us to check whether the decomposition results change in the right direction. Moreover, the tax schedule did not change in a major way between 2005/2006 and 2007/2008, so that the contribution of tax changes in the decomposition should essentially remain constant.

The results for the case 1999/2000 to 2007/2008 are given in tables 9 and 10. The contributions of changes in household structures and household characteristics are slightly higher than those in the decomposition for 1999/2000 to 2005/2006 (columns 1 and 2 of tables 9 and 10). The likely reason is that the change of these factors is relatively slow so that its effects are stronger over the longer term. As expected, the contribution of employment changes is much reduced when compared to the decomposition 1999/2000 to 2005/2006. However, the fact that there is still a contribution to the trend of rising inequality although unemployment in 2007/2008 was as low as in 1999/2000 implies that it was not so much rising unemployment but other changes in employment that contributed to the inequality increase between 1999/2000 and 2005/2006. This fits with the fact, that those other changes in employment structures – increasing part-time and marginal part-time work – accelerated after 2006 (see figure 12), probably partly offsetting inequality reducing effects of falling unemployment.

Table 9:

Ceteris paribus effects 1999/2000 – 2007/2008

| | Percentage of the overall inequality increase explained by ceteris paribus change of | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Household structure** | | **Household characteristics** | | **Employment outcomes** | | **Labour Market returns** | | **Transfer system** | | **Tax system** | |
| | **(1)** | | **(2)** | | **(3)** | | **(4)** | | **(5)** | | **(6)** | |
| **Q90/Q10** | 14.22 | (3.81) | 5.96 | (3.92) | 25.62 | (7.63) | 28.68 | (10.23) | 7.46 | (8.42) | 38.32 | (7.73) |
| **Q90/Q50** | 17.39 | (12.81) | 1.16 | (10.22) | 29.04 | (23.54) | -9.23 | (25.35) | 19.71 | (33.04) | 41.91 | (33.84) |
| **Q50/Q10** | 13.11 | (4.14) | 8.54 | (4.34) | 24.62 | (8.69) | 48.51 | (13.88) | 1.70 | (9.60) | 37.39 | (7.57) |
| **Theil** | 10.58 | (2.30) | 1.42 | (2.33) | 12.20 | (3.53) | 23.60 | (7.21) | 7.20 | (6.02) | 22.76 | (4.06) |
| **MLD** | 8.36 | (2.44) | 1.05 | (2.81) | 13.52 | (4.44) | 30.68 | (8.73) | 10.16 | (7.70) | 21.00 | (4.65) |
| **Gini** | 10.00 | (2.65) | 0.32 | (3.05) | 13.97 | (4.56) | 19.13 | (8.50) | 11.26 | (6.78) | 22.55 | (4.87) |
| **FGT(0)** | 12.96 | (3.76) | 8.28 | (4.03) | 22.76 | (7.41) | 44.47 | (12.17) | 4.08 | (8.11) | 37.52 | (6.98) |
| **FGT(1)** | 9.24 | (4.18) | 12.86 | (5.96) | 28.08 | (11.21) | 78.70 | (23.26) | -0.81 | (20.11) | 38.42 | (10.34) |

Source: SOEP, own calculations. The numbers in parentheses are bootstrap standard errors taking into account the longitudinal sample design, stratification, and the clustering of individuals in households.

Table 10:

Decomposition 1999/2000 – 2007/2008

| | Percentage of the overall inequality increase explained by sequential change of | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Household structure (1)** | | **Household characteristics (2)** | | **Employment outcomes (3)** | | **Labour Market returns (4)** | | **Transfer system (5)** | | **Tax system (6)** | | **Rest** |
| **Q90/Q10** | 14.22 | (3.81) | 6.06 | (4.03) | 24.65 | (7.23) | 36.63 | (12.35) | -5.67 | (9.94) | 22.46 | (8.90) | 1.65 |
| **Q90/Q50** | 17.39 | (12.81) | 0.00 | (10.95) | 28.05 | (24.64) | -5.87 | (25.89) | 18.82 | (37.80) | 39.22 | (32.41) | 2.39 |
| **Q50/Q10** | 13.11 | (4.14) | 9.20 | (4.45) | 23.27 | (8.17) | 58.09 | (16.50) | -18.22 | (10.75) | 13.34 | (6.59 ) | 1.22 |
| **Theil** | 10.58 | (2.30) | 0.65 | (2.40) | 11.05 | (3.28) | 28.61 | (7.99) | 1.48 | (10.01) | 10.52 | (10.67) | 37.10 |
| **MLD** | 8.36 | (2.44) | 3.22 | (2.78) | 15.10 | ( 4.17) | 41.88 | (9.81) | -2.53 | (13.41) | 9.83 | (14.29) | 24.14 |
| **Gini** | 10.00 | (2.65) | 2.36 | (2.98) | 15.54 | (4.29) | 24.88 | (24.88) | 4.14 | (13.28) | 8.83 | (14.32) | 34.25 |
| **FGT(0)** | 12.96 | (3.76) | 8.26 | (4.01) | 21.20 | (6.94) | 45.32 | (13.07) | -3.72 | (7.62) | 13.85 | (13.85) | 2.14 |
| **FGT(1)** | 9.24 | (4.18) | 13.95 | (6.21) | 25.87 | (10.24) | 97.84 | (28.08) | -31.70 | (16.49) | 20.29 | (9.08) | -35.49 |

Source: SOEP, own calculations. The numbers in parentheses are bootstrap standard errors taking into account the longitudinal sample design, stratification, and the clustering of individuals in households.

The effects of changes in the transfer system (which now include the introduction of the new parental leave benefit) are still relatively small (slightly above zero in table 9, and slightly below zero in table 10). The contribution of tax changes remains about the same as in the decomposition 1999/2000 to 2005/2006, showing that the introduction of the 'rich tax' (which only affects a tiny fraction of the population) did not have major distributive effects. Taken together, the results for the comparison 1999/2000 to 2007/2008 are consistent with our previous results, and all changes go into the expected directions.

## 2.6 CONCLUSION

This chapter analyzed which of different possible factors were behind the recent increase in personal income inequality in Germany. Our contribution lies in the fact that, although possible reasons for rising income inequality in different countries have been discussed and documented many times, little is known about their quantitative importance. Such information is relevant from a policy point of view. For example, it is essential to know whether growing income inequality is a consequence of employment changes, or of inequality in labour market returns. Our results suggest that the conspicuous increase in inequality and poverty over the period 1999/2000 to 2005/2006 was mostly due to increasing dispersion in labour market incomes, which has been attributed to skill-biased technical progress, deunionization and supply side effects (Dustmann et al. (2009), Antonczyk et al. (2010)). Other considerable effects come from changes in employment and changes in the tax system. Together, these three factors explain about 80 percent of the overall increase, where about one half is contributed by increasing inequality in labour incomes, and the other half is equally shared by employment changes and changes of the tax system.

By contrast, changes in household structures, household characteristics, and changes in the transfer system seem to have played a much smaller role. The latter is in sharp contrast to the widespread view that the changes in the transfer system, which were introduced in the year 2005 within a larger set of labour market reforms ('Hartz-reforms'), led to a drastic increase in inequality. In line with the results reported in Arntz et al. (2007), our results suggest that the changes in benefits due to the 'Hartz-reforms' were slightly inequality reducing rather than inequality increasing. Another interesting and policy-relevant conclusion of our analysis is that although the unprecedented inequality increase between 1999/2000 and 2005/2006 was accompanied by a steep increase in unemployment, this increase was probably only to a smaller extent responsible for the inequality increase. Our results rather suggest that other changes in employment patterns, especially the growth of part-time and marginal part-time work have contributed to the overall increase. This fits with the fact that overall inequality remained at its high level after 2005, while unemployment drastically fell but part-time, marginal part-time, and general employment grew at accelerated rates. The fact that we also measure a substantial contribution coming from tax changes shows that tax reforms – which were carried out in many countries – may also have an important effect on the income distribution. Often, the motivation for such tax reforms is to avoid 'fiscal drag', i.e. the automatically rising tax burden under a progressive tax schedule in the presence of inflationary income growth. Our results suggest that the tax reforms carried out in Germany did not only fight fiscal drag but changed the progression structure of the tax schedule in a substantial way.

We believe that our results contribute to the understanding of recent changes in the German income distribution and possibly that of other countries. As many other approaches, our approach has some limitations that should be borne in mind when interpreting the results. Apart from the many simplifying assumptions that have to be made when modeling the complex mechanisms of household arrangements, employment structures, labour market incomes, tax- and transfer rules, the methodological setup used here is not suited to address equilibrium effects and possibly complex interactions between the different factors considered. For example, changes in the tax and transfer system may also influence labour supply decisions and therefore eventually have an impact on the distribution (via changing employment outcomes) that goes beyond the direct effect. Similarly, increasing wage differentials and more wage flexibility may eventually lead to more employment, possibly countering their direct, inequality increasing effects. These aspects should be borne in mind when interpreting our results.

## 2.7 APPENDIX

### 2.7.1 *Details on the simulation of transfer changes*

*Unemployment benefit II*

Our approach to simulating unemployment benefit II is similar to that used in Becker and Hauser (2006), but an important difference is that we also simulate unemployment benefit II for former recipients of social assistance who are in principle able to take part in the labor market (rather than simulating it only for former recipients of unemployment assistance).

The steps for the simulation are as follows:[59]

1) We first determine the need of a household as the sum of personal allowances including housing costs and special allowances for lone parents and persons with disabilities. Housing costs are determined as actual housing costs, capped at 60 percent of household net income.

2) From the need, one has to subtract all other household income including labour income, capital income and certain transfers such as child allowances. Exempt from subtraction are parental leave benefits. There are also certain allowances for labour income which do not have to be subtracted. We implement these allowances in detail (100 percent of the first 100 Euros labour income, 20 percent of labour income between 200 and 800 Euros, 10 percent of labour income between 800 and 1200 Euros, 1500 Euros instead of 1200 Euros if children are present, everything per household member). Finally, there are other allowances such as for insurance contributions which we also implement.

3) Subtracting the items in 2) from the gross need in 1) yields the net need of the household.

4) Households are only eligible for unemployment benefit II if household wealth (assets, real estate, cash etc.) does not exceed certain limits. We

59 Further details are available on request. In most cases, we follow Becker and Hauser (2006).

implement the rules for these limits in as much detail as possible. They include allowances for savings for retirement (depending on age) as well as other allowances such as for necessary expenditures. We calculate household wealth from the information on capital income assuming an interest rate of 4 percent. Household can only receive unemployment benefit II if household wealth (after deductions) does not exceed the net need of the household. If a household receives income from rents, we exclude it from receiving unemployment benefit II.

5) Potential recipients of unemployment benefit II who received unemployment I before may receive an additional benefit for up to two years which aims to reduce the gap between the usually higher unemployment benefit I (which depends on former income) and the basic level of unemployment benefit II (§24 SGB II). This additional benefit amounts to 2/3 of the gap between former unemployment benefit I and the level of unemployment benefit II in the first year after unemployment benefit I has run out, and 1/3 in the second year.

6) If a household has passed the eligibility test, unemployment benefit II for this household is equal to its net need plus the extra benefit for former recipients of unemployment benefit I if applicable.

*Reduction of the maximum entitlement period for unemployment benefit I*

We identify spells of unemployment benefit I receipt and cut them to 12 months for persons below 55 years of age, and to 18 months for persons who are 55 years or older. For the remaining unemployment duration we simulate unemployment benefit II if applicable.

*Parental leave benefit*

From 2007 onwards, we simulate parental leave benefits ('Elterngeld') for mothers. We ignore parental leave benefits for fathers as only a tiny fraction of fathers took up this benefit. The old parental leave allowance ('Erziehungsgeld') which was replaced by the new parental leave benefit ran up to two years after the birth of a child, but it was much lower. The new parental leave benefit is paid for one year and amounts to 67 percent of the previous net labour income of the mother, but to at least 300 and at most 1800 Euros. For our simulation, we delete the old parental leave allowance in the data and impute the new parental leave benefit which we calculate according to the rules above.

### 2.7.2 *Additional Tables*

Table 11: Variable names

| | |
|---|---|
| hhemp | household employment outcome<br>(0 = 'no ft/no pt', 1 = 'no ft/at least 1 pt',<br>2 = '1 ft/no pt', 3 = '1 ft/at least 1 pt',<br>4 = 'at least 2 ft') |
| hhemp_d0...d4 | household employment category dummies |
| hhadult | number of adults in the household |
| f_ad_uni | fraction of adult HH-members with university degree |
| f_ad_abv | fraction of adult HH-members with high school degree and/or vocational training |
| f_ad_fem | fraction of adult HH-members female |
| f_ad_for | fraction of adult HH-members foreigner |
| f_ad_mar | fraction of adult HH-members married |
| f_ad_dis | fraction of adult HH-members with disabilities |
| f_ad_03 | fraction of HH-members aged 0-3 years |
| f_ad_11 | fraction of HH-members aged 4-11 years |
| f_ad_17 | fraction of HH-members aged 12-17 years |
| f_ad_30 | fraction of HH-members aged 18-30 years |
| f_ad_50 | fraction of HH-members aged 31-50 years |
| f_ad_65 | fraction of HH-members aged 51-65 years |
| f_ad_99 | fraction of HH-members aged 65 years or older |
| e | East Germany |

Table 12: Ordinal logit models

| | *Household Type 2* | | *Household Type 3* | | *Household Type 4* | | *Household Type 5* | | *Household Type 6* | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | 99/00 | 05/06 | 99/00 | 05/06 | 99/00 | 05/06 | 99/00 | 05/06 | 99/00 | 05/06 |
| adults | 2.347 | 1.787 | | | .488 | .647 | | | 1.137 | 1.195 |
| | (.299) | (.374) | | | (.059) | (.064) | | | (.107) | (.111) |
| f_ad_fem | | | | -.428 | -.913 | | -1.237 | | | |
| | | | | (.119) | (.250) | | (.512) | | | |
| f_ad_age50 | | | .859 | 1.090 | .603 | 1.257 | | | .488 | .444 |
| | | | (.146) | (.122) | (.135) | (.109) | | | (.158) | (.227) |
| f_ad_age64 | | | -.553 | | -1.340 | | | | -.464 | -.771 |
| | | | (.168) | | (.137) | | | | (.309) | (.446) |
| f_ad_age99 | -2.961 | -3.435 | | | | | | | -2.777 | -3.740 |
| | (.468) | (.366) | | | | | | | (1.106) | (.978) |
| f_ad_uni | 1.293 | 1.063 | .890 | 1.619 | 2.104 | 2.251 | 1.620 | 1.526 | 1.627 | 2.389 |
| | (.411) | (.389) | (.217) | (.213) | (.172) | (.188) | (.380) | (.402) | (.191) | (.233) |
| f_ad_abv | | .508 | .468 | .876 | 1.627 | 1.991 | .380 | .953 | 1.405 | 2.191 |
| | | (.358) | (.167) | (.174) | (.136) | (.144) | (.264) | (.320) | (.163) | (.200) |
| f_ad_dis | -.838 | -.852 | -1.062 | -.861 | -1.299 | -1.744 | | | | -.686 |
| | (.420) | (.313) | (.217) | (.188) | (.175) | (.191) | | | | (.328) |
| f_ad_mar | | | | .293 | | | | | 1.035 | 1.083 |
| | | | | (.220) | | | | | (.202) | (.195) |
| f_ad_for | | | | -.488 | .634 | | 1.192 | .861 | | |
| | | | | (.274) | (.205) | | (.467) | (.443) | | |
| f_ch_11 | | | | | | | 1.726 | 2.885 | .672 | .940 |
| | | | | | | | (.525) | (.646) | (.138) | (.139) |
| f_ch_17 | | | | | | | 2.205 | 3.633 | 1.301 | 1.302 |
| | | | | | | | (.527) | (.653) | (.142) | (.149) |
| e | | -.939 | -.809 | -.723 | -.209 | -.372 | | | .368 | |
| | | (.207) | (.136) | (.128) | (.089) | (.096) | | | (.123) | |
| /cut1 | 4.478 | 2.886 | -.889 | -.262 | -.641 | 1.008 | .266 | 3.468 | 2.515 | 3.740 |
| /cut2 | 5.31 | 3.933 | -.384 | .359 | -.044 | 1.841 | 1.949 | 5.193 | 3.288 | 4.596 |
| /cut3 | | | | | 1.487 | 3.293 | | | 5.785 | 6.614 |
| /cut4 | | | | | 2.359 | 4.402 | | | 7.596 | 8.780 |
| pseudo $R^2$ | .197 | .136 | .127 | .113 | .107 | .084 | .098 | .110 | .079 | .094 |
| clusters | 1345 | 1826 | 2103 | 2211 | 3750 | 3782 | 421 | 460 | 3489 | 2925 |

Source: SOEP, own calculations. Standard errors account for clustering of observations in households.

