USING SURVEY DATA IN ECONOMETRIC ANALYSES: DETERMINANTS OF WELL-BEING AND MEASUREMENT PROBLEMS

Inaugural-Dissertation

zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

Volkswirtschaftliche Fakultät

2012

vorgelegt von

Gregor Tannhof

Referent:	Prof. Dr. Joachim Winter
Korreferent:	Prof. Alexander Danzer, Ph.D.
Promotionsabschlussberatung:	7. November 2012

Berichterstatter: Prof. Dr. Joachim Winter

Prof. Alexander Danzer, Ph.D.

Prof. Dr. Uwe Sunde

Mündliche Prüfung: 22. Oktober 2012

To my mother

Acknowledgements

This thesis was written from 2009 to 2012 at the *Seminar für empirische Wirtschaftsforschung* (Department of Economics, Ludwig-Maximilians-Universität (LMU) in Munich) where I worked as a research and teaching assistant to Prof. Joachim Winter.

First and foremost I would like to thank my supervisor Joachim Winter whose support was crucial for the success of this project. His creative ideas and sharp comments led to an early and efficient start of my thesis and helped me remain focused and determined throughout its completion. His outstanding doctoral classes sharpened my understanding of econometrics and taught me what a thorough analysis actually is. Moreover, he was always available for valuable comments and suggestions during my own and our joint work for which I am very thankful.

Furthermore, I am also grateful to the participants of the "Research Workshop Empirical Economics" for their numerous and helpful comments on my work.

I am especially thankful to Rebecca Wasser for her constant support and motivation and to Christian Beermann for all the fruitful discussions we had.

Finally, my greatest gratitude goes to my mother for everything she taught me and for never losing faith. This work is dedicated to her.

Munich, June 15, 2012 Gregor Tannhof

Contents

Chapter 1

1 Introduction

Whenever we run a project we need to evaluate what has happened so far, where we are now and what is yet to come in order to determine if we are on track and if we have gotten where we wanted to go at all. To do so, any credible statement must always be based on reliable sources of information. No matter if it is about the evaluation of a political reform, a momentary snapshot of what is going on in society or a what-if consideration – the source of credibility is data.

Research questions are diverse and so are the demands for data. Long before economists stepped into laboratories to generate data, other scientific researchers felt the same need to validate their hypotheses and had to collect data. Whenever administrative records were not accessible or available this was of course a seemingly insuperable barrier. Only a few were lucky and were asked to collect data. This is far from where we are today with nearly unlimited data access for every researcher. An interesting and well documented example is the evolution of the Panel Study of Income Dynamics (PSID) in the United States. Duncan (2002) or (more recently) McGonagle and Schoeni (2006) outline its history in their papers: Initially the study was set up to assess the impact of the president's "War on Poverty" and data was collected from low income families in 1966 and 1967. In 1968 a representative sample of households was added and this became known as the first wave of the PSID which is still active and widely used in research today. This was a great leap forward and must have also been a true inspiration for others. And in fact the pioneers of the PSID and its early staff were a natural source of expertise and advice when it came to setting up a panel study. However, 10 years passed before plans for this kind of micro-panel became concrete in

Germany and additional six years passed before the inception of the German counterpart of the PSID: The first wave of the German Socio-Economic Panel Study (GSOEP) was fielded in 1984. Its beginning and early history is concisely described in Krupp (2008). In his paper he explicitly mentions the importance of previous work on panel studies and the support by those researchers who had already worked on this subject. According to Krupp (2008), Duncan's (one of the earlier PSID staff) talk during the final approval process for the GSOEP was very helpful. But of course the story does not end here and all the lessons learned were again helpful when different researchers started other panel studies in their countries. As laid out by Anger *et al.* (2008) this includes the Australian Household, Income and Labour Dynamics in Australia (HILDA) Survey which started in 2001 and the new UK Household Longitudinal Study (UKHLS) which is supposed to extend the British Household Panel Survey (BHPS). This discussion could be continued but for this work I only want to mention one further panel, namely the Dutch CentER panel which started in 1991, to show that the amount of available panel studies is steadily increasing and that more and more data becomes available.

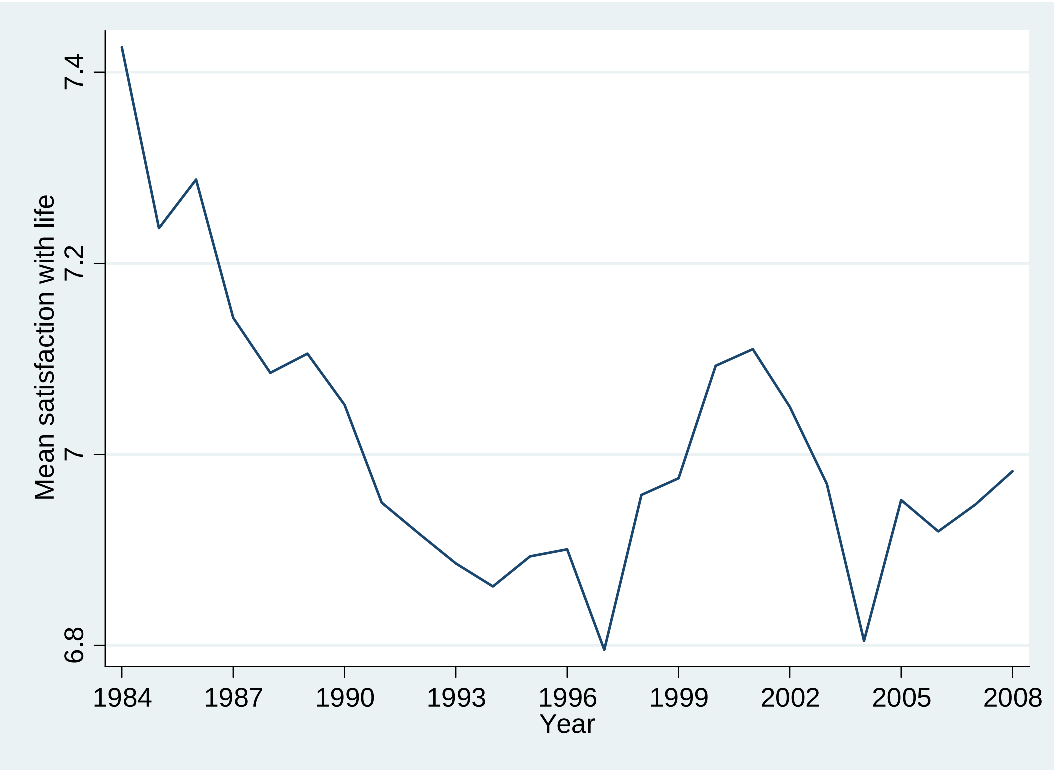
With increasing use of data and manifold interdisciplinary research questions addressed to it the panels become larger and more complex. Recently, the field of economics especially focuses on well-being and health (often also referred to as a part of the very general concept of well-being) related research. This can be seen exemplarily within the GSOEP by looking at the inclusion of new measures such as the self-rated health status (SRHS) which was asked for the first time in 1992 and is surveyed on a regular basis since 1994.

It is also very important to mention that the academic debate has reached international politics. In 2007 the European Commission in cooperation with the European Parliament, Club of Rome, WWF, and OECD launched the "Beyond GDP" initiative which aims at improving measures of progress, wealth and well-being.[1] Meanwhile, the main objectives are defined and the initiative works on complementing GDP with other aspects of sustainable development (i.e. social and environmental indicators) and on extending national accounts in the same way by considering social and environmental issues. The idea behind all of this is actually an interdisciplinary enrichment of traditional economic thinking. It says that we have to think further than the total value of goods and services produced –

[1] http://www.beyond-gdp.eu

we have to care about how sustainable our progress is and how the primary stakeholders in all of this, the people, value it. Additionally, we have to see and understand the non-monetary costs of our development, no matter if we think of them as ecological destruction or reduced well-being (or the interactions). Figure 1.1 uses GSOEP data to show the importance. Satisfaction with life is (at least in the raw data) anything but constant over time. Again, this is not only an interesting research question but also politicians must understand what drives well-being and its components in order to make successful, sustainable and people-oriented politics. Additionally, one must make sure that in the end we have measured what we wanted to measure and therefore all statistics using (survey) data must be calculated and interpreted very carefully.

Figure 1.1: Mean satisfaction with life over the years



All this is good news as more and more data becomes available, datasets become richer due to new and interdisciplinary research and some of these new research directions are recognized by leading political institutions and NGOs. Another positive aspect of the increased use of large household surveys is the growing awareness of potentially biased or inaccurate analyses when survey data is used. This literature is not really new and for example psychologists have been examining these issues for a while as collecting data by conducting (smaller) surveys is typical for this discipline. However, using large micro datasets made many economists concentrate their attention on this matter. In the beginning I stated that every credible analysis needs reliable sources of information and having more and better data surely helps but it is not an end in itself. Therefore, this thesis has two dimensions: On the one hand, it deals with determinants of well-being and on the other hand, it discusses measurement problems in surveys. By doing so, this thesis does not only enrich the fast growing and widely discussed well-being literature but also deals with commonly ignored quality aspects in survey research that – in order to present results which are as convincing as possible – should always be on the agenda of researchers in every discipline applying these methods.

The first part of this work is concerned with a substantial determinant of well-being and analyzes how well-being is influenced by overtime work. Afterwards, this thesis highlights methodological pitfalls in survey-based research. Firstly, it investigates transient effects on well-being (and risk attitude as another general concept) which can be both an interesting determinant of well-being itself and also a measurement quality issue, i.e. a threat to internal validity. Secondly, it focuses on problematical survey response behavior when bracketed answer formats are used to measure consumption expenditures. At the end this thesis takes an international perspective on well-being and analyzes its relationship with a country's wealth at the European level.

In chapter 2, the first essay of my dissertation "Differences in well-being between overtime and non-overtime workers" takes up the discussion by asking the question whether there is more that influences well-being than the typically investigated income variables. Recent contributions to the latter aspect are for example Blanchflower and Oswald (2004) who use American and British data and find a positive relationship between income and happiness as well as ethical and gender differences. Deaton (2008) analyzes Gallup World Poll data and describes a positive relationship between per capita GDP and life satisfaction as well as cross-country differences. Frijters *et al.* (2004) use German data and show a positive relationship between income and life satisfaction. But working overtime is also a particularly interesting source of variation in well-being as the potential effects are severe and can range from increased fatigue to suicide (Starrin *et al*. (1990)). Therefore, overtime work can have an immediate impact on society, for example by increasing health care costs.

Except for the work of Golden and Wiens-Tuers (2006) most research is conducted in the fields of medicine and psychology and there is not much that the literature agrees on. Of course this does not mean that overtime itself has not been interesting to economists before – it was just analyzed in different contexts, such as overtime as a device for workers to signal their value (e.g. Anger (2008)) or as an explanatory variable where occupational injuries were considered (e.g. Ruser (1991)).

The typical problems or differences in the literature regarding the relationship between overtime and well-being often arise due to small and highly specific samples, the definition used for overtime and the inclusion or omission of important control variables. To overcome these issues, I use data from the large German SOEP panel which offers a huge variety of personal and household information. I analyze differences in well-being between overtime and non-overtime workers as well as underemployed persons whereas I use deviations from the individually preferred workload to assign observations to the appropriate group (of overtime, non-overtime and underemployed workers) rather than using total hours of work. Using these deviations has the huge advantage that the effects of overtime can be analyzed when there is heterogeneity in unobservable characteristics like stress-bearing capabilities and that it is therefore not necessary to restrict the sample to those who work more than e.g. 70 hours per week. More specifically, I define two competing measures as the reference point: Contracted and desired working hours. I argue why it is not satisfactory to restrict the sample to excessive overtime work or to simply use the total amount of hours worked as an approach to uncover a relationship between high workload and well-being. The data shows that there exists almost no real underemployment (meaning that only few people work less than they have to) but it can also be seen that there is in fact underemployment when it comes to how many hours people actually want to work. And as this is a different situation than working more than desired or contracted and also different to working as much as desired or contracted it should not be ignored from the start. The regressions also include lifestyle variables to account for correlations between overtime work and adverse health behavior which is one of the prominent findings in the related literature (e.g. Johnson and Lipscomb (2006) provide a summary). The empirical set-up is determined by the ordinal scales that are used to measure most of the variables in the survey and consequently I use ordered probit models and a random-effects ordered probit model for the panel dimension. The latter has the advantage of controlling for time-constant and individual-specific heterogeneity. The relationship between overtime work and well-being is considered in different ways and I show that deviations from contracted and desired hours are associated with worse self-rated health status, lower satisfaction with health, lower satisfaction with life and lower satisfaction with work. It is interesting to see that the results differ depending on which overtime measure is used but that the baseline results are unchanged as both measures share a large fraction of the same information. The results obtained in the main part of the essay are found to be robust to restricting the sample and omitting those who have only few contracted working hours and they are also robust to exploiting the panel dimension which uses a lot more information and controls for time-constant and individual-specific heterogeneity.

The main conclusion of the second chapter is that not achieving the personally preferred amount of work is associated with lower well-being and by using deviations from desired hours we can rule out that people choose to work overtime voluntarily. However, this result has difficult policy implications: On the one hand, overtime creates costs for the entire society (e.g. health problems), but on the other hand, firms need the flexibility overtime provides and it would not make any sense to hire and train additional staff for say, a two month period. Policy actions would have to balance these two aspects.

While chapter 2 discusses the relationship of well-being and fundamental personal living conditions, chapter 3 introduces transient effects on well-being. General statements about well-being such as life satisfaction are supposed to be quite stable over time as there would have to be major and long-lasting changes in one’s life to make an adjustment to the personal evaluation. Permanently working overtime would be an example. Unfortunately, people are not always (or maybe they are never) perfectly rational and it is quite likely that having a “bad day” can influence the answers to different kinds of supposedly stable well-being questions. Joachim Winter’s and my essay “Let the sunshine in: The effect of weather on survey measures of health, well-being, and preferences” is based on and inspired by the early work of Cunningham (1979) and Schwarz and Clore (1983). The former finds that weather influences helping behavior and the explanation is that changes in weather induce changes in mood meaning “good weather” is associated with good mood and ”bad weather” is associated with bad mood. Schwarz and Clore (1983) also use weather as a mood induction and find that people in a good mood (i.e. on sunny days) report more happiness and higher life satisfaction than respondents in a bad mood (i.e. on rainy days). These results are interesting in a way that they contradict the idea of perfectly measuring stable personal evaluations of well-being by asking for them in a survey. In our work we also use current weather conditions as mood induction but instead of using a small controlled experiment, we exploit a large natural experiment by matching weather data from the *German Weather Service* (DWD) with information from the GSOEP including additional information (under a separate contract) on the respondent’s residence. Using these data sources not only increases the amount of available data but also has the advantage that we can obtain and control for a larger variety of weather variables than typically done in the literature. Our main interest lies in investigating the influence of current mood on self-stated measures of well-being and risk attitude – all of which are typically considered as not being influenced by transient effects even though others like Yardley and Rice (1991) have already pointed out that well-being can be both stable over time and prone to transient effects. There is another important difference in our analysis compared to recent work on mood effects: While others typically focus on cloud cover and include rain as a continuous variable we construct a rain dummy because we do not think that the actual amount of rain matters but rather believe that only the fact whether it rains or not is relevant in the mood induction. Additionally, a rain dummy captures at least some of the effects of other weather variables as for example rain and cloudiness are correlated. Furthermore, such a dummy is very easy to obtain, even in large surveys. Like in the last chapter, the dependent variables of interest are measured on an ordinal scale and we consequently use ordered probit models throughout the essay. It turns out that interviewees report worse health, lower satisfaction with health and life and also a lower willingness to take risks when the interview is conducted on a rainy day. We also find that the results are robust over both waves we took data from.

The essay concludes that researchers must be careful when self-reported measures from surveys are used in their analyses as they could be measured with error. As a result the error-term variance could be inflated or, even worse, parameters could be estimated inconsistently. Regarding laboratory experiments, the worst-case scenario is probably a correlation of weather and the treatment because then weather could mimic a treatment effect. This is not unlikely, especially when the control and the treatment group visit the lab on different days. However, as most weather characteristics can easily be measured without expert knowledge, we are confident that these factors can be controlled for without increasing the effort level too much. Finally, weather provides an interesting instrument when researchers want to work with e.g. self-stated mood but for whatever reason think that it is endogenous.

So far two different sources of variation in well-being have been discussed and the overall conclusions are supported by previous findings in the literature, for example by Yardley and Rice (1991). The first one is in line with the hypothesis that generally stable evaluations like satisfaction with life can change with important long-term or fundamental influences – an example in the context of this thesis is overtime work. The second one shows that there are in fact transient influences on the evaluation of one's well-being (or risk preferences) such as mood effects – which can be induced by weather as outlined previously. The latter can seriously harm inference.

Going a little bit further and away from well-being as described above, chapter 4 takes a look at measurement problems in consumption expenditures. Even though it may be necessary to move beyond GDP when measuring societal development, GDP and consumption (as a part of it) are likely to influence well-being and therefore are of particular interest. The essay "The influence of self-perception on bracket choice" extends previous work by Winter (2002) who conducted an experiment in the 2001 CentER panel where respondents had to answer bracketed expenditure questions. In this experiment one group received high bracket values, i.e. the values were shifted to the right of the expected population distribution, hence we expect more respondents to check the leftmost bracket as it contains the interval from 0 to the relatively high starting value of the next bracket, and another group which received low bracket values where the reasoning is the other way around. Winter (2002) finds that the median of self-reported consumption expenditures varies with the bracket values that are given to the respondents, a phenomenon also referred to as "bracket effect". More precisely, those who received the high values report higher consumption expenditures than those who received the low values. Now the interesting question is why such an effect occurs. The idea is based on Sudman *et al.* (1996) who argue that the brackets shown to respondents provide more information than just the (obvious) bracket values, namely contextual information. Thus, respondents interpret the brackets as a representation of the underlying population distribution with the mean lying somewhere in the middle. Respondents may be uncertain about how much they actually spent but as they have an idea of how much they spend relative to others, i.e. their typicality, they use the contextual distribution information and their self-perceived typicality to answer the question. Unlike the relationship between well-being and overtime or the transient influence of mood the mechanism described above is fundamentally different. It is a result of the question design and it is neither a long-term nor a transient effect but a heuristic that is used by the respondents because they lack exact knowledge of the quantity they are supposed to report on.

An important aspect of this essay is that Joachim Winter and I use the same data as in Winter (2002) but exploit the fact that the respondents also answered a question about their self-perceived typicality. Given that there is uncertainty about the exact amount spent on consumption we can analyze whether respondents use their self-perceived typicality to answer the question or not. We start by replicating the bracket effect in an interval regression model but then (using ordered probit) we find that those in the group with high bracket values also tend to choose lower bracket positions which works in the opposite direction of the bracket effect. But as soon as we include self-perceived typicality in our analysis we can see that its effect works in the direction of the bracket effect and (partially) offsets the (correct) tendency of choosing lower bracket positions when high values are presented (and the other way around). Therefore, the respondents with a higher typicality also tend to pick a higher bracket position and vice versa implicating that individuals actually use the bracket position for orientation, i.e. they apply the typicality heuristic. Our results shed additional light on the mechanism in the response process and offer an explanation for the bracket effects found by Winter (2002).

So far this work always exploited data from a single country. This can often be advantageous because these specialized (panel) datasets are typically very large and collect much demographic and personal information. In terms of consistent estimation this is a very important aspect. On the other hand, external validity is always a matter of opinion: Even though it is likely that results obtained in the UK can be transferred to e.g. Germany no one knows for sure unless the study has been replicated. The easiest way to overcome the external validity issue is to collect data simultaneously in many countries. As this is more expensive, data providers sometimes reduce costs by fielding only short questionnaires but this is not necessarily true for all international surveys.

Chapter 5 takes up these thoughts by analyzing the well-being complex on an international level from a macro rather than a microeconometric perspective. The starting point here is the work of Easterlin (1973, 1974) and what became known as the "Easterlin paradox". In brief he finds that within a country, richer people are happier than poorer people but that a rising income for all does not increase happiness on the aggregate level. He also finds that richer countries are not happier than poorer ones. Joachim Winter, Melanie Lührmann and I use a large international dataset covering the entire European Union (EU) during the world economic (financial) crisis (five waves fielded between July 2009 and October 2010) which was kindly provided by *The Gallup Organisation* in Brussels, Belgium. Over the years there was an intense debate whether this paradox holds or not but as the availability of (international) data grows, there seems to be consent that there actually are cross-country differences in well-being (e.g. Stevenson and Wolfers (2008) present extensive work on this matter using many different data sources). Nevertheless, we can still contribute in several ways. Besides the fact that we use current data from a period of severe economic turmoil, it is also interesting that the questions asked cover different aspects of well-being compared to the literature that has dealt with the Easterlin paradox so far. While most of the literature (with good reason) investigates the relationship between happiness or life satisfaction and a macro indicator (typically GDP), we work with existential fears (financial and professional expectations for the future, i.e. worries about income in old age, confidence in keeping the job and the subjective probability of finding a job if one was to be laid off) as determinants of well-being and we are convinced that this broadens the horizon of current macro level well-being research. Additionally, the dataset allows us to use both subjective and objective wealth measures and compare the well-being / wealth gradients. Finally, we have information on perceived poverty and can test for correlation with subjective income (at the country level).

We start with an Easterlin-type analysis and find that worries about income in old age and confidence in keeping the job are strongly related to GDP and even to a subjective wealth measure (which asks for personal living standards). These relationships also hold for the subjective probability of finding a job if one was to be laid off but they are weaker (we use simple OLS regressions throughout this chapter). We also show that the distribution of well-being at the EU level differs by age: Compared to the youngest age group, the oldest age group shows slightly less worries about income in old age and also a slightly higher confidence in keeping their job whereas their self-stated probability of finding a job if laid off is much lower. Next we show that the cross-country relationship of worries about income in old age and GDP also depends on age as it gets stronger the older the interviewees are. As we have data from five waves we are also able to exploit the time dimension. We show that even though country level well-being varies over time, its overall relationship with GDP remains stable during the crisis. In these above mentioned analyses we have found some patterns, i.e. northern countries always perform much better than countries from eastern Europe and it also becomes obvious that those countries especially distressed by the crisis (e.g. Greece) are typically amongst the bad performing ones. At the end of chapter 5 we show (at the country level) that the subjectively perceived poverty rate is highly correlated with subjective wealth and we interpret this as an indication that people use an internationally standardized reference point to compare themselves to when they are asked to answer such point scale questions.

Taken together, chapter 5 enriches the previous single-country analysis in this thesis by extending it to the international level. New data and new dimensions of well-being as well as alternative wealth measures are used to contribute to the literature that has been concerned with the Easterlin paradox so far. Nevertheless, (and as we will point out) further research is needed to derive clear-cut policy implications in order to help governments find a way of achieving the best possible life for their people. GDP is one determinant but well-being measures can be others, depending on how much of an increase in overall well-being is attributable to an increase in wealth.

My dissertation highlights the importance of the relatively modern well-being research and contributes to the literature by pointing towards the relationship between overtime work and well-being. This field of research should receive more attention in economics than it currently does, especially because the effects of changes in personal circumstances on well-being can be severe and because they can affect the entire society. I also argue that measurement problems in surveys due to transient mood effects or the use of heuristics are something that researchers should be concerned about. Nevertheless, the conclusion of every chapter is positive in a way that I either argue how problems can be avoided or by the fact that we now better understand how surveys are actually answered. Finally, I take the well-being analysis to the European level and show that there are correlations of subjective expectations (existential fears as a part of well-being) and a country's wealth. This shows that well-being can be defined by many aspects where some are more and some are less correlated with wealth at the country level and that we need more research in order to establish causality. The latter is very important because the more aggregate well-being is driven by wealth (e.g. GDP) the less we would need other measures to accompany it. Nevertheless, if we could consider all possible aspects and dimensions of well-being, we can imagine that some are mainly driven by wealth whereas others are probably not. This is also an important question that future research should address as we need to understand what a country should actually try to maximize in order to make their people's life as good as possible.

As seen above there is of course a lot of work ahead for the survey research community and therefore I point out additional open research questions at the end of every chapter. Looking at the existing literature and the broad interest in it I am convinced that current survey research is already doing a great job in boosting the quality of survey analysis and I am optimistic that it will find ways to steadily increase the quality of measurement and data analysis as well as political relevance and impact.

Chapter 2

2 Differences in well-being between overtime and non-overtime workers

2.1 Introduction

The economics literature often deals with overtime. Part of the literature focuses on overtime in classic microeconomic models and tries, for example, to understand why people would choose to work overtime at all (e.g. Anger (2008)) and other parts deal with the relationship of overtime, payment and employment (e.g. Trejo (1991)) or control for overtime when modeling occupational injuries (e.g. Ruser (1991)). Interestingly, the economics literature does not care too much about the interaction of overtime and well-being. Somehow it is common sense that working too much is not beneficial to one's health or life satisfaction and medical studies find different types of evidence ranging from fatigue to suicide (Starrin *et al.* (1990)). Nevertheless, it is not really clear how to measure the impairments for the employees. Typically this is done by considering overtime and long working hours (e.g. 70 and more) simultaneously. The motivation for doing so is not necessarily straightforward as the effects could not only depend on the amount of hours worked but also or alternatively result from a deviation from the individually preferred amount of hours. The latter approach has the immediate advantage that larger samples can be analyzed as focusing on excessive hours only naturally changes the sample and reduces its size.

As mentioned above one could think of deviations from the individually preferred working hours rather than total hours worked to explain adverse effects of overtime. If so, the only reason to believe that there should be an effect is that we implicitly assume some sort of perfect labor market matching equilibrium where everybody receives a contract with the desired working time. If this is not the case we should not primarily compare actual and contracted hours but rather actual and desired hours. Consequently, this implies that working less than actually desired might also be related to different levels of well-being.

Even though there is a lot of literature on overtime and health effects in clinical and psychological journals, these studies are usually based on small, highly specific samples, do not control for important covariates and are often biased towards excessive working hours. By discussing the differences between various types of overtime measures and applying econometric models to a large and representative sample, I close this gap.

By using GSOEP[2] data I will show that working more (or even less) than the preferred workload has an adverse effect on well-being. The results differ to some extent depending on the overtime measure used in the analysis. I will also discuss the assumptions needed for a meaningful interpretation of the findings.

The remainder of this chapter is structured as follows: Section 2.2 gives a literature overview, then section 2.3 discusses the potential overtime measures and section 2.4 describes the dataset. Afterwards, section 2.5 outlines the empirical strategy, section 2.6 gives the results and section 2.7 discusses the findings. Finally, section 2.8 concludes.

2.2 Literature review

First of all, it is necessary to understand why a deviation from preferred working hours should affect individuals. For example, why should working extra hours have adverse effects on your health? Especially this question has been of major interest in medicine and psychology but not so much in economics. Therefore, the most recent articles are published in medical and psychological journals.

[2] German Socio-Economic Panel (*German Institute for Economic Research* – DIW, Berlin).