Table 13: Regression of log labour incomes on household characteristics (× denotes interaction effects)

| *Household Type 1* | | | | | | |
|---|---|---|---|---|---|---|
| (no regression because labour income negligible) | | | | | | |

| *Household Type 2* | | | | | | |
|---|---|---|---|---|---|---|
| Variable | 1999/2000 | | 2005/2006 | | 2007/2008 | |
| hhemp_d1 | .589 | (.233) | | | -1.038 | (.734) |
| hhemp_d2 | 1.995 | (.220) | 1.467 | (.129) | -.423 | (.188) |
| hhemp_d3 | 2.125 | (.224) | 1.836 | (.209) | | |
| hhemp_d4 | 2.833 | (.360) | 2.026 | (.227) | | |
| hhadult | | | .306 | (.148) | | |
| f_ad_uni | .558 | (.263) | .606 | (.203) | | |
| f_ad_fem | .558 | (.263) | .606 | (.203) | -3.739 | (.850) |
| f_ad_dis | -.501 | (.252) | | | -.238 | (.188) |
| f_ad_age99 | | | -1.109 | (.302) | -.885 | (.224) |
| e | -.280 | (.185) | -.453 | (.144) | | |
| hhemp× | | | | | | |
| c.f_ad_fem | | | | | | |
| 1 | | | 2.512 | (1.484) | | |
| 2 | | | 3.831 | (.961) | | |
| 3 | | | 3.117 | (.320) | | |
| 4 | | | 3.707 | (.623) | | |
| _cons | 8.242 | (.216) | 8.613 | (.374) | 11.031 | (.433) |
| $R^2$ | 0.434 | | 0.388 | | 0.442 | |
| Number of clusters | 283 | | 480 | | 511 | |

| *Household Type 3* | | | | | | |
|---|---|---|---|---|---|---|
| Variable | 1999/2000 | | 2005/2006 | | 2007/2008 | |
| hhemp_d1 | -.698 | (.242) | -.685 | (.372) | .579 | (.129) |
| hhemp_d2 | .864 | (.168) | 1.247 | (.202) | 1.728 | (.111) |
| hhemp_d3 | | | 1.114 | (.191) | | |
| f_ad_fem | | | | | -.111 | (.043) |
| f_ad_for | | | .720 | (.340) | .802 | (.354) |
| f_ad_age50 | | | | | | |
| f_ad_age64 | -.509 | (.226) | | | | |
| f_ad_uni | | | | | .403 | (.043) |
| f_ad_abv | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| f_ad_dis | -.885 | (.333) | -.797 | (.271) | -.215 | (.089) |
| e | -.394 | (.047) | -.579 | (.175) | -.374 | (.050) |
| hhemp × | | | | | | |
| c.f_ad_age50 | | | | | | |
| 1 | .032 | (.249) | -.249 | (.212) | | |
| 2 | .698 | (.143) | .433 | (.181) | | |
| 3 | .234 | (.040) | .092 | (.033) | | |
| hhemp × | | | | | | |
| c.f_ad_age64 | | | | | | |
| 1 | .807 | (.271) | | | | |
| 2 | .678 | (.231) | | | | |
| hhemp × | | | | | | |
| c.f_ad_uni | | | | | | |
| 0 | .138 | (.297) | .293 | (.294) | | |
| 1 | 1.203 | (.214) | 1.144 | (.319) | | |
| 2 | .466 | (.066) | .593 | (.087) | | |
| hhemp × | | | | | | |
| c.f_ad_abv | | | | | | |
| 0 | -.223 | (.171) | .104 | (.203) | | |
| 1 | .413 | (.193) | .637 | (.275) | | |
| 2 | .164 | (.060) | .253 | (.084) | | |
| hhemp × | | | | | | |
| c.f_ad_dis | | | | | | |
| 1 | .652 | (.415) | .143 | (.447) | | |
| 2 | .770 | (.333) | .835 | (.280) | | |
| hhemp × | | | | | | |
| c.f_ad_fem | | | | | | |
| 0 | | | -.254 | (.175) | | |
| 1 | | | .217 | (.173) | | |
| 2 | | | -.104 | (.034) | | |
| 3 | | | -.060 | (.000) | | |
| hhemp × | | | | | | |
| c.f_ad_for | | | | | | |
| 1 | | | -1.363 | (.491) | -1.092 | (.493) |
| 2 | | | -.714 | (.352) | -.770 | (.362) |
| hhemp × c.e | | | | | | |
| 1 | | | .502 | (.234) | | |
| 2 | | | .167 | (.185) | | |
| _cons | 9.181 | (.151) | 8.866 | (.191) | 8.615 | (.110) |
| $R^2$ | 0.501 | | 0.515 | | 0.479 | |
| Number of clusters | 1715 | | 1838 | | 1701 | |

| *Household Type 4* | | | | | | |
|---|---|---|---|---|---|---|
| Variable | 1999/2000 | | 2005/2006 | | 2007/2008 | |
| hhemp_d1 | | | | | .550 | (.143) |
| hhemp_d2 | .813 | (.095) | | | 1.435 | (.137) |
| hhemp_d3 | .886 | (.098) | | | 1.700 | (.140) |
| hhemp_d4 | 1.148 | (.089) | .655 | (.082) | 1.887 | (.139) |
| hhadult | .154 | (.015) | | | .103 | (.026) |
| f_ad_fem | -.165 | (.090) | -.159 | (.089) | | |
| f_ad_abv | .101 | (.051) | .126 | (.062) | | |
| f_ad_uni | | | .642 | (.081) | .468 | (.050) |
| f_ad_for | -.133 | (.064) | -.316 | (.081) | | |
| f_ad_dis | -1.160 | (.346) | -.337 | (.075) | -.329 | (.089) |
| f_ad_mar | -.502 | (.189) | | | | |
| f_ad_age50 | .244 | (.041) | .314 | (.048) | .294 | (.063) |
| f_ad_age64 | .109 | (.044) | .258 | (.049) | .217 | (.060) |
| f_ad_age99 | -.744 | (.155) | -.393 | (.123) | 1.699 | (.525) |
| e | -.3792 | (.027) | -.373 | (.034) | | |
| hhemp× | | | | | | |
| c.f_ad_mar | | | | | | |
| 1 | .103 | (.213) | | | | |
| 2 | .368 | (.202) | | | | |
| 3 | .578 | (.202) | | | | |
| 4 | .565 | (.191) | | | | |
| hhemp× | | | | | | |
| c.f_ad_age99 | | | | | | |
| 1 | | | | | -1.632 | (.634) |
| 2 | | | | | -2.046 | (.538) |
| 3 | | | | | -2.277 | (.581) |
| 4 | | | | | -2.538 | (.634) |
| hhemp× | | | | | | |
| c.hhadult | | | | | | |
| 0 | | | -.313 | (.049) | | |
| 1 | | | -.146 | (.038) | | |
| 2 | | | .185 | (.029) | | |
| 3 | | | .247 | (.026) | | |
| 4 | | | .100 | (.020) | | |
| hhemp× | | | | | | |
| c.f_ad_dis | | | | | | |
| 1 | .911 | (.439) | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | .987 | (.353) | | | | |
| 3 | 1.193 | (.368) | | | | |
| 4 | 1.110 | (.362) | | | | |
| hhemp× | | | | | | |
| c.f_ad_uni | | | | | | |
| 0 | -.209 | (.287) | | | | |
| 1 | 1.171 | (.212) | | | | |
| 2 | .576 | (.090) | | | | |
| 3 | .636 | (.078) | | | | |
| 4 | .501 | (.057) | | | | |
| _cons | 9.322 | (.119) | 9.881 | (.114) | 8.578 | (.142) |
| $R^2$ | 0.516 | | 0.465 | | 0.448 | |
| Number of clusters | 3370 | | 3536 | | 3193 | |

*Household Type 5*

| Variable | 1999/2000 | | 2005/2006 | | 2007/2008 | |
|---|---|---|---|---|---|---|
| hhemp_d1 | | | | | -2.65 | (.692) |
| hhemp_d2 | 1.494 | (.255) | 1.323 | (.186) | | |
| f_ad_abv | .503 | (.276) | | | | |
| f_ad_uni | .497 | (.244) | | | | |
| f_ad_for | 1.088 | (.240) | | | | |
| f_ad_dis | .629 | (.163) | -.660 | (.515) | 1.184 | (.357) |
| f_ad_age50 | | | | | -.824 | (.309) |
| f_ad_age64 | .683 | (.168) | -1.521 | (.441) | -2.743 | (1.185) |
| e | -.874 | (.347) | | | | |
| hhemp× | | | | | | |
| c.f_ad_fem | | | | | | |
| 0 | | | | | -.809 | (.422) |
| 1 | | | | | .619 | (.328) |
| 2 | | | | | .002 | (.160) |
| hhemp× | | | | | | |
| c.f_ad_for | | | | | | |
| 1 | -.623 | (.401) | | | | |
| 2 | -1.106 | (.278) | | | | |
| hhemp× | | | | | | |
| c.f_ad_abv | | | | | | |
| 0 | | | | | -.048 | (.377) |
| 1 | .346 | (.263) | | | 1.233 | (.355) |
| 2 | -.382 | (.327) | | | .226 | (.128) |
| hhemp× | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| c.f_ad_uni | | | | | | |
| 0 | | | | | -.103 | (.487) |
| 1 | | | | | 1.484 | (.393) |
| 2 | | | | | .812 | (.159) |
| hhemp× | | | | | | |
| c.f_ad_dis | | | | | | |
| 1 | -.023 | (.250) | 1.314 | (.533) | -1.106 | (.526) |
| 2 | -1.07 | (.220) | .287 | (.529) | -1.307 | (.367) |
| hhemp× | | | | | | |
| c.f_ch_age50 | | | | | | |
| 1 | | | | | 1.470 | (.560) |
| 2 | | | | | 1.145 | (.410) |
| hhemp× | | | | | | |
| c.f_ch_age64 | | | | | | |
| 1 | .914 | (.586) | | | 2.388 | (1.195) |
| 2 | 1.746 | (.516) | | | 3.236 | (1.226) |
| hhemp× | | | | | | |
| c.f_ch_age17 | | | | | | |
| 0 | -1.484 | (.323) | | | | |
| 1 | .726 | (.196) | | | | |
| 2 | -.017 | (.153) | | | | |
| hhemp×c.e | | | | | | |
| 1 | .973 | (.412) | | | | |
| 2 | .604 | (.359) | | | | |
| _cons | 8.637 | (.163) | 8.863 | (.148) | 9.557 | (.335) |
| $R^2$ | 0.390 | | 0.345 | | 0.461 | |
| Number of clusters | 347 | | 364 | | 323 | |

*Household Type 6*

| Variable | 1999/2000 | | 2005/2006 | | 2007/2008 | |
|---|---|---|---|---|---|---|
| hhemp_d1 | | | | | | |
| hhemp_d2 | | | .515 | (.196) | | |
| hhemp_d3 | | | .598 | (.200) | | |
| hhemp_d4 | .311 | (.191) | .877 | (.207) | | |
| hhadult | .557 | (.159) | .169 | (.036) | .170 | (.035) |
| f_ad_fem | -2.910 | (.681) | | | -2.258 | (.301) |
| f_ad_abv | .177 | (.067) | | | | |
| f_ad_uni | .678 | (.081) | .566 | (.044) | 1.724 | (.465) |
| f_ad_for | -.230 | (.066) | -.367 | (.058) | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| f_ad_dis | | | | | -1.879 | (.694) |
| f_ad_mar | -.832 | (.419) | -.788 | (.238) | | |
| f_ad_age50 | 1.064 | (.294) | 1.111 | (.227) | .585 | (.067) |
| f_ad_age64 | | | .591 | (.111) | .639 | (.109) |
| f_ad_age99 | 1.878 | (.846) | | | | |
| f_ch_age11 | -1.104 | (.342) | | | | |
| f_ch_age17 | -.961 | (.381) | -1.000 | (.233) | | |
| e | -.423 | (.034) | -.439 | (.047) | | |
| hhemp× | | | | | | |
| c.hhadult | | | | | | |
| 1 | -.426 | (.203) | | | | |
| 2 | -.446 | (.160) | | | | |
| 3 | -.397 | (.160) | | | | |
| 4 | -.489 | (.161) | | | | |
| hhemp× | | | | | | |
| c.f_ad_fem | | | | | | |
| 1 | 1.966 | (.776) | | | 1.494 | (.275) |
| 2 | 2.615 | (.676) | | | 1.783 | (.252) |
| 3 | 2.743 | (.678) | | | 2.236 | (.253) |
| 4 | 2.827 | (.689) | | | 2.349 | (.259) |
| hhemp× | | | | | | |
| c.f_ad_for | | | | | | |
| 0 | | | | | -1.031 | (.335) |
| 1 | | | | | -.958 | (.294) |
| 2 | | | | | -.175 | (.059) |
| 3 | | | | | -.354 | (.094) |
| 4 | | | | | .132 | (.090) |
| hhemp× | | | | | | |
| c.f_ad_mar | | | | | | |
| 1 | .162 | (.459) | .469 | (.281) | | |
| 2 | .783 | (.418) | .893 | (.246) | | |
| 3 | .875 | (.420) | .886 | (.249) | | |
| 4 | .798 | (.421) | .874 | (.254) | | |
| hhemp× | | | | | | |
| c.f_ad_age50 | | | | | | |
| 1 | -.675 | (.350) | -.270 | (.220) | | |
| 2 | -.796 | (.299) | -.655 | (.234) | | |
| 3 | -.914 | (.297) | -.505 | (.233) | | |
| 4 | -1.007 | (.298) | -.649 | (.243) | | |
| hhemp× | | | | | | |
| c.f_ad_age99 | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | -3.457 | (1.948) | | | | |
| 2 | -2.427 | (.847) | | | | |
| 3 | -2.227 | (.881) | | | | |
| 4 | -2.419 | (.895) | | | | |
| hhemp × | | | | | | |
| c.f_ad_uni | | | | | | |
| 1 | | | | | -1.430 | (.508) |
| 2 | | | | | -1.156 | (.471) |
| 3 | | | | | -1.201 | (.469) |
| 4 | | | | | -1.269 | (.476) |
| hhemp × | | | | | | |
| c.f_ch_age11 | | | | | | |
| 1 | 1.298 | (.427) | | | | |
| 2 | 1.084 | (.346) | | | | |
| 3 | 1.062 | (.345) | | | | |
| 4 | 1.239 | (.357) | | | | |
| hhemp × | | | | | | |
| c.f_ch_age17 | | | | | | |
| 1 | 1.223 | (.440) | .911 | (.251) | | |
| 2 | .990 | (.383) | .975 | (.240) | | |
| 3 | 1.049 | (.385) | 1.026 | (.238) | | |
| 4 | 1.229 | (.392) | 1.075 | (.244) | | |
| hhemp × | | | | | | |
| c.f_ad_dis | | | | | | |
| 1 | | | | | .712 | (.740) |
| 2 | | | | | 1.678 | (.712) |
| 3 | | | | | 1.543 | (.709) |
| 4 | | | | | 1.586 | (.706) |
| _cons | 10.197 | (.140) | 9.193 | (.207) | 9.859 | (.149) |
| $R^2$ | 0.427 | | 0.419 | | 0.408 | |
| Number of clusters | 3416 | | 2859 | | 2433 | |

Standard errors account for clustering of observations in households.

Source: SOEP, own calculations.

Part IV

# THE INCOME DISTRIBUTION AND THE BUSINESS CYCLE IN GERMANY - A SEMIPARAMETRIC APPROACH

The question of how the level of incomes and their dispersion behave in different macro-economic conditions (as given by the GDP, inflation, government expenditure and unemployment, among others) goes back at least to the 1950s, when Kuznets stated an inverse u-shaped relation between the inequality of incomes and the state of the economy (Kuznets (1955)). He argued that emerging economies that transform from agricultural societies to industrial ones become increasingly unequal at first, and, when the transformation shifts towards a service and knowledge oriented society, human capital becomes more important relative to physical capital so that the inequality of incomes decreases. Since then, many analyses have been conducted on this question and a large theoretical and empirical literature has developed. While the theoretical work generally identified a diversity of functional channels about how the income distribution and the business cycle are connected to each other, the overall effect remained unclear and most likely depends on the economy being analyzed.[60]

So far, the majority of the literature has used summary measures, i.e different quantiles (e.g. Blinder and Esaki (1978), Jäntti (1994)), point measures of inequality, or shape parameters under distributional assumptions (e.g. Thurow (1970), Salem and Mount (1974), Jäntti and Jenkins (2010)) to establish a connection between macroeconomic states and the income distribution. While there was a strong development in the econometric strategies involved, starting with (system) regression approaches (e.g. Blinder and Esaki (1978)) to time series settings involving cointegration of the potentially non-stationary time series of macroeconomic variables and inequality measures (e.g. Parker (1998), Jäntti and Jenkins (2010)), these papers discarded detailed information on the income distribution, when they collapsed the income distribution into summary statistics (e.g. of inequality).

When household panel data is available, we have the option to make use of detailed panel information to analyze the macroeconomic influence on incomes in tandem with microeconomic factors. While numerous papers on the relationship

60 See e.g. Easterly and Fischer (2001), Berthold et al. (2010).

between household level variables and the income structure exist, it also makes sense to follow a hybrid approach that uses both micro- and macro-level information and their possible interactions to analyze their impact on the distribution of incomes at the micro-level. The assumption that the effect of macroeconomic variables may be household specific is not new. For example, Blank (1989) found that elderly households and households with a female head were less responsive to the business cycle than others.

In this chapter, we adapt a hybrid-level approach following the strategy in Farré and Vella (2008), who analyze the impact of micro- and macro-level variables jointly with a semi-parametric double index model (Ichimura and Lee (1991)), using Spanish data. With several adaptations in technical and data related regards, we apply the method to German data between 1996 and 2010 extracted from the German Socio-Economic Panel (SOEP), to analyze potential mechanisms between the GDP, inflation, government expenditure and unemployment on the one hand and the income distribution on the other.

Making use of cross-sectional information, the double-index model allows for arbitrary functional relationships between indices of micro- and macro-level variables to describe their joint influence on the income distribution and allows us to employ a counterfactual approach to simulate different macroeconomic states and to analyze their impacts on the distribution of incomes. We find that the impact of changing macroeconomic conditions on the distribution of incomes is small, but in parts significant, drawing a comprehensible picture on potential mechanisms. In particular, we find that the level of government expenditure has a positive impact on the income levels, while the level of inflation is inversely related to it. We find further that both a negative GDP shock and a lower inflation are in line with lower inequality, and that higher government expenditure is also associated with lower inequality. As regards unemployment, a significant effect on the income distribution could not be revealed. We also find indications that qualitatively negative macroeconomic shocks go in line with relatively higher effects, while the short-term effects of positive shocks are vanishingly small.

The rest of this chapter is organized as follows. In Section 2, we summarize the theoretical role of macroeconomic relationships regarding the shape of the income distribution and give a summary over existing econometric approaches. In Section 3, we describe the double-index model in detail, while Section 4 gives an overview over the data to be used for estimation. In Section 5 we report our detailed results on the estimated relationship between the income distribution and the business cycle and discuss some counterfactual experiments. Section 6 concludes.

## 3.2 LITERATURE

### 3.2.1 *The Theoretical Role of Macroeconomic Conditions*

In this subsection, we partly follow Parker (1998) and Berthold et al. (2010), who summarize theoretical impacts of macroeconomic variables on the income distribution. In particular, our research concentrates on four factors to describe the state of the macro-economy: GDP, inflation, government final consumption expenditure[61], and unemployment.

General theories on a possible relationship between the distribution of incomes and macroeconomic circumstances go back at least to the 1950s. In his 1955 paper, Kuznets states an inverse u-shaped relationship between income inequality and the countries' general level of wealth, e.g. as measured by the gross domestic product. He argues that emerging economies transforming from agricultural societies to industrial ones become increasingly unequal as profits belong to those who provide capital for the transformation. When the transformation shifts towards a service and knowledge oriented society, human capital becomes increasingly important relative to real capital, leading to a more equal society. Nevertheless, these mechanisms are highly idealized. Above all, they do not take into account effects of globalization and of adaptation to international structures. The overall effect of changing GDP remains unclear.

61 "General government final consumption expenditure consists of expenditure incurred by government in its production of non-market final goods and services [...] and market goods and services provided as social transfers in kind." (OECDiLibrary).

Blinder and Esaki (1978) is an early study investigating how inflation and unemployment may relate to income inequality. Inflation is historically referred to as the "the cruelest tax" affecting classes with fixed incomes like retirees in countries where pensions are not adjusted for inflation or only adjusted for inflation with delay,[62] employees and workers with non-adjusted and non-renegotiated contracts and the poor, depending on possibly static welfare payments. Furthermore, as stated in Palmer and Barth (1977), under a progressive tax regime (as in Germany), poor income earners bear the greatest proportional increase in the tax burden, when cold progression occurs. Nevertheless, this secondary effect of inflation should have an equivalizing effect in general as the tax burden generally increases, leading to higher income redistribution.[63] On the other hand, there are also beneficiaries of non-anticipated positive inflationary shocks, namely borrowers with fixed interest rate contracts, making their lenders the losers of this complementary relationship. From the capitalists' point of view, as pointed out by Schultz (1969) and Buse (1982), there are two kinds of production related effects. The first is the so called demand-pull-inflation, where the rise of the prices exceeds the rise in production costs. For example, this may be the case for firms with fixed contracts that disconnect the firm from the effect of a higher price level on the production side. In this case, capitalists benefit from inflation. The other kind is the cost-push-inflation, the opposite of the former one, leading to decreasing profits. Due to the existence of manifold channels, an overall effect remains a-priori undetermined.

When turning to unemployment, we can identify different structural impacts on the income distribution in theory. The fact that unemployment benefits are lower than ordinary income does not imply a general statement with respect to inequality. Higher unemployment among the high earners for example, would lead to lower inequality, while the same circumstance in lower income groups has the opposite effect. The empirical fact that the risk of unemployment is higher for lower income groups leads to the assumption that an unemployment shock results in higher inequality. An additional secondary labor market effect is that, with higher unemployment, the income structure of the lower income recipients

62 See Minarik (1979).
63 See Heer and Süssmuth (2009).

also shifts downwards due to a shift in the supply-demand equilibrium on the job market, while higher income groups may be less affected. This secondary effect amplifies the overall impact on inequality. Another issue that points in the opposite direction is the fact that, in absolute terms, lower income recipients benefit much more from the social safety net of unemployment insurance. When analyzing incomes at the household level, a further effect may occur. As microeconomic units with potentially more that one earning member, households have the option of in-unit compensation in the event of employment loss. As pointed out in Parker (1996), members of household units compensate for this event by activating earning potential that Parker refers to as 'secondary worker effect'. This effect may lower unemployment effects considerably. The overall effect though might still be a tendency to higher inequality when unemployment rises.

Finally, the effect of government expenditure is somewhat more straight-forward. In a welfare state the state's redistribution mechanism is central to achieve social compensation and is an active tool for keeping the effect of disequivalizing forces within limits. The effect of higher government expenditure in welfare states should lower the income inequality.

### 3.2.2 *Empirical Approaches*

Empirical studies to reveal relationships between income inequality and different aspects of the macroeconomic business cycle had existed at least since the late 1960s.[64] The prominent approach of Blinder and Esaki (1978) uses a quantile income share technique. They regress the quantile shares of the income distribution on the aggregate rate of unemployment and inflation. By additionally controlling for general time effects, they identify the cyclical effects in a multiple regression setting. To allow for a more general error structure, Jäntti (1994) uses an extension of this model based on a generalized least squares (GLS) approach. As an alternative, other authors use different summary measure approaches and regress measures of inequality, such as the Gini coefficient, on determinants

64 For more details on the history of such studies see Parker (1998).

of the business cycle. Additionally, there exists a large number of papers using different shape parameters as dependent variables, after making a parametric distributional assumption on the size distribution of incomes, such as the Beta-, Gamma- or Sing-Maddala distribution. Here, authors derive the impact of macroeconomic variables by using functional relationships between inequality measures and distribution parameters.[65]

As Parker (1998) and Parker (2000) point out, despite their initial popularity, the above models have important theoretical drawbacks. The first one is that, in many cases, only few explanatory variables were included in the models. Although this strategy led to more powerful inference when data was sparse, it is likely that such models also suffer from omitted variable bias, making results questionable. In addition to this, these models are also likely to be affected by the non-stationarity of the variables involved. For example, inequality and different macroeconomic variables, such as GDP, experienced upward shifts in developed countries and may be trend stationary processes or may have unit roots. In either case, an OLS approach ignoring these properties is likely to lead to spurious correlation and incorrect, upwards biased test results.

If variables involved in the estimation have unit roots, it is appropriate to use a cointegration setting as pioneered by Engle and Granger (1987) or Johansen (1988). This approach is applied to investigate the relationship between the Gini coefficient of income and several macroeconomic variables by Parker (1998) for UK data and by Ashworth (1994) and Mocan (1999) using US data, among others. Following the argumentation in Jäntti and Jenkins (2010), there is still one caveat that remains in these approaches. They state that because of the boundedness of inequality measures, such as the Gini coefficient or quantile shares, it may be inappropriate to involve them in a cointegration context. While variables with unit roots exploit an unlimited variation by definition, bounded variables cannot have an unlimited variation and thus cannot possess a unit root, even if standard unit-root tests indicate its existence. Jäntti and Jenkins (2010) sidestep this problem by using a parametric model, namely the Singh-Maddala distribution, to approximate the size distribution of incomes, and apply a cointegration approach to the shape parameters that have an unlimited support.

65 The most recent example is Jäntti and Jenkins (2010).

In general, while time series analysis is a very important tool to show relationships between income inequality and macroeconomic variables for different time frames, these approaches also have limitations by construction, that, depending on the information available, may lead to strong inefficiencies. One source of such an inefficiency, given a sizeable panel data source, is to disregard unit-level information on the size distribution of incomes and to collapse it to point estimates like inequality measures, quantiles or shape parameters. When sufficient micro-data is available, methods from panel data analysis can be used as an alternative. A great advantage of these methods is to allow for different impacts by the observed macroeconomic circumstances on different micro-level units. The idea that different income domains are affected differently by changed macroeconomic environments was already captured to some extent by the above models, e.g. by looking at potentially different impacts on different quantiles or allowing for different impacts on different shape parameters that depend on some distribution moments of a higher order. But those models are not able to incorporate selective effects on micro-units depending on vectors of micro-unit characteristics. In a panel context with dominating cross sectional information though, it is possible to incorporate such dependencies and to identify more detailed (short-term) relationships. Such a data setting generally circumvents the problems as mentioned in the settings with dominating time dimensions as time series persistence does not lead to biased estimates with dominating cross sectional dimension.[66] To approach the question from this direction, Farré and Vella (2008) use a semi-parametric double-index model of Ichimura and Lee (1991) to simultaneously model the relationship between household income, individual characteristics and the business cycle for Spanish data. As the differential effect of a changed macroeconomic environment on different households is a-priori unclear, the model they use allows for an arbitrary relationship between micro- and macroeconomic indices. This chapter follows their strategy and applies the double-index model to micro-level data from the German Socio-Economic Panel (SOEP).

66 See Wooldridge (2002), p.175.

Following Farré and Vella (2008), we use a special case of semi-parametric multiple-index models by Ichimura and Lee (1991) that allows for unrestricted relationships between two indices representing micro and macroeconomic variables. Let

$$y_{it} = g(I_{1it}, I_{2c(i,t)}) + u_{it} \tag{3.1}$$

be the relationship between the household income $y_{it}$ of individual $i$ at time $t$, belonging to some cluster $c(i,t)$ (in our application: month of the questionnaire $\times$ federal state $\times$ year), and let $g$ be an unknown function, satisfying some regularity conditions as given in Ichimura and Lee (1991). Let further be

$$I_{1it} = x'_{it}\beta \tag{3.2}$$

a linear combination of micro-level (i.e. household-level) variables $x_{it} = (x_{1it}, x_{2it}, \dots, x_{mit})'$ and an unknown column parameter vector $\beta \in \mathbf{R}^m$, and

$$I_{2c(i,t)} = z'_{c(i,t)}\gamma \tag{3.3}$$

a linear combination of macro-level variables $z_{c(i,t)} = (z_{1c(i,t)}, z_{2c(i,t)}, \dots, z_{nc(i,t)})'$ and an unknown column parameter vector $\gamma \in \mathbf{R}^n$ and let $u_{it}$ be an additive, zero-mean, idiosyncratic error term, uncorrelated with $I_{1it}$ and $I_{2c(i,t)}$ for $i \in \{1, \dots, N\}$ and $t \in \{1, \dots, T\}$.

*Model Estimation*

To estimate the parameter vectors $\gamma$ and $\beta$ we minimize the objective function

$$Q(\beta,\gamma) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{n=1}^{N} w_{it}\left\{y_{it} - E\left[y_{it}|x'_{it}\beta, z'_t\gamma\right]\right\}^2 \tag{3.4}$$

with observation weight $w_{it}$ and conditional expectation $\mathrm{E}\left[y_{it}|x_{it}'\beta, z_t'\gamma\right]$.[67] As we do not make any assumptions on the particular functional form of the conditional expectation, we estimate it with a two dimensional Nadaraya-Watson (Nadaraya (1964), Watson (1964)) style estimator

$$\hat{\mathrm{E}}\left[y_{it}|x_{it}'\beta, z_t'\gamma\right] = \frac{\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl} K_{2,H}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right) y_{jl}}{\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl} K_{2,H}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right)}. \tag{3.5}$$

Here, in particular, we choose the kernel function to be given by

$$K_{2,H}(\zeta_1, \zeta_2) = \frac{1}{2\pi}|H|^{-\frac{1}{2}} e^{-\frac{1}{2}(\zeta_1,\zeta_2)H^{-1}(\zeta_1,\zeta_2)'} \tag{3.6}$$

for a bandwidth matrix $H$.