Dahlgren *et al.* (2006) use an experimental field study and a sample of 15 workers which work one regular and one overtime week and find shorter sleep time, an increase in sleepiness and more symptoms of fatigue. Dembe *et al.* (2005) analyze responses from 10,793 NLSY (National Longitudinal Survey of Youth) participants while accounting for a variety of control variables such as personal and workplace characteristics. They find long working hours to be related to an increasing risk of injury and mention that the findings are in line with the hypothesis (amongst others) that long working hours induce fatigue and stress. Johnson and Lipscomb (2006) evaluate different types of papers and find that long working hours (together with other organizational aspects) are associated with stress, fatigue, adverse health behavior (smoking) and cardiovascular and musculoskeletal disorders. Nakamura *et al.* (1998) find a positive correlation between overtime work and change in BMI and waist circumference (over three years). Caruso *et al.* (2004) provide an overview of the literature on overtime and extended work shifts and state that (amongst many other findings) the literature they studied relates overtime work to poorer perceived general health, more illnesses, or increased mortality. Other findings also relate overtime to increased injury rates, weight gain and adverse health behavior such as smoking (one of two studies) and drinking (two of three studies). They also report a connection between extended shifts and increased fatigue. Kawada and Ooya (2005) find a relationship between workload and health complaints based on a sample of 109 male workers at a car manufacturer workplace in Japan. Kawakami and Haratani (1999) provide a literature overview and find that for Japanese male workers 50 or more hours of work are associated with an increase in psychological distress but also state that the magnitude (of the association) appears to be weak. By using a meta-analytic approach Sparks *et al.* (1997) show that long working hours are correlated with overall, physiological and psychological health symptoms. They also discuss the idea of different types of people having different energy and effort levels. Beckers *et al.* (2004) find no association between overtime hours and fatigue and also state that overtime workers seem to be happy and work in attractive jobs. Golden and Wiens-Tuers (2006) even find that overtime is associated with better (self-rated) health and main satisfaction in life coming from work but also with feeling used up at the end of the day, more work-family interference and more stress.

Having reverse causality in mind it is important to understand the effect of health (and health shocks) on hours worked. Cai *et al.* (2008) conclude that lower health results in a reduction of working hours and that health shocks additionally reduce working time. An article of the DGB, the *Confederation of German Trade Unions* (2010), discussed recent findings with respect to unemployment and mental and physical health problems. One of their examples is that the group with the highest amount of antidepressant prescriptions is unemployed women. Unemployment is not considered here but underemployment could have similar effects as for example Starrin *et al.* (1990) find that the level of employment is an important contributory factor of suicide rates (besides overtime work).

Taken all together, the literature described here differs largely in sample type and size, the definition of overtime or high workload, well-being measures, the underlying time period and frequency of excessive work, the evaluation method and the voluntariness of participation, the econometric approach and the inclusion of covariates. All this gives good reason to tackle the overtime and well-being nexus in a more general way without imposing too many restrictions but with a large sample size and the inclusion of the most important moderating factors.

2.3 Details on the differences between overtime measures

There are several potential overtime measures available in the dataset (GSOEP – details follow later on) which are discussed here briefly.

The first question explicitly asking for overtime is very general: *"Do you work overtime?"* In this case it is not clear what the answer really means as it could be "yes" even if the individual only works some overtime, e.g. a couple of hours every three months. On the other hand, some people could think that the few overtime hours per week are not worth mentioning and answer "no". It occurs that many respondents tend to state that they are overtime workers even though their actual weekly working time equals the contracted time and only a few classify themselves as non-overtime workers whilst working more than actually contracted. Using only this general question in order to split the sample into groups of regular and overtime workers is therefore very imprecise and we cannot learn much from the subsequent comparison of the groups with respect to well-being.

The second question that is directly related to overtime is:

"How was your situation with regards to overtime last month? Did you work overtime? If yes, how many hours?"

Here, the question refers to a specific time period and, additionally, it is asked for the exact amount of hours. On the one hand, some of those who work overtime only every other month are categorized as non-overtime workers if they did not work overtime in the last month. On the other hand, the main disadvantage is that the question only refers to the last month and it could be true that individuals truthfully answer "no" because it was an exceptionally calm month or in other words a very rare event. This means that those who usually work overtime except for the last month are in the reference group. It can also happen that someone only works overtime once a year and answers "yes" because it was in the last month. If the sample is split into overtime and non-overtime workers we could have a bias towards contemporaneous differences and could also miss those who would contribute most to the differences as they constantly work overtime (besides in the last month). Again, there is not much that we can learn about the differences in well-being between overtime and non-overtime workers.

In this chapter I will focus on two other measures of overtime (and present results for the first very general overtime measure as a reference). The first one is close to the general overtime question but more precise. It is derived from the average of actual weekly working hours and the contracted weekly hours. A positive difference is defined as overtime, a negative difference is real underemployment which rarely occurs. The main advantage is that it is more likely that overtime workers are categorized correctly as we would only expect those who constantly work extra hours to have a positive difference. Additionally, those are also the ones of major interest as long-term effects are more important than current effects. As a next step it is important to think about why total working time should not be used. For example: Think about a scenario with two individuals, both working 40 hours per week. One has agreed on working 40 hours and the other one has agreed on working 30 hours. In this case only the second individual is working overtime. If we simply focused on hours worked (as this is often done) we would treat them the same way (or even exclude the second one). Is this plausible? Not necessarily. We have to think of how overtime can influence well-being. One typical example is stress due to time pressure which leads to other symptoms such as fatigue. Clearly, individuals can differ in terms of their intrinsic stress-bearing capabilities and therefore looking at hours worked does not reveal much about differences in well-being. An analogy is running: Two runners start together but the first one is a long distance runner and he would choose to run 15 miles but runner 2 is a short distance runner and would want to run only 10 miles. After 10 miles we will see that runner 2 gets exhausted and every additional mile becomes increasingly painful for him (and he might stop after 15 miles) while runner 1 is still doing well. After 15 miles runner 1 will also start getting tired and exhausted and then every extra mile becomes painful for him as well, so he will also stop at some point. If we ask them how they feel after they have stopped we will probably get similar answers even though runner 1 did more miles. Had we asked both after 15 miles we would have gotten different answers. If we think of employees trying to find a job that (perfectly) matches their preferred workload, 40 hours are no problem for the first person but maybe for the second one as he deviates from his preferred hours. This approach should reduce the risk that results are driven by unobserved underlying individual characteristics such as stress-bearing capabilities (which is reflected in their individually preferred workload). Thinking of our runners this would be the case where we ask them how they feel when they stop. Both have deviated from their ideal distance and are therefore tired and exhausted. Furthermore, consider the following example: Now person 1 works 50 hours and person 2 works 40 hours – so now both of them work overtime. If both have a job that reflects their preferred workload they are both treated in the same way while looking at total hours worked would make a difference here. Again, if we think that a deviation from preferred hours matters, this should not necessarily be the case. This is also analogous to asking the runners once they stop and not after a certain amount of miles.

Medical and psychological papers typically focus on excessive working hours which is the same as evaluating the total distance of the runners instead of the deviation from their preferred distance. The problem with this approach is the underlying assumption. Does your contracted working time really reflect your truly desired hours? There are good arguments why this could not be the case. One example is that contracts are typically standardized and many employees either work full time (approx. 40 hours) or part time (approx. 20 hours). So it might be difficult to find an employer that offers 30 hours just because you want to. This is especially true for challenging jobs with complex tasks. Also, some workers may even want to work more in order to earn more money or because they like their job so much. Those would be underemployed and that is where an alternative overtime measure comes in.

The next overtime measure I present here is derived from actual hours of work and the desired workload. Respondents are explicitly asked for their desired working hours:

"If you could choose your own number of working hours, taking into account that your income would change according to the number of hours: How many hours would you want to work?"

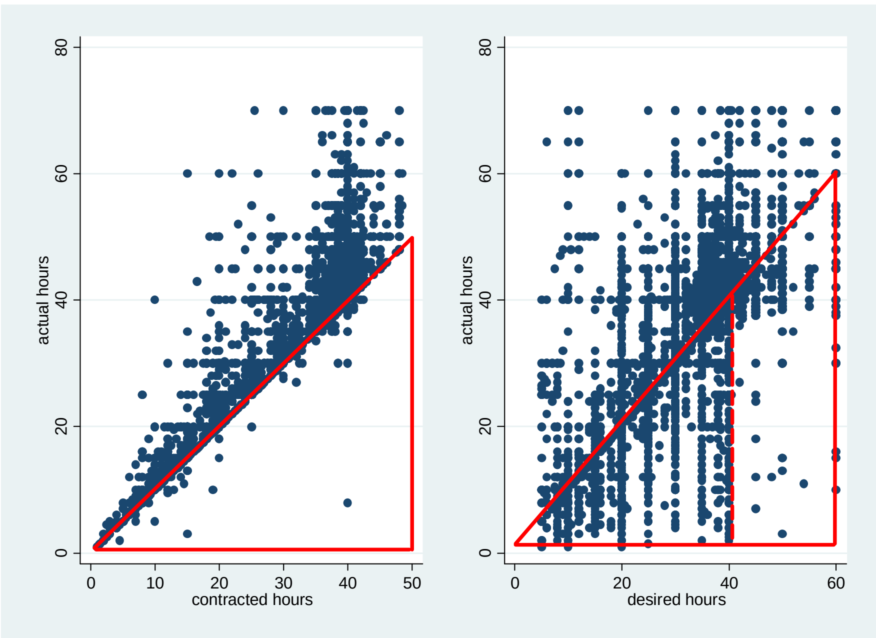
The advantage is that self-stated preferences should be closer to the truly desired working hours and therefore rule out voluntary overtime. In terms of identification, the question has to be understood correctly. People are asked to choose their amount of working hours so they should not state that they want to work more if they are in fact unable to do so because of e.g. mental or physical health problems. Additionally, the question is embedded in questions asking about the current occupation and therefore it asks for desired hours of work in the same company and job. All in all this question aims at the truly desired personal workload, considering all that comes with it. If a person has already reduced working time down to the natural bound (contracted hours) he might still want to work less to reach his desired working time and restore well-being. In the long run, this person will probably try to find a new job but in the meantime the additional hours cause further problems as the individual is kept from restoring well-being by the gap between actual and desired hours. So the gap either causes the problems directly by creating them (someone works more hours than would be good for him or less than he wants to) or indirectly by preventing a necessary adjustment of working hours. The underlying idea of this overtime measure is the same as before: People know their desired workload and deviating from it is associated with different levels of well-being compared to those who work their desired or contracted amount of time.

To summarize the idea of the overtime measures described above: People that are not working desired or contracted hours are worse off (e.g. by not achieving work-life balance). Getting back to the favored schedule solves the problems and people will become better off.

If you feel overworked and reduce hours you are back on track and start to feel better again. It is not necessary that the problems are entirely solved as these people move into the base-group and all that is required is that moving towards contracted or desired hours is on average better than not going back. In terms of contracted hours, problems are caused when you are pushed away from your contracted working time and in terms of desired hours, they are caused by either being pushed away from the desired workload in the one way or the other (you have to work more or less, i.e. actual hours are being moved) or by being kept away from working the favored amount of time (you cannot reduce your hours, i.e. desired hours are being moved further than a reduction of actual hours can compensate because of contracted hours being a lower bound). Of course combinations can occur, e.g. someone contracts more than she initially wants to and then she additionally has to work even more than contracted.

The following illustrations (with the yet unrestricted data) help to better understand the differences between the last two overtime measures. The left plot in figure 2.1 refers to the penultimate overtime measure and plots actual against contracted working hours and everybody on the diagonal is not working overtime.

Figure 2.1: Actual vs. contracted and desired working hours

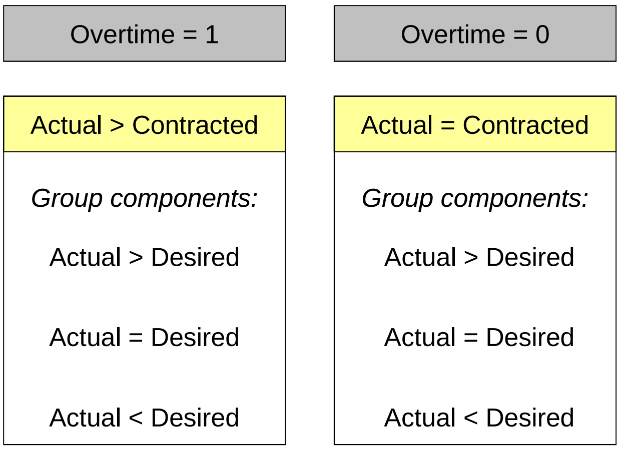


It is interesting to see that the lower triangle is almost empty: Working less than contracted is extremely uncommon. This is good news as those could be persons who work less because of low well-being (e.g. bad health). All observations above the diagonal are working overtime, which is a substantial part of the sample.

The last overtime measure is shown in the right plot of figure 2.1 where actual working hours are plotted against desired working hours. Everybody who is not working overtime is represented by the main diagonal and overtime or positive deviations from contracted hours are above. Unlike in the left plot of figure 2.1 the lower triangle is not empty so there are workers who would want to work more than they do. As the dashed line implies, these are mainly employees who work less than 40 hours per week.

The next step is to think about the changes when we use the one measure or the other. Who is considered to be an overtime worker and who is not if we consider deviations from the contracted hours and who is in which group if we consider deviations from the desired workload? Figure 2.2 shows what a sample would look like if we defined overtime as a dummy variable which equals 1 if we observe a positive difference between actual and contracted hours and 0 if both are equal (base group). Accordingly, three cases per state have to be considered.

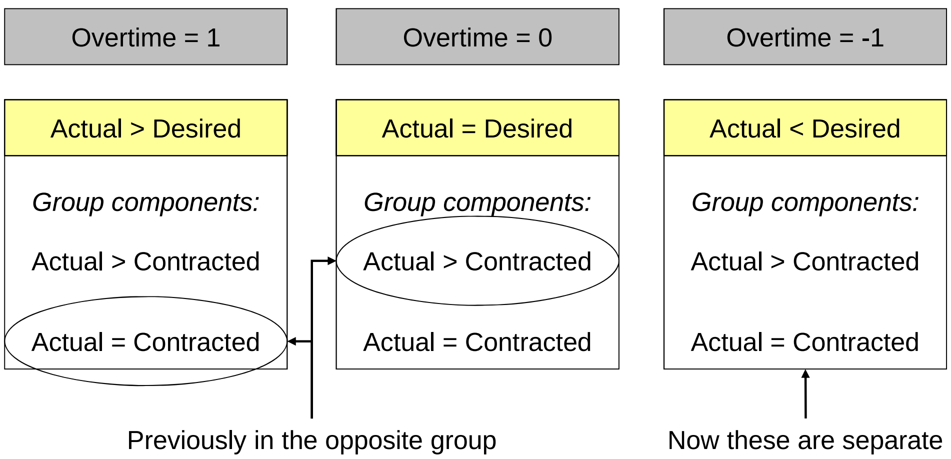
Figure 2.2: Overtime and non-overtime workers (contracted working time)



If in fact desired hours were the true reference point then the overtime group would also contain those who work their preferred schedule and the non-overtime group would also contain those who work more than they actually want to. Those who work less than desired add additional complication as they might increase or decrease differences in e.g. well-being.

For a better understanding of this discussion figure 2.3 shows a hypothetical sample if desired and actual hours were compared. A first look immediately reveals the difference to the overtime measure with contracted hours: Now it is necessary to divide the population into three groups. The first group works overtime, the second one works the amount of hours desired and people in the last group work less than they want to. The advantage is that those in the special case of working less than desired are separated. We can also see that in comparison to considering actual and contracted working hours, part of the overtime and the non-overtime workers switch groups. If the other process was true the differences between groups would in the best case be too small because a fraction of overtime and non-overtime workers would be in the wrong group.

Figure 2.3: Overtime and non-overtime workers (desired working time)



Note that comparing columns in one model is always the same as comparing rows in the other model.

2.4 Data

The data used here is provided by the DIW (*German Institute for Economic Research*) – mainly by the 2008 wave of its GSOEP (German Socio-Economic Panel) database.[3] The GSOEP started in 1984 and is a large and representative study of the German population that collects data on a wide range of topics that is relevant to current research. Most importantly, the panel contains individual and household-specific data on income, job characteristics and different measures of well-being.

This analysis is restricted to workers between the ages of 18 and 45 because in this group chronic illnesses are less likely and because there might be different attitudes in general and towards work in particular for those who grew up after World War II and in the times of the German economic miracle. Plots of different outcome variables and demographic factors could show these irregular patterns over the age groups. Self-employed persons are also excluded as they typically do not work on a fixed schedule and therefore would drop out when looking at workers with overtime defined as a deviation from contracted hours. As we have seen previously there exists almost no real underemployment, i.e. people who work less than they are supposed to are very rare and are therefore excluded. The group is too small to think of a meaningful selection problem and also too small to derive plausible results from. Additionally, cells with few observations are excluded (e.g. a certain industry or those whose marital status is widowed) in order to avoid statistical problems when computing cluster-robust standard errors. Besides this, some potentially influential outliers have been removed. These are observations in the highest percentile with respect to income, height and weight (BMI) as well as actual and contracted hours. Finally, observations in the lowest and highest percentile of desired hours are also excluded from the data.

As the goal is to find differences in well-being among overtime workers and non-overtime workers the dependent variables of main interest are self-rated health status (SRHS), satisfaction with health, life (at present, in a year and in five years) and work. Besides the overtime measure the obvious control variables gender, age, marital status, nationality,

[3] Waves 2004 to 2007 are added later as a robustness check but the main discussion refers to the 2008 wave.

logarithm of net income, years of education and a dummy variable for the presence of children under the age of 16 are included as explanatory variables. In terms of job related variables the model controls for the time an individual has spent with the firm, whether the workplace is located in Eastern Germany, several industry types and the number of employees in the firm. Additionally, the levels of contracted, actual and desired hours as well as indicators for irregular daily working hours and irregular starting times capture cross-job characteristics. Moreover, it is controlled for the occupational status (white and blue-collar, public servant or others, such as military and civilian service or apprenticeship, collectively called "training"). Another important determinant of individual well-being is the personal lifestyle and this could plausibly be correlated with overtime work. Consequently, BMI and indicators for unhealthy food (1 if someone does not care about healthy food at all), frequent sports (1 if someone works out every day), smoking and drinking behavior (regular beer or wine consumption) try to mimic it. Health status itself is of course included in models assessing satisfaction with life and work. Classical health stock models (e.g. Grossman (1972) and Cropper (1977)) in economics capture the idea of lifestyle influencing well-being: In these models a change in the health stock is determined by an investment in health and a constant decay of health over time. This was also discussed as an important aspect in the literature section. The models considered in this work allow the investment in health to be positive or negative. Moreover, dynamics are not modeled here but it is implicitly assumed that a certain kind of (regular) behavior is already reflected in well-being.

As an example for investments in health figure 2.4 shows the differences in awareness for healthy food for different age groups and the weight of the particular group in the sample. It can be seen that among the youngest, those who work more than contracted are also those who state less often that they do not care about healthy nutrition at all.

As another example consumption of alcoholic beverages is considered. Looking at drinking behavior, figure 2.5 shows that those who state that they drink wine regularly are typically also those who work overtime.

The interesting point is really the proportion stating a regular consumption of alcohol and not the beverage itself. The intention of the figures is merely to show that there are (ex ante) differences in behavior that can plausibly affect well-being and that it is therefore important to control for lifestyle in the models.

Figure 2.4: Unhealthy food by overtime status and age

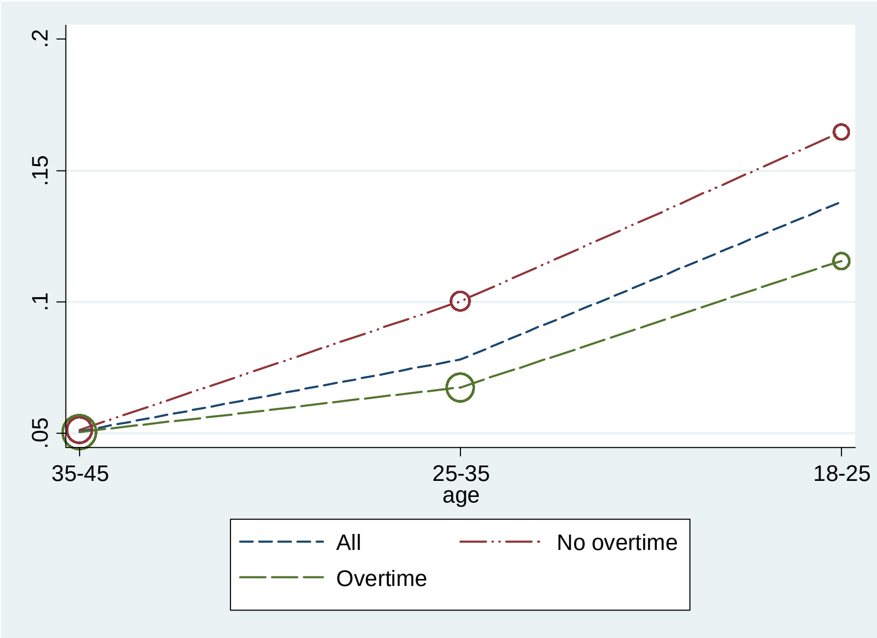
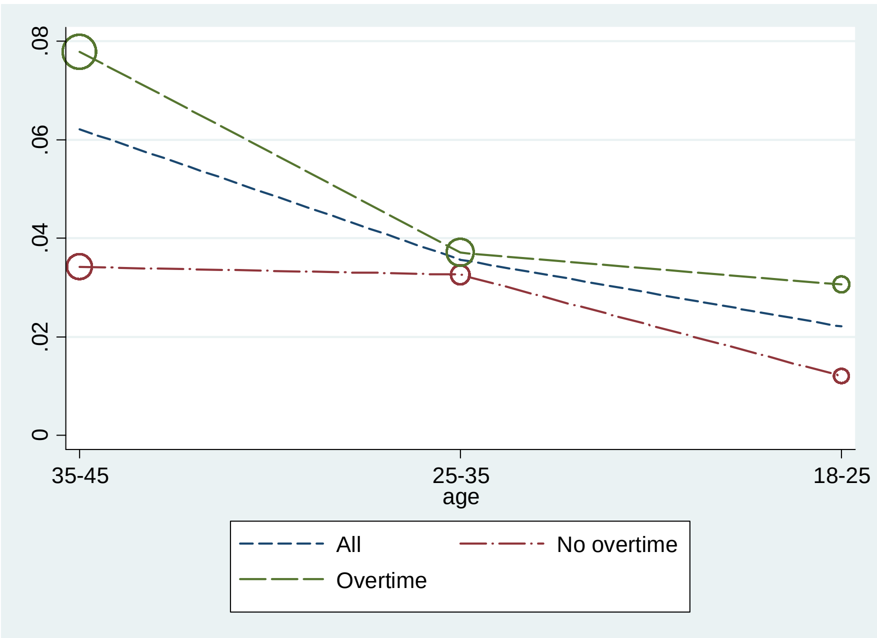


Figure 2.5: Regular wine consumption by overtime status and age



In order to characterize the overtime and non-overtime groups table 2.1 presents descriptive statistics (sample means and standard deviations) of the explanatory variables (except for industry, firm size and health variables – the latter can be seen in the next table).

Table 2.1: Means and standard deviations of the explanatory variables

	Actual vs. Desired			Actual vs. Contracted	
	Overtime (N=2350)	No overtime (N=943)	Underemployment (N=676)	Overtime (N=2531)	No overtime (N=1438)
Female (B)	0.454	0.478	0.595	0.447	0.548
	(0.498)	*(0.500)*	*(0.491)*	*(0.497)*	*(0.498)*
Age	35.016	34.925	34.664	35.157	34.542
	(7.232)	*(7.613)*	*(7.401)*	*(7.081)*	*(7.793)*
Years of education / training	12.933	12.128	12.196	13.001	11.938
	(2.648)	*(2.488)*	*(2.321)*	*(2.648)*	*(2.322)*
Job tenure	7.707	7.645	6.609	7.745	7.084
	(6.700)	*(6.882)*	*(6.590)*	*(6.758)*	*(6.680)*
Job in Eastern Germany (B)	0.196	0.213	0.195	0.199	0.202
	(0.397)	*(0.410)*	*(0.397)*	*(0.399)*	*(0.401)*
Unhealthy food (B)	0.064	0.083	0.084	0.064	0.086
	(0.245)	*(0.276)*	*(0.278)*	*(0.245)*	*(0.280)*
Sports (B)	0.069	0.056	0.072	0.066	0.067
	(0.253)	*(0.230)*	*(0.259)*	*(0.248)*	*(0.251)*
Smoker (B)	0.339	0.347	0.399	0.353	0.347
	(0.473)	*(0.476)*	*(0.490)*	*(0.478)*	*(0.476)*
Beer drinker (B)	0.136	0.135	0.114	0.136	0.125
	(0.343)	*(0.342)*	*(0.318)*	*(0.343)*	*(0.331)*
Wine drinker (B)	0.057	0.035	0.036	0.058	0.030
	(0.231)	*(0.184)*	*(0.185)*	*(0.234)*	*(0.170)*
BMI	24.965	25.080	25.189	25.056	24.985
	(3.885)	*(4.101)*	*(4.278)*	*(3.904)*	*(4.181)*
German nationality (B)	0.941	0.923	0.935	0.944	0.920
	(0.236)	*(0.267)*	*(0.247)*	*(0.229)*	*(0.271)*
Children under age 16 (B)	0.442	0.502	0.561	0.457	0.510
	(0.497)	*(0.500)*	*(0.497)*	*(0.498)*	*(0.500)*
Log of net personal income	7.266	6.989	6.690	7.256	6.830
	(0.566)	*(0.660)*	*(0.767)*	*(0.586)*	*(0.705)*
Married (B)	0.495	0.520	0.504	0.497	0.513
	(0.500)	*(0.500)*	*(0.500)*	*(0.500)*	*(0.500)*
Single (B)	0.423	0.398	0.368	0.412	0.398
	(0.494)	*(0.490)*	*(0.483)*	*(0.492)*	*(0.490)*
Divorced (B)	0.065	0.066	0.098	0.070	0.073
	(0.247)	*(0.248)*	*(0.297)*	*(0.254)*	*(0.260)*
Blue-collar worker (B)	0.254	0.331	0.330	0.248	0.352
	(0.436)	*(0.471)*	*(0.471)*	*(0.432)*	*(0.478)*
Public servant (B)	0.068	0.058	0.044	0.067	0.051
	(0.251)	*(0.234)*	*(0.206)*	*(0.250)*	*(0.221)*
Training (B)	0.059	0.086	0.053	0.046	0.097
	(0.236)	*(0.280)*	*(0.225)*	*(0.210)*	*(0.296)*
Irregular hours of work (B)	0.108	0.086	0.149	0.117	0.097
	(0.311)	*(0.280)*	*(0.357)*	*(0.321)*	*(0.297)*
Varying starting time (B)	0.272	0.346	0.339	0.273	0.350
	(0.445)	*(0.476)*	*(0.474)*	*(0.446)*	*(0.477)*
Actual hours of work	42.422	35.977	28.665	41.898	32.651
	(8.386)	*(9.491)*	*(12.170)*	*(9.427)*	*(10.363)*
Contracted hours of work	37.254	34.445	26.865	36.048	32.651
	(6.120)	*(8.684)*	*(11.466)*	*(7.431)*	*(10.363)*
Desired hours of work	34.799	35.977	36.217	36.508	33.230
	(7.616)	*(9.491)*	*(10.699)*	*(8.075)*	*(9.331)*

Note: All values with respect to the largest common sample. Standard deviations of the variables given in parentheses. (B) indicates a binary variable.