*Fixed Bandwidth Estimation*

As suggested e.g. in Härdle/Müller (1997), we choose an initial

$$H_1 = \Sigma^{-\frac{1}{2}} \cdot \tilde{h}_j = \Sigma^{-\frac{1}{2}} \cdot \left(\frac{4}{q+2}\right)^{1/(q+4)} n^{-1/(q+4)}\sigma_j. \tag{3.7}$$

with $\Sigma^{-\frac{1}{2}}$ as the variance-covariance matrix of $(\zeta_1, \zeta_2)$ and with $\tilde{h}_j$ derived from the generalized rule of thumb plug-in bandwidth of Scott (1992) with $\sigma_j$ as j-th diagonal element of $\Sigma$ and $n$ as the sample size. A standard estimation of the variance-covariance matrix delivers

$$\hat{H}_1 = \hat{\Sigma}^{-\frac{1}{2}} \cdot n^{-1/6}\hat{\sigma}_j, \tag{3.8}$$

67 From now on, we will refer to the cluster index $c(i,t)$ simply by $t$ for notational convenience.

where we used the fact that the dimension is $q = 2$.[68]

*Local Adaptive Bandwidth Estimation*

While this bandwidth is a good choice asymptotically, we use an additional multi-step procedure to estimate locally adapted bandwidths to reduce the finite sample bias of the Nadaraya-Watson estimator, as suggested by Klein/Vella (2006).

Abramson (1982) suggests the local bandwidth adaptation

$$\hat{\lambda}_{it}^{1} = \sqrt{\frac{G\left(\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl} K_{2,\hat{H}_1}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right)\right)}{\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl} K_{2,\hat{H}_1}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right)}} \tag{3.9}$$

with $G()$ as the geometric mean over the estimated densities based on the bandwidth matrix $\hat{H}_1$. Defining now

$$\hat{H}_2(i,t) = \hat{H}_1 \cdot \hat{\lambda}_{it}^{1}, \tag{3.10}$$

we receive a locally adapted bandwidth matrix. In a second and final step, we estimate

$$\hat{\lambda}_{it}^{2} = \sqrt{\frac{G\left(\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl} K_{2,\hat{H}_2(j,l)}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right)\right)}{\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl} K_{2,\hat{H}_2(j,l)}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right)}} \tag{3.11}$$

based on $\hat{H}_2(i,t)$ and receive

$$\hat{H} := \hat{H}(i,t) = \hat{H}_1(i,t) \cdot \hat{\lambda}_{it}^{2} \tag{3.12}$$

68 This strategy is equivalent to a Mahalanobis transformation of the data and using a diagonal bandwidth matrix.

which is used in the final estimator in (3.5).After estimating the link function $\hat{g}$ and the coefficients $(\hat{\beta}, \hat{\gamma})$ we receive

$$\hat{y}_{it} = \hat{g}(\hat{I}_{1it}, \hat{I}_{2t}) = \hat{g}(x_{it}'\hat{\beta}, z_t'\hat{\gamma}). \tag{3.13}$$

*Estimated Conditional Heteroskdasticity*

As regards the disturbance term $u_{it}$, we can either estimate

$$\hat{u}_{it} = u_{it}, \tag{3.14}$$

or estimate $u_{it}$ by the Nadaraya-Watson estimator applied to the second central moment as

$$\begin{aligned} \hat{\sigma}^2\left(y_{it}|I_{1it}, I_{2t}\right) :&= \widehat{Var}\left[y_{it}|x_{it}'\beta, z_t'\gamma\right] = \hat{E}\left[y_{it}^2|x_{it}'\beta, z_t'\gamma\right] - \hat{y}_{it}^2 \\ &= \frac{\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl}K_{2,H}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right)\cdot y_{jl}^2}{\sum_{j=1}^{T}\sum_{l=1}^{N} w_{jl}K_{2,H}\left(x_{it}'\beta - x_{jl}'\beta, z_t'\gamma - z_l'\gamma\right)} - \hat{y}_{it}^2\,. \end{aligned} \tag{3.15}$$

Once the conditional standard deviations are know, we can simulate $\tilde{u}_{it}$ as conditional (and even as counterfactual) heteroskedasticity under normality assumption and receive

$$\tilde{y}_{it} = \hat{y}_{it} + \tilde{u}_{it},$$

our final model fit.

*Counterfactual Analysis*

Once the model fits are estimated, the resulting distribution can be derived, e.g., with standard weighted kernel density estimation as

$$\hat{f}_t(\tilde{y}_{it}) = \frac{1}{N}\sum_{j=1}^{N}\frac{w_{jt}K_h(\tilde{y}_{it} - \tilde{y}_{jt})}{\sum_{l=1}^{N} w_{lt}} \tag{3.16}$$

with a Kernel function $K$ and a bandwidth $h$ for any point of time $t \in \{1, \ldots, T\}$. If we are interested in a counterfactual macroeconomic condition, we may simply change the particular value(s) and calculate the model fit

$$\hat{y}_{it}^{c} := \hat{E}^{c}\left[y_{it} | I_{1it}, I_{2t}^{c}\right] = \frac{\sum_{j=1}^{T} \sum_{l=1}^{N} w_{jl} K_{2,H}\left(\hat{I}_{1it} - \hat{I}_{1jl}, \hat{I}_{2t}^{c} - \hat{I}_{2l}\right) y_{jl}}{\sum_{j=1}^{T} \sum_{l=1}^{N} w_{jl} K_{2,H}\left(\hat{I}_{1it} - \hat{I}_{1jl}, \hat{I}_{2t}^{c} - \hat{I}_{2l}\right)} \tag{3.17}$$

and analogously the counterfactual error term as shown in (3.15). Subsequently, based on $\tilde{y}_{it}^{c}$, we can estimate the counterfactual distribution and derive arbitrary measures of counterfactual location or inequality.

## 3.4 DATA AND DESCRIPTIVE ANALYSIS

### *Household Level Variables*

In order to obtain microeconomic variables, we use the German Socio-Economic Panel (SOEP) provided by the German Institute for Economic Research (DIW) in Berlin.[69] The SOEP is a representative yearly panel study of private households in Germany that has been running since 1984. We focus our analysis on data between 1996 and 2010. The dependent variable of our model is the equivalized monthly household income at survey time in prices of 2005. As equivalence scale we use the so called 'new OECD scale' to account for household economies of scale. This scale assigns the weight 1 to the first household member, 0.5 to any further member above 14 years of age and 0.3 to children under 14. For our estimation strategy, we assume that there may be structural differences between different household types regarding the effects to be estimated. For this reason we define six different household types. The first household type is the one of single elderly, the second one of multiple elderly, allowing for one spouse to be 55 or older. The third household type is singles without children, where children are family members with at most 17 years of age, and the fourth type is multiple adults without children. The fifth and the sixth group are single and non-single households with children. In order to obtain sufficiently large numbers of ob-

69 For details see Haisken-DeNew and Frick (2005) or Wagner et al. (2007).

servations for our nonparametric estimation, we merge the above six types into three, labeling them as 12, 34 and 56. We chose this notation to clarify that we will still differentiate among the merged types by including a dummy variable in our estimations.

Following the above structure, we define slightly different sets of micro-level explanatory variables for each household type. We use household level variables because the income variable is derived from the household level. In all household types we control for the share of adult household members with labor market participation, the share of members with high school degree, the share of members with university degree, and the share of females. Additionally, we use a dummy variable indicating the membership to the first of the two merged household types. Finally, we include a dummy for East Germany. For household type 12, we use in addition the share of family members in the age group of below 65 years of age and the share of persons with disability. For household types 34 and 56, we use the share of adults in the age category 31 to 50 and the age category 51 to 64 and take 18 to 30 years of age as reference category.[70]

*Sample Selection*

For reasons of numerical stability, we do not make use of 2.5% of the lowest and highest incomes. As incomes of self-employed household members are often misreported, we do not take into account households with self-employment either. The final full sample consists of 282,018 observations with 41,286 different individuals in 17,860 households. For reasons of computational feasibility, we can only use an 80% random subsample of the data.

*Macroeconomic Variables*

As macroeconomic variables we use the logarithmic, lagged, real GDP on the level of the federal state (Bundesland), the lagged inflation rate, the government final consumption expenditure and the month-specific male ILO unemployment

70 It is our understanding that the choice of control variables of relevance in this setting is not exhaustive. The limit of using 8 micro-level and 4 macro-level control variables is, as described below, a computational one.

rate for the years 1996 to 2010. These variables were taken from the Genesis Data Base of the German Federal Statistical Office and the World Bank Database, respectively. The choice of these variables is quite similar to the one suggested by Farré and Vella (2008), but there are some differences. While we are aware of the fact that this choice is not exhaustive and that other choices of time and functional transformations may also make sense, we limited the number of macro-economic variables to four to keep numerical calculations in reasonable time limits. Also, we based the specific time and functional pre-transformations of the variables on their goodness-of-fit in linear pre-estimations. In the case of unemployment, we particularly concentrate on male unemployment as it usually depicts the labor market situation better than general unemployment.

Figure 17 below shows the time series of macro-economic variables between 1991 and 2010. After German reunification in 1991, the time series of inflation shows a particularly strong fall until 1996. The time period before 1996 is also characterized by a rise in unemployment and government expenditure. While we generally welcome variation in the macro-economic data, we must assume that these time series were related to a unique structural change following the German Reunification so that we do not include the years 1991 to 1995 into our estimations.


Figure 17: Time series: **(left)** Male ILO unemployment and inflation rate. **(right)** GDP per capita and government expenditure per capita without pensions. Source: Genesis data base, Federal Statistical Office.

The period of 1996 to 2010 was a period with little change in inflation rates. They were stable at around 2% and somewhat lower after the 2008 crisis. The unemployment rate was around 8% with the exception of the years 2003 to 2007 where they peaked at somewhat below 12% describing a total variational range of about 4%. After the crisis, the unemployment level in Germany stayed exceptionally stable. While the GDP was steadily rising until 2008, it fell strongly by about 5% during the crisis. The total range of the time series was around 10% of its starting value. The level of government expenditure on the other hand was relatively stable until 2005 and fell slowly before the crisis. Afterwards it rose by around 6% onto the level before 2005.

Turning now to the distribution of incomes in the same time period, figure 18 shows relevant summary measures to analyze the underlying dynamics. The time series can be divided into four sub-periods. In the first one, between 1991 and 1996, inequality fell substantially as mean income in East Germany converged to that in West Germany. The years 1996 to 2000 were mainly characterized by a stagnation in income inequality, while the period from 2000 to 2006 was characterized by a steep rise. Around the crisis year 2008 inequality was about constant.[71]

While there are some parallels between the macro-economic time series and the ones describing income inequalities, there are at least as many differences, meaning that statistical relationships, if they exist, are not straight forward enough to be seen on a descriptive level and thus require a more rigorous statistical approach.

71 Here we refer to the time series based on our sample selection. As we dropped the highest 2.5% of observations, our picture slightly differs from the one including it. Including the highest 2.5% of incomes shows a rather falling inequality after the 2008 crisis that strongly affected the right end of the income distribution.

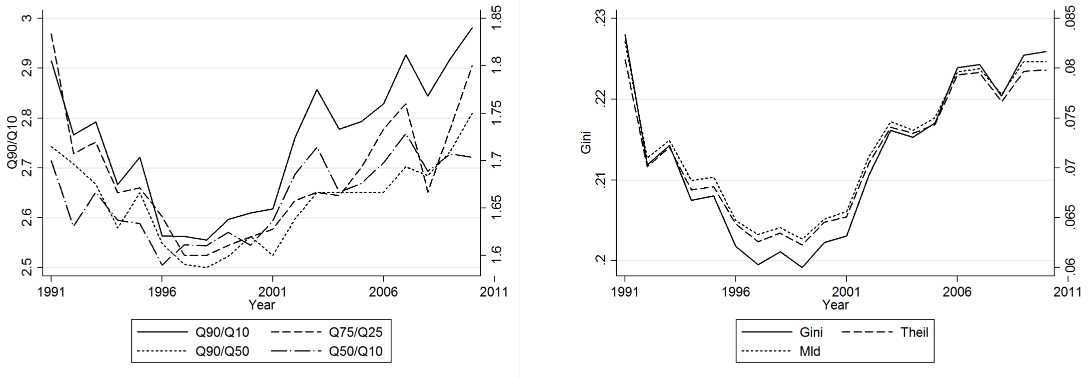



Figure 18: Time series: **(left)** Quantile ratios of the income distribution. **(right)** Gini, Theil measure and the mean log deviation. Own calculations based on data after sample selection.

## 3.5 RESULTS

### 3.5.1 *The Income Distribution and Macro-economic states*

Before we turn to the interpretation of the outcomes in detail, we begin with some general remarks on the interpretation of the results, which is more difficult than in linear models. For the following reasons, the interpretation has to be more cautious. Firstly, all variables are standardized,[72] so the interpretation refers to standard deviations. Furthermore, all microeconomic coefficients are to be understood relative to the size of the coefficient of the (standard deviation of the) share of adult household members in employment. All macro-economic coefficients are measured relative to the coefficient of GDP, because, in each index, one coefficient has to be normalized to one for identification. In addition to this, there are further restrictions. For example, if the coefficient of high school degree is, e.g., 1.5, we cannot interpret its magnitude as 1.5 times higher than the one of the constrained coefficient, because the marginal effect still depends of the shape of the function g. While this is less important for the microeconomic variables showing a monotonic behavior for the overall income level as depicted in figure 19, the behavior in macro-economic direction is not always monotonic. Thus in some cases, a seemingly higher coefficient may also come along with a

72 See identification restrictions of the Double Index Model, Ichimura and Lee (1991).

| | HH-Type 12 | | HH-Type 34 | | HH-Type 56 | |
|---|---|---|---|---|---|---|
| Dependent Variable: Income | coeff. | P-value | coeff. | P-value | coeff. | P-value |
| **Micro-Level Variables** | | | | | | |
| Employed$^{(a)}$ | 1 | | 1 | | 1 | |
| High School$^{(a)}$ | 4.491*** | 0.005 | 0.428*** | 0.000 | 0.5200*** | 0.000 |
| University$^{(a)}$ | 8.413*** | 0.003 | 0.851*** | 0.000 | 1.3867*** | 0.000 |
| East Germany | -4.176*** | 0.003 | -0.344*** | 0.000 | -0.1465** | 0.035 |
| Age 31-50$^{(a)}$ | | | 0.343*** | 0.000 | 0.2831*** | 0.000 |
| Age 51-64$^{(a)}$ | 0.185 | 0.499 | 0.431*** | 0.000 | 0.1696*** | 0.000 |
| Disabled$^{(a)}$ | 0.466 | 0.296 | | | | |
| HH-Type 1/3/5$^{(b)}$ | -0.510 | 0.438 | -0.578*** | 0.000 | -0.5965*** | 0.000 |
| Female$^{(a)}$ | -1.136* | 0.057 | -0.115*** | 0.000 | -0.1424*** | 0.004 |
| **Macro-Level Variables** | | | | | | |
| ln(GDP)$^{(c)}$ | 1 | | 1 | | 1 | |
| Inflation$^{(c)}$ | -0.018 | 0.654 | -0.045 | 0.267 | -0.1952*** | 0.000 |
| Government Exp.$^{(d)}$ | -0.063* | 0.094 | 0.080 | 0.131 | 0.1535*** | 0.002 |
| Unemployment$^{(e)}$ | 0.012 | 0.763 | 0.009 | 0.817 | -0.0228 | 0.621 |
| $R^2$ | 0.234 | | 0.382 | | 0.377 | |

* p<0.1 ** p<0.05 *** p<0.01

$^{(a)}$ share of adults in the HH, $^{(b)}$ according to HH pool, $^{(c)}$ lagged

$^{(d)}$ share of GDP, $^{(e)}$ male ILO

All variables are standardized.

Table 14: Double Index Model estimation results: Micro- and Macro-Index components. P-values are based on clustered subsampling with sample size correction and estimated rate of convergence. HH-Type 12: Elderly, HH-Type 34: non-elderly w/o children, HH-Type 56: non-elderly with children. Own calculations.

changed sign in the overall partial effect. So, what we may say generally in the above example is that, ceteris paribus, one standard deviation of the fraction of adult high school graduates in the family has the same effect as 1.5 standard deviations change in the fraction of employed adults in the family.

*Index Coefficients*

Table 14 shows the estimated results based on the double-index model. To analyze the results of microeconomic variables, we take an exemplary look at the results for elderly households. We see that this household type shows positive and highly significant[73] coefficients of schooling. The fact that the coefficients are so high, relative to the coefficient of the share of employed household members, shows that employment income plays a subordinated role in elderly households. The effect of living in the eastern part of Germany comes with a coefficient that has the size of high school degree with a negative sign and is also highly significant, in line with the fact that pensioners are financially less well off in the eastern part of Germany. While the share of females has a significantly negative coefficient, the other variables are insignificant. Households without children and households with children show a qualitatively similar picture, but there, all coefficients are highly significant. Qualitatively, signs are in line with a-priori expectations.

As regards macro-economic effects on the income distribution, the coefficient of inflation is highly insignificant for elderly households. This can be explained by the fact that the dominating source of income is pension payment, which is inflation adjusted. Following the same argument, incomes are also unaffected by changes in the unemployment rate. If we take a look at the surface plots in figure 19, we see that GDP may have an implicitly negative effect as constrained variable. While we cannot be sure about the significance of GDP in the model, it is likely that the influence is not significant and cannot be distinguished from small positive values. This point is important however for the interpretation of the effect of government expenditure. At first glance, we see a surprising negative sign, but adding this to the fact that the surface mainly falls with a higher macro-economic index, the resulting effect of government expenditure is positive and significant at the 10% level.

As regards households without children, the signs of the coefficients of inflation and government expenditure are in line with theoretical expectations, even if the

73 For more details on the calculation of the confidence intervals see 3.7.1.

results turn out to be statistically insignificant. The sign of the former result is negative, and we may assume that higher inflation is of disadvantage to an important part of the population, e.g. due to more expensive access to the capital market or smaller real asset returns. On the other hand, households benefit from higher government expenditure, which is in line with the positive coefficient. As regards the unemployment rate, the coefficient is almost zero and statistically not different from zero.

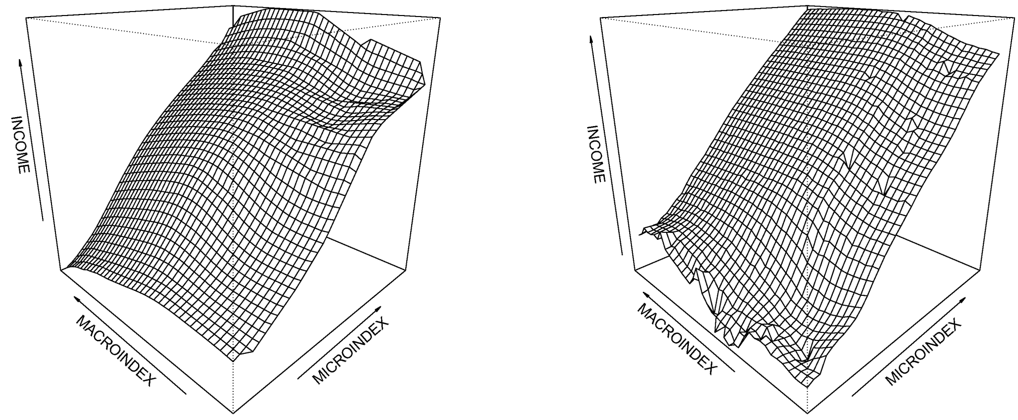
Turning to households with children we see that both the coefficient of inflation and the coefficient of government expenditure are significant at the 1% level. While the coefficient of inflation is negative, the one of government expenditure is positive. This is in line with our former results and with economic theory. The fact that we can show a significant influence here may be explained by the higher dependency of this group on fixed employment contracts as their degree of freedom to renegotiate might be limited due to the fact that they have dependent children. Also, a change in government expenditure may affect households with children more, as they are more frequently recipients of social transfers. Finally, the effect of employment is small and insignificant like in the other two cases.

As regards the general insignificance of unemployment in our model, we give three possible explanations. The first one is that a great part of the population is really unaffected by unemployment, never entering the state of unemployment, while the coefficient is fitted globally, making a lack of effect an understandable compromise. The second one is that Germany has a strong social net especially for those with lower income, who are at risk of becoming unemployed. In these cases, and especially in cases of shorter unemployment, unemployment payments compensate the loss of income to a large extent, while our dependent variable does not differentiate between unemployment payments and employment income as income sources. And the third explanation lies in the developments on the German labor market, where even despite the financial crises, the unemployment as measured with the ILO definition stayed surprisingly stable. Still, there was also generally a strong rise in part-time employment over the whole estimation period, and especially after the 2008 crisis, also a rise in temporary employment and labor leasing (Bundesagentur für Arbeit (2012)) with lower hourly payment. In numbers, the part-time share in employment rose from the level of 12.9% in 1996 to 19.4% in 2010 (Bundesagentur für Arbeit (2011)). An important part of this development was a substitute for regular full time employment. While these effects probably strongly influenced the distribution of household incomes, they were not captured in the unemployment definition. So mechanically, the two variables are to a large extent unrelated.

Furthermore, our first point above also discloses an important limitation of the use of aggregate macro-economic variables with the double-index model, that is more a model problem than a data-driven one. Although the double-index model is quite flexible in estimating household-specific impacts of macro-variables on the household level due to its nonparametric link function, it does not distinguish between households given the same micro-index-level. If the index is not capable of dividing the population into subgroups that are e.g. influenced by unemployment and in those who are generally not, estimation will rather result in a trade-off. If the size of the subgroup of households, who are not affected by the unemployment dominates, this trade-off will indicate a non-response to unemployment at the population level.

*The Link Function*

After analyzing the estimated coefficients, we take a closer look at the link function that was estimated nonparametrically along with the coefficients. Figure 19 shows regression surfaces that depict the overall shape of the function $g$ for the three different household types. The upper right and the lower left surfaces of the household types 34 and 56, respectively, increase in the direction of the micro-index with minimal local exceptions. For example, a household with higher characteristic values of variables with positive coefficients, let's say a higher share of university degree holders in the household, also has a higher mean income as the overall micro-index value increases. The same is true for the macro-index direction. For example, if social expenditure increases ceteris paribus, the surface shows mostly a higher income level, although there are exceptions. In general though, the macro-direction has a much weaker increase and a much weaker role for household income than microeconomic characteristics. It is interesting to see



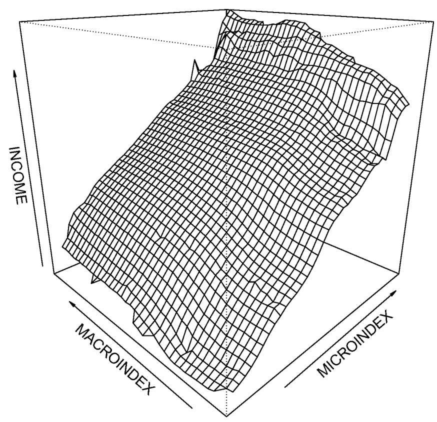
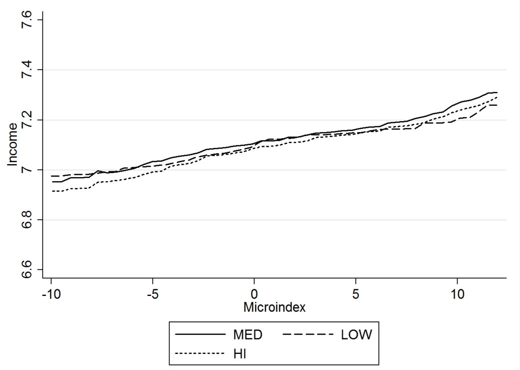
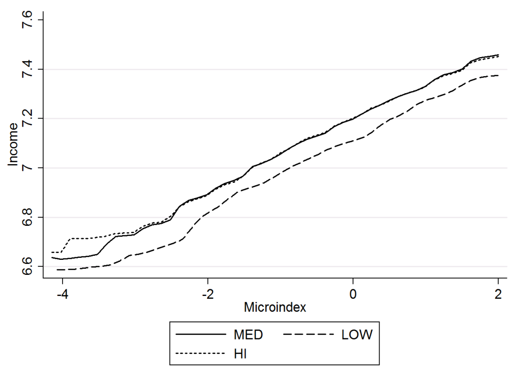




**Figure 19** - Regression Surfaces based on Double Index Model estimation for household types 12 of the elderly **(upper left)**, 34 **(upper right)** and 56 **(lower left)** of the non-elderly without and with children, respectively. Own calculations.

that the coefficients in table 14 do not provide this information, as this behavior is a sole property of the function g. Turning finally to the household type 12 in the upper left, we see that while the surface is leveling up in the micro-index dimension, it is rather leveling down with higher macro-index. When looking at the signs of the estimated macro-level coefficients, the sign of the significant coefficient of government expenditure is not as expected, either. This may be due to the fact that GDP itself, the constrained variable, has a statistically insignificant negative effect (and a true positive one), changing all other signs as they are measured in a relative manner.

To analyze interesting properties of $g$ further, we take a look at different cross-sections of the surface. Figure 20(a)-(c) shows the cross sections derived from figure 19(a)-(c) at different levels of the macro-economic index. While the dashed lines show the behavior of the function below the 25th centile of the macro-index for each of the household types, the solid line shows the behavior around the median and the dotted line the one above the 75th centile. As the lines for each of the household types are different across centiles, we may conclude that the macro-economic state plays a non-trivial role for the conditional distribution of incomes, so that $g(x, y) \not\equiv g(x)$ for the link function $g$.



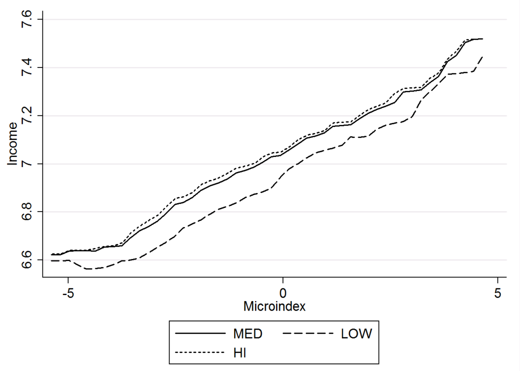




**Figure 20** - Average cross sections of the double-index regression fits below the 25th (LOW), between the 25th and 75th (MED) and above the 75th (HI) centile of the macro-indices. For household type 12 **(upper left)**, household type 34 **(upper right)**, household type 56 **(lower left)**, own calculations.

An additional conclusion that we may derive is that the lines of different levels are non-parallel. If link function was of a some simpler form $g(x, y) = g_1(x) + g_2(y)$, then the lines were parallel, meaning that the micro- and macroeconomic arguments of the link function were additively separable. But as the lines show a non-parallel behavior, we expect that the two indices truly interact with each other. In a practical sense, this means that the magnitude of the macro-level impact on households varies with household characteristics. Additionally, we observe that the shapes of the cross-sections differ throughout household types, which gives empirical support to our differentiated approach.

As regards elderly households, we see in figure 20 that both the micro-index and the macro-index have a limited effect on the expected income. Additionally, we see that the lines resulting from the levels of the macro-index cross each other, suggesting a strongly non-linear behavior or a generally rather insignificant one (despite the significance of the sole component government expenditure). The other two household types show a clearer pattern. As regards the levels of income resulting from the levels of the macro-index, we observe that while above-average macro states have not much additional effect, negative shocks are associated with higher deviations from the median level. Furthermore, if we take a look at the micro-indices, we observe that, while lowest micro-indices go along with nearly constant income levels, in its other parts the relationship is not very far from linear.

### 3.5.2 *Counterfactual Analysis*

While interpreting partial effects directly is not straight forward in this context, estimation results may be used to analyze effects on the unconditional distribution level in 2010. The information given on the distribution level allows us to analyze arbitrary measures of location and dispersion. Similar to the results in the last section though, we do not find pronounced changes in the values of these measures under counterfactual circumstances. More interesting is the fact that despite their low magnitudes, the changes we find are mostly in line with a-priori expectations and theory. We approach this point by taking a look at distributional effects when putting counterfactual macro-economic states into the estimated model and analyze the counterfactual fits. To receive clear effects, we use emphasized magnitudes of macro-economic states.

The counterfactual macro-economic settings are as follows. To analyze the behavior of inequality measures for changed GDP states, we adjust the GDP counterfactually by ±20%. In case of inflation we use ±3%, in the one of government expenditure ±20%, and we set unemployment to zero and +10% as compared to the rate in 2010 in absolute terms. We emphasize here that this counterfactual model is only able to capture ceteris paribus effects, giving a mechanical approach to analyze the link between the income distribution and the state of the world described by macro-economic factors. We will not be able to incorporate secondary effects and forces towards economic equilibria. Results should be interpreted in the light of these limitations.

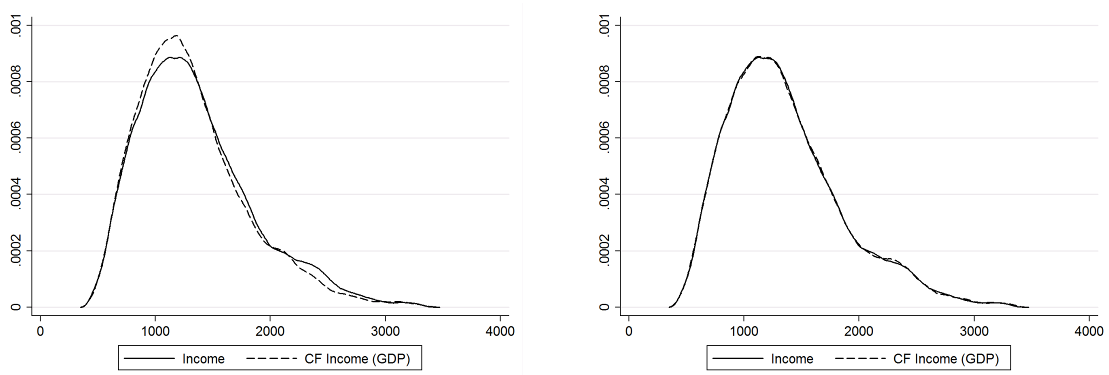



Figure 21: Unconditional income distributions and their counterfactual (cf.) counterparts for lower GDP (-20%) **(left)** and for higher GDP (+20%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations.