Table 2.2 does the same for the dependent variables (descriptive statistics without differentiation by overtime status can be found in the appendix: Tables 2.10 and 2.11).

As one would expect, the overtime group of the first overtime measure has more years of education than both other groups. Besides that, the underemployed group consists of more women which is also not surprising as they are more likely to reduce hours worked because they have to and not because they want to (e.g. due to childcare because this group also has children at home more often). Comparing the groups of the second overtime measure reveals that more men work overtime and that the average years of education are also higher in this group. Interestingly, the groups that deviate from desired or contracted hours are not necessarily those with the unhealthiest lifestyle. The underemployed group for example has the highest mean for frequent sports activities and the overtime workers care more about healthy food than those not working overtime. With respect to the dependent variables the underemployed group is always worse off compared to the others (satisfaction with work as the exception) and for both measures overtime workers are worse off than non-overtime workers whereas the differences are smaller for the second measure.

Table 2.2: Means and standard deviations of the dependent variables

	Actual vs. Desired			Actual vs. Contracted	
	Overtime (N=2313)	No overtime (N=920)	Underemployment (N=655)	Overtime (N=2499)	No overtime (N=1389)
Health status	2.292	2.174	2.405	2.305	2.243
	(0.795)	*(0.768)*	*(0.871)*	*(0.805)*	*(0.804)*
Satisfaction with health	7.224	7.459	6.934	7.166	7.346
	(1.811)	*(1.772)*	*(1.979)*	*(1.827)*	*(1.853)*
Satisfaction with life (now)	7.200	7.335	6.979	7.178	7.224
	(1.465)	*(1.456)*	*(1.605)*	*(1.473)*	*(1.524)*
Satisfaction with life (1y)	7.501	7.622	7.348	7.501	7.510
	(1.444)	*(1.418)*	*(1.593)*	*(1.448)*	*(1.499)*
Satisfaction with life (5y)	7.641	7.718	7.518	7.638	7.639
	(1.648)	*(1.588)*	*(1.809)*	*(1.665)*	*(1.660)*
Satisfaction with work	6.945	7.325	7.015	6.975	7.176
	(1.936)	*(1.762)*	*(1.929)*	*(1.904)*	*(1.888)*

Note: All values with respect to the largest common sample. Standard deviations of the variables given in parentheses.

2.5 Methodological aspects

Measurement of well-being in large surveys is usually achieved by presenting different types of scales to the respondent which creates ordinal variables. On the one hand, this approach is probably the only feasible way to collect data from a large population but on the other hand, this has an immediate impact on the variety of econometric models that can be used. To model the ordered outcomes ordered probit models are used here.

As described above, the groups of overtime workers and non-overtime workers (and underemployed workers in the case of desired hours) are separated by deviations from contracted or desired hours. The levels of each variable are also included as controls. Even though people might know quite precisely how much working time is contracted – the reason why variation in the amount of overtime is not exploited is due to potentially severe measurement error problems in actual and desired hours. Interviewees are asked to state their average working time per week which is a very difficult, if not impossible, task. Desired hours are also difficult to state as a reduction or increase in income has to be considered simultaneously and because one has to think really hard and take a lot of things into account to understand what a desirable and achievable amount of work is. Nevertheless, the assumption that people know whether they work more than contracted on average or if their desired hours are lower than their actual hours is much less problematic. Of course people may know whether they work a lot of overtime on average or not but this would still not justify the calculation of the marginal effect of one additional hour of overtime.

2.6 Results

First of all, the question whether overtime work is associated with differences in health in comparison to those who do not work overtime is analyzed. The typical health measures are available in the GSOEP: (Self-rated) health status (SRHS) and satisfaction with health. For the sake of readability only the two main measures of overtime (deviations from desired and actual hours – both measures equal 1 if there is a positive deviation from the respective reference point and 0 otherwise) are discussed in detail for health. Subsequently the presentation will be briefer but also with information on a third overtime measure, namely overtime in general (equals 1 if someone answers yes to the question *“Do you work overtime?”*). As discussed previously, the latter is not an ideal measure but as it is used frequently it might be interesting to see it. Table 2.3 shows the results.

Table 2.3: Health and overtime work

		Parameter estimates			
	Sample means	Health status	Health status	Satisfaction with health	Satisfaction with health
Actual vs. Desired (B)	0.592	0.196 ***	na	-0.138 ***	na
		(0.048)	na	*(0.045)*	na
Actual vs. Contracted (B)	0.638	na	0.162 ***	na	-0.163 ***
		na	*(0.047)*	na	*(0.043)*
Female (B)	0.483	0.181 ***	0.178 ***	-0.130 ***	-0.125 ***
		(0.046)	*(0.046)*	*(0.042)*	*(0.042)*
Age	34.934	0.030 ***	0.029 ***	-0.023 ***	-0.022 ***
		(0.004)	*(0.004)*	*(0.003)*	*(0.004)*
Years of education / training	12.616	-0.022 **	-0.021 **	0.004	0.003
		(0.009)	*(0.009)*	*(0.008)*	*(0.008)*
Job tenure	7.505	-0.001	-0.001	-0.001	-0.001
		(0.003)	*(0.003)*	*(0.003)*	*(0.003)*
Job in Eastern Germany (B)	0.200	0.012	0.011	-0.088 **	-0.088 **
		(0.046)	*(0.046)*	*(0.042)*	*(0.042)*
Unhealthy food (B)	0.072	0.087	0.094	-0.027	-0.036
		(0.070)	*(0.070)*	*(0.068)*	*(0.068)*
Sports (B)	0.067	-0.215 ***	-0.199 **	0.185 ***	0.171 ***
		(0.080)	*(0.080)*	*(0.063)*	*(0.064)*
Smoker (B)	0.351	0.118 ***	0.121 ***	-0.115 ***	-0.118 ***
		(0.039)	*(0.039)*	*(0.037)*	*(0.037)*
Beer drinker (B)	0.132	0.016	0.017	-0.054	-0.055
		(0.051)	*(0.051)*	*(0.051)*	*(0.050)*
Wine drinker (B)	0.048	-0.089	-0.090	0.086	0.090
		(0.089)	*(0.089)*	*(0.078)*	*(0.079)*
BMI	25.031	0.038 ***	0.039 ***	-0.031 ***	-0.032 ***
		(0.005)	*(0.005)*	*(0.004)*	*(0.004)*
German nationality (B)	0.936	0.111	0.111	-0.167 **	-0.165 **
		(0.078)	*(0.077)*	*(0.074)*	*(0.074)*
Children under age 16 (B)	0.476	-0.064	-0.065	0.043	0.044
		(0.044)	*(0.044)*	*(0.042)*	*(0.042)*
Log of net personal income	7.102	-0.106 **	-0.115 **	0.092 *	0.103 **
		(0.051)	*(0.051)*	*(0.048)*	*(0.047)*
Married (B)	0.502	0.067	0.053	0.135	0.151
		(0.121)	*(0.123)*	*(0.111)*	*(0.112)*
Single (B)	0.407	0.056	0.043	0.102	0.115
		(0.124)	*(0.125)*	*(0.113)*	*(0.114)*
Divorced (B)	0.071	0.115	0.115	0.072	0.073
		(0.135)	*(0.136)*	*(0.125)*	*(0.126)*
Blue-collar worker (B)	0.285	0.060	0.074	-0.090 *	-0.104 **
		(0.049)	*(0.049)*	*(0.047)*	*(0.047)*
Public servant (B)	0.061	0.052	0.046	-0.059	-0.055
		(0.086)	*(0.086)*	*(0.074)*	*(0.075)*
Training (B)	0.064	-0.065	-0.068	0.061	0.062
		(0.100)	*(0.099)*	*(0.096)*	*(0.095)*
Irregular hours of work (B)	0.110	0.096 *	0.104 *	-0.066	-0.071
		(0.058)	*(0.058)*	*(0.056)*	*(0.056)*
Varying starting time (B)	0.301	-0.069 *	-0.073 *	0.077 **	0.079 **
		(0.041)	*(0.041)*	*(0.038)*	*(0.039)*
Actual hours of work	38.548	0.001	-0.005	-0.005	0.003
		(0.004)	*(0.005)*	*(0.004)*	*(0.004)*
Contracted hours of work	34.817	0.006	0.007	-0.002	-0.004
		(0.005)	*(0.006)*	*(0.005)*	*(0.005)*
Desired hours of work	35.320	-0.001	0.001	0.003	-0.001
		(0.004)	*(0.003)*	*(0.003)*	*(0.003)*
Number of observations		3969	3969	3969	3969
Log (pseudo)likelihood		-4468.976	-4479.876	-7479.734	-7487.987
Pseudo R-squared		0.037	0.035	0.017	0.016

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses. (B) indicates a binary regressor. Industry and firm size indicators included in all regressions but estimates not reported. Base category for marital status variables is "married but separated" and "white-collar" for occupational status.

Note that health status is measured on a 5-point scale (1: Very good – 5: Bad) and satisfaction with health is measured on an 11-point scale (0: Completely dissatisfied – 10: Completely satisfied) and the different signs simply reflect the different coding. Most importantly, it can be seen that both overtime measures produce similar results for both health measures in terms of significance and magnitude. The signs are as expected and state that overtime work is generally associated with worse SRHS and lower satisfaction with health.

The parameter estimates of the indicator for working less than desired are not reported here but are substantially larger than those for overtime work and also highly significant. The typical findings in health economics are confirmed:[4] Higher education and higher income are both associated with a better SRHS and income is also associated with a higher satisfaction with health. The lifestyle variables have signs as expected and smoking as well as higher BMI are associated with lower (satisfaction with) health and frequent sports have the opposite effect. The worse SRHS and lower satisfaction with health of women and the decline in these variables with age is also noteworthy. Workers with a varying starting time seem to feel better. Maybe they can compensate adverse effects with more flexible schedules.

Table 2.4: Comparison of parameter estimates (health and overtime work)

	Sample means	N	Parameter estimates	
			Health status	Satisfaction with health
Model 1:				
Overtime in general (B)	0.780	3969	0.130 ***	-0.122 ***
			(0.049)	*(0.046)*
Model 2:				
Actual vs. Desired (B)	0.592	3969	0.196 ***	-0.138 ***
			(0.048)	*(0.045)*
Model 3:				
Actual vs. Contracted (B)	0.638	3969	0.162 ***	-0.163 ***
			(0.047)	*(0.043)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses. (B) indicates a binary regressor.

Table 2.4 adds the parameter estimate for overtime in general. In comparison to the overtime measures used before, using overtime in general returns smaller parameter

[4] The causality debate related to these variables is not part of this work.

estimates but the conclusion remains unchanged. It can also be seen that there are quite large changes in the groups that are considered as overtime workers: While 78% are considered as being exposed to overtime when overtime in general is used this is only true for roughly 60% when the other two overtime measures are applied.

Besides health, another aspect of well-being is satisfaction with life. The latter is asked by the GSOEP with respect to current satisfaction with life and anticipated satisfaction in one and in five years. All three variables are measured on an 11-point scale where 0 means completely dissatisfied and 10 means completely satisfied. As the personal health status is very likely to affect one's evaluation of life it is included in the following models by two dummy variables where the first indicator equals one if SRHS is very good or good and the second indicator represents poor or bad, i.e. a satisfactory SRHS is the base category. Table 2.5 compares the estimated parameters of all three overtime measures.

Looking at (current) satisfaction with life the overtime coefficients are close to each other and all of them are highly significant. The situation changes as the discrepancies increase with the reference period. When people are asked how they think their satisfaction with life is going to be in five years, the very general first overtime measure is significant whereas a deviation from desired hours is only marginally significant and a deviation from contracted hours does not result in significant differences between the groups. The coefficient for working less than desired (not reported) is about the same size as the overtime coefficient and has the same sign but it is only significant at the 5% level and insignificant when anticipated satisfaction in five years is considered.

Table 2.5: Comparison of parameter estimates (satisfaction with life and overtime work)

	Sample means	N	Parameter estimates: Satisfaction with life (now)	Satisfaction with life (1y)	Satisfaction with life (5y)
Model 1:					
Overtime in general (B)	0.782	3915	-0.124 ***	-0.122 ***	-0.106 **
			(0.047)	*(0.047)*	*(0.046)*
Model 2:					
Actual vs. Desired (B)	0.594	3915	-0.134 ***	-0.150 ***	-0.083 *
			(0.049)	*(0.048)*	*(0.048)*
Model 3:					
Actual vs. Contracted (B)	0.640	3915	-0.130 ***	-0.108 **	-0.073
			(0.045)	*(0.045)*	*(0.045)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses.
(B) indicates a binary regressor.

As the last part of well-being, satisfaction with work (11-point scale as before) is additionally analyzed. Table 2.6 shows the results which differ largely. Overtime in general yields no significant differences between the satisfaction of overtime and non-overtime workers, the two other measures do – although the coefficient of working more than contracted is only significant at the 5% level. In the case of working fewer hours than wanted, the parameter estimate is considerably smaller and significant at the 10% level.

Table 2.6: Comparison of parameter estimates (satisfaction with work and overtime work)

	Sample means	N	Satisfaction with work
Model 1:			
Overtime in general (B)	0.783	3942	-0.075
			(0.047)
Model 2:			
Actual vs. Desired (B)	0.594	3942	-0.173 ***
			(0.047)
Model 3:			
Actual vs. Contracted (B)	0.641	3942	-0.103 **
			(0.044)

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses. (B) indicates a binary regressor.

2.7 Discussion

As shown in the previous section, there are differences between overtime and non-overtime workers with respect to health, satisfaction with life and satisfaction with work. The estimated parameters (naturally) depend on the overtime measure used but taken all together, the same picture arises no matter what kind of overtime model is used. The reason for this is that the variables share a large degree of the observations and thereby carry a lot of the same information even though the fraction of employees that is defined as overtime-workers changes. Remember that comparing columns in one measure means comparing rows in the other measure. Table 2.7 shows the fraction of observations that does not change groups, no matter if contracted or desired hours are considered. In any case over two-thirds of the observations are not affected by changing the overtime measure.

Table 2.7: Fraction of observations that does not change groups

	N	Not changing group	%
Health	3969	2676	67%
Satisfaction with life	3915	2647	68%
Satisfaction with work	3942	2663	68%

Simultaneity could be a problem and needs to be discussed. The findings of Cai *et al.* (2008) suggest that health events lead to a reduction in hours worked and that lower health status is associated with less working hours. This is important when deviations from desired or contracted hours are considered. Even though this means that health affects working hours, it does not mean that the differences found here are too large or that bad health could not have been a result of overtime work. More work might lead to a decrease in health and afterwards the decreased health leads to a decrease in working hours. The important message is that the sicker people are, the less likely they become or remain overtime workers and that therefore bad health does not cause overtime work. The data also showed that real underemployment (working less than contracted) rarely occurs and that therefore working no overtime is the lower bound of actual hours. This would mean that well-being in the base group is rather too low than too high even if those with the worst health symptoms would drop out of the labor force.

It was suggested by Beckers *et al.* (2004) that overtime workers are happier than others. Even though this finding is contrary to the working hypothesis here and in other papers in the literature and even though these studies differ largely, it still means that the negative effect found in this work could be biased – but it would rather be too small than too large (in absolute terms) if indeed happier people would choose to work overtime more often. These potential simultaneity issues are another reason not to rely on the marginal effect of overtime hours but rather focus on a simple comparison of groups which is not or at least less affected.

As said previously, the group of underemployed workers was also found to be worse off than those working desired hours. This is plausible as underemployment could also cause personal distress and therefore lead to lower well-being. The underlying assumption here is that people take the overall situation into account when they choose their desired amount of work. Additionally, desired hours are more credible as a reference point because imperfect labor market matching could lead to contracts that do not reflect the desired workload. Those working more than desired are worse off because they are either pushed away from their desired amount of work or kept away from it if desired hours are adjusted disproportionately.

Now one could think that especially those in jobs that are associated with only few hours of work drive these results. As a robustness check I restrict the sample to those who have at least 35 hours of contracted working time which is above the average of the non-overtime group in the full sample (i.e. the sample is reduced by roughly 24%) and present results for the health and satisfaction with life and work variables.[5] If the previous results had been driven by those who work in jobs with few contracted hours, the differences between overtime and non-overtime workers would disappear. Table 2.8 gives the results, where each row reports the specific overtime coefficients (each one from a separate regression) and the corresponding dependent variable is shown on the left.

Table 2.8: Robustness check of overtime measures (restricted sample)

		Model type		
	N	Overtime in general	Actual vs. Desired	Actual vs. Contracted
Health status	3002	0.170 ***	0.246 ***	0.173 ***
		(0.059)	*(0.058)*	*(0.054)*
Satisfaction with health	3002	-0.170 ***	-0.192 ***	-0.209 ***
		(0.055)	*(0.053)*	*(0.051)*
Satisfaction with life (now)	2969	-0.139 **	-0.154 ***	-0.147 ***
		(0.056)	*(0.058)*	*(0.052)*
Satisfaction with life (1y)	2969	-0.140 **	-0.160 ***	-0.139 ***
		(0.056)	*(0.057)*	*(0.052)*
Satisfaction with life (5y)	2969	-0.130 **	-0.086	-0.081
		(0.056)	*(0.057)*	*(0.053)*
Satisfaction with work	2987	-0.052	-0.120 **	-0.090 *
		(0.057)	*(0.054)*	*(0.051)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses.

[5] Note that using only those who have at least 40 hours of contracted working time would reduce the sample by roughly 56%.

All previous results remain unchanged. The only exception is the insignificant difference between overtime and non-overtime workers when desired hours are used in the model for satisfaction with life in five years. It can also be seen that standard errors and parameters increase to some extent but not where satisfaction with work is considered as the results tend to be less significant than before. The results for the group working less than desired (not reported) also remain unchanged. With respect to health and satisfaction with work the coefficients are higher than those for the overtime group but are also as significant. When current satisfaction with life is evaluated the coefficient for the underemployed group has about the same magnitude as the coefficient of the overtime group and is significant at the 5% level. However, it is insignificant in the other two life satisfaction models.

The last robustness check exploits the panel structure of the underlying data and uses the 2004 to 2008 waves which results in roughly 24,000 observations. The problem here is that not all waves contain the same set of questions in the respective questionnaire. With respect to the dependent variables, this affects life satisfaction in one year and in five years. Unfortunately, some explanatory variables are also lost: Firstly, the indicator for varying starting time and secondly the lifestyle variables (BMI, drinking behavior, etc.) cannot be used as controls anymore, while the latter would have been the most interesting ones to take a look at. The calculation of the parameter estimates was performed by using Frechette's (2001a and b) "reoprob" command (for the STATA software package) which estimates a random-effects ordered probit model that allows to control for time-constant and individual-specific heterogeneity. Additionally, it is controlled for a time trend in every specification. Outliers were removed in the same way as before but as there are enough observations it was not necessary to exclude the (previously) small groups of e.g. widowed persons. Self-employed persons are excluded for the same reasons as before. With respect to age, everyone between the ages of 18 and 45 (throughout the five years, which is a little bit less restrictive than in the original set-up) is considered in the analysis. Table 2.9 reports the parameter estimates for each overtime measure and the proportion of the panel level variance (rho). The dependent variables are shown on the left. Again, the conclusion remains unchanged: Overtime work is associated with worse SRHS, lower satisfaction with life and health and also with lower satisfaction with work compared to non-overtime work. It also has to be mentioned that the only coefficient that is not more than marginally significant is the one for overtime in general, when satisfaction with work is considered, which is also consistent with the previous findings. The proportion of the panel level variance is quite high and around .5 – if all variables from the preferred ordered probit specification (as in the cross-section) would be available in the panel dimension, likelihood-ratio tests would reveal more of its importance. Finally, it is noteworthy that the underemployment coefficients have the same sign and significance level as the corresponding coefficients of the overtime indicator when deviations from desired hours are used as the overtime measure.

Table 2.9: Robustness check of overtime measures (panel data)

	N	Model type: Overtime in general	Actual vs. Desired	Actual vs. Contracted
Health status				
Overtime coefficient	23,903	0.094 ***	0.093 ***	0.096 ***
		(0.027)	*(0.027)*	*(0.026)*
Rho		0.555 ***	0.555 ***	0.555 ***
		(0.008)	*(0.008)*	*(0.008)*
Satisfaction with health	23,903			
Overtime coefficient		-0.095 ***	-0.078 ***	-0.093 ***
		(0.025)	*(0.024)*	*(0.024)*
Rho		0.545 ***	0.545 ***	0.545 ***
		(0.007)	*(0.007)*	*(0.007)*
Satisfaction with life (now)	23,914			
Overtime coefficient		-0.103 ***	-0.112 ***	-0.142 ***
		(0.024)	*(0.024)*	*(0.023)*
Rho		0.486 ***	0.485 ***	0.485 ***
		(0.008)	*(0.008)*	*(0.008)*
Satisfaction with work	23,694			
Overtime coefficient		-0.039 *	-0.096 ***	-0.066 ***
		(0.024)	*(0.023)*	*(0.023)*
Rho		0.448 ***	0.446 ***	0.447 ***
		(0.008)	*(0.008)*	*(0.008)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Standard errors given in parentheses.

2.8 Conclusions

This work analyzed differences in well-being by comparing overtime and non-overtime workers in a large and representative German sample. The main underlying dataset was the 2008 wave of the GSOEP. It was argued that using the total amount of hours worked is not a satisfactory measure as the responses given are supposedly prone to measurement error (which could or could not be classical) and because people might differ in unobservable personal characteristics (e.g. stress-bearing capabilities). With the data at hand an alternative approach was suggested and it was assumed that people that work contracted or desired hours are better off than those who do not, which does not depend on the total hours worked per se. The first measure used contracted hours as the relevant reference point but the problem is that it is unclear whether this truly reflects the preferred amount of work as contracts in a labor market are typically not that flexible. Therefore, the second measure used desired hours as the individual reference point. The advantage is that these self-stated preferences should be closer to what people actually want as their personal workload. The underlying idea of both measures is that people may not be able to tell the exact amount of actual or desired hours but they do know whether they work more than they would have to or if they would want to work more or less that they currently do. As both measures shared a large proportion of the same information, the results differed to some extent but the baseline interpretation was always the same: Overtime workers are worse off than non-overtime workers with respect to health and satisfaction with life and work. Additionally, it was found that underemployed persons are also worse off than those who work desired hours. Therefore, it is not necessarily a fixed amount of time spent at work that makes people feel better but rather an amount of time that equals their personally preferred workload. In other words: Working nine-to-five is only desirable if that fits the personal needs – otherwise working more or less could also be the right choice for other individuals. As suggested by related literature, the models controlled for lifestyle characteristics, such as smoking and drinking behavior.

The results were replicated in a restricted sample using only those observations with at least 35 hours of contracted working time and in an extended (panel) sample (with unfortunately not all variables being available).

Simultaneity is an issue for both measures and the necessary assumptions for a meaningful interpretation of the results were discussed. Literature plausibly suggests that lower health is associated with fewer working hours so there is no selection into overtime by those that are already worse off. Where desired working hours are considered, the assumption is similar but expressed differently as it is necessary that moving back towards desired hours makes people (at least to some extent) better off. In these cases the effects found here would rather be too small than too large. In comparison to other work this puts more structure on the overtime model and narrows down the complications rather than just mentioning the issue. Experimental studies and existing literature support the assumptions made here.

Reducing overtime is often brought forward as an argument in order to reduce unemployment. The findings from this work serve as another important argument why a reduction of overtime hours could be beneficial for society itself. Making use of deviations from desired hours rules out that people choose to work overtime intentionally and more restrictive laws would help to prevent them from harming themselves and increasing for example health care costs. On the other hand, an implementation of such an idea is very complicated as firms depend on flexible working schedules to compensate for business volatility when it would make no sense to employ and train people just for a couple of months when there is a busy season. It is therefore neither straightforward nor currently possible to recommend a policy action to be taken as the two opposing effects mentioned above must be balanced.

Nevertheless, different survey questions are needed for a better understanding of the dynamics. It is yet unclear if there is a turning point with respect to well-being. People with above average well-being might indeed be more likely to work overtime and after a while this additional work might well lead to adverse effects. However, the mechanism and its determinants are unknown. Other job-specific characteristics might play an important role as they could be related to both well-being and overtime work. These could be soft factors such as team dynamics. Cooperation within the team as well as trust and loyalty amongst co-workers are examples.

2.9 Appendix

Table 2.10: Means and standard deviations of the explanatory variables

Female (B)	0.483
	(0.500)
Age	34.934
	(7.352)
Years of education / training	12.616
	(2.585)
Job tenure	7.505
	(6.736)
Job in Eastern Germany (B)	0.200
	(0.400)
Unhealthy food (B)	0.072
	(0.258)
Sports (B)	0.067
	(0.249)
Smoker (B)	0.351
	(0.477)
Beer drinker (B)	0.132
	(0.339)
Wine drinker (B)	0.048
	(0.214)
BMI	25.031
	(4.006)
German nationality (B)	0.936
	(0.246)
Children under age 16 (B)	0.476
	(0.500)
Log of net personal income	7.102
	(0.664)
Married (B)	0.502
	(0.500)
Single (B)	0.407
	(0.491)
Divorced (B)	0.071
	(0.257)
Blue-collar worker (B)	0.285
	(0.452)
Public servant (B)	0.061
	(0.240)
Training (B)	0.064
	(0.246)
Irregular hours of work (B)	0.110
	(0.313)
Varying starting time (B)	0.301
	(0.459)
Actual hours of work	38.548
	(10.739)
Contracted hours of work	34.817
	(8.762)
Desired hours of work	35.320
	(8.694)

Note: All values with respect to the largest common sample (N = 3969). Standard deviations of the variables given in parentheses. (B) indicates a binary variable.

Table 2.11: Means and standard deviations of the dependent variables

Health status	2.283
	(0.805)
Satisfaction with health	7.231
	(1.838)
Satisfaction with life (now)	7.195
	(1.491)
Satisfaction with life (1y)	7.504
	(1.467)
Satisfaction with life (5y)	7.639
	(1.663)
Satisfaction with work	7.047
	(1.901)

Note: All values with respect to the largest common sample (N = 3888). Standard deviations of the variables given in parentheses.

A2.1 Question wording as in the English version of the questionnaire

Question asking for contracted hours:

How many hours are stipulated in your contract (excluding overtime)?

__,_ hours per week // No set hours

Question asking for actual hours:

And how many hours do your actual working-hours consist of including possible overtime?

__,_ hours per week

Chapter 3

3 Let the sunshine in: The effect of weather on survey measures of health, well-being, and preferences[6]

3.1 Introduction

Self-reported survey measures of health, well-being, and preferences are an important source of data in applied economics. A large and growing literature shows that such subjective measures have predictive power for economic choices and outcomes of individuals and households. Nevertheless, self-reported survey data can be subject to measurement error, with potentially severe consequences for the estimation of econometric models and the conclusions we draw from them.

As an example of research questions that rely on self-reported, subjective data, consider the relationship between income and well-being or happiness (Frey and Stutzer (2002)). Using American and British data, Blanchflower and Oswald (2004) find a positive relationship between income and happiness as well as differences with respect to covariates such as gender. Deaton (2008) uses Gallup World Poll data and shows a positive relationship between per capita GDP and life satisfaction as well as cross-country differences. For

[6] This chapter is based on joint work with Joachim Winter.

We would like to thank Christian Traxler and participants of the SFB/TR 15 Workshop on Natural and Field Experiments in Holzhausen for helpful comments.

Germany, Frijters *et al.* (2004) find that higher income is associated with higher life satisfaction. Subjective measurements are also used to study the determinants and effects of preferences, for example risk attitudes. Dohmen *et al.* (2011) show that such variables as height and parental background have a positive effect on the willingness to take risks. Despite their value for applied research, subjective measurements are subject to potentially severe measurement error, and there is a lively discussion about their validity (see, *inter alia*, Bertrand and Mullainathan (2001), or Krueger and Schkade (2008)).

In this chapter, we address one source of measurement error in self-reported measures of health, well-being, and preferences: The respondent's mood at the time of a survey interview. We investigate the impact of current weather conditions, which as we discuss in section 3.2 are known to affect mood, on such self-reports. Using representative data from the German Socio-Economic Panel (GSOEP)[7], we first confirm earlier findings on this association which, however, were typically obtained with smaller samples from more specific populations. Specifically, we confirm and extend the evidence that current weather conditions affect self-reported health and well-being. Moreover, we show – to our knowledge, for the first time – that weather has an impact on the measurement of preferences, specifically on risk attitudes. Self-reported willingness to take risks is lower when the interview is conducted on a rainy day.

We provide a literature review in section 3.2, with specific focus on the underlying mechanisms that connect weather, mood, and survey response. Section 3.3 provides details on our data and discusses methodological aspects. In section 3.4 we present some descriptive statistics. Our main results are reported in section 3.5 and discussed in section 3.6. Section 3.7 concludes with a discussion of the implications of these findings for applied research that uses self-reported survey data.

[7] In addition to the GSOEP public use data, we obtained information on the residence of the respondents (under a special contract with the DIW). This data is merged with a variety of location-specific measures of current weather conditions on the interview date as provided by the *German Weather Service* (*Deutscher Wetterdienst*, DWD). The data is described in section 3.3 below.

3.2 Literature review

Following the increased use of survey data on individuals, households, and firms that have become available in the last three decades, economists have also become interested in survey response behavior and its implications for measurement and econometric estimation (e.g. Bound *et al*. (2001), McFadden *et al*. (2005)). This research builds on a vast body of research in the social sciences, particularly in survey research and social psychology (as reviewed among others by Sudman *et al*. (1996) and Tourangeau *et al*. (2000)). A key insight of this line of research is that unless survey respondents can easily recall a response from memory, the heuristics they use to form a response lead to measurement error that is potentially correlated with other model variables. That is, measurement error is not classical, with potentially severe consequences for econometric estimation (e.g. Hoderlein and Winter (2010)). In this section, we provide a selective review of this literature, concentrating on subjective measures and the effects of mood and weather on survey response behavior.

Krueger and Schkade (2008) investigate the test-retest reliability of well-being measures and find that life satisfaction has a lower reliability ratio than variables such as income (but is still suitable for research if measurement error is properly accounted for). They also argue that even though the evaluation of life satisfaction should not change much in a relatively short time period, it could be that measures of this kind are affected by momentary influences such as one's mood or the weather. Along these lines, Yardley and Rice (1991) show that (subjective) well-being is both stable over time and influenced by current mood. Helliwell and Wang (2011) make use of high frequency US data to show that respondents report more positive and less negative emotions on weekends but that there is no such day-of-week effect when it comes to the evaluation of one's life.

An influential contribution on the relation between weather and mood was made by Cunningham (1979). He investigates the effect of current weather on helping behavior and argues that the relationship between sunshine and helping behavior can best be explained by changes in mood. He also discusses in depth how this relationship might work. Possible channels are connections with pleasant events (e.g. picnics), spectral characteristics of sunlight and stimulating illumination of the environment (e.g. enhanced colors and sharpened detail) or the connection between sunlight and physiological processes (e.g. adrenal corticosteroid or Vitamin D production). Given this evidence, a direct analysis of the interaction between weather, mood and well-being can be found in Schwarz and Clore (1983). The authors conduct telephone interviews in an experimental setting and show that rainy (or sunny) weather affects individual statements about general well-being, i.e. current mood is used in the evaluation of one's happiness and life satisfaction. Riener and Traxler (2011) use changes in mood induced by the weather to better explain payments in a "pay-what-you-want" restaurant. They find that (controlling for temperature) sunshine has a season-specific effect on payments and that therefore mood is positively related to payments. In an experimental study of individual contribution to climate change mitigation Diederich and Goeschl (2011) find that the contribution probability increases with mean ambient temperature.

Kliger and Levy (2003) use financial market data to examine the effects of (weather induced) variation in mood on risk attitudes and find that investors in a good mood are less willing to take risks than investors in a bad mood. More general analyses in finance, for example in Saunders (1993) or Hirshleifer and Shumway (2003) relate good weather (less cloudiness / more sunshine) to higher stock returns. They also discuss differences in information processing and present literature evidence which shows that people in a good mood process information less critically and that people in a bad mood process information more carefully. We consider this as an alternative result to Kliger and Levy (2003) who argue that bad mood makes people hastier and good mood makes them more cautious. In contrast, Ruder and Bless (2003) find that in their studies happy participants rely more on the ease of retrieval heuristic and that sad participants rely more on the activated content. Additionally, and as discussed above, good mood is believed to be related to remembering positive events and therefore a lower risk aversion seems more plausible in the context of our work.

The effect of weather on decision making has been studied in other contexts as well. A recent paper by Simonsohn (2010) shows (projection) bias in the decision about whether one wants to attend a certain college or not. The main finding is that increased cloud cover increases the probability of enrolment. The reasoning here is similar to before: Bad weather induces sad moods and that makes studying more attractive whereas good weather (i.e. good mood) makes outdoor activities more favorable. A similar contribution is from

Redelmeier and Baxter (2009) who find that in medical school admission interviews, candidates who are interviewed on a rainy day receive lower scores than candidates who are interviewed on a sunny day.

Taken together, the relationship between mood and weather is well established, and so is the relationship between (current) mood and the evaluation of self-reported general concepts and personal attitudes. However, to our knowledge, there are no studies that use large national social surveys to study the effects of weather and mood in subjective survey measures. In this chapter, we close this gap. Using data from a large survey also allows us to control for a large variety of potentially confounding variables.

3.3 Data

All respondent-specific data in this study is taken from the 2007 and 2008 wave of the German Socio-Economic Panel (GSOEP), a large and representative survey of the German population which offers extensive information not only on typical personal or household characteristics like age, employment or occupational status and income but also on variables describing family background, e.g. parental education. Additionally, the questionnaire includes general well-being questions, i.e. the respondent’s satisfaction with life or health and the self-rated health status (SRHS). We also have information on the residence of the interviewees (obtained by a separate DIW license) and therefore know which of the (nearly) 100 regional policy regions each individual lives in. These regional policy regions are constructed by the Federal Office for Building and Regional Planning (BBR) by considering aspects like commuter flows and such.

We use two cross-sections for the following two reasons. Firstly, the year 2007 data can be used as a robustness check for the results obtained with the 2008 data (and vice versa) and secondly, the question on personal willingness to take risks is included in the 2008 questionnaire but not in the corresponding version from 2007.

The data on current (daily) weather is freely available for both years and taken from the website of the *German Weather Service*.[8] It is collected from 46 weather stations throughout Germany and the data provider also assigns a quality level ranging from 0 to 10 to each observation period (a specific day at a certain measurement point including all measures). All data used in our work has the highest possible quality level (10) meaning that all data is systematically checked and corrected if necessary. The set of measures provided by the DWD includes the most common weather measures used in the literature: Cloud cover and sum of sunshine duration. Besides the free availability, our data has the advantage that weather measures that were used in the seminal work of Cunningham (1979) and that are often omitted in other work are also available, e.g. barometric pressure, wind speed and relative humidity. Even though it is not clear (in a psychological or behavioral sense) how these variables interact, we rather control for them than initially omitting the information. Table 3.1 summarizes the weather variables and how they are measured.

Table 3.1: Details on the weather variables

	Unit of measurement	Observation period
Minimum temperature	Celsius	23:51 on t-1 to 23:50 on t
Maximum temperature	Celsius	23:51 on t-1 to 23:50 on t
Average relative humidity	Percent	Average of 24 values
Average wind force	Bft	Average of 24 values
Maximum wind speed	m/sec	23:51 on t-1 to 23:50 on t
Sum of sunshine duration	Hours	23:51 on t-1 to 23:50 on t
Average cloudiness	0/8 - 8/8	Average of 24 values
Rain (precipitation)	mm	05:51 on t to 05:50 on t+1
Average barometric pressure	hPa	Average of 24 values

Note: All time specifications refer to the Coordinated Universal Time (UTC).

Merging weather with the survey data is not straightforward. Even if we knew where the individuals lived exactly, we do not have weather data of that granularity. Therefore, we must assume that weather is not purely local but spreads over a wider area and as a consequence we have to consider information from a particular weather station as

[8] http://www.dwd.de – In some sections the language can be switched to English.

representative for the entire policy region. In order to make this a plausible approach, we restrict our dataset to those regions where we expect most people to live in a reasonable distance to the weather station, i.e. policy regions around Germany's largest cities with the weather station being close by. This procedure makes us confident that the weather we actually measure is indeed representative for most people living in that particular policy region and leaves us with observations for Berlin, Dusseldorf, Dresden, Frankfurt, Hamburg, Hanover, Leipzig, Munich and Stuttgart.

Let's consider Munich as an example. The measurement point is located at Munich Airport (48° 22' latitude and 11° 49' longitude) which is roughly 28 kilometers (17.4 miles) away from the city center. The weather measured at the airport is considered to be representative for Munich and the surrounding area.

As the purpose of this chapter is to investigate the influence of weather induced mood variation on well-being and risk attitudes, our four dependent variables are: Self-rated health status, satisfaction with health, satisfaction with life and willingness to take risks. SRHS is measured on a 5-point scale ranging from "very good" (1) to "bad" (5) and satisfaction with life as well as satisfaction with health are measured on an 11-point scale with 0 representing "completely dissatisfied" and 10 representing "completely satisfied". The willingness to take risks is also measured on an 11-point scale where 0 means "not willing to take risks at all" and 10 means "fully prepared to take risks".

All weather variables besides rain are seasonally adjusted by subtracting the respective monthly mean. Rain is the key variable of interest in our analysis. Rain should capture most effects that come with bad weather, such as less sunshine, lower temperature, etc. We construct a dummy variable that takes on the value 1 if it rained during the observation period on the interview date and 0 otherwise. This procedure of course does not imply that it was raining during the interview. Even though this measure is somewhat imprecise, it should capture the main weather effects. For instance, rain before the interview might already have influenced the mood, and if it was raining after the interview, this was probably foreseeable and therefore might already have influenced the mood. Some measures (e.g. sum of sunshine duration and average cloudiness) are highly correlated. Nevertheless, as these measures are based on 24 hours of observation, they might capture different effects.

The sets of explanatory variables are similar in all specifications. In the model for the individual willingness to take risks, we add the respondent's height and an education dummy variable for each parent, following Dohmen *et al.* (2011). The education variables indicate whether the parent completed technical or secondary school (the necessary qualification for attending a technical college or university, respectively). Additionally, we control for SRHS in the models for life satisfaction and risk attitude as health is likely to be an important factor in these settings. Therefore, SRHS is split into an indicator for good health (SRHS is either very good or good) and another for bad health (SRHS is either poor or bad) with a satisfactory SRHS as the reference group. With respect to the financial situation of the respondents, we rather control for household income than for personal income as we would otherwise exclude all unemployed persons.

In every specification of the subsequent analysis we include month and city dummies to account for month and city-specific differences as there might be for example less rain in May and also a generally better mood because individuals anticipate summer and all the positive things that come with it.

In order to keep the results in our models as general as possible we use information from the entire labor force (ages 18 to 65) including unemployed persons, i.e. only retirees and people that do not participate in the labor market are excluded.

Since all of our dependent variables are ordinal, we use ordered probit models (with clustering at the household level) throughout this chapter. Finally, all observations in the highest percentile with respect to income are removed.

3.4 Descriptive statistics

The following tables present summary statistics for the most important variables in our samples. Month and city dummies are not reported and weather variables are reported prior to seasonal adjustments.

Table 3.2: Descriptive statistics for the dependent variables

	2007 (N = 2897)		2008 (N = 2434)	
	Mean	SD	Mean	SD
Health status	2.437	0.862	2.412	0.866
Satisfaction with health	6.971	1.972	6.897	1.925
Satisfaction with life	7.052	1.669	7.077	1.624
Willingness to take risks	n/a	n/a	4.793	2.200

Note: All values with respect to the largest common sample.

Table 3.3: Descriptive statistics for the explanatory variables

	2007 (N = 2897)		2008 (N = 2434)	
	Mean	SD	Mean	SD
Rain (B)	0.494	0.500	0.527	0.499
Average cloudiness	5.137	2.196	5.346	1.962
Average wind force	2.785	0.901	2.803	0.942
Minimum temperature	5.375	5.248	4.219	5.955
Maximum temperature	14.266	6.829	13.053	7.094
Average relative humidity	74.858	12.099	73.562	10.301
Maximum wind speed	11.120	4.061	11.519	4.546
Average barometric pressure	992.346	20.125	994.038	21.231
Sum of sunshine duration	5.117	4.541	4.638	3.939
Height	n/a	n/a	173.235	9.624
Secondary school father (B)	n/a	n/a	0.206	0.404
Secondary school mother (B)	n/a	n/a	0.122	0.328
Age / 10	4.192	1.190	4.198	1.188
Female (B)	0.484	0.500	0.488	0.500
German nationality (B)	0.917	0.276	0.929	0.258
Years of education / training	13.000	2.850	13.202	2.875
Married (B)	0.552	0.497	0.548	0.498
Single (B)	0.307	0.461	0.312	0.464
Divorced (B)	0.104	0.305	0.099	0.299
Widowed (B)	0.012	0.111	0.014	0.119
Log of household income	7.959	0.592	7.990	0.577
Children under age 16 (B)	0.325	0.468	0.323	0.468
Unemployed (B)	0.081	0.273	0.067	0.251
Blue-collar worker (B)	0.171	0.376	0.160	0.367
Self-employed (B)	0.100	0.301	0.098	0.298
Public servant (B)	0.064	0.244	0.067	0.251
Training (B)	0.071	0.258	0.067	0.250

Note: All values with respect to the largest common sample. (B) indicates a binary variable.

Table 3.2 gives summary statistics for the dependent variables and table 3.3 for the explanatory variables. It can be seen that the years 2007 and 2008 do not differ dramatically but 2007 was slightly warmer. Note that all values refer to the largest common sample and therefore do not necessarily contain all observations used in the following estimations as there are additional missing values when willingness to take risks and the additional explanatory variables are considered.

3.5 Results

For the interpretation of the results one must keep in mind that weather may have direct effects on SRHS as weather conditions might for example cause migraine. Nevertheless, if this was the only channel then the effect of weather on satisfaction with health should be at least very weak as this is a more general evaluation and not just a different question wording. Additionally, weather would then have no effect where satisfaction with life and willingness to take risks are considered – firstly, because it is very difficult to imagine how weather could drive changes in these variables directly and secondly, because we explicitly control for SRHS. Table 3.4 gives the results for the 2007 cross-section and table 3.5 for 2008.

The most striking result is that there are only two explanatory variables with significant parameter estimates for every model. The first one is rain and the second one is income, whereas the rain coefficients are more significant in 2008. If we ignore the willingness to take risks, we can add the unemployment dummy to this list. Surprisingly, all other weather variables do not seem to have an important impact. In 2007 we can find no significant coefficients at all and only a few in 2008.

With respect to both health measures we can find significant associations with age, education, income and unemployment. In 2008 this also holds for being German and being self-employed. A gender difference is only found in 2007 for SRHS. The negative effect of cloudiness in 2008 (satisfaction with health) is consistent with findings in the literature. The sign of the marginally significant coefficient of sunshine duration is in contrast to our prior expectations but it is the only regression where we can see this.

Table 3.4: Results for 2007

	Parameter estimates 2007		
	Health status	Satisfaction with health	Satisfaction with life
Good health	n/a	n/a	0.670 ***
	n/a	n/a	*(0.046)*
Bad health	n/a	n/a	-0.525 ***
	n/a	n/a	*(0.071)*
Rain (B)	0.118 *	-0.124 **	-0.131 **
	(0.063)	*(0.061)*	*(0.064)*
Average cloudiness (adj.)	-0.011	-0.017	0.014
	(0.025)	*(0.023)*	*(0.023)*
Average wind force (adj.)	-0.008	-0.006	-0.032
	(0.046)	*(0.044)*	*(0.046)*
Minimum temperature (adj.)	-0.013	0.011	0.016
	(0.012)	*(0.012)*	*(0.013)*
Maximum temperature (adj.)	0.013	-0.006	-0.006
	(0.011)	*(0.011)*	*(0.011)*
Average relative humidity (adj.)	-0.005	0.002	0.002
	(0.004)	*(0.004)*	*(0.004)*
Maximum wind speed (adj.)	0.004	0.004	0.000
	(0.010)	*(0.009)*	*(0.009)*
Average barometric pressure (adj.)	0.002	0.003	-0.003
	(0.003)	*(0.003)*	*(0.003)*
Sum of sunshine duration (adj.)	-0.021	-0.010	0.002
	(0.014)	*(0.014)*	*(0.013)*
Age / 10	0.248 ***	-0.240 ***	-0.035
	(0.025)	*(0.025)*	*(0.027)*
Female (B)	0.100 **	-0.042	0.071 **
	(0.040)	*(0.037)*	*(0.036)*
German nationality (B)	0.024	-0.019	-0.090
	(0.085)	*(0.083)*	*(0.081)*
Years of education / training	-0.041 ***	0.036 ***	0.010
	(0.009)	*(0.008)*	*(0.008)*
Log of household income	-0.126 ***	0.137 ***	0.359 ***
	(0.046)	*(0.046)*	*(0.045)*
Children under age 16 (B)	-0.014	0.029	-0.049
	(0.053)	*(0.052)*	*(0.053)*
Unemployed (B)	0.223 **	-0.226 ***	-0.641 ***
	(0.090)	*(0.086)*	*(0.091)*
Blue-collar worker (B)	0.080	-0.039	-0.126 **
	(0.061)	*(0.060)*	*(0.063)*
Self-employed (B)	-0.101	0.042	-0.039
	(0.075)	*(0.072)*	*(0.072)*
Public servant (B)	0.090	-0.110	0.048
	(0.084)	*(0.081)*	*(0.079)*
Training (B)	-0.188 *	0.079	0.110
	(0.104)	*(0.091)*	*(0.092)*
Number of observations	2897	2897	2897
Log (pseudo)likelihood	-3405	-5600	-4842
Pseudo R-squared	0.052	0.030	0.083

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses. (B) indicates a binary regressor. Indicators for month, city and marital status included in all regressions but estimates not reported. Base category for occupational status is "white-collar".

Table 3.5: Results for 2008

	Parameter estimates 2008			
	Health status	Satisfaction with health	Satisfaction with life	Willingness to take risks
Good health	n/a	n/a	0.626 ***	0.052
	n/a	n/a	(0.047)	(0.048)
Bad health	n/a	n/a	-0.617 ***	-0.192 **
	n/a	n/a	*(0.071)*	*(0.075)*
Rain (B)	0.123 **	-0.161 ***	-0.170 ***	-0.145 ***
	(0.054)	*(0.053)*	*(0.059)*	*(0.053)*
Average cloudiness (adj.)	0.031	-0.050 **	0.045 *	-0.022
	(0.024)	*(0.024)*	*(0.024)*	*(0.024)*
Average wind force (adj.)	-0.032	0.025	-0.115 **	-0.019
	(0.050)	*(0.047)*	*(0.051)*	*(0.047)*
Minimum temperature (adj.)	-0.003	0.013	-0.004	0.007
	(0.013)	*(0.012)*	*(0.013)*	*(0.013)*
Maximum temperature (adj.)	-0.004	0.005	-0.006	-0.005
	(0.012)	*(0.011)*	*(0.011)*	*(0.011)*
Average relative humidity (adj.)	-0.005	0.004	-0.005	0.003
	(0.004)	*(0.004)*	*(0.004)*	*(0.004)*
Maximum wind speed (adj.)	-0.001	0.002	0.033 ***	0.011
	(0.010)	*(0.009)*	*(0.010)*	*(0.009)*
Average barometric pressure (adj.)	-0.001	-0.001	-0.004	-0.003
	(0.003)	*(0.003)*	*(0.003)*	*(0.003)*
Sum of sunshine duration (adj.)	0.011	-0.022 *	0.001	-0.004
	(0.013)	*(0.013)*	*(0.013)*	*(0.013)*
Height	n/a	n/a	n/a	0.001
	n/a	n/a	n/a	*(0.003)*
Secondary school father (B)	n/a	n/a	n/a	0.024
	n/a	n/a	n/a	*(0.060)*
Secondary school mother (B)	n/a	n/a	n/a	0.003
	n/a	n/a	n/a	*(0.071)*
Age / 10	0.253 ***	-0.215 ***	0.001	-0.069 ***
	(0.026)	*(0.027)*	*(0.028)*	*(0.026)*
Female (B)	0.041	-0.011	0.067 *	-0.372 ***
	(0.042)	*(0.039)*	*(0.037)*	*(0.057)*
German nationality (B)	0.187 **	-0.167 **	-0.004	-0.035
	(0.089)	*(0.085)*	*(0.089)*	*(0.098)*
Years of education / training	-0.041 ***	0.031 ***	-0.002	0.004
	(0.009)	*(0.008)*	*(0.009)*	*(0.009)*
Log of household income	-0.222 ***	0.240 ***	0.314 ***	0.102 **
	(0.048)	*(0.049)*	*(0.054)*	*(0.046)*
Children under age 16 (B)	0.000	0.014	0.058	-0.055
	(0.053)	*(0.052)*	*(0.053)*	*(0.054)*
Unemployed (B)	0.284 ***	-0.196 **	-0.440 ***	-0.003
	(0.095)	*(0.093)*	*(0.098)*	*(0.097)*
Blue-collar worker (B)	0.014	0.017	-0.169 ***	-0.140 **
	(0.065)	*(0.061)*	*(0.063)*	*(0.069)*
Self-employed (B)	-0.192 **	0.187 ***	-0.064	0.377 ***
	(0.076)	*(0.072)*	*(0.072)*	*(0.071)*
Public servant (B)	0.026	-0.011	0.046	-0.117
	(0.084)	*(0.080)*	*(0.076)*	*(0.084)*
Training (B)	-0.074	0.129	0.034	0.005
	(0.104)	*(0.097)*	*(0.097)*	*(0.097)*
Number of observations	2757	2757	2757	2436
Log (pseudo)likelihood	-3278	-5325	-4624	-5169
Pseudo R-squared	0.053	0.030	0.075	0.021

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses. (B) indicates a binary regressor. Indicators for month, city and marital status included in all regressions but estimates not reported. Base category for occupational status is "white-collar".

When we look at life satisfaction as the dependent variable, we find a significant association with health, gender, income, being unemployed and being a blue-collar worker in both years. In 2008 the cloudiness coefficient is again marginally significant but the sign is the opposite of what theory predicts and again it is the only model where this occurs. As this is also the only model with significant wind variables it is unclear what kind of variation they capture here but it could be for example that people are happy to be inside on especially stormy days.

The last model in our analysis is the individual's willingness to take risks. Interestingly, there are no significant parameter estimates for height and parental education as in Dohmen *et al.* (2011) but one has to keep in mind that we use a different wave of the GSOEP, a much smaller sample, a (to some extent) different set of controls and a different model (ordered probit instead of interval regression). In terms of the other control variables we can see that (bad) health, age, gender, income, being a blue-collar worker and being self-employed are significantly associated with risk attitude.

The importance of the rain variable can be seen even better when average marginal effects are calculated (not reported). Consider the risk model as an example: The average marginal effects of rain are always stronger than those of age (for all categories) and the latter variable is typically a major driver in explaining individual decisions.

3.6 Discussion

Our results show a consistent picture of the influence of weather on the individual evaluation of general concepts (well-being) and attitudes (willingness to take risks) but they also differ from previous findings as typically cloudiness or sunshine are described as the most influential variables. However, this is not necessarily surprising as our design is different. For example the work by Cunningham (1979) has no observations on rainy or stormy days and as a second example the paper by Simonsohn (2010) measures precipitation in inches. In contrast, our approach lets a rainy day capture most variation in the dependent variables as we think that the amount of rain does not matter, but more importantly, whether it was raining at all. Therefore, it is reasonable that sunshine or cloudiness do not contain much additional information once we control for the occurrence of a rainy day which typically comes with less sunshine and more wind, etc. As table 3.6 shows, specifications that use the rain dummy as the only weather variable (in the same samples) also yield a significant rain coefficient (except for satisfaction with life). Unfortunately, our data is by far not perfect. While Cunningham (1979) has weather information for the exact time of the interview we use daily values that could be a good proxy for weather during the interview although the weather could also have changed. Nevertheless, we argue that considering rainy days as a way of mood induction can still lead to plausible results as rain might still influence mood after it has stopped or even influence mood beforehand.

Table 3.6: Parameter estimates for the rain dummy as the only weather variable

	2007		2008	
	Parameter estimate	Robust std. error	Parameter estimate	Robust std. error
Health status	0.090 *	0.046	0.085 *	0.046
Satisfaction with health	-0.100 **	0.043	-0.107 **	0.044
Satisfaction with life	-0.053	0.046	-0.063	0.047
Willingness to take risks	n/a	n/a	-0.089 **	0.045

*, **, *** indicate significance at the 10%, 5%, 1% levels.

Another important caveat is that we use measures from a specific weather station and declare it representative for the entire region the respective respondents live in. Cunningham (1979) is able to gather information more specifically as weather measures are collected on-site. Besides measuring weather it is also important to obtain reliable information on control variables. By using data from the GSOEP we have the huge advantage that we can control for important covariates such as age while Cunningham (1979) had to estimate the approximate age.

Taken all together, there are some advantages and disadvantages to our approach but the results make us confident that it worked out very well. If weather was not representative for most respondents, we would probably not see any effects as our variables would be no more than random noise. As the interviews are conducted at home weather probably has a weaker effect on mood than if people were asked outside but the effect is obviously strong enough to be reflected in our data. Consequently, we think that our weather variables are measured precisely enough and thereby the variation in mood induced by weather conditions is a valid natural experiment for testing whether changes in mood can affect the evaluation of one's health, life or the personal willingness to take risks.

3.7 Conclusions

In this work we used variation in weather as a large natural experiment to show that variation in current mood affects the individuals' responses to subjective survey questions, i.e. their current health, their satisfaction with health and life in general and preferences such as the willingness to take risks. We showed that on rainy days, people tend to report that they are less healthy, less satisfied with their health and also less satisfied with their life even though current mood should not systematically influence such global evaluations. Additionally, we showed that a bad mood, i.e. a rainy day, also leads to a lower willingness to take risks. We also presented evidence that a simple rain dummy (which is easy to obtain) can capture most of the effects of current weather – unfortunately, this did not work with life satisfaction as the dependent variable.