Taking a look at figure 21 and at Tables 15 to 18, we see that with a lower GDP the distribution shifts to the left (lower median) and becomes less dispersed (lower quantile ratios and inequality measures), while when we elevate the level of the GDP, there is a slight right slide of the distribution and a slightly higher dispersion, but the effects are much smaller. In more detail, when the GDP is lowered, the very left part of the distribution remains almost unchanged, while the rest of the distribution experiences a stronger effect. A possible explanation is that

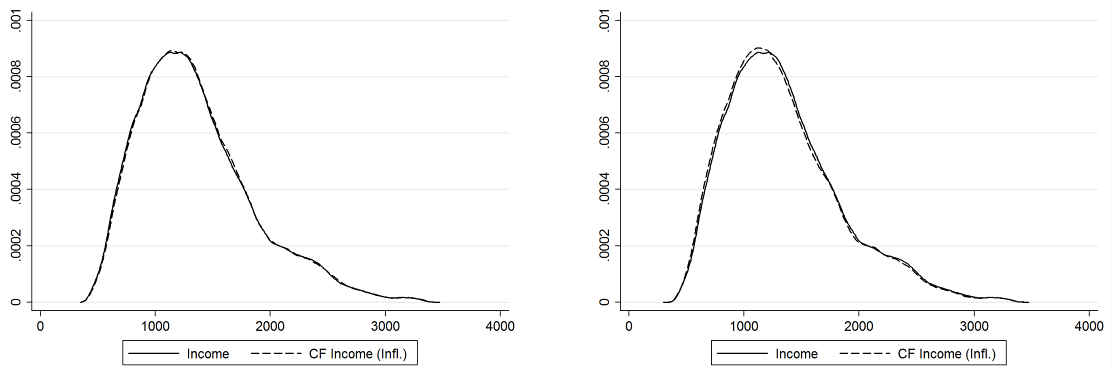



Figure 22: Unconditional income distributions and their counterfactual (cf.) counterparts for lower inflation (-3%) **(left)** and for higher inflation (+3%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations.

the loss of full-time income in the event of an economic downturn (e.g. working time reduction or salary cuts in case of performance-linked payment) is less pronounced for recipients of low incomes in absolute terms, given a strong financial compensation of the social security system in Germany. As we see, households in the very left part of the distribution with probably mostly independent income sources of the labor and capital market (e.g. social benefits) are not influenced by the state of economy, while the influence is especially pronounced in the middle section. As for high income earners, people with higher incomes tend to have a higher degree of freedom to reallocate their income sources according to the actual state of economy. In general, we observe that while short-term negative economic shocks have stronger negative effects on the income distribution, positive shocks are either more neutral or have a more delayed effect, underpinning the non-linearity of macro-economic influence. In addition, the effects over the distribution are strongly heterogenous, meaning that differential effects of changes in macro-economic states on different households exist.

As regards inflation, we see in figure 22 that lower inflation goes along with a slight right shift of the income distribution and as the measures of dispersion in Tables 15 and 17 indicate, inequality becomes slightly lower. We observe the opposite, but more pronounced effect in case of higher inflation, leading to a general left shift of the income distribution and to higher inequality.

The effect of changed government expenditure in figure 23 is in line with general expectations. A lower government expenditure leads to a slight left shift of

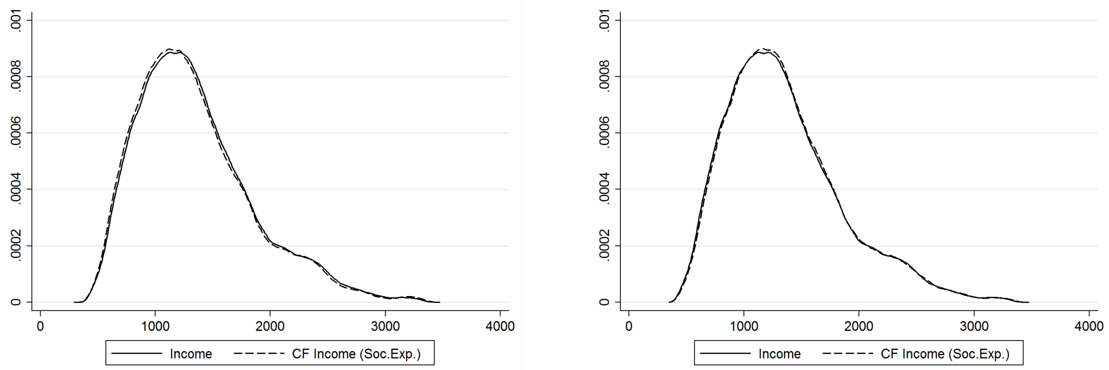



Figure 23: Unconditional income distributions and their counterfactual (cf.) counterparts for lower government expenditure (-20%) **(left)** and for higher government expenditure (+20%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations.

the income distribution, while a higher expenditure leads to a slight right shift. Additionally, as measures of inequality indicate, lower government expenditure is in line with higher inequality and vice versa. It might surprise at first glance that also the right part of the distribution benefits from government expenditure though (see e.g. the shift of the 90th quantile). The reason is that only one half of the government expenditure is allocated to the social security sector. Other components are spendings with more equally distributed beneficiaries like education, free time and sports, health and others as summarized in table 19 in 3.7.2.

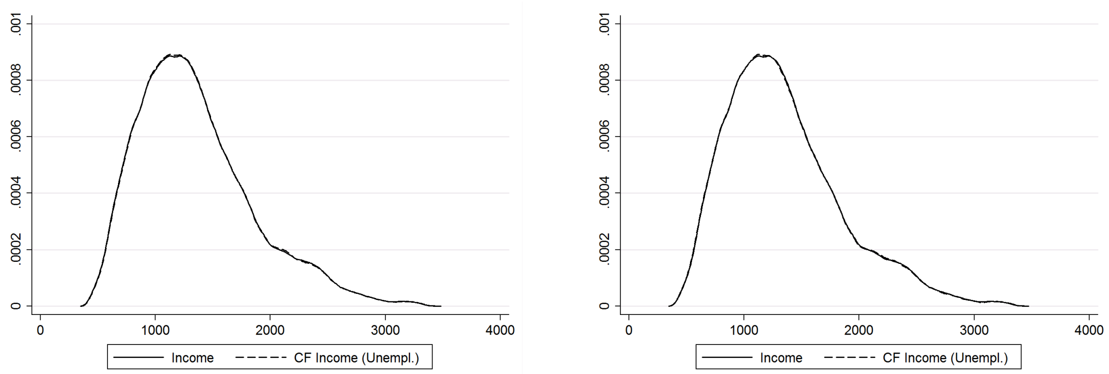



Figure 24: Unconditional income distributions and their counterfactual (cf.) counterparts for no unemployment **(left)** and for higher unemployment (+10%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations.

Finally, figure 24 shows the behavior of the income distribution when changing the unemployment rate to zero and when elevating it by 10%, respectively. We see almost no effect. This observation is in line with the very small coefficients of unemployment.[74]

| Income | 10th Cent. | Median | 90th Cent. | QR 90/10 | QR 50/10 | QR 90/50 |
|---|---|---|---|---|---|---|
| Factual | 764.794 | 1217.005 | 1946.051 | 1181.256 | 452.210 | 729.045 |
| GDP (-20%) | 763.798 | 1215.375 | 1943.486 | 1179.688 | 451.576 | 728.110 |
| Inflation (-3%) | 764.410 | 1218.018 | 1947.935 | 1183.525 | 453.607 | 729.917 |
| Gov. Exp. (-20%) | 763.430 | 1216.288 | 1947.104 | 1183.674 | 452.857 | 730.816 |
| Unemployment (0%) | 764.064 | 1217.400 | 1947.935 | 1183.871 | 453.335 | 730.535 |

Table 15: Quantiles and quantile ratios (QR) of the income distribution for counterfactual, lowered macro variables for the year 2010. Own calculations.

| Income | 10th Cent. | Median | 90th Cent. | QR 90/10 | QR 50/10 | QR 90/50 |
|---|---|---|---|---|---|---|
| Factual | 764.794 | 1217.005 | 1946.051 | 1181.256 | 452.210 | 729.045 |
| GDP (+20%) | 764.062 | 1217.266 | 1948.823 | 1184.761 | 453.204 | 731.556 |
| Inflation (+3%) | 763.430 | 1216.110 | 1947.177 | 1183.747 | 452.679 | 731.066 |
| Gov. Exp. (+20%) | 764.690 | 1218.018 | 1948.823 | 1184.133 | 453.327 | 730.805 |
| Unemployment (+10%) | 763.928 | 1217.265 | 1948.188 | 1184.259 | 453.335 | 730.923 |

Table 16: Quantiles and quantile ratios (QR) of the income distribution for counterfactual, elevated macro variables for the year 2010. Own calculations.

74 We have already given possible explanations for this result on page 117.

| Measure of Inequality | Income Factual | GDP (-20%) | Inflation (-3%) | Gov. Exp. (-20%) | Unemployment (0%) |
|---|---|---|---|---|---|
| Coefficient of variation | 0.3731 | 0.3671 | 0.3703 | 0.3776 | 0.3720 |
| Gini coefficient | 0.2038 | 0.1991 | 0.2021 | 0.2057 | 0.2030 |
| Theil index | 0.0665 | 0.0640 | 0.0655 | 0.0679 | 0.0661 |
| Mean Log Deviation | 0.0676 | 0.0645 | 0.0665 | 0.0688 | 0.0670 |

Table 17: Inequality measures of the income distribution for counterfactual, lowered macro variables for the year 2010. Own calculations.

| Measure of Inequality | Income Factual | GDP (+20%) | Inflation (+3%) | Gov. Exp. (+20%) | Unemployment (+10%) |
|---|---|---|---|---|---|
| Coefficient of variation | 0.3731 | 0.3743 | 0.3762 | 0.3695 | 0.3737 |
| Gini coefficient | 0.2038 | 0.2044 | 0.2053 | 0.2016 | 0.2042 |
| Theil index | 0.0665 | 0.0669 | 0.0675 | 0.0652 | 0.0667 |
| Mean Log Deviation | 0.0676 | 0.0680 | 0.0685 | 0.0662 | 0.0679 |

Table 18: Inequality measures of the income distribution for counterfactual, elevated macro variables for the year 2010. Own calculations.

## 3.6 CONCLUSION

In this chapter a double-index model is employed to analyze the impacts of macro-economic variables on the size distribution of incomes in Germany between 1996 and 2010. Making use of cross-sectional information, the model allows for arbitrary functional relationships between indices of micro- and macro-level variables in order to trace out their joint influence on the income distribution. After estimating the relationship, we employ a counterfactual approach to simulate different macro-economic states and to analyze their impacts on the distribution of incomes.

The results suggest that while the measurable impact of changed macro-economic variables on micro-level incomes is small, it is in parts significant and draws a comprehensive picture of potential links between micro- and macro-variables.

In particular, we find that the level of government expenditure has a significant positive impact on the income levels of elderly households and on those of households with children. The level of inflation is highly significant and negatively associated with the income level of households with children. We also find that qualitatively negative macro-economic shocks are associated with relatively higher effects and that the short-term effects of positive shocks are almost non-measurable. A negative GDP shock and a lower inflation are associated with a lower inequality, and higher government expenditure corresponds to lower inequality. As regards unemployment, we find no evidence for an effect on the general level of incomes or on their inequality. This seems to be driven by the fact that the unemployment rate fails to reflect the diversity of structural changes that took place on the German labor market.

The fact that the impact of macro-economic states was found to be very small is somewhat unsatisfactory, but it has a manifold of possible explanations. Firstly, as regards the model, the interaction of micro- and macro-level variables was only given at the level of indices, leading to a potential mixture of heterogenous effects. This problem is particularly pronounced in the case of unemployment, where only a small part of the population is strongly affected. This situation results in an absence of an overall effect in our estimation results. Another point, and perhaps a cautious implication driven from the results, might be that the income distribution in Germany is in general not as responsive to changed macro-economic conditions as e.g. the one in Spain. To take a single example, while the share of short-term labor contracts in Germany was around 11% in 2004, it was around 30% in Spain,[75] indicating a different organization of the labor market with a possibly stronger response to the business cycle. In general, differences in state legislatures, government systems and market structures are sources of heterogeneity of possible response mechanisms and may lead to very different results across countries.

75 See Zachert (2004).

### 3.7.1 *Subsampling and speed of convergence*

For the calculation of the standard errors we use the method of subsampling. Subsampling is an alternative method to the ordinary boostrap. Its advantage lies in the fact that the sample size used for the re-sampling may be lower than the original sample size. This is particularly important in the given case because non-parametric estimation in greater samples sizes is highly time-consuming and conventional bootstrap calculations would be far beyond reasonable time constraints. For more on subsampling see Politis et al. (1999).

For our model and data we used a subsample size of 12,000 observations for each household type and repeated the draw 200 times for each household type. As a direct estimation of the confidence intervals based on the observed standard deviation under normality assumption would strongly overestimate the size of the confidence intervals, we additionally corrected the outcome for the sample size that we actually used for our full estimations. For this purpose, we estimated the overall convergence speed of our model as follows. For further details, see Geyer (2006).

Assuming that the speed of convergence follows some rate $\tau_n = n^\rho$ with some fixed $\rho$, Politis et al. (1999), Chapter 8 shows that

$$G_b^{*-1}(t) \approx b^{-\rho} F^{-1}(t)$$

with

$$\tau_n(\theta_b^* - \hat{\theta}_n) \sim G_b^*(t), \ \tau_n(\hat{\theta}_n - \theta) \sim F(t),$$

where $\theta_b^*$ is the subsampling estimator for some sample size $b < n$, $\hat{\theta}_n$ the estimator based on the full sample of size $n$ and $\theta$ the true parameter. Furthermore, $G_b^*(t)$ is the distribution of the subsampling estimator and $F(t)$ the (unknown) limiting distribution function of the parameter estimate $\hat{\theta}_n$.

Now assuming that G and F are smooth and applying this approximation to two different points t and s with $t > s$ plus taking the natural logarithm,[76] we receive

$$\log\left[G_b^{*-1}(t) - G_b^{*-1}(s)\right] \approx -\rho \log(b) + \log\left[F_b^{*-1}(t) - F_b^{*-1}(s)\right].$$

For several sample sizes $b_j$ and pairs of quantiles $(s_i, t_i)$, we can estimate $\rho$ with simple OLS after taking the average over the pairs of quantiles for each sample size. The expression then simplifies to

$$\bar{y}_j \approx -\rho \log(b_j) + c(F)$$

with a constant depending on F. Independent of F, $\rho$ can be estimated as

$$\hat{\rho} = -\frac{\operatorname{cov}\{\bar{y}, \log(b)\}}{\operatorname{var}\{\log(b)\}},$$

which is the final estimator proposed.