What are the consequences for research that uses measures of variables where the answer is likely to be influenced by transient effects? As to data taken from surveys (even from large and representative ones) researchers must consider the possibility that their variables are prone to measurement error. Consequences can range from inflated error variance to biased estimates. The magnitude of the potential problems naturally depends on how strongly the dependent variable or the regressors are actually influenced by current mood. On the other hand, our findings also have constructive implications. For instance, one might use weather as an instrumental variable. The idea is of course not new as a leading example is the estimation of supply and demand elasticities but that is also a very different setting. While we were investigating subjective measures that are supposed to be very stable (they probably are – at least in the long run) others might be interested in measures that are assumed to be very transient. For example consumer research is interested in buying decisions and mood might plausibly play a role. If one thinks that self-stated mood is endogenous then current weather could credibly capture the exogenous variation. We only mentioned surveys so far, but laboratory experiments could also be easily affected, especially when the sessions are held on different days. Here, weather could even mimic a treatment effect if it is indeed correlated with the treatment.

As basic weather variables such as rain and temperature during the interview are easily obtainable, it would be helpful to collect this information at least in all computer-assisted personal interviews (CAPI) and laboratory experiments so that researchers can easily control for these factors or make use of them as described above.

Another path for future research is a better understanding of the connection between weather, mood and for example well-being. Even though this is not feasible in large surveys, a simultaneous collection of biomarkers (e.g. the above mentioned adrenal corticosteroid or Vitamin D production) could shed more light on the underlying mechanisms.

Chapter 4

4 The influence of self-perception on bracket choice[9]

4.1 Introduction

Various research questions in economics involve consumption expenditures either as the dependent variable or as an important regressor in cross-sectional, panel data or time series models. The most important examples would be measurement of household wealth and forecasting the growth of an economy. Therefore, one would be interested in having correct and error free data – which is of course rare in practice. A large field in statistics, economics, psychology and related sciences is working on measurement error (correction) models (e.g. Battistin *et al*. (2003)), on how the answers given to various questions are biased and – most importantly – why.

There are several ways to measure consumption expenditures, each with its specific advantages and disadvantages. The most commonly used ones are diaries, which provide relatively reliable and exact data but impose a heavy respondent burden, brackets, which can be answered more easily but are less precise, open format questions which are also easier to answer and also less precise (the disadvantage of knowing only intervals when brackets are used vanishes but this is offset by rounding or heaping in open questions) and categorized answer formats which are more precise than brackets but, depending on the

[9] This chapter is based on joint work with Joachim Winter.

number of sub-categories, also impose a substantial respondent burden (Browning *et al.* (2003)). It can be shown that the results differ significantly depending on which way one asks for consumption expenditures (e.g. Winter (2002, 2004)). In the context of this work we will concentrate on bracketed answer formats.

There is some literature on whether a middle alternative (also referred to as the neutral position or fence-sitting option) should be included or not. In these cases, people would choose the middle position either because they are truly neutral (indifferent or low activation of the positive and negative evaluations) or because of ambivalence (Nowlis *et al.* (2002) provide a recent overview). The situation in measuring consumption expenditures is different. Facing bracketed answer formats to expenditure questions is not about whether the one or the other end of the scale is correct but about picking the bracket that contains the correct value.

Initially, this task sounds easy but respondents can have problems because they either simply do not know the answer, are uncertain about the exact amount or are even unsure which quantity they are supposed to report on (Schwarz and Oyserman (2001)). That is why respondents use various types of heuristics, such as rounding, heaping and guessing or contextual information to derive their answer (Browning *et al*. (2003) for an overview and Ruud *et al.* (2009) in the case of rounding / heaping).

In the case of bracketed response formats, the choice of bracket values (i.e. the position of the brackets relative to the expected distribution of true expenditures) influences, i.e. biases, the outcome. So, responses in this case are subject to bracket effects. In an experiment Winter (2002) shows that the median of the reported expenditures depends on the bracket values presented to the respondents, i.e. the group that receives higher bracket values reports higher expenditures. The question that arises is why such a bias occurs. If respondents do know their consumption expenditures why don't they report it by checking the appropriate bracket? The idea is that respondents are unsure and use some (or a combination) of the heuristics mentioned above to come up with an answer to the expenditures question.

An experiment (set up by Norbert Schwarz and Joachim Winter) administered in the 2001 Dutch CentER panel is set up to learn more about the question why bracket effects occur and how they can be explained. The advantage of this experiment is that it is conducted in the same wave as the experiment with the bracket bounds and therefore we can use the same data to analyze the effect of using heuristics on the bracket choice as Winter (2002).

We will show that presenting higher bracket values is initially associated with choosing lower bracket values but also that a higher self-perceived typicality leads to choosing higher bracket values. We will discuss these findings as a potential explanation for the occurrence of the bracket effects observed by Winter (2002).

The remainder of this chapter is set up as follows: Section 4.2 describes the original experiment by Winter (2002) and the idea of how self-perceived typicality can lead to biased results when asking for consumption expenditures. Section 4.3 gives details on the experimental set-up of this work and section 4.4 gives the results. Section 4.5 discusses potential validity concerns and section 4.6 concludes.

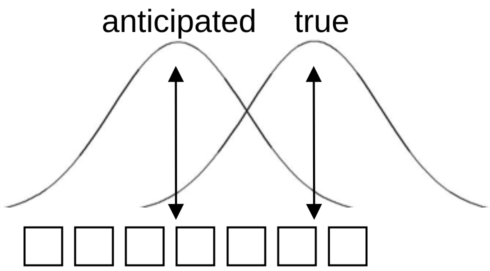
4.2 The idea of how bracket effects occur

We use Winter (2002) as a starting point. He shows that the bracket values presented to the respondents change the median of the outcome they have to report on, namely consumption expenditures. Participants were randomly assigned to groups which were presented different bracket values. One group received low bracket values (the values were shifted to the left of the expected population distribution, i.e. we expect more responses in the rightmost bracket as this is the open ended interval containing all the responses above the censoring threshold) and the other group received high bracket values (the values were shifted to the right of the expected population distribution, i.e. we expect more responses in the leftmost bracket as this is the interval containing all the responses between zero and the relatively high starting value of the second bracket). The interesting result is that the second group reported higher expenditures. So far we see that the answers depend on the bracket values but we do not know why, yet. If respondents knew the exact amount they spend every month, we would not see this shift in the distribution and if they were just picking random brackets or values we would not see this effect either. Respondents may still use several heuristics to answer bracketed questions. If someone is uncertain about how much he spends on total consumption, then simply guessing a rough value is the most obvious approach but here we want to put more structure on how this is done: Respondents probably do not just pick a random number that falls somewhere in the interval from zero to infinity but may use a somewhat more sophisticated procedure to answer. We believe that a combination of self-perception and contextual information is used.

Firstly, the bracketed answer format presented to the respondents not only contains the obvious numerical bracket values but also contextual information. The latter can be described as follows: The given set of brackets can be interpreted as a representation of the underlying population distribution (Sudman *et al*. (1996)), i.e. the expected value lies somewhere in the middle and observations in the tails are less common and therefore less frequent. Next, the respondents have an idea of how typical they are regarding their consumption expenditures, i.e. when they are asked to compare themselves with others they can state whether their expenditures are lower than, in line with or higher than the average of the relevant peer group. A combination of these two concepts can be used as a heuristic to derive an answer to a question with a bracketed answer format. Let's assume that a respondent is unsure of how much he actually spends on consumption. In this case, he would not just check a random interval but might think of the given range as the population distribution and compare himself to others. Let's assume further that he thinks of his own consumption as a little bit higher than that of his comparison group. He will be less likely to choose an interval to the left of the middle alternative(s), more likely to choose an interval somewhere in the middle or a little bit to the right and less likely to choose one of the extreme brackets to the right, representing a very high consumption.

Figure 4.1 illustrates this idea and it becomes easier to see why bracket effects occur: If the brackets that are supposed to represent the population distribution are artificially shifted, the heuristic of using self-perceived typicality in combination with assumptions on how the brackets represent the population distribution leads to biased results as the position of the underlying brackets does not represent the position of the interval, i.e. its observed frequency, in the population distribution. Of course it is unlikely that someone has absolutely no idea of how much is spent on consumption and takes all bracket values into consideration, as the set of possible answers is usually bounded, for example by monthly income. Nevertheless, uncertainty could still lead to a combination of best guess and inference from contextual information.

Figure 4.1: Inference from a set of brackets to the population distribution and vice versa

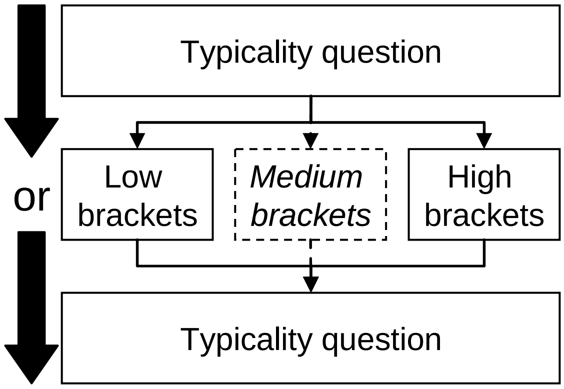


4.3 Set-up of the experiment

The experiment described here was administered by Schwarz and Winter in the 2001 CentER panel, a large and representative Dutch household survey. The experimental set-up and the different treatment groups are the same as in Winter (2002) but here we make use of an additional question that the respondents were asked: The typicality question. Here, all respondents were asked to compare the consumption of their household to the consumption of an average Dutch household and then state whether it is much lower, somewhat lower, pretty much in line, somewhat higher or much higher. Depending on the group the individual was in, he received this question before or after being asked what his consumption expenditures actually are. The latter is asked by the same question Winter uses in his analysis. It is a bracketed question asked differently for sub-groups: Each sub-group received different bracket positions (the exact wording of the questions can be found in appendix A4.1). The low bracket values are shifted to the left of the expected population distribution, the middle bracket values are centered on the middle of the expected population distribution and the high brackets are shifted to the right (a detailed summary of the bracket values in the different sub-groups can be seen in the appendix, table 4.10). Note that low and high brackets have overlapping bracket values and are therefore used in the subsequent analysis. The medium brackets are constructed differently and therefore we cannot use them in some of our models. For interpretation purposes we leave out these observations. Figure 4.2 provides a summary.

In summary, what we have is one group that is asked the typicality question first (hereafter referred to as experiment 1 with 197 observations) and one that is asked the typicality question afterwards (hereafter referred to as experiment 2 with 183 observations). Within each of these groups respondents were randomly assigned one of the two versions of the expenditures questions. Basic descriptive statistics with respect to the entire sample can be found in the appendix, table 4.11.

Figure 4.2: Set-up of the experiment



How can these experiments help us learn more about why bracket effects occur and how the self-perceived typicality is used as a heuristic?

Experiment 1 focuses on the question whether self-perceived typicality is a heuristic respondents use to derive an answer to the expenditures question. It could also be possible that there is a difference in the bracket effects between those who were asked the typicality question first and those who were asked afterwards. This would mean that the typicality heuristic has to be activated by asking the question and having people think about their typicality. Then it would not be a real heuristic though, because it is not naturally used. If we do not see a difference in the bracket effect between both groups (and show that the heuristic is used) this would mean that it is a heuristic that does not need to be activated.

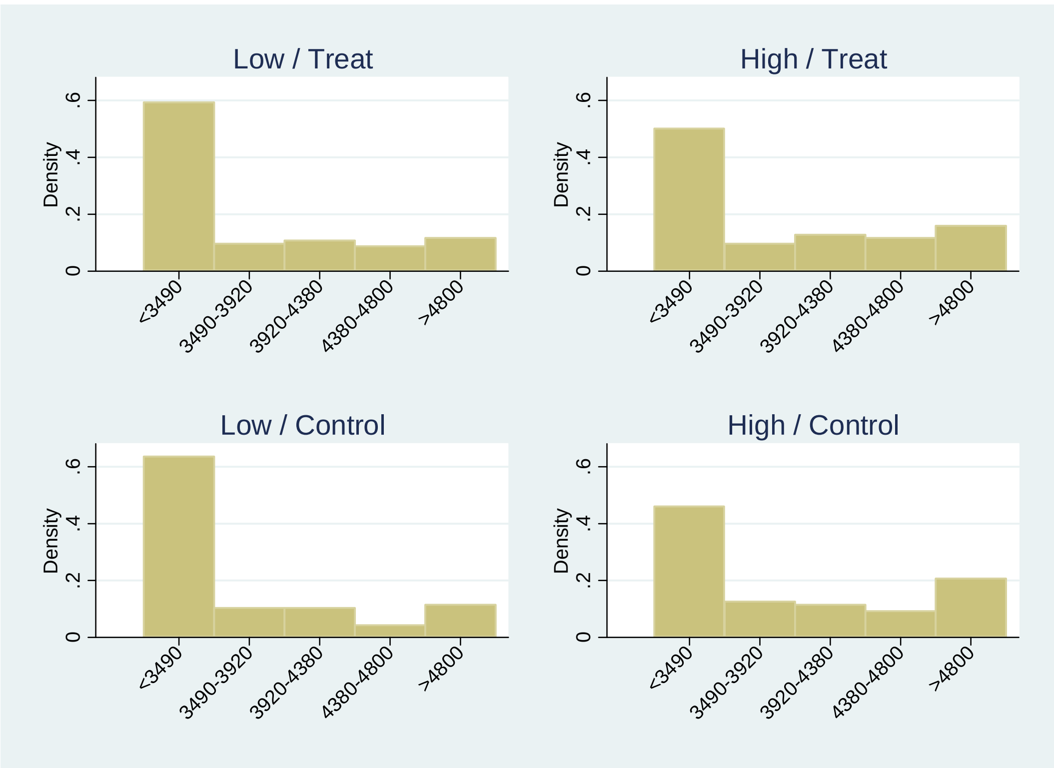
Experiment 2 focuses on the question whether respondents, once they have stated their consumption expenditures, use their answer to derive a different perspective on their self-perceived typicality, depending on the treatment group they are in. In this case we would see different distributions of answers to the typicality question in both groups which would mean that self-perceived typicality is not a stable concept.

4.4 Results

For illustrative purposes we start by presenting a short summary of the bracket effects shown by Winter (2002). For consistency with the remainder of this chapter we use a regression-based approach instead of a chi-squared test for equality of the distributions.

Figure 4.3 shows the distribution of the stated consumption expenditures by sub-group (*treat* refers to experiment 1 and *control* refers to experiment 2). To ensure comparability, the first four brackets in the groups with the low bracket values and the last four brackets in the groups with high bracket values were each combined. As all other brackets have the same interval bounds, we can compare results across groups (details on the construction of the bracket bounds are given in the appendix, table 4.10).

Figure 4.3: Bracket effects by group



The bracket effect itself can be seen relatively easily as the values below 3490 guilders[10] are chosen less often and values above 4800 guilders are chosen more often among those in the groups with high bracket values. Whether the bracket effect is different between treatment and control group cannot be seen as easily – we will therefore proceed with a regression-based analysis.

The regression is set up as follows: The dependent variable is the interval chosen (note that in this case the four lower brackets are combined in the groups with the low bracket values and the four highest brackets are combined in the groups which were subject to the high bracket values, respectively). As the dependent variable is discrete with a natural ordering but with known thresholds, we use an interval regression model. The explanatory variables in specification 1 consist of dummy variables and an interaction term. The variable *treat* is a dummy that takes on the value 1 if respondent *i* was asked the typicality question first and 0 otherwise. The variable *high* is a dummy that takes on the value 1 if respondent *i* received the answer format with high bracket values and 0 otherwise. The product *treat*high* is the corresponding interaction term. Using only these dummies, specification 1 is comparable to the classical difference-in-difference approach and makes the results very easy to interpret. In specification 2 we add a set of control variables even though we have an experiment with random assignment. These are: Square root of household size, a dummy for the presence of children, a dummy for the partner living in the household, logarithm of household net income (per month), sex, age and age squared, a set of dummies to characterize education (low secondary, high secondary and high education, with primary education being the base group) and a set of dummies to characterize the current occupation (housekeeper, retired and paid employment). In our analysis, we focus on characteristics of the household head instead of the respondent as the questions are always aimed at the household rather than at the personal expenditures. The main idea of this analysis is to replicate the identification of the bracket effect but as we distinguish between question orders we can add some remarks to the analysis that was previously done.

Table 4.1 shows the results and we can see the bracket effect by looking at the highly significant and positive *high* coefficient, which is of pretty much the same magnitude in both specifications. What is new is that we can analyze the effect of having asked the typicality

[10] 1 guilder was approx. 0.5 U.S. dollars (USD).

question first and we see that this does not change the results, neither in the group with low bracket values nor in the group with high bracket values, which is shown by the insignificance of the treatment dummy and the interaction term. This result is consistent with our idea that the typicality heuristic (if used) is always present and does not necessarily need to be activated. The last possibility, that typicality has simply no influence at all will be confuted in the remainder of this chapter.

Table 4.1: The bracket effect shown in an interval regression model

	Spec. 1	Spec. 2
treat	135.30	-35.56
	(208.75)	*(182.21)*
high	553.05 **	549.50 ***
	(216.72)	*(190.88)*
treat*high	-264.96	40.76
	(298.88)	*(266.80)*
N	380	346
Controls	no	yes

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses.

When we look at the control variables we can see that root household size, the partner dummy, logarithm of net income, high secondary and high education are significantly positive and that the dummy for children is significantly negative (results not shown).

As a next step we want to analyze where the bracket effect comes from. When we started describing the idea of a typicality heuristic, we referred to the bracket position rather than the bracket value. Therefore, we cannot proceed with looking at the intervals but with analyzing the bracket position respondents chose.

We start by using a similar approach as with the interval regression but some important changes are necessary. First of all, we need a different estimation method as brackets have different interval bounds in the low and the high treatments. In order to model the actual choice of the bracket position we do not want to collapse either the lower or the upper brackets as we are really only interested in the position of the bracket the respondents chose. This is necessary because we want to show that the typicality heuristic works in a way that bracket positions are considered representative of the expected population distribution and are therefore chosen due to their position. We also need to account for the ordinal structure of the outcome variable.

For these reasons we will use an ordered probit approach to model bracket choice. Again, our dependent variable is the bracket position chosen by the respondent (a number between 1 and 8). As explanatory variables (in specification 1) we use the dummy variable *treat* to indicate if one was asked the typicality question first and the dummy variable *high* to indicate whether the respondent received the high bracket values.

Another complication that arises is that using interaction terms in non-linear models is not straightforward – a relatively new discussion also known as the Ai and Norton (2003) problem which was originally about interaction terms in logit and probit models. In our case this is not an issue as there is no theoretical justification for why there should be different decision processes in the treatment groups, i.e. there is no reason to believe – and that was additionally shown in the previous model – that there are differences in the bracket effect between those who were asked the typicality question first and those who were asked the question after having stated their expenditures. Therefore, we do not use an interaction between *high* and *treat*. In summary, we allow for different expenditures among those with different bracket values and also account for the question order. In specification 2 we add the same control variables as before.

Table 4.2 gives the results and table 4.3 and table 4.4 give details on the marginal effects in specification 1 and 2, respectively.

Table 4.2: Bracket choice shown in an ordered probit model

	Spec. 1	Spec. 2
treat	0.03	0.01
	(0.11)	*(0.12)*
high	-0.73 ***	-0.75 ***
	(0.11)	*(0.12)*
N	380	346
Controls	no	yes

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses.

Table 4.3: Average marginal effects in specification 1

Bracket no.:	1	2	3	4	5	6	7	8
treat	-0.009	-0.001	0.000	0.001	0.002	0.002	0.001	0.004
	(0.036)	*(0.005)*	*(0.001)*	*(0.004)*	*(0.006)*	*(0.008)*	*(0.006)*	*(0.015)*
high	0.252 ***	0.033 ***	-0.011 **	-0.035 ***	-0.049 ***	-0.059 ***	-0.039 ***	-0.092 ***
	(0.039)	*(0.009)*	*(0.005)*	*(0.009)*	*(0.011)*	*(0.013)*	*(0.011)*	*(0.018)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Delta-method standard errors given in parentheses.

Table 4.4: Average marginal effects in specification 2

Bracket no.:	1	2	3	4	5	6	7	8
treat	-0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	(0.030)	*(0.005)*	*(0.001)*	*(0.004)*	*(0.005)*	*(0.007)*	*(0.005)*	*(0.013)*
high	0.200 ***	0.035 ***	-0.005	-0.029 ***	-0.040 ***	-0.051 ***	-0.032 ***	-0.078 ***
	(0.033)	*(0.007)*	*(0.004)*	*(0.008)*	*(0.010)*	*(0.012)*	*(0.009)*	*(0.016)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Delta-method standard errors given in parentheses.

The most interesting fact is that the probability that brackets 1 and 2 are chosen in the group with the high bracket values is positive (in both specifications) and highly significant. This is surprising because it works in the opposite direction of the bracket effect: If respondents are more likely to choose the lowest bracket and less likely to choose the higher brackets (from bracket # 3 onwards) this means that they are giving answers that are more truthful than we had previously expected. As we know that the bracket effect exists, there must be another effect that drives bracket choice and explains why higher brackets are chosen in the high treatment.

Many of the covariates and their marginal effects are also significant (results not shown), these include root household size, dummy for partner living in the household, dummy for children, logarithm of net income, age and age squared, high secondary as well as high education, and the occupation dummies for paid employment, retired and housekeeper – so again, no gender effect.

As we want to explain where the bracket effect comes from we need an additional model using the self-perceived typicality as an explanatory variable.

In order to make appropriate statements about the direction of the effect, we do not use the observations from experiment 2 where the typicality question was asked afterwards but only those from experiment 1. We will use experiment 2 later on to show that the self-perceived typicality is a stable concept. By doing so, the model is consistent with our hypothesis that a heuristic is driving the results and we can exclude the possibility of reverse causality and do not detect simple correlation.

The set-up for the next model is as follows: The dependent variable is the bracket position chosen by the respondent. This is as before but our explanatory variables have changed. In specification 1 we do not have a dummy variable for the treatment group as we are only looking at the treatment group and because we have already seen that the outcome does not change between groups. We still include our dummy variable *high* but now we also include the self-perceived typicality "*typi*". In specification 2 we add the same covariates as before and in both cases we stick to the ordered probit approach.

Table 4.5 gives the results and table 4.6 and table 4.7 give details on the marginal effects in specification 1 and 2, respectively.

Table 4.5: Bracket choice and self-perceived typicality

	Spec. 1	Spec. 2
high	-0.89 ***	-0.87 ***
	(0.16)	*(0.18)*
typi	0.74 ***	0.62 ***
	(0.11)	*(0.12)*
N	197	180
Controls	no	yes

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses.

Table 4.6: Average marginal effects in specification 1

Bracket no.:	1	2	3	4	5	6	7	8
high	0.237 ***	0.040 ***	-0.003	-0.039 ***	-0.049 ***	-0.061 ***	-0.044 ***	-0.081 ***
	(0.046)	*(0.011)*	*(0.006)*	*(0.013)*	*(0.014)*	*(0.017)*	*(0.015)*	*(0.021)*
typi	-0.189 ***	-0.028 ***	0.002	0.026 ***	0.034 ***	0.044 ***	0.034 ***	0.077 ***
	(0.021)	*(0.008)*	*(0.004)*	*(0.007)*	*(0.009)*	*(0.012)*	*(0.011)*	*(0.015)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Delta-method standard errors given in parentheses.

Table 4.7: Average marginal effects in specification 2

Bracket no.:	1	2	3	4	5	6	7	8
high	0.193 ***	0.037 ***	0.004	-0.025 **	-0.041 ***	-0.052 ***	-0.040 ***	-0.075 ***
	(0.040)	*(0.010)*	*(0.005)*	*(0.011)*	*(0.013)*	*(0.016)*	*(0.015)*	*(0.021)*
typi	-0.137 ***	-0.021 ***	0.001	0.018 ***	0.024 ***	0.030 ***	0.025 ***	0.060 ***
	(0.024)	*(0.006)*	*(0.003)*	*(0.006)*	*(0.007)*	*(0.010)*	*(0.009)*	*(0.013)*

*, **, *** indicate significance at the 10%, 5%, 1% levels. Delta-method standard errors given in parentheses.

Now we are able to see where the bracket effect comes from. After controlling for the fact that the two groups received different bracket values, we can see that respondents that state a higher typicality also tend to choose higher brackets.[11] The result is twofold: Respondents that were presented with higher bracket values tended to choose lower bracket positions (which makes sense) but those with a higher self-perceived typicality tended to choose a higher bracket position. The latter works in the opposite direction of the first effect and therefore offsets it to some extent. Regarding the control variables (and their marginal effects) it is interesting to see that now only net income and the partner dummy are significant (results not shown) – the rest is no longer significant – so this is not driving the results.

4.5 Validity concerns

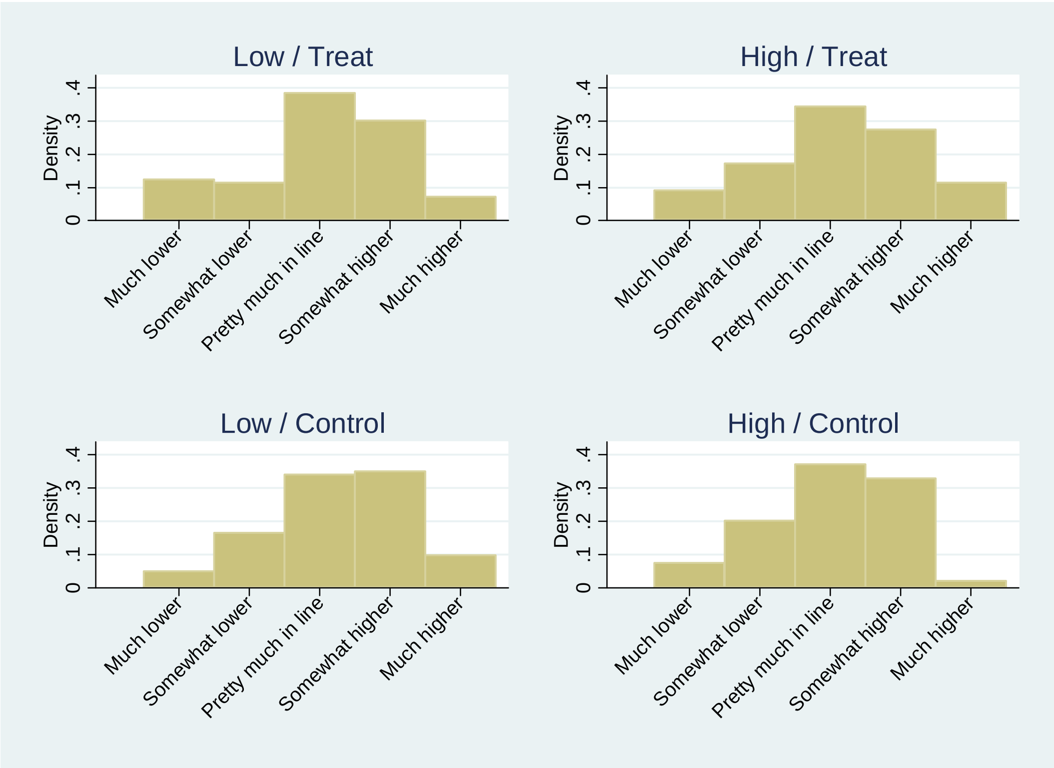
For a better interpretation of the results we would want the self-perceived typicality to be a stable concept. That is, we would like to see that respondents who answered the expenditures question first do not systematically report a different typicality than those who stated their typicality first and then answered the expenditures question. Additionally, we would like to see that those who received the high bracket values do not report a different typicality than those who received the lower bracket values, i.e. there should be no

[11] We replicated this finding in a similar laboratory experiment at LMU Munich's Experimental Laboratory for Economic and Social Sciences (MELESSA). Students were asked their self-perceived typicality with respect to cell phone use and cell phone expenditures and then their own actual use and expenditures (N=79). Again, a higher self-perceived typicality is positively (and significantly) associated with choosing higher brackets.

equivalent to the bracket effect in the self-perceived typicality. In order to address these questions we use experiment 2 as the robustness check.

Figure 4.4 shows the distribution of self-perceived typicality in the different sub-groups. There seems to be little change within treatment and control group – but some between these groups.

Figure 4.4: Distribution of self-perceived typicality



We continue by setting up an ordered probit model with self-perceived typicality as the dependent variable. In specification 1 we have a dummy variable for the treatment group, i.e. those who had to answer the typicality question afterwards. In specification 2 we add the same covariates as before. This is a first step to check whether the results differ. Table 4.8 gives the results. In both specifications the treatment dummy and its marginal effects are insignificant, i.e. there is no difference between groups on average.

We have not added the indicator for *high* here as only the treatment group answered the typicality question afterwards and therefore we would only expect to see an effect there. In the previous analysis this was different as we were looking at the bracket chosen by the respondent which is an immediate effect of the brackets presented. Additionally, reverse causality is not very intuitive as you might not think of bracket values first when you are asked for your typicality.

Table 4.8: Self-perceived typicality by experimental group

	Spec. 1	Spec. 2
treat	-0.03	-0.01
	(0.11)	*(0.12)*
N	380	346
Controls	no	yes

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses.

To check whether the bracket values given play a role, we have to restrict our sample to those who were in the treatment group. The problem with this analysis is that as soon as we add the dependent variable from the previous analyses (bracket position or value chosen by the respondent) as an explanatory variable, we would only identify some correlation, not causality and would therefore not be able to interpret these results and this explanatory variable would also contain measurement error as shown before. Thus, as the last step, we consequently use only one dummy variable, i.e. the dummy variable for *high*. Table 4.9 summarizes the results.

Table 4.9: Self-perceived typicality by bracket treatment

	Spec.1	Spec.2
high	0.07	0.02
	(0.16)	*(0.18)*
N	183	166
Controls	no	yes

*, **, *** indicate significance at the 10%, 5%, 1% levels. Cluster-robust standard errors given in parentheses.

As we had expected, there is no significant difference (neither by parameter nor by marginal effect) between the two groups with the different bracket values. We consider this to support our hypothesis that there is no equivalent to the bracket effect and that therefore the self-perceived typicality is a stable concept.

Nevertheless, another type of experiment is needed to rule out the reverse causality that might potentially be present: One would need a questionnaire where people are asked to state their typicality first, then answer (some) consumption expenditures questions and then are asked the typicality question again. The format of the (bracketed) expenditure questions should hereby be manipulated in the same way as before, with overlapping values but different bracket positions. It would be interesting to see if people stick to their original answer about their typicality.

4.6 Conclusions

We contributed to the survey research literature by showing that respondents use their self-perceived typicality as a heuristic when there is uncertainty and that this helps to understand and explain the bracket effects found in previous research. Respondents think of the brackets given (in a survey) as a representation of the underlying population distribution and also know what their typicality is. These two things taken together help them to derive an answer to complicated questions, such as consumption expenditures, that require a high level of concentration to recall and aggregate the required amounts. In this chapter we have shown that persons that receive higher bracket values also tend to choose lower bracket positions (which is correct) but also that persons tend to choose higher bracket positions when their self-perceived typicality is high. Furthermore, we presented arguments to rule out simple correlation in favor of causality. There is no difference in bracket choice (and consumption) between those who received the typicality question before and those who received it afterwards which can be interpreted in a way that the typicality heuristic does not need to be activated. Additionally, belonging to one of the experimental groups and the bracket format given to the interviewee do not lead to differences in self-perceived typicality and we therefore conclude that it is a stable concept.

Future research should try to prove causality with additional experiments. Also, larger samples would help to underline the findings in this chapter. It would also be interesting to artificially vary the level of uncertainty to see how this changes the use of heuristics and therefore the outcome.

4.7 Appendix

A4.1 Question wording as in the English version of the questionnaire

Lead-in:

Think about how much you and your household spent on everything in the past month. Please think about all bills such as rent, mortgage loan payments, utility, insurance and other bills, as well as all expenses such as food, clothing, transportation, entertainment and any other expenses you and your household may have.

Typicality question:

When you compare your own monthly household consumption with the consumption of an average Dutch household, do you think your own consumption is pretty much in line with the consumption of the average household, or is it very different from the consumption of the average household? Please check one of the following answers. Would you say that your household consumption is (please check one):

__ *Much lower than the typical consumption*

__ *Somewhat lower than the typical consumption*

__ *Pretty much in line with the typical consumption*

__ *Somewhat higher than the typical consumption*

__ *Much higher than the typical consumption*

Consumption expenditures question:

Roughly, how much would your household consumption last month amount to? Please indicate the range in which the amount most likely falls:

__ *Less than* $d_1(x)$

__ *Between* $d_1(x)$ *and* $d_2(x)$

__ *Between* $d_2(x)$ *and* $d_3(x)$

__ *Between* $d_3(x)$ *and* $d_4(x)$

__ *Between* $d_4(x)$ *and* $d_5(x)$

__ *Between* $d_5(x)$ *and* $d_6(x)$

__ *Between* $d_6(x)$ *and* $d_7(x)$

__ *More than* $d_7(x)$

Note: The values d1 to d7 differ with respect to the bracket values the individuals received and are given in the next table.

Table 4.10: Bracket values

	Group with low bracket values	Group with high bracket values
d1	1890	3490
d2	2590	3920
d3	3080	4380
d4	3490	4800
d5	3920	5340
d6	4380	6130
d7	4800	7710

Note: All values in guilders. Values for the medium bracket group are not reported as they are not used in our analysis. This is because the medium bracket group intervals do not overlap with the intervals in the low and the high bracket groups.

Table 4.11: Basic descriptive statistics

	Experiment 1		Experiment 2	
	Mean	Std. dev.	Mean	Std. dev.
Household size	2.422	1.406	2.440	1.507
Children (B)	0.356	0.480	0.367	0.484
Partner (B)	0.678	0.469	0.657	0.476
HH net income	8081.333	13950.340	5116.584	6305.386
Sex (B)	0.239	0.428	0.229	0.421
Age	46.728	14.571	44.717	14.102
Low secondary education (B)	0.172	0.379	0.175	0.381
High secondary education (B)	0.517	0.501	0.506	0.501
High education (B)	0.161	0.369	0.187	0.391
Housekeeper (B)	0.039	0.194	0.018	0.134
Retired (B)	0.161	0.369	0.127	0.333
Paid employment (B)	0.678	0.469	0.741	0.439
N	180		166	

Note: (B) indicates a binary variable. Sex is 1 if female and 0 if male.

Chapter 5

5 A European perspective on expectations about personal living conditions as determinants of subjective well-being[12]

Lessons learned from the Flash Eurobarometer

5.1 Introduction

Nowadays research on subjective well-being (SWB) is widely accepted as being relevant – not only for a specialized survey-research community but in economics in general. We can see papers concerned with this matter being published throughout specialized and general interest journals. The motivation here seems quite obvious: At the end of the day people want to live their lives in dignity and researchers try to understand what influences and determines a happy life. Helliwell (2003) traces this motivation back to Aristotle and gives a brief summary of the ancient view. The importance of SWB has even reached real world politics: A group of non-governmental and governmental organizations, including the European Commission, launched the “Beyond GDP” initiative in 2007. This initiative aims at adding social indicators (e.g. climate change, poverty, health and quality of life) to GDP in

[12] This chapter is based on joint work with Joachim Winter and Melanie Lührmann.
We would like to thank Femke De Keulenaer, Andrzej Pyrka, Aurélien Renard, Nicolas Scharioth and Jan Sonnenschein from *The Gallup Organisation* for their helpful comments. Part of this work was conducted during a visit at *The Gallup Organisation* in Brussels, Belgium.

order to achieve a better and more adequate measurement of a country's wealth and progress.[13]

This of course does not mean that there is no dissent left in well-being research, especially when it comes to measurement. As one of the more recent examples, Hamermesh (2004) is very critical with respect to regressions that contain subjective measures on the left-hand and on the right-hand side of the regression equation.

One prominent part of the SWB literature explores the role of wealth and income. This discussion goes back to the well-known "Easterlin paradox" which describes that higher income is related to more happiness at the individual level but that rising income does not lead to more happiness at the aggregate level and that richer countries are not necessarily happier than poorer countries (e.g. Easterlin (1973, 1974, 1995)). Without going into great detail, there is also literature that suggests that there actually is a correlation between (national) income and SWB. Deaton (2008) finds that high per capita GDP is associated with high life satisfaction. This is still in accordance with Easterlin's reasoning of a relative comparison with others as people could potentially use a global standard for comparing themselves with others. Stevenson and Wolfers (2008) show that economic growth is also related to increased happiness. As laid out later, the question is not so much if income is related to well-being at all. It is more important to see how strong this relationship actually is and whether we can see an effect only at the individual or also at the country level and if so, whether it is true in general or only for a certain group of countries. It is also heavily discussed whether it is absolute or relative income that matters – unfortunately, this is a question that needs to be based on data (e.g. time series) that we do not have here.

Besides the literature that is concerned with international comparisons, the probably most influential part of the well-being literature exploits variation within a single country. Reasons could be that external shocks and exogenous variation in the regressor of interest are more likely to be found by using data from a specific country and that literature that relies on one of the high quality American datasets receives more attention.

Nevertheless, if we want to move beyond GDP as the standard indicator for development and if we want to learn more about heterogeneity in the European Union, there is no way

[13] http://www.beyond-gdp.eu

around international datasets. In other words: We need international data for external validation of hypotheses such as the relationship between income and well-being to make far-reaching policy actions and initiatives to move beyond GDP more credible and meaningful.

We use a large and recent dataset from the European Union – a set of countries that can plausibly be grouped together as they share individual characteristics, political understanding and dependence on central policy actions at least to a larger extent than countries from different continents with different ethnicities. This dataset mainly consists of five consecutive waves of the Flash Eurobarometer survey (kindly provided by *The Gallup Organisation*). The fieldwork was conducted between July 2009 and October 2010 and this makes the dataset even more interesting as this period falls into the world economic crisis (or financial crisis). Compared to the majority of previous literature, our focus is on different aspects of subjective well-being. As we do not have data on typical measures like life satisfaction, we concentrate on existential fears such as financial and professional expectations for the future as important determinants of subjective well-being. Besides that, we make use of the micro-structure of the underlying dataset and try to point out variation across countries, over time and within the European Union as well as heterogeneity between groups across countries. We do not present evidence on the relationship between our well-being measures and wealth on the micro level as this is something that never really was doubted and because (as we will explain later) our dataset only contains a few and mostly subjective control variables so that it is not possible to present a state-of-the-art microeconometric analysis.

The remainder of this chapter is structured as follows: Section 5.2 gives an overview of findings in previous work and highlights other potential data sources. Afterwards, section 5.3 provides details on our data. This is followed by section 5.4 where we analyze the relationship between well-being (worries about income in old age, confidence in keeping the job and the self-rated probability of finding a job if one was to be laid off) and wealth (measured by GDP and subjective living standard). We find that wealth explains a lot of the cross-country differences while the results remain very stable over time for the first two variables. Furthermore, we show that the overall distribution of well-being differs by age group (especially for finding a job if laid off) and that the gradient of the relationship between worries about income in old age and GDP also depends on age. In section 5.5 we highlight a close relationship between the average perceived poverty rate in a country and its average subjective living standard. We argue that this supports the idea that people use an internationally comparable standard for answering such questions. Finally, section 5.6 concludes.

5.2 Previous findings in related literature

The seminal work by Easterlin and the formulation of what became known as the Easterlin paradox has had a huge impact on economic research. The paradox consists of two seemingly incompatible findings: Firstly, in a given country at a given point in time he finds that richer people are happier, secondly he finds that this association does not hold between rich and poor countries and within a country over time (there are many papers but Easterlin (1974) is his earliest data-based contribution even though it was published one year after the famous Easterlin (1973) paper which in fact relates back to the original work). Easterlin argues that the reason behind all this is a relative perception of wealth. Accordingly, richer individuals are happier than poorer ones but if the reference point rises for all over time there is no increase in happiness. In a later contribution Easterlin (1995) states that there is indeed an association of income and well-being in cross-country comparisons.

Afterwards, a huge amount of related literature followed so we restrict ourselves to some relevant examples that describe its development. Early contributions are of course somewhat limited with respect to international (and comparable) data but the amount of useful sources has grown over time (e.g. Andrews and Inglehart (1979) use data from only nine (western) countries). Some scholars even questioned the idea that well-being is based on relative comparisons. Veenhoven (1991), as an example, argues that happiness is at least to some extent not relative when it depends on the satisfaction of basic needs. Diener *et al.* (1993) find that income is indeed associated with well-being within a country and also between countries and that relative standards have no influence. They also favor the idea that individuals (if at all) compare themselves to their peer-group rather than other groups (which gets harder to test the more general the group is). At the end they conclude that their results do not support Easterlin's "relative" approach and that they also do not fully support Veenhoven's explanation of (basic) need gratification. The analysis was based on two studies, one with US national probability sample data and another one with data from students in 39 countries. Later on, a literature review by Diener and Biswas-Diener (2002) states that the correlation of wealth and SWB between nations is large and that it is small within countries and also that economic growth in developed countries is only weakly related to an increase in SWB. However, literature in favor of the reference level has also evolved over the years. Clark and Oswald (1996) use British Household Panel Survey (BHPS) data to show that well-being (job satisfaction) depends on reference rather than actual income – which is obviously a finding at the country level and does not tell us anything about cross-country differences. Deaton (2008) exploits Gallup's 2006 World Poll which consists of data from over 100 countries. His main findings are that richer countries are more satisfied than poorer countries and that this relationship also holds for a comparison of countries with a high income (in contrast to the basic needs theory). His explanations are consistent with the Easterlin Paradox as he argues that there could be a global standard for comparison (due to the globalization of information). Based on Vogel's (1999) work, Christoph and Noll (2003) have a slightly different tenor in their analysis of European data (Eurobarometer and European Community Household Panel (ECHP) data) as they investigate the similarity of country clusters that share certain characteristics. This seems to be a logical consequence to all the discussions about whether or not there are differences in well-being between countries. Maybe there are differences between clusters but not within a certain cluster (e.g. clustering by welfare systems results in a Nordic, Southern and a Central European cluster). There seems to be a tendency that the Nordic cluster is better off than the Central cluster which is in turn better off than the Southern cluster. Nevertheless, there are too many exceptions to that rule to conclude that this is a general pattern. Additionally, sub-group (e.g. age) differences in well-being hardly follow any pattern here which does not work in favor of the common characteristics hypothesis. On a global level Helliwell (2003) proceeds in a similar way with the World Values Survey data by grouping countries under the assumption of similarity of neighboring countries and finds at least some support for this in the data. The most extensive work in assessing the Easterlin paradox is probably the paper by Stevenson and Wolfers (2008). They analyze many different datasets (including World Values Survey, Gallup's World Poll, Eurobarometer and many more). Not only do they find a well-being / income relationship within and between countries but also over time. Interestingly, the magnitude of the coefficient is quite robust over these specifications. With respect to the ongoing discussion it is important to mention that their findings do not work in favor of the satiation point theory (basic needs) as the relationship between income and well-being not only holds for poor, but also for rich countries. Additionally, they interpret their results as (at least) a hint towards relative income not being the only or even most important driver behind well-being but that absolute income does in fact also matter.

It is quite obvious that most of the relationships being established in this literature measure correlations which are definitely interesting per se. Fortunately, the general tendency in the findings is also supported by studies that try to establish a causal relationship between income and SWB. For example Frijters *et al.* (2004) use the German reunification as a source of exogenous variation and find that 35-40% of the increase in life satisfaction was due to the increase in household income.

This summary shows that there is already a large body of literature. The more recent the literature is, the less there seems to be a debate whether there is a cross-country relationship between SWB and income or not. It rather seems that the results actually depend on the data, i.e. the data source, the survey date or the set of countries used in the analysis. We contribute by extending previous work in terms of more recent data and new methodological aspects.

5.3 Data

In this work our main data comes from the Flash Eurobarometer surveys, with kind permission of *The Gallup Organisation*. As these surveys are less known (and set up differently) than other household surveys such as the GSOEP or SHARE, it is worth outlining the most important facts here. In general, the Flash surveys are conducted upon request of the European Commission and their content varies broadly – always depending on the specific topic of interest. Examples of these topics include “Views on the European Union Enlargement” or “Euro Attitudes”. For the purpose of our analysis we use the Flash Eurobarometer survey “Monitoring the social impact of the crisis: Public perceptions in the

European Union" which consists of five consecutive waves.[14] A very important aspect of this dataset is that it contains representative information from all 27 member states of the European Union (EU). The interviews were conducted throughout the (later) financial crisis in July 2009 (wave 1), November / December 2009 (wave 2), March 2010 (wave 3), May 2010 (wave 4), and October 2010 (wave 5) by various local fieldwork organizations in the respective language. The overall sample in each wave contains more than 25,500 observations.

As interesting as this is per se, there is a (natural) limitation in this kind of ad-hoc survey in terms of the number of questions that can be included. Here, each wave asks eight questions regarding some demographics and 19 questions on individual perceptions or the personal situation. Even though we know, for example, the respondent's sex and age we have to rely on subjective statements when it comes to the personal living standard or the area a person lives in. In the Flash surveys, living standard is measured by the question:

"On a scale from 1 to 10, where would you place the current living standards of your household? Please choose one number from 1 to 10, where "1" stands for "very poor", and "10" stands for "very wealthy", while the remaining numbers indicate something in between these two positions."

One could argue that this is an inadequate wealth measure but we will be able to check that hypothesis on an aggregate level with additional data. Moreover, a single income or wealth measure is not necessarily better (as a related example see Winter (2004) who shows that consumption expenditures stated in a one-shot question differ from those derived from many disaggregated questions) and Diener and Biswas-Diener (2002) outline what is actually needed as a wealth measure when it comes to the evaluation of subjective well-being. In their opinion income is just a small part of material well-being as it sheds no light on the individual access to subsidized goods, spending efficiency, local prices, etc. and that it therefore does not measure the material quality of life. Maybe such a measure of self-perceived living standard does the job even better than a simple undifferentiated income measure.

[14] Questionnaires, analytical reports, and fieldwork information are available online: http://ec.europa.eu/public_opinion/archives/flash_arch_en.htm

Additionally, we retrieved 2009 GDP per capita (measured at purchasing power parity in constant 2005 international dollars) from the World Bank's online database and added it as an objective country level economic indicator.[15]

5.4 Results

The graphs presented in this section typically show conditional means plotted against a macro measure (logarithm of GDP per capita measured at purchasing power parity in constant prices). Conditional means are calculated by using regressions of the dependent variable (e.g. worries about income in old age) on the most important demographics (i.e. age, sex, education, job-type, wealth, household size and housing area) in order to extract a plausible sample (that does not include respondents with missing values for these demographics). The regressions that do not focus on a certain age group or wave include observations from all five waves and the only observations that are excluded are persons outside the age interval of 18 to 80 years and those who are still in education. Job market questions additionally exclude all respondents without a professional activity. All regressions (and the calculation of the means of the dependent variables as well as the cumulative distribution functions) are weighted using either country or EU weights.[16] The graphs shown also present the means of the dependent variable (and the corresponding macro measure) for the entire EU but these values were not used in the cross-country regressions that follow. In our subsequent analysis we concentrate on the following three determinants of well-being (the exact wording of these questions is provided in appendix A5.1):

1. Worries about income in old age: 1 (not worried at all) – 10 (very worried)
2. Confidence in keeping the job: 1 (very confident) – 4 (not at all confident)
3. Subjective probability of finding a job if one was to be laid off: 1 (not at all likely) – 10 (very likely)

---

[15] http://data.worldbank.org

[16] Details on the construction of weights can be found in the analytical reports: http://ec.europa.eu/public_opinion/archives/flash_arch_en.htm

We focus on different types of variation in our variables: Across countries (with all waves pooled), across sub-groups (at the European level with all waves pooled), between sub-groups across countries (with all waves pooled), and across countries over time. Additionally, we compare our subjective wealth measure with GDP as the standard macro indicator and also present the correlation of such a subjective measure with people’s perception of poverty at the country level.

5.4.1 Cross-country comparisons

Following Easterlin’s idea, we start with the relationship between the above mentioned determinants of well-being and GDP (as our macro indicator) across countries and show what happens when subjective wealth statements are used instead of GDP.

Figure 5.1: Worries about income in old age / GDP (waves pooled)

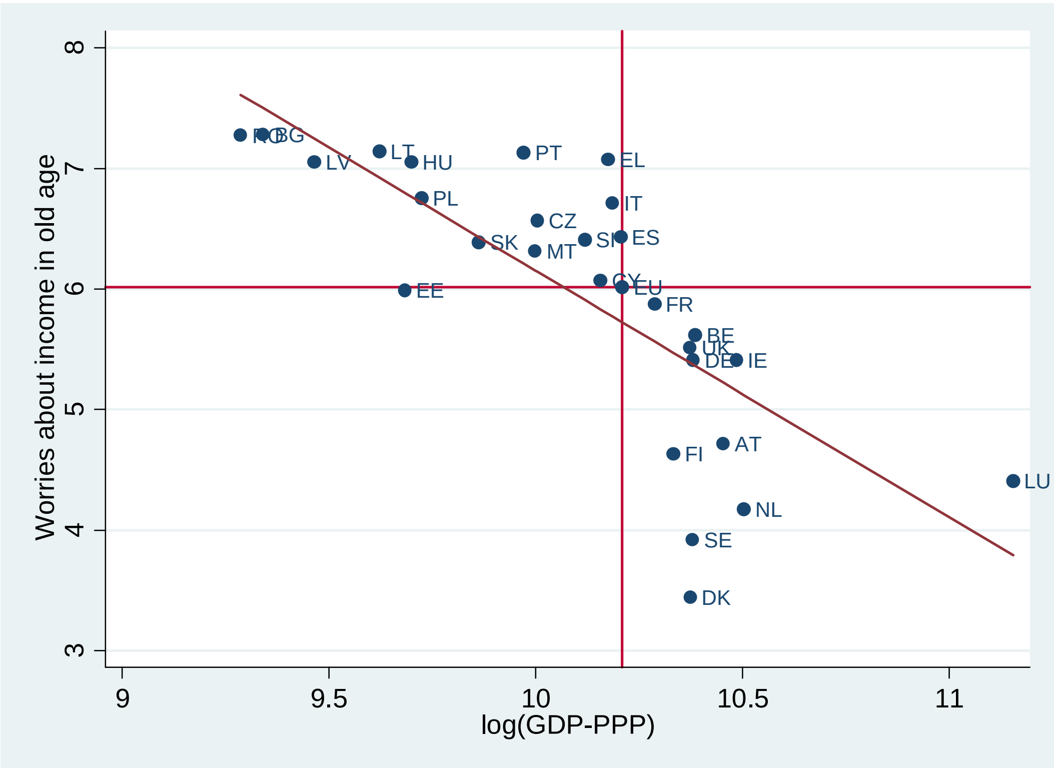


Figure 5.1 plots the country's average of the individual worries about income in old age (pooled over all five waves) against the logarithm of GDP (per capita, measured at purchasing power parity (PPP) and in constant prices). The horizontal and vertical reference lines on the y- and x-axis indicate the EU average, which makes it easier to distinguish well-performing countries from worse ones. A list of country names and their abbreviations is shown in the appendix, table 5.10. The line drawn in the graph shows the fitted values from a regression of the country averages on GDP (excluding the EU average).

Two things can immediately be seen. Firstly, average worries decline with rising GDP so there are clear cross-country differences. Secondly, the upper left and the lower right quadrant contain (almost) all data points which means that all countries with below average GDP have above average worries and vice versa. The figure gives little evidence that neighboring countries necessarily form clusters as for example Sweden, Finland and Denmark have a similar GDP but large differences in their average worries, whereas Romania and Bulgaria lie closer to each other. Nevertheless, something else becomes evident here: All eastern countries and those especially distressed by the financial crisis (Portugal, Italy, Greece and Spain) lie in the worst possible quadrant (Ireland is the exception) and all other large industrial western countries lie in the best possible quadrant. The exceptionally high per capita GDP of Luxembourg is still visible in the figure but as GDP is used in logarithms, the problem is weakened. The quantitative analysis strongly supports the relationship described so far: The slope of the regression line in figure 5.1 is -2.05. With a $|t|$-statistic of 7.44 (robust standard errors), the relationship is highly significant at the 1% level. The correlation between average worries and GDP is -0.76 (p-value < 0.01).

Figure 5.2 presents the same set-up but in this case the country average of subjective living standard is used as the explanatory (x-axis) variable. The conclusions drawn from this second figure are close to those before with one notable exception. The relationship is now much more linear, especially Luxembourg lies closer to the estimated relationship shown by the regression line. Additionally, it is no longer true that all countries lie either in the upper left or lower right quadrant (e.g. France) but this is just a minor discrepancy. Again, the quantitative analysis supports what we have described so far. The slope of the regression line is -1.53 and significant at the 1% level (with a $|t|$-statistic of 8.62). The correlation between average worries and subjective living standard is -0.88 (p-value < 0.01).