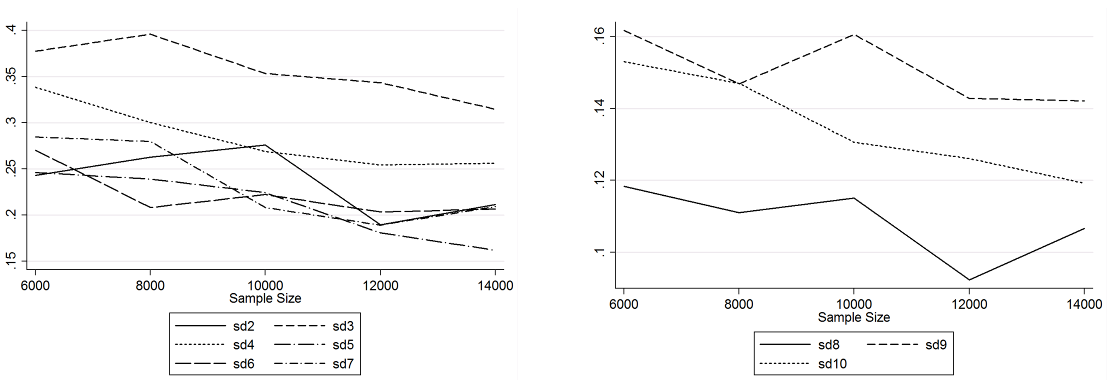



Figure 25: Micro-Index **(left)**, Macro-Index **(right)** coefficient estimators convergence for different sample sizes measured by the 90/10 quantile ratio for the 7 micro- and 3 macro-coefficients respectively. Own calculations.

To receive more stable estimates, we modified this approach slightly by using an additional mild assumption. We assume that the rate of convergence is the same for all micro- and for all macro-coefficient estimators respectively

76 We have to take the difference first, so that the logarithm is always defined.

and pool the estimated quantile differences together over all entries of the coefficients $\beta$ and $\gamma$ as defined in Section 2. As demonstrated by the empirical convergence estimators in Figures 25(a) and 25(b), our assumption is in line with empirical findings.[77] Using 120 observations for each of the sample sizes $6000, 8000, 10000, 12000$ and $14000$ and calculating the quantile differences of $q75 - q25, q70 - q30, q65 - q35, q60 - q40, q55 - q45$, we estimated a convergence rate of $\hat{\rho}_{\text{micro}} = 0.306$ and $\hat{\rho}_{\text{macro}} = 0.399$ with clustered confidence interval $[0.278; 0.333]$ and $[0.376; 0.422]$ respectively over the coefficient pool. The difference comes from the different nature of the two variable sets as macro-variables have much less volatile outcomes. The rates are significantly lower than the theoretical rate $\tilde{\rho} = 0.5$ with an optimal weighting matrix based on a known conditional heterogeneity structure. For more details on the efficient estimation of index models, see Härdle et al. (1993), Newey and Stoker (1993) or Horowitz (2009).

77 The dispersions were measured by the 90-10 quantile ratio to minimize the effect of very few extreme values that were probably caused by local optima of the iterations.

### 3.7.2 *Complementary Information*

| Government Expenditure | Billion Euro | % of Total |
|---|---|---|
| General Public Administration | 155.20 | 13.03% |
| Defense | 26.23 | 2.20% |
| Public Security and Order | 40.20 | 3.38% |
| Economic Matters | 118.90 | 9.98% |
| Environment Protection | 17.04 | 1.43% |
| Housing and Municipal Infrastructure | 15.72 | 1.32% |
| Health Care | 178.48 | 14.99% |
| Free time, Sport, Culture and Religion | 20.39 | 1.71% |
| Education | 107.00 | 8.98% |
| Social Security | 511.81 | 42.97% |
| Total | 1190.97 | 100.00% |

Table 19: The components of German final government expenditure in 2010. German Federal Statistical Office, Wiesbaden 2013, Version: 22. March 2013.

# Part V

# DISSERTATION SUMMARY AND CONCLUSION

# Dissertation Summary and Conclusions

This thesis presented three studies of related but separate aspects of the distribution of income and subjective welfare in Germany.

In its first part, a novel, data-driven poverty line definition, called the Satisfaction-Driven Poverty Line, was introduced. It was shown that the SDPL is time-stable and that it coincides with the OECD's poverty line definition for Germany. It was also shown that there is indeed a discontinuous phenomenon such as a poverty line rather than a smooth transition area which separates the poor from the non-poor. Furthermore, the calculation of country-specific poverty lines revealed that poverty line structures are heterogeneous across European countries. Nevertheless, differences were shown to have a systematic character, depending on the countries' overall wealth and the inequality aversion of the population.

In the second part, the analysis of rising income inequality in Germany identified several factors behind the change in the income distribution around the turn of the millennium. Sequential and ceteris paribus analyses of the potential sources showed that while the changes in the household structure, changes in the household characteristics and the changes in the transfer system only explained a small part of the rise in income inequality, the changes in employment outcomes, changes in labor market returns and changes in the tax system explained around 80% (e.g. as measured by the Gini coefficient). About one half of this share was contributed by increasing inequality in labor income, and the other half was equally shared by employment outcome changes and the tax system. Among others, the research demonstrated that changes in the labor market structure were not the only driving forces behind the change in inequality. Furthermore, detailed results, that were also displayed graphically on the distribution level, provided a rich analysis of the development of the German income distribution in general. In this context, the question of what factor led to changes in what area of the income distribution could also be answered.

The third part of the thesis analyzed possible connections between the shape of the German income distribution and macroeconomic circumstances in a flexible, semi-parametric model. The analysis identified a link function that is non-trivial and non-additive, meaning that the influence of macroeconomic conditions (GDP, inflation, government expenditure or unemployment) is likely to vary with the individual characteristics of the income recipients. While the results were in parts statistically significant and almost always qualitatively plausible, the model found only a weak direct influence of the macroeconomic variables on the income distribution, suggesting that microeconomic forces are more important than macroeconomic developments in shaping the income distribution.

# BIBLIOGRAPHY

Abrams, M. (1973). Subjective social indicators. *Social Trends*, 4:35–50.

Antonczyk, D., DeLeire, T., and Fitzenberger, B. (2010). Polarization and rising wage inequality: comparing the US and Germany. IZA Discussion Paper No. 4842, Bonn.

Arntz, M., Clauss, M., Kraus, M., Schnabel, R., Spermann, A., and Wiemers, J. (2007). Arbeitsangebotseffekte und Verteilungswirkungen der Hartz-IV-Reform. IAB-forschungsbericht, Nuremberg.

Ashworth, J. (1994). An empirical examination of macroeconomic activity and income distribution in the united states. Economics working paper 130, University of Durham, Durham.

Atkinson, A. B. and Bourguignon, F. (2001). Poverty and inclusion from a world perspective. In L, P. P. and Bruni, L., editors, *Economics and Happiness*, page 65. Oxford University Press.

Bargain, O. (2012a). Decomposition analysis of distributive policies using behavioural simulations. *International Tax and Public Finance*, 19(5):708–731.

Bargain, O. (2012b). The distributional effects of tax-benefit policies under new labour: A decomposition approach. *Oxford Bulletin of Economics and Statistics*, 74(6):856–874.

Bargain, O. and Callan, T. (2010). Analysing the effects of tax-benefit reforms on income distribution: a decomposition approach. *The Journal of Economic Inequality*, 8(1):1–21.

Becker, I. and Hauser, R. (2006). Verteilungseffekte der Hartz-IV-Reform: Ergebnisse von Simulationsanalysen.

Belhadj, B. (2011). New fuzzy indices of poverty by distinguishing three levels of poverty. *Research in Economics*, 65(3):221–231.

Berthold, N., Brunner, A., and Zenzen, J. (2010). Makroökonomische rahmenbedingungen und die einkommensverteilung – welchen einfluss hat die finanzkrise? Wirtschaftswissenschaftliche beiträge des lehrstuhls für volkswirtschaftslehre, insbes. wirtschaftsordnung und sozialpolitik, nr. 108.

Biewen, M. (2000). Income inequality in germany during the 1980s and 1990s. *Review of Income and Wealth*, 46(1):1–19.

Biewen, M. (2001). Measuring the effects of socio-economic variables on the income distribution: An application to the east german transition process. *Review of Economics and Statistics*, 83(1):185–190.

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics*, 108(2):317–342.

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

Blanchflower, D. G. and Oswald, A. J. (2008). Is well-being u-shaped over the life cycle? *Social Science & Medicine*, 66(8):1733–1749.

Blank, R. M. (1989). Disaggregating the effect of the business cycle on the distribution of income. *Economica*, 56:141–163.

Blinder, A. S. and Esaki, H. Y. (1978). Macroeconomic activity and income distribution in the postwar united states. *The Review of Economics and Statistics*, 60(4):604–609.

Blundell, R., Gosling, A., Ichimura, H., and Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, 75(2):323–363.

Boes, S. and Winkelmann, R. (2010). The effect of income on general life satisfaction and dissatisfaction. *Social Indicators Research*, 95(1):111–128.

Bound, J. and Johnson, G. (1992). Changes in the structure of wages in the 1980's: An evaluation of alternative explanations. *The American Economic Review*, 82:371–392.

Brickman, P., Coates, D., and Janoff-Bullman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 36(8):917–927.

Bundesagentur für Arbeit (2011). Analyse Arbeitsmarkt Deutschland, Zeitreihen 2011. http://statistik.arbeitsagentur.de/Statischer-Content/Statistische-Analysen/Analytikreports/Zentrale-Analytikreports/Jaehrliche-Analytikreports/Generische-Publikationen/Arbeitsmarkt-Deutschland-Zeitreihen/Analyse-Arbeitsmarkt-Deutschland-Zeitreihen-2011.pdf. Accessed April 2nd 2013.

Bundesagentur für Arbeit (2012). Arbeitsmarkt Deutschland, Zeitarbeit Aktuelle Entwicklung. http://statistik.arbeitsagentur.de/Statischer-Content/Arbeitsmarktberichte/Berichte-Broschueren/Arbeitsmarkt/Generische-Publikationen/Arbeitsmarkt-Deutschland-Zeitarbeit-Aktuelle-Entwicklung-1HJ2012.pdf. Accessed April 2nd 2013.

Buse, A. (1982). The cyclical behaviour of the size distribution of income in canada: 1947-78. *Canadian Journal of Economics*, 15:189–204.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47:225–238.

Checchi, D. and García-Pñalosa, C. (2010). Labour market institutions and the personal distribution of income in the oecd. *Economica*, 77(307):413–450.

Clark, A. E. and Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of Public Economics*, 61(3):359–381.

Clark, A. E. and Senik, C. (2010). Who compares to whom? the anatomy of income comparisons in europe. *The Economic Journal*, 120(544):573–594.

Cowell, F. A. (2000). Measurement of inequality. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 1. Elsevier.

Daly, M. C. and Valletta, R. G. (2006). Inequality and poverty in united states: the effects of rising dispersion of men's earnings and changing family behaviour. *Economica*, 73(289):75–98.

D'Ambrosio, C. and Frick, J. R. (2007). Income satisfaction and relative deprivation: An empirical link. *Social Indicators Research*, 81(3):497–519.

Danziger, S., van der Gaag, J., Taussig, M. K., and Smolensky, E. (1984). The direct measurement of welfare levels: How much does it cost to make ends meet? *The Review of Economics and Statistics*, 66(3):500–505.

De Vos, K. and Garner, T. I. (1991). An evaluation of subjective poverty definitions: Comparing results from the u.s. and the netherlands. *Review of Income and Wealth*, 37(3):267–285.

Di Tella, R. and MacCulloch, R. (2008). Happiness adaptation to income beyond" basic needs". Nber working papers 14539, National Bureau of Economic Research.

Diener, E., Suh, E. M., Lucas, R. E., and Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125(2):276–302.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

Dustmann, C., Ludsteck, J., and Schönberg, U. (2009). Revisiting the German wage structure. *Quarterly Journal of Economics*, 124(2):843–881.

Easterly, W. and Fischer, S. (2001). Inflation and the poor. *Journal of Money, Credit and Banking*, 33:160–178.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55:251–276.

Farré, L. and Vella, F. (2008). Macroeconomic conditions and the distribution of income in spain. *Labour*, 22(3):383–410.

Federal Government of Germany (2008). *Lebenslagen in Deutschland. Der 3. Armuts- und Reichtumsbericht der Bundesregierung*. Berlin.

Ferrer-i Carbonell, A. (2011). Happiness economics. *Els Opuscles del CREI*, (28).

Ferrer-i Carbonell, A. and Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497):641–659.

Fortin, N., Lemieux, T., and Firpo, S. (2010). Decomposition methods in economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4. Elsevier.

Frank, R. H. (2005). Does absolute income matter? In Stiglitz, J. and Muet, P.-A., editors, *Economics and Happiness*, page 65. Oxford University Press.

Frenette, M., Green, D. A., and Milligan, K. (2007). The tale of the tails: Canadian income inequality in the 1980s and 1990s. *Canadian Journal of Economics*, 40(3):734–764.

Frey, B. S. and Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, 40(2):402–435.

Frick, J. R. and Grabka, M. M. (2009). Gestiegene Vermögensungleichheit in Deutschland. DIW Wochenbericht 4/2009, DIW Berlin.

Frick, J. R. and Grabka, M. M. (2010). Weiterhin hohes Armutsrisiko in Deutschland. DIW Wochenbericht 7/2010, DIW Berlin.

Fuchs, V. (1967). Redefining poverty and redistributing income. *The Public Interest*, 8:88–95.

Fuchs-Schündeln, N., Krueger, D., and Sommer, M. (2010). Inequality trends for germany in the last two decades: A tale of two countries. *Review of Economic Dynamics*, 13(1):103–132.

German Council of Economic Experts (2009). Securing the future through responsible economic policies. Annual report, Wiesbaden.

Gernandt, J. and Pfeiffer, F. (2007). Rising wage inequality in germany. *Journal of Economics and Statistics*, 227:358–380.

Geyer, C. J. (2006). 5601 notes: The subsampling bootstrap. Unpublished manuscript.

Goedhart, T., Halberstadt, V., Kapteyn, A., and Van Praag, B. (1977). The poverty line: concept and measurement. *Journal of Human Resources*, pages 503–520.

Gosling, A., Machin, S., and Meghir, C. (2000). The changing distribution of male wages in the uk. *The Review of Economic Studies*, 67(4):635–666.

Grabka, M. M. (2009). Codebook for the PEQUIV file 1984-2008: CNEF variables with extended income information for the SOEP. Data documentation no. 45, DIW Berlin.

Hagenaars, A. J. and van Praag, B. (1985). A synthesis of poverty line definitions. *Review of Income and Wealth*, 31(2):139–154.

Haisken-DeNew, J. P. and Frick, J. R. (2005). Desktop companion to the german socio-economic panel study. Technical report, DIW Berlin. Version 8.0.

Härdle, W., Hall, P., and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, 21(1):157–178.

Hauser, R. and Becker, I., editors (2002). *Reporting on Income Distribution and Poverty: Perspectives from a German and a European Point of View*. Springer.

Headey, B. and Wooden, M. (2004). The effects of wealth and income on subjective well-being and ill-being. *The Economic record*, 80:24–33.

Heer, B. and Süssmuth, B. (2009). Cold progression and its effects on income distribution. CESifo working paper no. 951.

Horowitz, J. (2009). *Semiparametric and nonparametric methods in econometrics*, volume 692. Springer.

Huppert, F. A. and Whittington, J. E. (2003). Evidence for the independence of positive and negative well-being: Implications for quality of life assessment. *British Journal of Health Psychology*, 8:107–122.

Hyslop, D. and Mare, D. (2005). Understanding new zealands changing income distribution 1983-98: A semiparametric analysis. *Economica*, 72(3).

Ichimura, H. and Lee, L. F. (1991). Semiparametric least squares (SLS) and weighted SLS estimation of multiple index models: Single equation estimation. In Barnett, W., Powell, J., and Tauchen, G., editors, *Nonparametric and Semiparametric Methods in Economics and Statistics*, volume 3–49. Cambridge University Press.

Jäntti, M. (1994). A more efficient estimate of the effects of macroeconomic activity on the distribution of income. *The Review of Economics and Statistics*, pages 372–378.