Figure 5.2: Worries about income in old age / subjective living standard (waves pooled)

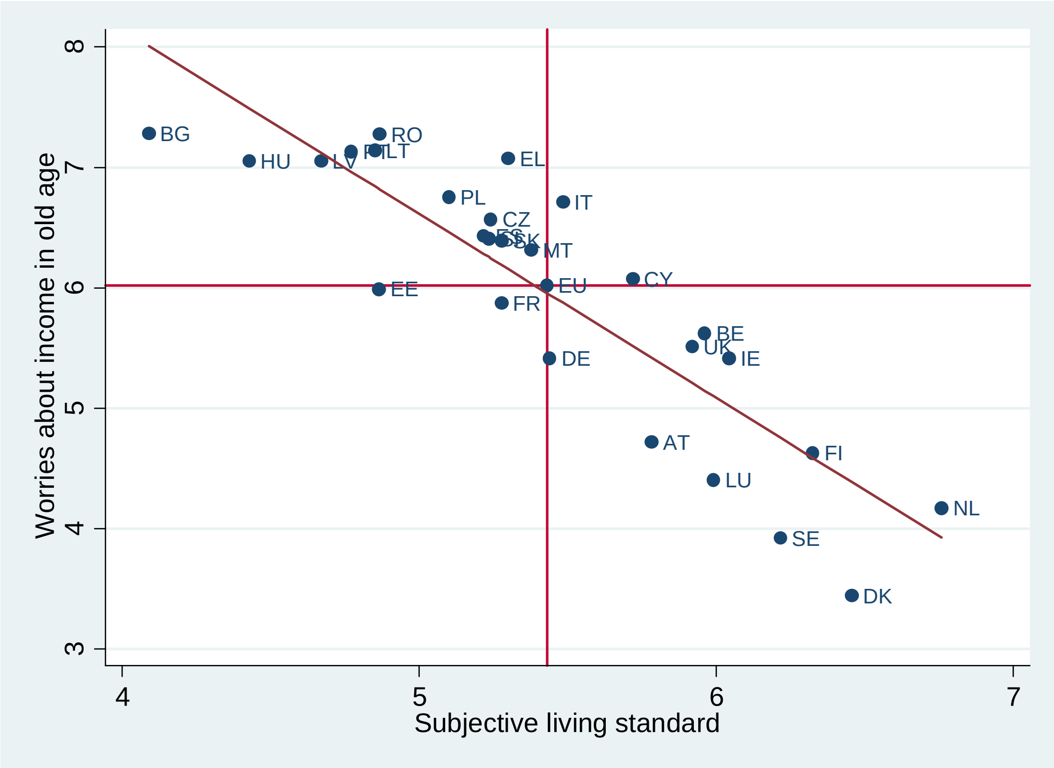
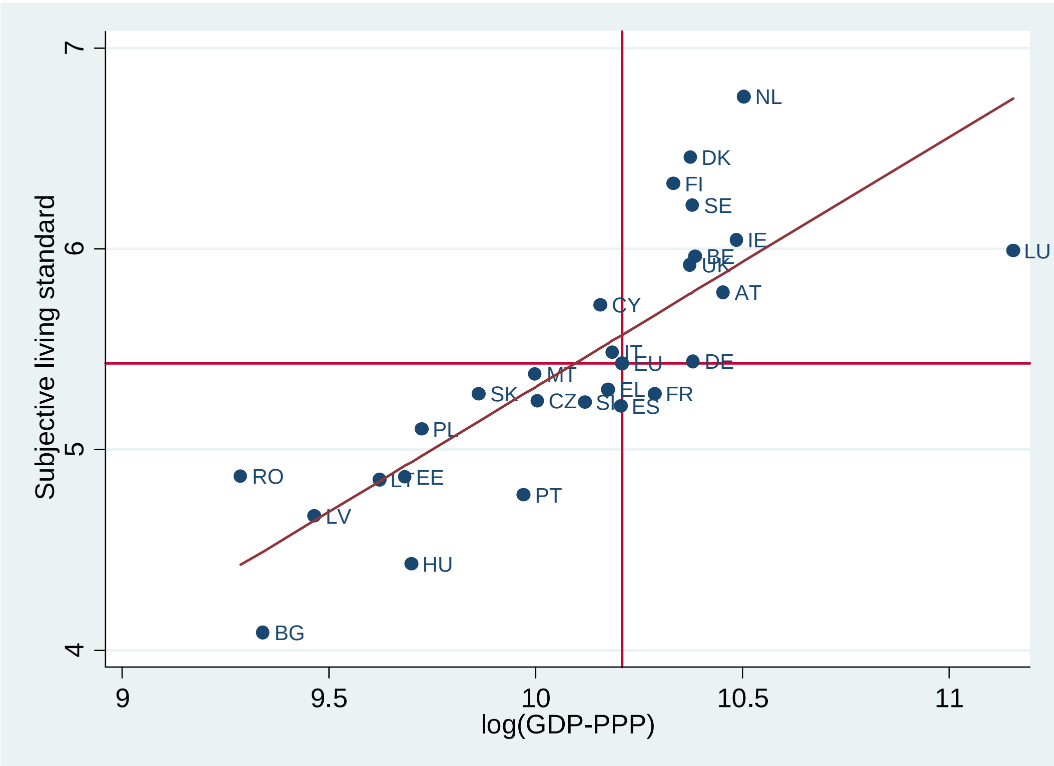


Figure 5.3: Correlation of subjective living standard and GDP (waves pooled)



Finding only small differences is not too surprising as the correlation of subjective living standard and GDP is 0.80 (p-value < 0.01) in this case. Figure 5.3 illustrates this relationship in a scatter plot with the according regression line (the slope is 1.24 and with a |t|-statistic of 5.09 it is significant at the 1% level). As expected, the relationship is very clear but, again, Luxembourg is an outlier due to the very high per capita value of GDP – even though the logarithm is used. It turns out that at least in this case we can actually learn a lot from explaining one subjective measure by another one. This does not mean that Hamermesh (2004), who states the opposite, is wrong. Here, we focus on simple correlations, not on consistent microeconometric estimation. So we follow Diener and Biswas-Diener (2002) and ask ourselves if subjective living standard could actually be an appropriate measure of material well-being.

The next dependent variables of interest outline another crucially important aspect of one's life. During a (financial) crisis people might ask themselves what would happen if they were to be laid off. The Flash Eurobarometer explicitly asks how individuals rate their confidence in keeping their job over the next twelve months and how respondents rate the probability of getting a new job within the next six months *if* they were to lose their job.

The relationship between the individual confidence in keeping the job (country average – a higher value indicates lower confidence) and GDP and subjective living standard can be seen in figure 5.4 a) and b), respectively. Higher GDP and higher subjective living standard are both associated with a higher confidence in keeping the job. Similar to before (with only some exceptions) the countries are nicely distributed over the quadrants. The worst possible (upper left) quadrant contains the eastern European countries as well as Spain, Greece and Portugal whereas Italy's position remains somewhat unclear as it lies in the middle and Ireland is a clear exception as it represents above average GDP (or living standard) but average confidence. Table 5.1 quantifies the graphical analysis of the relationship of confidence in keeping the job (country average) and the two explanatory variables logarithm of GDP (at purchasing power parity) and subjective living standard. Slopes of the regression lines and the correlation coefficients are presented. It can be seen that both the macro indicator and the subjective measure are negatively and significantly (1% level) related to average confidence, i.e. the wealthier a country is, the more confidence people have in keeping their job – even during a severe global crisis.

Figure 5.4: Confidence in keeping job / GDP and subj. living standard (waves pooled)

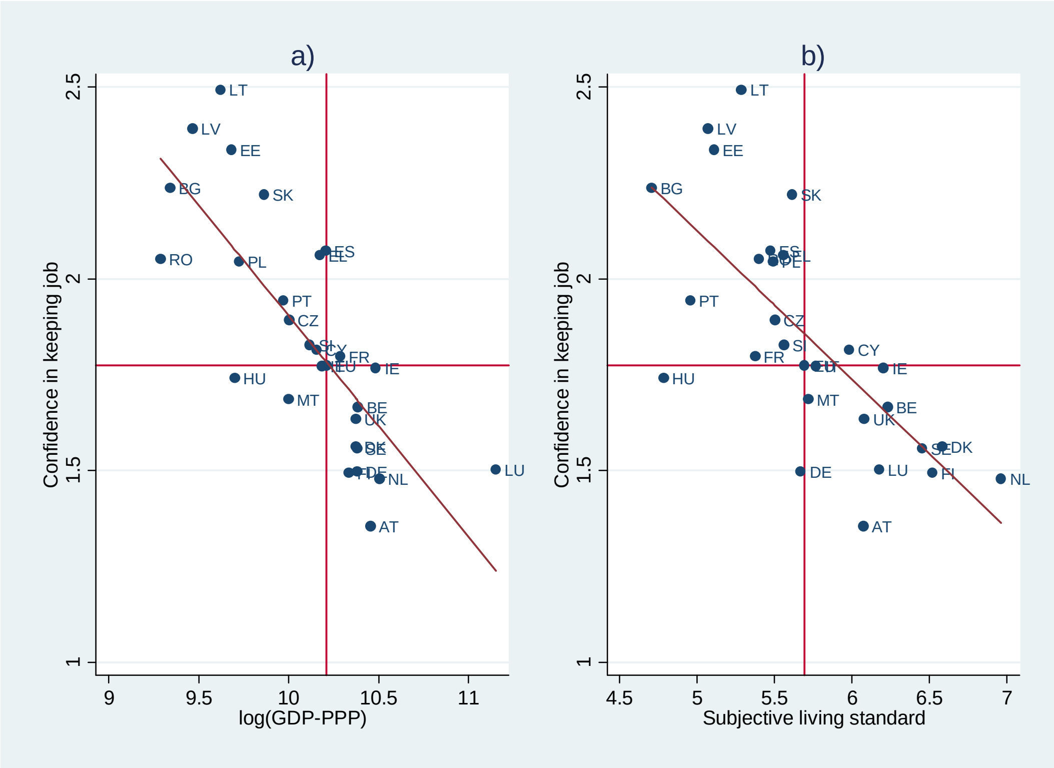


Table 5.1: Regression results and correlations (confidence in keeping job)

Dependent variable: Confidence in keeping job

Explanatory variable	Parameter estimate	|t|-statistic	Correlation	p-value
log(GDP)	-0.58 ***	5.23	-0.78	< 0.01
Living standard	-0.39 ***	5.13	-0.72	< 0.01

*, **, *** indicate significance at the 10%, 5%, 1% levels.

As a next step we are interested in the cross-country relationship of the self-perceived probability of finding a job if laid off and GDP (figure 5.5 a)) as well as subjective living standard (figure 5.5 b)). In this case the relationship is less clear-cut than in the previous cases. First of all, the countries are not simply split into a group that is contained in the worst possible quadrant (lower left) and another one that is contained in the best possible quadrant (upper right). Instead there are many ambiguous cases (e.g. Poland) with below average GDP (and living standard) but an above average self-perceived probability of finding a job if laid off. Nevertheless, some patterns remain consistent, e.g. Baltic countries lie at the lower end and northern countries (e.g. Denmark and Sweden) lie at the upper end of the regression line.

Figure 5.5: Find a job if laid off / GDP and subjective living standard (waves pooled)

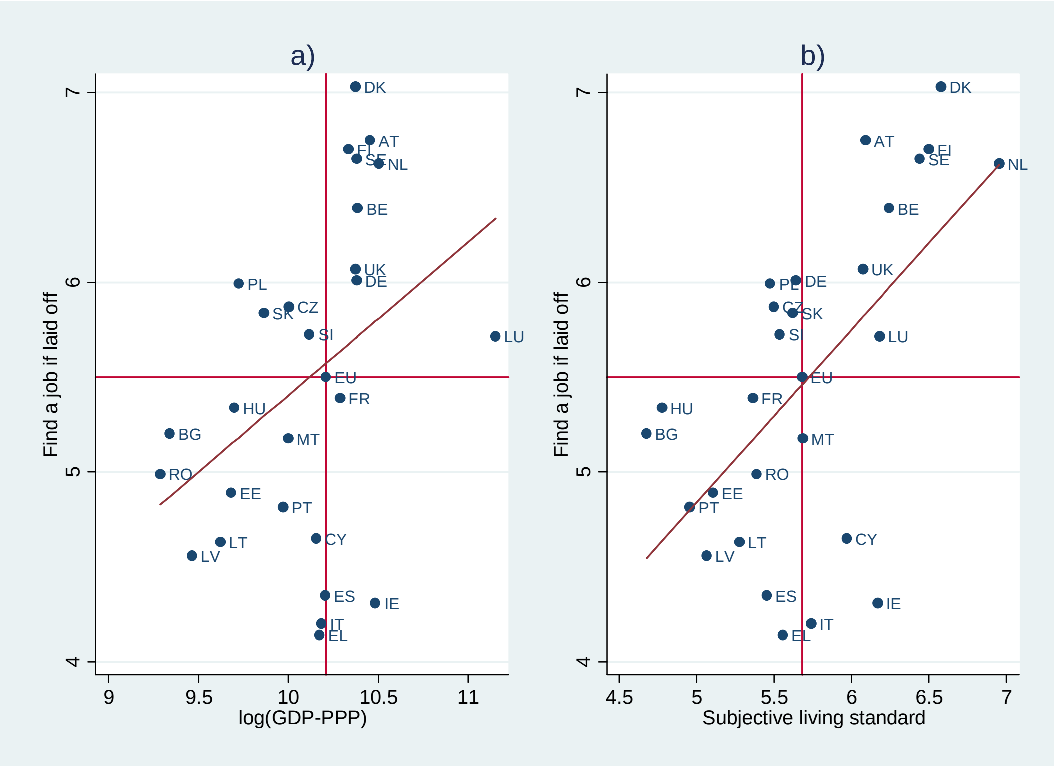


Table 5.2: Regression results and correlations (find a job if laid off)

Dependent variable: Find a job if laid off				
Explanatory variable	**Parameter estimate**	**|t|-statistic**	**Correlation**	**p-value**
log(GDP)	0.81 **	2.68	0.38	0.048
Living standard	0.91 ***	4.55	0.59	< 0.01

*, **, *** indicate significance at the 10%, 5%, 1% levels.

Table 5.2 gives the quantitative results for the previous figure and the weaker relationship becomes obvious. In the case of GDP as the explanatory variable, the regression coefficient is (only) significant at the 5% level and the correlation is comparatively low with a large p-value. When living standard is used as the explanatory variable, the slope parameter is highly significant and the correlation increases substantially with a p-value that is again below 0.01. Taken all together, the relationship of the self-perceived probability of finding a job if laid off and GDP on the one hand and subjective living standard on the other hand remains similar no matter what measure is used but the relationship (in the case of GDP) is weaker than in the previous cases and we do not observe a splitting of countries into two quadrants.

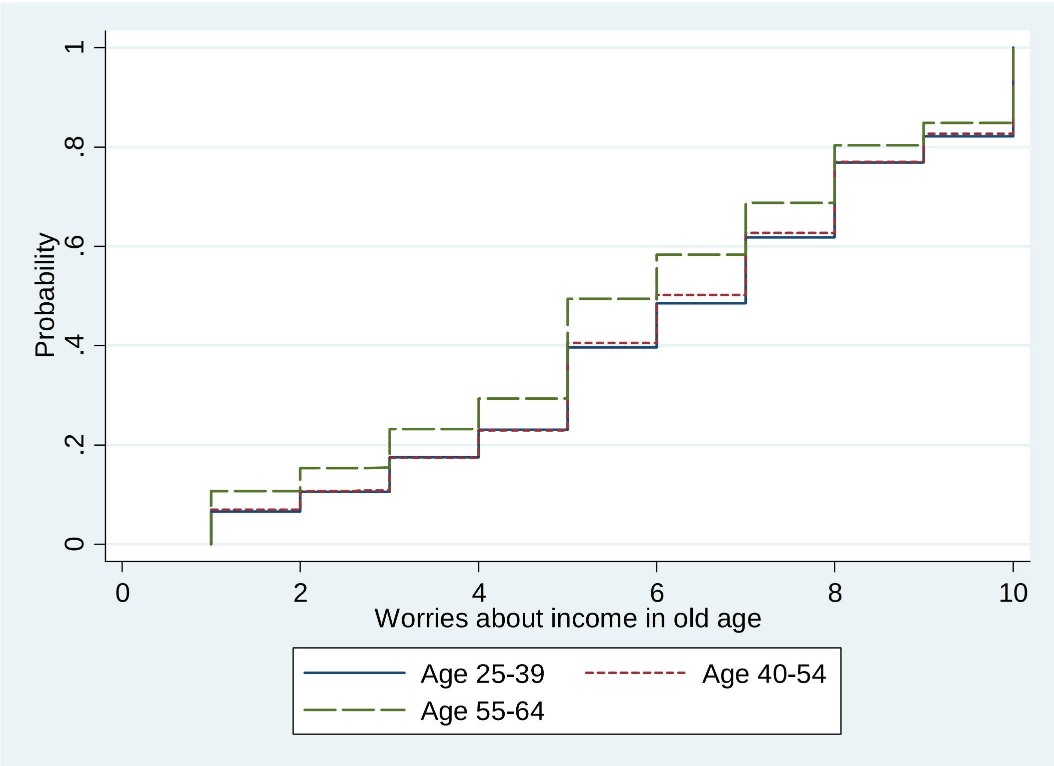
This section replicates a central tendency in the literature and adds new aspects at the same time. We have shown a clear relationship between three different determinants of well-being and wealth and it has become evident that this central result remains robust no matter if we use GDP or subjective living standard as the explanatory variable.

5.4.2 Group comparisons

The next part of our analysis is concerned with differences between groups in the EU. While the last subchapter tried to point towards variation across countries, we now want to focus on variation across groups. For our EU-wide analysis, we again pool data over the five waves but this time also over all countries. Understanding group differences is as important as understanding cross-country differences as this could provide another useful starting point for policy actions and their evaluation (e.g. implementation of job market programs for young people and their subsequent evaluation). As mentioned earlier, most of the demographic variables are subjective statements (which is not necessarily a bad thing as shown previously). Because it is of particular relevance for policy actions (see above), we focus on differences between the cumulative distributions of the same self-assessed variables as before but now differentiate between three age groups: 25-39, 40-54, and 55-64.

Figure 5.6 shows the cumulative distribution functions (CDFs) for the question on worries about income in old age for the three age groups. Even though the distributions for the age groups 25-39 and 40-54 are similar, the group of people between 55 and 64 years of age is less worried than the rest. This is intuitive as older people should have accumulated enough money in order to live through retirement age in dignity (assuming that individuals are at least to some extent rational and forward looking). One has to keep in mind that people in this age group have little or no time left to accumulate more money for retirement whereas the opposite is true for younger individuals. This is especially relevant during and after external shocks like the financial crisis that happened during the fieldwork periods. At this point we cannot rule out that worries were not disproportionately increased in the oldest group but we can see that they are at least not more worried than the younger citizens. The difference between the oldest and the youngest group is significant at the 1% level in a chi-squared test using the unweighted distribution.

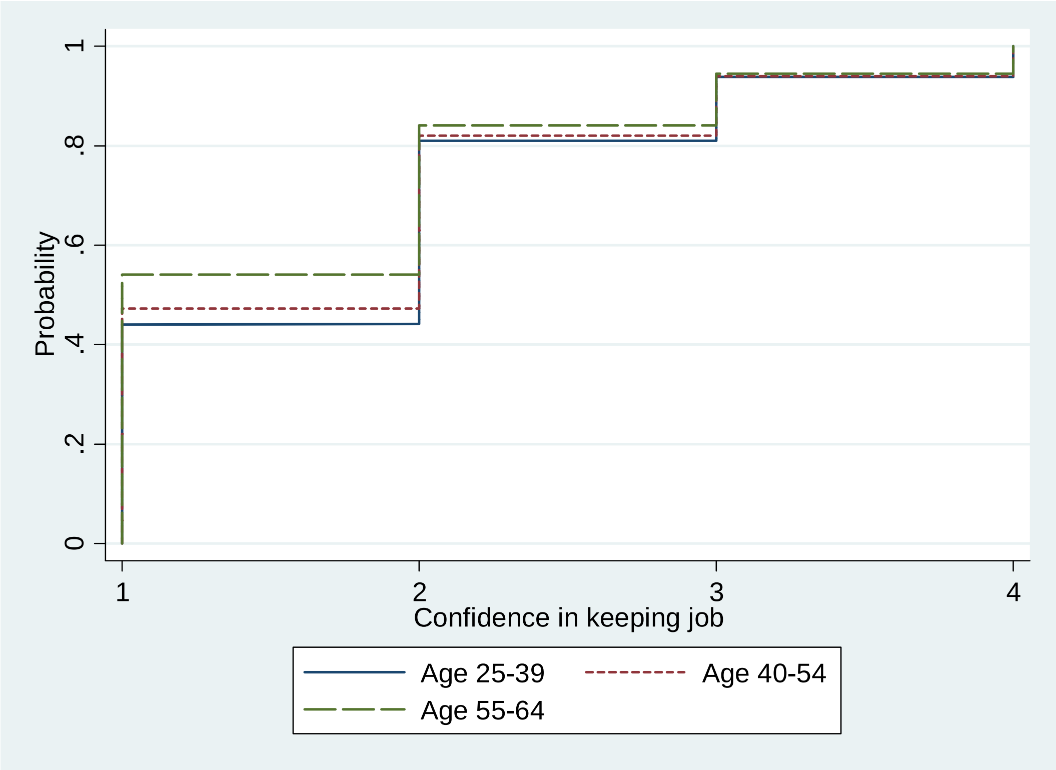
Figure 5.6: Worries about income in old age by age (waves and countries pooled)



The next step is to look at age group differences when it comes to confidence in keeping the job in the next 12 months – a determinant of how much money one can accumulate for retirement.

Figure 5.7 shows the CDFs by age. We can see that older people are more confident than younger people (the difference between the oldest and the youngest group is significant at the 1% level in a chi-squared test as described above). A possible explanation is that less experienced (i.e. younger) employees are laid off first and that it is also cheaper to lay off younger people because they get paid lower redundancy payments.

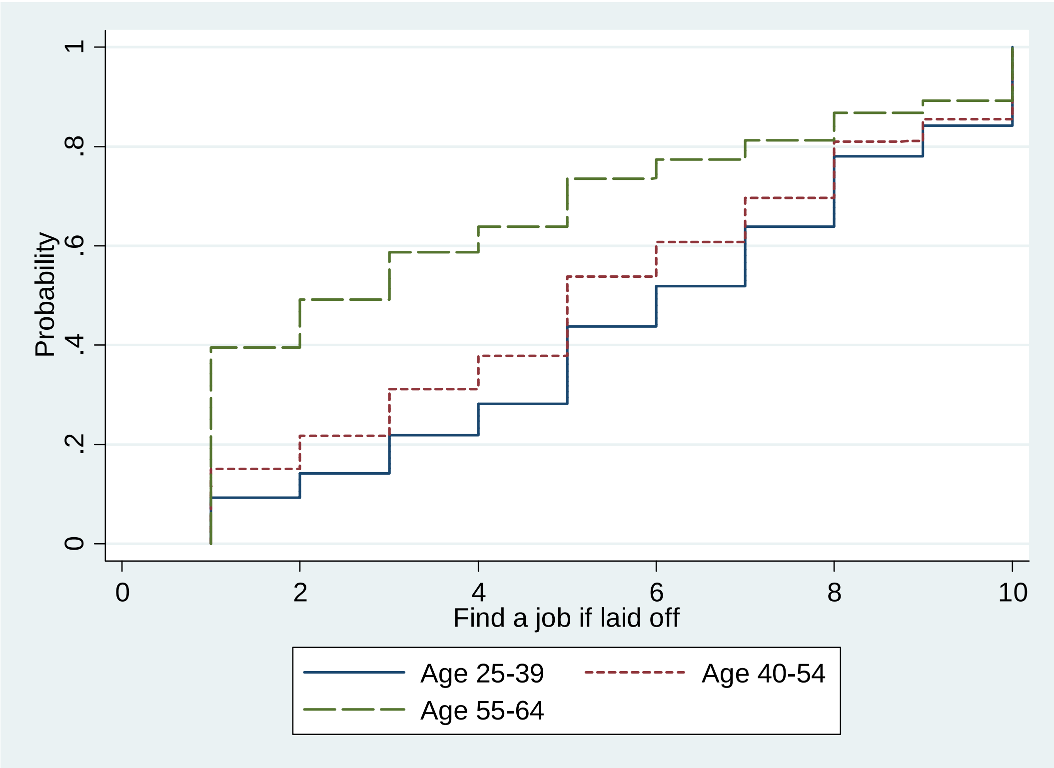
Figure 5.7: Confidence in keeping job by age (waves and countries pooled)



Given that one has to imagine being laid off, we would expect the oldest to state the lowest probability of finding a new job as firms would probably want younger and cheaper employees that are easier to integrate than the more expensive and more qualified older workers. Besides that, older people are more likely to be laid off from senior positions that are harder to find simply because there are fewer senior positions than positions in the lower hierarchy.

Figure 5.8 shows the CDFs for the self-stated probability of finding a new job in the next six months if laid off. The age differences in the distribution become immediately obvious. The oldest group has the lowest and the youngest group has the highest self-rated chance of success (the difference is again significant at the 1% level in a chi-squared test) while the middle age group lies in between. The difference is most striking for those who rate their chance of finding a new job with 1, i.e. “not at all likely” where we can see a difference of roughly 0.30.

Figure 5.8: Find a job if laid off by age (waves and countries pooled)



This section shows that the distributions of worries about income in old age and job market perspectives depend on age. We have shown that the differences between the CDFs are small when it comes to worries about income in old age and confidence in keeping the job but large when the subjective probability of finding a job if laid off is considered.

5.4.3 Cross-country comparisons by age

So far we have seen considerable variation in determinants of subjective well-being across European countries and also across age groups in the EU as a whole. In this section we want to analyze whether the cross-country relationships we have found are different for the age groups. We proceed in the same order as in the previous subchapters.

Figure 5.9: Worries about income in old age by age group (waves pooled)

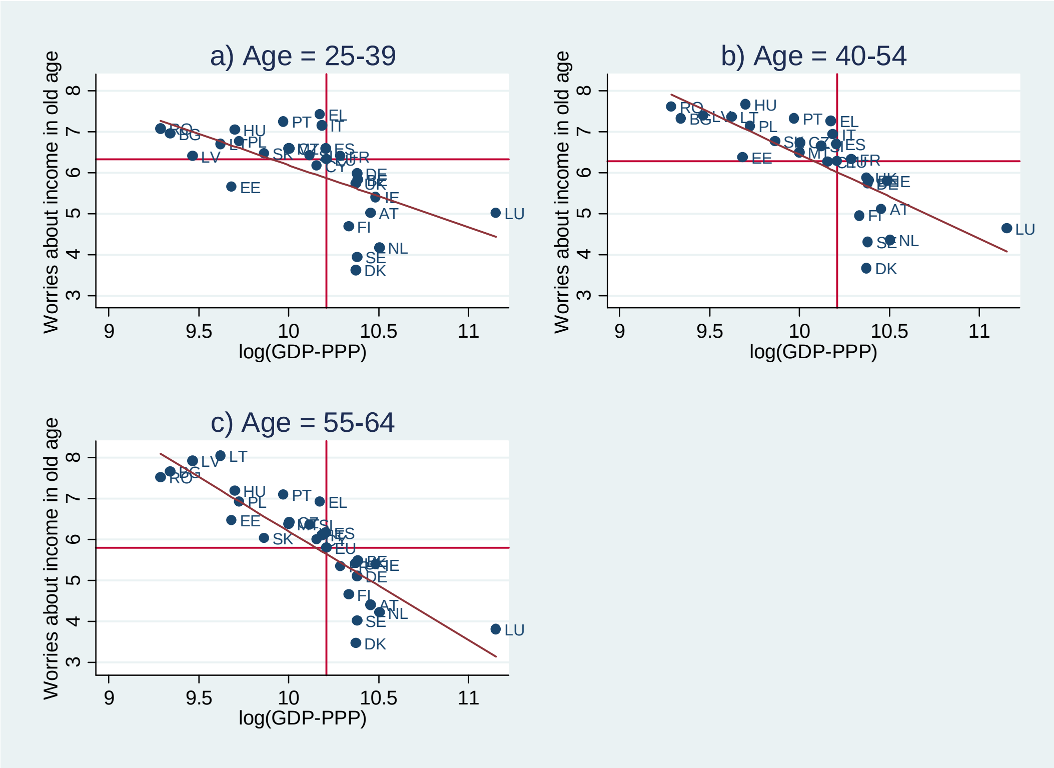


Table 5.3: Regression results and correlations (worries about income in old age)

	Parameter estimate	|t|-statistic	Correlation w/ dep. variable	p-value
Age 25 - 39				
log(GDP)	-1.52 ***	4.97	-0.60	< 0.01
Constant	21.35 ***	7.11		
Age 40 - 54				
log(GDP)	-2.06 ***	7.34	-0.77	< 0.01
Constant	27.01 ***	9.79		
Age 55 - 64				
log(GDP)	-2.65 ***	9.05	-0.86	< 0.01
Constant	32.74 ***	11.20		

*, **, *** indicate significance at the 10%, 5%, 1% levels.