Jäntti, M. and Jenkins, S. P. (2010). The impact of macroeconomic conditions on income inequality. *The Journal of Economic Inequality*, 8(2):221–240.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2):231–254.

Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, pages 410–442.

Kahneman, D. (1999). Objective Happiness. In Kahneman, D., Diener, E., and Schwarz, N., editors, *Well-Being: Foundations of Hedonic Psychology*, pages 3–25. Russell Sage Foundation Press.

Kapteyn, A., Kooreman, P., and Willemse, R. (1988). Some methodological issues in the implementation of subjective poverty definitions. *Journal of Human Resources*, pages 222–242.

Kohn, K. (2006). Rising wage dispersion, after all! the german wage structure at the turn of the century. Iza discussion paper no. 2098, Bonn.

Krämer, W. (1994). *Statistische Probleme bei der Armutsmessung: Gutachten im Auftrag des Bundesministeriums für Gesundheit*. Nomos Verlagsgesellschaft, Baden-Baden.

Kroh, M. (2005). Surveying the left-right dimension: The choice of a response format. DIW Berlin discussion papers 491, DIW Berlin.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45(1):1–28.

Levy, F. and Murnane, R. J. (1992). Us earnings levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of Economic Literature*, 30(3):1333–1381.

Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *Quarterly Journal of Economics*, 120(3):963–1002.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press, Cambridge.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons.

Mentzakis, E. (2011). Allowing for heterogeneity in monetary subjective well-being valuations. *Health Economics*, 20(3):331–347.

Minarik, J. J. (1979). The size distribution of income during inflation. *Review of Income and Wealth*, 25(4):377–391.

Mocan, H. N. (1999). Structural unemployment, cyclical unemployment, and income inequality. *Review of Economics and Statistics*, 81(1):122–134.

Murphy, K. M. and Welch, F. (1992). The structure of wages. *The Quarterly Journal of Economics*, 107(1):285–326.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Newey, W. K. and Stoker, T. M. (1993). Efficiency of weighted average derivative estimators and index models. *Econometrica*, pages 1199–1223.

OECD (2008). *Growing unequal? Income distribution and poverty in OECD countries*. Paris.

OECD (2009). *National Accounts at a Glance 2009*.

OECD (2011). *Divided We Stand: Why Inequality Keeps Rising*. Paris.

OECD (2012). OECD StatExtract. Technical report. Accessed January 2012.

Palmer, J. L. and Barth, M. C. (1977). The distributional effects of inflation and higher unemployment. In Moon, M. and Smolensky, E., editors, *Improving Measures of Economic Well-Being*, volume 201–239. Academic Press.

Parker, S. C. (1996). Explaining the determinants of us income inequality, 1948-90. Economics working paper 163, University of Durham, Durham.

Parker, S. C. (1998). Income inequality and the business cycle: a survey of the evidence and some new results. *Journal of Post Keynesian Economics*, 21(2):201–225.

Parker, S. C. (2000). Opening a can of worms: the pitfalls of time-series regression analyses of income inequality. *Applied Economics*, 32(2):221–230.

Peichl, A., Pestel, N., and Schneider, H. (2012). Does size matter? the impact of changes in household structure on income distribution in germany. *Review of Income and Wealth*, 58(1):118–141.

Politis, D., Romano, J., and Wolf, M. (1999). *Subsampling*. Springer.

Prasad, E. S. (2004). The unbearable stability of the german wage structure: Evidence and interpretation. Imf staff papers no. 51, International Monetary Fund.

Ravallion, M. (2008). Dollar a day revisited. Policy research working paper 4620, World Bank.

Ravallion, M. (2010). Poverty lines across the world. Policy research working paper 5284, World Bank.

Ravallion, M. and Chen, S. (2011). Weakly relative poverty. *The Review of Economics and Statistics*, 93(4):1251–1261.

Salem, A. and Mount, T. (1974). A convenient descriptive model of income distribution: the gamma density. *Econometrica*, pages 1115–1127.

Sandvik, E., Diener, E., and Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non-self-report measures. *Journal of Personality*, 61(3):317–342.

Schultz, T. P. (1969). Secular trends and cyclical behaviour of income distribution in the united states: 1944-1965. In Soltow, L., editor, *Six Papers on the Size Distribution of Wealth and Income*. Columbia University Press/NBER.

Singh, S. and Maddala, G. (1976). A function for size distribution of incomes. *Econometrica*, 44:963–970.

Steiner, V. and Wagner, K. (1996). Has earnings inequality in Germany changed in the 1980's? *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 118:29–59.

Steptoe, A. and Wardle, J. (2005). Positive affect and biological function in everyday life. *Neurobiology of Aging*, 26(1):108–112.

Terza, J. V. (1985). Reduced-form trinomial probit: A quantal response model without a priori restrictions. *Journal of Business & Economic Statistics*, 3(1):54–59.

Thurow, L. C. (1970). Analyzing the american income distribution. *American Economic Review Papers and Proceedings*, 60(2):261–269.

Urry, H. L., Nitschke, J. B., Dolski, I., Jackson, D. C., Dalton, K. M., Mueller, C. J., Rosenkranz, M. A., Ryff, C. D., Singer, B. H., and Davidson, R. J. (2004). Making a life worth living neural correlates of well-being. *Psychological Science*, 15(6):367–372.

Van Praag, B. M. S. (1991). Ordinal and cardinal utility: an integration of the two dimensions of the welfare concept. *Journal of Econometrics*, 50(1):69–89.

Wagner, G., Frick, J., and Schupp, J. (2007). The german socio-economic panel study (SOEP) - scope, evolution and enhancements. In *Schmollers Jahrbuch 127*, number 1, pages 139–169.

Wagner, G. G. (2007). Wie die 11er-Skala in das SOEP kam – Ein Beitrag zu den Problemen und Möglichkeiten multidisziplinärer Forschung und zugleich eine Fußnote zum Design der SOEP-Stichprobe. In et al., S., editor, *Arbeitsmarkt-und Sozialpolitikforschung im Wandel – Festschrift für Christof Helberger zum 65. Geburtstag*, volume 65, pages 40–62.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.

Winkelmann, L. and Winkelmann, R. (1998). Why are the unemployed so unhappy? evidence from panel data. *Economica*, 65(257):1–15.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

World Bank (1990). *World Development Report: Poverty*. Oxford University Press, New York.

Zachert, U. (2004). Kündigungsschutz, Befristung und Leiharbeit in Europa. WSI Mitteilungen 3/2004.

# LIST OF FIGURES

*A Satisfaction-Driven Poverty Indicator - A Bustle Around The Poverty Line* 11

Abb. 1 Distribution of answers to the SOEP income satisfaction question in 2009. Dashed lines separate the three sections outlined by the questionnaires formulation and the missing values. Own calculations. 17

Abb. 2 Illustrative example of the indicator smoothing concept applied to the relationship between income satisfaction and income based on real SOEP data of the year 1985. Here the jump function is combined with a third degree polynomial. Own calculations. 27

Abb. 3 Grid search results in 5 Euro steps using the linear probability model. Baseline is set as the $R^2$ of regressions without dummy. The highest value is normalized to 1. Top left (a) result for 1985, top right (b) for 1995, bottom left (c) for 2005 and bottom right (d) for 2009. Own calculations. 30

Abb. 4 Grid search results in 5 Euro steps using the 3-years fixed effects panel model. Baseline is set as the $R^2$ of regressions without dummy. The highest value is normalized to 1. Top left (a) result for 1985-1987, top right (b) for 1995-1997, bottom left (c) for 2005-2007 and bottom right (d) for 2007-2009. Own calculations. 30

Abb. 5 Probability plots for not being dissatisfied with income based on fixed effects estimation with and without the income dummy variable. The plots are valid for a reference person with zero valued control variables as in table 2, models (2) and (3). Own calculations. 34

Abb. 6 Time series: SDPLs based on LPM, FE, NLS and FENLS estimations and the 60% poverty line. Left (a) in Euros, right (b): as median percentage. Own calculations. 37

Abb. 7 Left (a): Time series of the (marginal) probability effect of the PTR transition based on the LPM, FE, NLS and FENLS estimations. A single outlier in 1989 was set to the bootstrap mean for the FENLS. Right (b): Time series of the length of the PTR based on NLS and FENLS estimations and their one sided percentile based bootstrap 90% confidence intervals. Outliers in 1987, 1990, 2001 for the NLS were set to their bootstrap means. Own calculations. 37
Abb. 8 Fixed Effects SDPL poverty lines based on the complete ECHP panel information. 95% confidence intervals are provided based on bootstrap estimation. Results based on the BHPS 1994-2001 and on the SOEP 1994-1996 are provided for comparison. Own calculations. 40

*Understanding Rising Income Inequality in Germany* 49

Abb. 9 Trends in inequality and poverty 1994-2008 (a) 52
Abb. 10 Trends in inequality and poverty 1994-2008 (b) 52
Abb. 11 Trends in employment and unemployment 1994-2008 54
Abb. 12 Share of individuals living in households with different employment outcomes 1994-2008 54
Abb. 13 Inequality in equivalized labour income 1994-2008 55
Abb. 14 Overall change in density from 1999/2000 to 2005/2006 68
Abb. 15 Density change if only one factor is changed 68
Abb. 16 Density change if only one factor is changed 70

*The Income Distribution and the Business Cycle in Germany - A Semiparametric Approach* 99

Abb. 17 Time series: **(left)** Male ILO unemployment and inflation rate. **(right)** GDP per capita and government expenditure per capita without pensions. Source: Genesis data base, Federal Statistical Office. 112

Abb. 18 Time series: **(left)** Quantile ratios of the income distribution. **(right)** Gini, Theil measure and the mean log deviation. Own calculations based on data after sample selection. 114
Abb. 19 Regression Surfaces based on Double Index Model estimation for household types 12 of the elderly **(upper left)**, 34 **(upper right)** and 56 **(lower left)** of the non-elderly without and with children, respectively. Own calculations. 118
Abb. 20 Average cross sections of the double-index regression fits below the 25th (LOW), between the 25th and 75th (MED) and above the 75th (HI) centile of the macro-indices. For household type 12 **(upper left)**, household type 34 **(upper right)**, household type 56 **(lower left)**, own calculations. 120
Abb. 21 Unconditional income distributions and their counterfactual (cf.) counterparts for lower GDP (-20%) **(left)** and for higher GDP (+20%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations. 122
Abb. 22 Unconditional income distributions and their counterfactual (cf.) counterparts for lower inflation (-3%) **(left)** and for higher inflation (+3%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations. 123
Abb. 23 Unconditional income distributions and their counterfactual (cf.) counterparts for lower government expenditure (-20%) **(left)** and for higher government expenditure (+20%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations. 124
Abb. 24 Unconditional income distributions and their counterfactual (cf.) counterparts for no unemployment **(left)** and for higher unemployment (+10%) **(right)** with estimated cf. heterogeneity for the year 2010. Own calculations. 124

Abb. 25 Micro-Index **(left)**, Macro-Index **(right)** coefficient estimators convergence for different sample sizes measured by the 90/10 quantile ratio for the 7 micro- and 3 macro-coefficients respectively. Own calculations. 129

# LIST OF TABLES

*A Satisfaction-Driven Poverty Indicator - A Bustle Around The Poverty Line* 11

Table 1 Year 1998, income satisfaction dummy as regressand. (1) LPM with the best income dummy fit, (2) 3-years FE panel model with the best income dummy fit, (3) as (2) without dummy and (4) as (2) without additional controls. Robust standard errors. Own calculations. 31

Table 2 Year 2008, income satisfaction dummy as regressand. (1) LPM with the best income dummy fit, (2) 3-years FE panel model with the best income dummy fit, (3) as (2) without dummy and (4) as (2) without additional controls. Robust standard errors. Own calculations. 32

Table 3 POLS country panel estimations with the SDPL as poverty line for 13 ECHP countries. Models (3) and (4) make use of multiple imputation. Standard errors are clustered. Own calculations. 42

*Understanding Rising Income Inequality in Germany* 49

Table 4 Changes in the German tax schedule 57

Table 5 Inequality and poverty indices 64

Table 6 Ceteris paribus effects 1999/2000 – 2005/2006 72

Table 7 Decomposition 1999/2000 – 2005/2006 76

Table 8 Results from 120 possible sequential decompositions for the decomposition 1999/2000 – 2005/2006 77
Table 9 Ceteris paribus effects 1999/2000 – 2007/2008 80
Table 10 Decomposition 1999/2000 – 2007/2008 81
Table 11 Variable names 87
Table 12 Ordinal logit models 88
Table 13 Regression of log labour incomes on household characteristics (× denotes interaction effects) 89

*The Income Distribution and the Business Cycle in Germany - A Semiparametric Approach* 99

Table 14 Double Index Model estimation results: Micro- and Macro-Index components. P-values are based on clustered subsampling with sample size correction and estimated rate of convergence. HH-Type 12: Elderly, HH-Type 34: non-elderly w/o children, HH-Type 56: non-elderly with children. Own calculations. 115
Table 15 Quantiles and quantile ratios (QR) of the income distribution for counterfactual, lowered macro variables for the year 2010. Own calculations. 125
Table 16 Quantiles and quantile ratios (QR) of the income distribution for counterfactual, elevated macro variables for the year 2010. Own calculations. 125
Table 17 Inequality measures of the income distribution for counterfactual, lowered macro variables for the year 2010. Own calculations. 126
Table 18 Inequality measures of the income distribution for counterfactual, elevated macro variables for the year 2010. Own calculations. 126

Table 19 The components of German final government expenditure in 2010. German Federal Statistical Office, Wiesbaden 2013, Version: 22. March 2013. 131

# ACRONYMS

BHPS British Household Panel Survey

CDF Cumulative density function

ECHP European Community Household Panel

EUR Euro

FE Fixed effects

FENLS Fixed effects nonlinear least squares

FGT Foster-Greer-Thorbecke metric

FT Full-time employment

GDP Gross domestic product

GLS Generalized least squares

HH Household

LPM Linear probability model

MI Multiple imputation

MLD Mean log-deviation

NLS Nonlinear least squares

OLS Ordinary least squares

POLS Pooled ordinary least squares

PT Part-time or marginal employment

PTR Poverty transition region

| | |
|---|---|
| QR | Quantile ratio |
| RSS | Residual sum of squares |
| SDPL | Satisfaction-Driven Poverty Line |
| SOEP | German Socio-Economic Panel |
| USD | US-Dollar |

# PUBLICATION NOTES

Some ideas and figures have appeared previously in the following publications:

Biewen, M., A. Juhasz (2011): "Can Employment Changes Explain Rising Inequality in Germany?", Journal of Applied Social Science Studies, Vol. 131, pp. 349 - 357.

Biewen, M., A. Juhasz (2012): "Understanding Rising Inequality in Germany, 1999/2000 - 2005/06", Review of Income and Wealth, Vol. 58, pp. 62-647.

The third chapter is based on joint work with Martin Biewen and has been published in a similar form in the Review of Income and Wealth, Vol. 58, pp. 622-647. Much of this work is related to a report we produced for the Federal Government of Germany for its "4th Report on Poverty and Richness in Germany", which was also published as:

Arndt, C., Biewen, M., B. Boockmann, M. Rosenmann et al. (2013): "Aktualisierung der Berichterstattung über die Verteilung von Einkommen und Vermögen in Deutschland für den 4. Armuts- und Reichtumsbericht der Bundesregierung", Gutachten für das Bundesministerium für Arbeit und Soziales.