Figure 5.9 a) - c) plots the relationship between average worries about income in old age and GDP by age group and table 5.3 quantifies the results by reporting parameter estimates of the explanatory variable (GDP) and the constant as well as the correlation of the y-axis and the x-axis variable.

For all age groups we find that (as before) almost all countries are contained in either the upper left (worst) or the lower right (best) quadrant. General patterns also remain stable as for example eastern or Baltic countries are typically in a situation with GDP far below and worries far above the EU average and vice versa for northern countries. Portugal, Greece, Spain and Italy are again amongst those in the upper left quadrant but (as before) Ireland is not. As suggested by the cumulative distribution functions of EU-wide worries in the previous subchapter, we can see that the horizontal line (indicating the EU average) in the oldest age group (figure 5.9 c)) is below that of the youngest group (figure 5.9 a)), indicating that the oldest group is less worried (the difference is significant at the 1% level). Table 5.3 reveals that the relationship between worries and GDP becomes stronger as age rises. We can see that the regression line gets steeper and ranges from -1.52 to -2.65 (the difference is significant at the 1% level). The slope estimated using overall country averages was -2.05 and comes close to the arithmetic average of the parameters in table 5.3. The correlation of the dependent and the explanatory variable also increases substantially from 0.60 amongst the youngest to 0.86 for the oldest members in this population. So although average worries decrease by age, the relationship to GDP gets stronger. An explanation is that countries with above average GDP are also those with the best social security systems that guarantee a more or less acceptable living standard in old age and vice versa.

For every country and each age group the relationship between average confidence in keeping one’s job and GDP is plotted in figure 5.10 a) - c) and the according regression results as well as the correlation coefficients are presented in table 5.4.

At a first glance we can see the split into countries in the upper left and in the lower right quadrant as it was shown before. Eastern European countries have below average GDP and below average confidence (a lower value indicates a higher confidence) and the opposite is true for northern countries (including Denmark and the Netherlands). The limitations are similar to before because Portugal, Spain, and Greece tend to be contained in the upper left quadrant but Italy is very close to the EU average and Ireland shows average confidence but above average GDP.

Figure 5.10: Confidence in keeping job by age group (waves pooled)

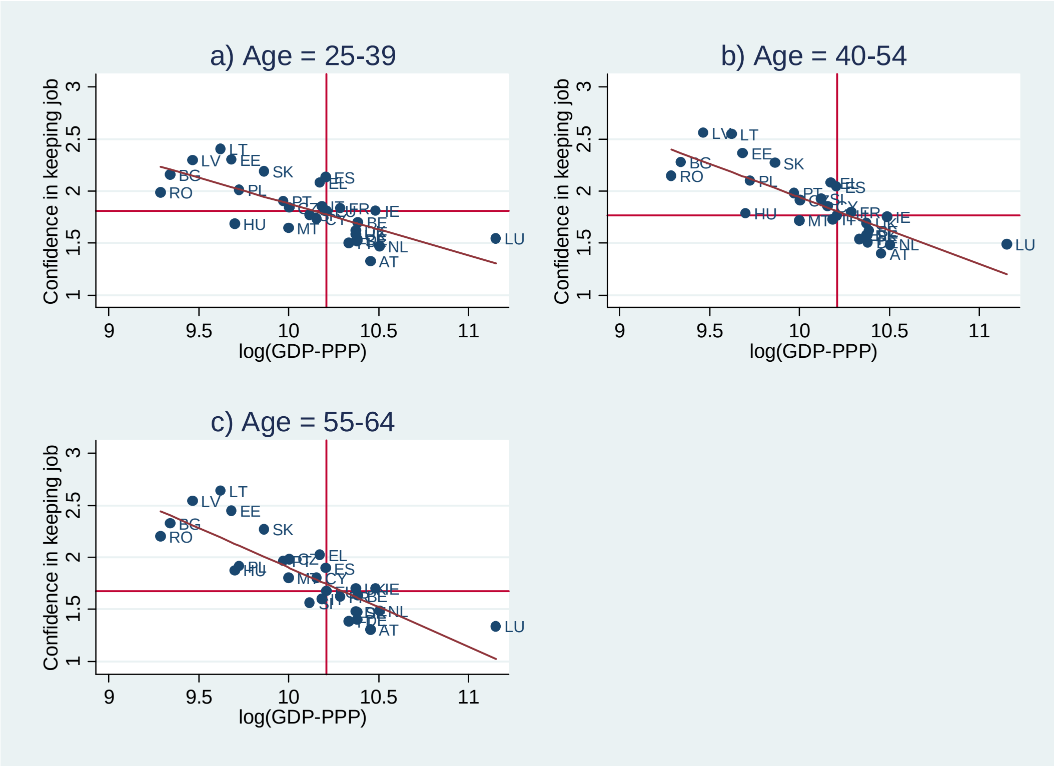


Table 5.4: Regression results and correlations (confidence in keeping job)

	Parameter estimate	|t|-statistic	Correlation w/ dep. variable	p-value
Age 25 - 39				
log(GDP)	-0.50 ***	4.73	-0.71	< 0.01
Constant	6.85 ***	6.43		
Age 40 - 54				
log(GDP)	-0.64 ***	5.52	-0.81	< 0.01
Constant	8.35 ***	7.08		
Age 55 - 64				
log(GDP)	-0.76 ***	6.22	-0.83	< 0.01
Constant	9.50 ***	7.65		

*, **, *** indicate significance at the 10%, 5%, 1% levels.

As shown in the previous subchapter (by plotting the distribution functions), average confidence rises with age (the horizontal reference line shifts downwards and the differences to the youngest group are significant at the 1% level) but now we can see an effect similar to the relationship between worries about income and GDP: The regression line gets steeper for higher age groups, i.e. it seems that amongst the oldest people in the sample, the same decrease in GDP between countries leads to a higher reduction in the average confidence in keeping the job than among younger individuals – but the difference between the parameters (compared to the youngest age group) is insignificant.

Again, the previously estimated slope parameter (not differentiating by age groups) is close to the average of the slope parameters estimated here. Additionally, the correlation coefficient in the oldest group is again higher than in the youngest group.

We have already shown that the overall relationship between the self-perceived likelihood of finding a job if one was to be laid off and GDP is rather weak (both in terms of significance of the regression parameter and correlation) but it also became evident that there are large differences between the cumulative distribution functions by age when the entire EU is considered. Now we get back to the cross-country comparison but differentiate by age.

Figure 5.11: Find a job if laid off by age group (waves pooled)

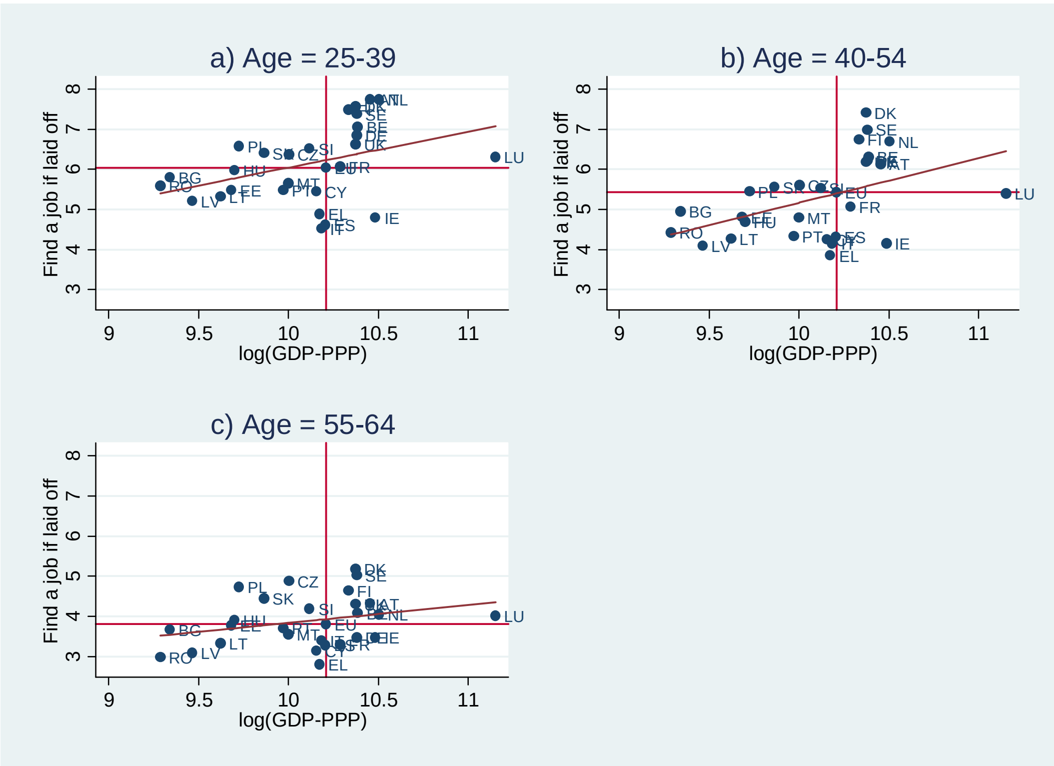


Table 5.5: Regression results and correlations (find a job if laid off)

	Parameter estimate	|t|-statistic	Correlation w/ dep. variable	p-value
Age 25 - 39				
log(GDP)	0.89 **	2.58	0.38	0.049
Constant	-2.89	0.85		
Age 40 - 54				
log(GDP)	1.11 ***	2.88	0.45	0.017
Constant	-5.96	1.57		
Age 55 - 64				
log(GDP)	0.44 *	1.93	0.28	> 0.10
Constant	-0.60	0.26		

*, **, *** indicate significance at the 10%, 5%, 1% levels.

Figure 5.11 a) - c) presents the scatter plots of the average self-rated probability of finding a job if laid off and a country's GDP as well as the corresponding regression lines. The latter are quantitatively described in table 5.5 with additional information on simple correlation coefficients.

Apparently, the relationship is still weak in every age group whereas the positive slope in the middle age group seems to be mainly driven by northern countries and in the oldest group it seems to disappear completely. But as mentioned earlier, the horizontal reference line (EU average) drops from more than 6 in the first group to less than 4 in the third group (the difference is significant at the 1% level). Compared to the analysis of the other subjective statements, grouping of countries into the worst and best possible quadrant is also no longer evident. Table 5.5 strongly supports the graphical analysis. Correlations are low in the youngest and oldest group and in the latter case far from statistical relevance. The situation is similar for the parameter estimates of the explanatory variable: In the oldest group the slope is flat and hardly significant whereas it is twice as high in the youngest group but still not significant at the 1% level (an exception compared to the other estimates presented in this work). Again, the difference between the parameters is insignificant. The middle age group is different with a comparatively large parameter estimate which is also significant at the 1% level. The slope estimated in the overall cross-country analysis (not differentiated by age) is again the average of the estimates presented in table 5.5. Taken together, variation across age groups in the EU is more important than variation across countries (by age group) in the case of the self-rated probability of finding a job if laid off.

Differentiating the cross-country analysis by age shows that the relationship between worries about income in old age and GDP gets stronger as people's age rises. On average, the difference in worries between a rich and a poor country is therefore largest in the oldest group. With respect to confidence in keeping the job we see the same age pattern but the slope differences are insignificant. When the subjective probability of finding a job if laid off is analyzed, the relationship with GDP is again weak and we also do not find the age pattern described above.

5.4.4 Cross-country comparisons over time

Now that we have seen variation across countries, groups and groups across countries, we turn to cross-country differences over time. So far we have pooled observations from all waves, i.e. we have reduced the whole time period to one point in time which one might refer to as "the crisis". This approach has the advantage that it ensures enough observations in the models, especially when sub-groups of particular interest, e.g. a certain age group in a specific country are considered. Yet we can also make use of the different waves we have available by comparing cross-country relationships over time. We will do so by running cross-country analyses for the same subjective determinants of well-being as before – each for the first and the last (fifth) wave.

Note that comparing country-specific sub-groups (e.g. by age) over time is not possible as the amount of observations per group and country at one particular point in time is too small to calculate meaningful group, country, and time-specific averages.

We start with worries about income in old age. Figure 5.12 plots the wave and country-specific mean of people's worries on the y-axis and GDP on the x-axis. Note that GDP does not change from wave to wave as we always use GDP from 2009. The solid line shows the fitted values from a regression using data from wave 1 and the dashed line refers to fitted values from a regression using data from wave 5. Solid circles indicate mean worries for a specific country and its GDP in wave 1 and the hollow circles show the same for wave 5.

Table 5.6 gives details on the regression coefficients and the correlation of worries with GDP for both waves.

Figure 5.12: Worries about income in old age by wave

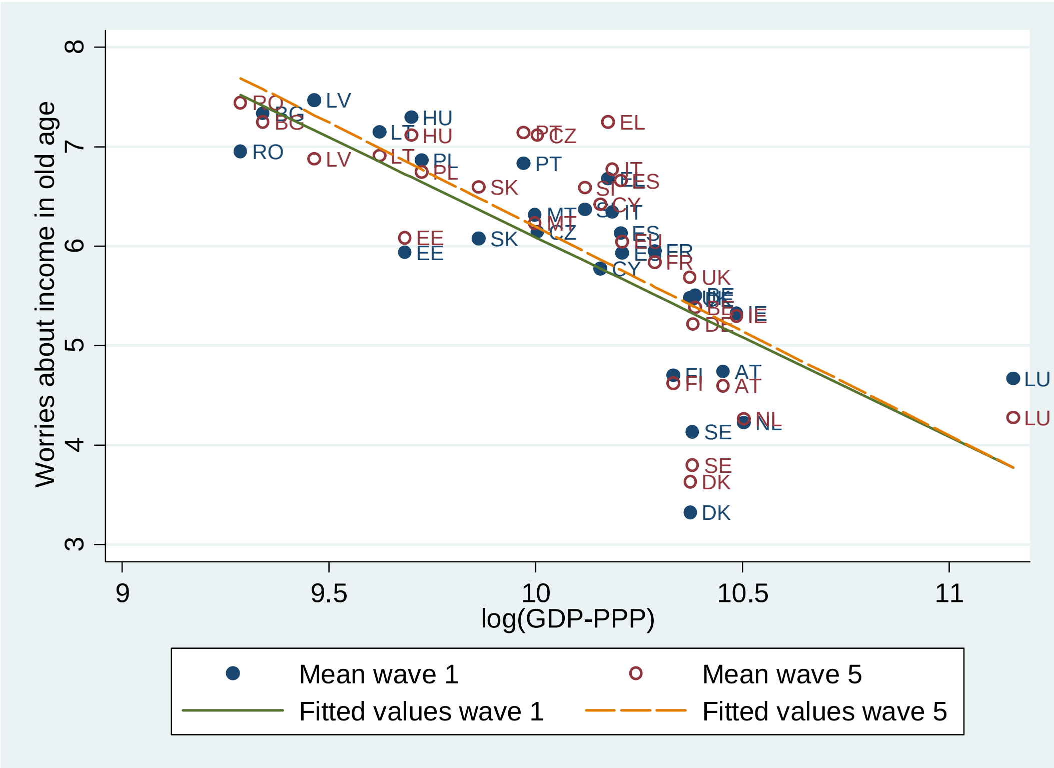


Table 5.6: Regression results and correlations (worries about income in old age)

Dependent variable: Worries about income in old age				
X-variable: log(GDP)	**Parameter estimate**	**|t|-statistic**	**Correlation**	**p-value**
Wave 1	-2.01 ***	6.15	-0.78	< 0.01
Wave 5	-2.10 ***	7.91	-0.75	< 0.01

*, **, *** indicate significance at the 10%, 5%, 1% levels.

At a first glance we can see that average worries are not constant over time as they increase in some countries and decrease in others, sometimes the change is large and sometimes it is small but as we had expected, they increase in Portugal, Greece, Spain, and Italy. Ireland seems to be less affected as average worries remain constant.

In eastern Europe worries increase in Romania but strongly decrease in Latvia. In western Europe worries decrease in Sweden and Luxembourg for example but there is an increase in Denmark and also a slight increase in the UK. So the pattern is as expected for the distressed countries and changes in eastern Europe are especially large in Latvia and Romania but overall countries do not seem to follow a consistent pattern and changes are typically small. The regression lines in figure 5.12 also do not change substantially from wave 1 to 5 but this still needs to be quantified. Table 5.6 shows how these changes over time affect the overall relationship of average worries and GDP. With all waves pooled, the regression coefficient was -2.05 and is -2.01 in wave 1 and -2.10 in wave 5 (both are significant at the 1% level but the difference is insignificant). As we cannot observe meaningful changes over time and because it does not make a large difference whether we pool the waves or not, variation over time is not an important source of heterogeneity in this case. The same holds for the correlation coefficients which are high and hardly change over time either. With pooled data, the correlation was -0.76 and is now -0.78 in wave 1 and -0.75 in wave 5 (all p-values are smaller than 0.01).

We continue with the country and wave-specific average of people's confidence in keeping their jobs. Figure 5.13 is set up according to figure 5.12 and the regression results together with the correlation coefficients are reported in table 5.7.

It can be seen that changes in average confidence over time seem to be larger than changes in average worries but remember that confidence is measured on a 4-point and not on a 10-point scale. Confidence falls slightly in Italy and Spain and heavily in Greece. Surprisingly, people in Ireland and Portugal have become more confident. Again, in general, changes are both positive and negative but as the regression line shifts downwards, there is a slight increase in people's confidence on average. It also becomes obvious that the overall relationship between average confidence and GDP does not change over time as the regression line from wave 5 parallels the regression line from wave 1. The slope we estimated using pooled data was -0.58 and is now (as shown in table 5.7) -0.60 in wave 1 and -0.58 in wave 5 (both are significant at the 1% level but the difference is insignificant). The correlation coefficient was -0.78 before and is now -0.77 in wave 1 and -0.75 in wave 5 (all p-values smaller than 0.01). Even though we can see an increase in people's confidence on average, the overall relationship between average confidence and GDP is virtually unchanged over time. In this case there is variation over time but it is small and does not affect the confidence-GDP gradient.

Figure 5.13: Confidence in keeping job by wave

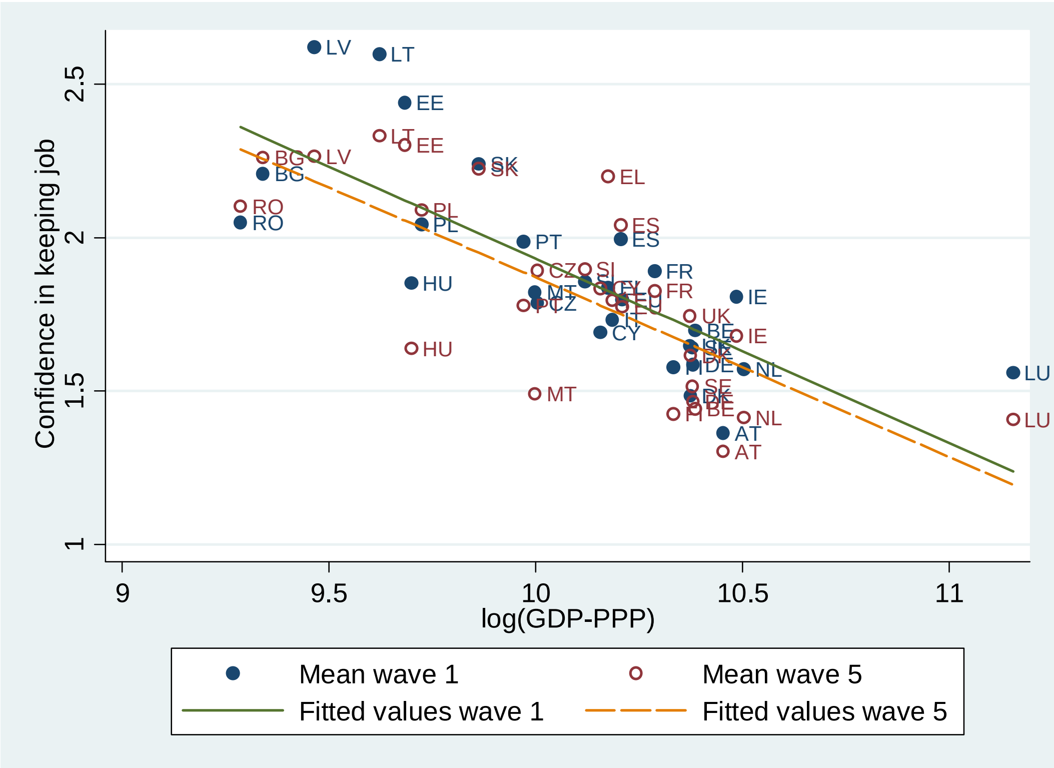


Table 5.7: Regression results and correlations (confidence in keeping job)

Dependent variable: Confidence in keeping job				
X-variable: log(GDP)	**Parameter estimate**	**|t|-statistic**	**Correlation**	**p-value**
Wave 1	-0.60 ***	4.46	-0.77	< 0.01
Wave 5	-0.58 ***	6.26	-0.75	< 0.01

*, **, *** indicate significance at the 10%, 5%, 1% levels.

The self-rated probability of finding a job if one was to be laid off is next and the corresponding scatter plot with regression lines for both waves is presented in figure 5.14. Table 5.8 gives details on the regression parameters and the correlation coefficients.

Figure 5.14: Find a job if laid off by wave

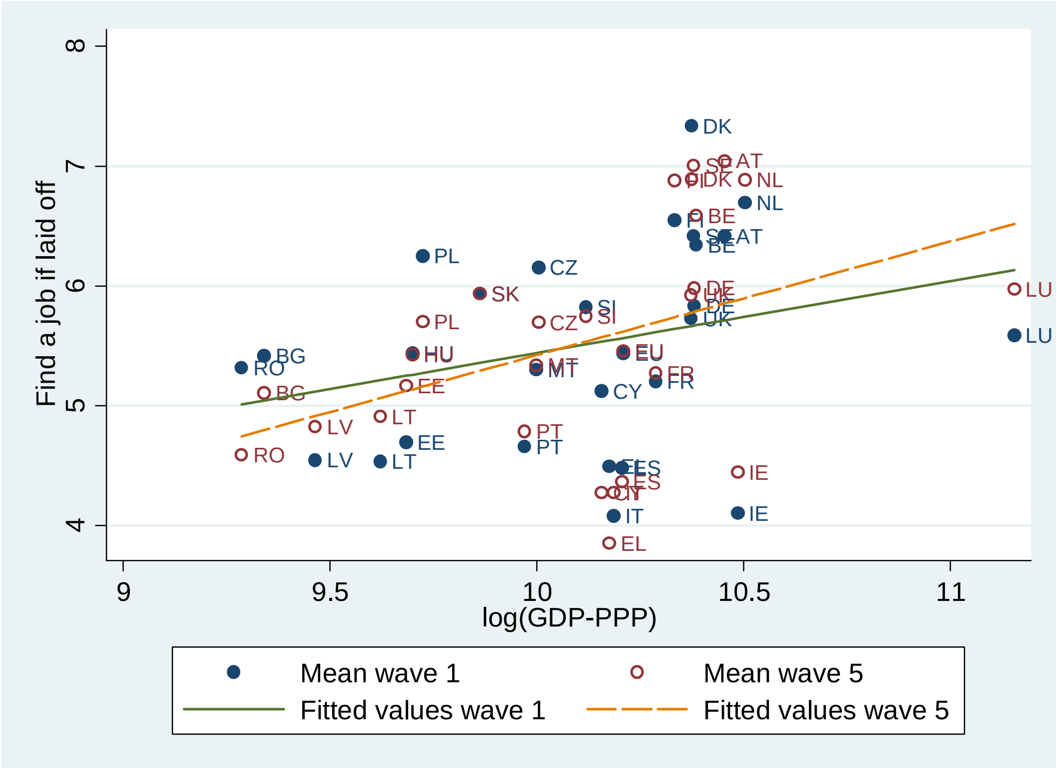


Table 5.8: Regression results and correlations (find a job if laid off)

Dependent variable: Find a job if laid off				
X-variable: log(GDP)	**Parameter estimate**	**|t|-statistic**	**Correlation**	**p-value**
Wave 1	0.60 *	1.87	0.29	> 0.10
Wave 5	0.95 ***	3.29	0.42	0.030

*, **, *** indicate significance at the 10%, 5%, 1% levels.

With respect to the distressed countries, changes go in both directions again but they are rather small (besides Greece). The self-rated probability increases in Italy, Ireland, and Portugal but decreases in Greece and Spain. Nevertheless, this group's self-rated probability is far below the European average even though GDP is at or over the European average (besides Portugal). Again, there is no overall pattern as the country averages increase for some countries (e.g. Netherlands) and decrease for others (e.g. Denmark). Interestingly, the Baltic countries Estonia, Latvia, and Lithuania all increase in average self-rated probability over time. Figure 5.14 also shows that there is a change in the regression line between waves, indicating that the relationship between the self-rated probability and GDP has increased from wave 1 to wave 5.

The slope estimated with pooled data was 0.81 but now it ranges from 0.60 in wave 1 to 0.95 in wave 5 (table 5.8) although the difference is insignificant. The correlation was 0.38 before and is now between 0.29 and 0.42. The relationship is not very strong in either case as in wave 1 the parameter estimate is only significant at the 10% level and the p-value of the correlation coefficient is above 0.10. The coefficient in wave 5 is significant at the 1% level but the correlation is still weak (in the context of results presented for other determinants of well-being) with a p-value of more than 0.01. However, it seems that in this case, variation over time does indeed play a more important role than in the regressions presented above.

Analyzing how the gradients estimated with pooled data change when we run separate regressions for the first and the last wave reveals interesting insights. The relationships of worries about income in old age and confidence in keeping the job with GDP both are very constant over time. In the latter case we see a vertical shift in the regression line indicating that overall confidence had increased. In the case of the self-rated probability of finding a job if laid off there is an increase in the parameter estimate over time but as the coefficient from the first wave is hardly significant (and as we see no statistically meaningful correlation as well as no significant difference between the slopes) we do not want to put too much weight on that finding.

5.5 Subjective wealth and subjective poverty across countries

In the last chapter we showed that using average subjective living standard (subjective wealth) as a measure to explain differences in the country averages of subjective statements (i.e. determinants of well-being) delivers similar results as when GDP is used. The motivation was that even though subjective statements are considered, these could actually measure material well-being that cannot be captured by a single income question. Consequently, the country-specific average of subjective wealth should determine the overall poverty rate that people expect in their country. We have this information as Eurobarometer interviewees are asked to estimate the fraction of poor people living in their country (question wording in appendix A5.1). They can choose between (1) 30%, (2) 20%, (3) 10%, (4) 5% or (5) less than 5% (i.e. a higher value indicates a lower expected poverty rate).

Figure 5.15: Subjective poverty and subjective living standard across countries

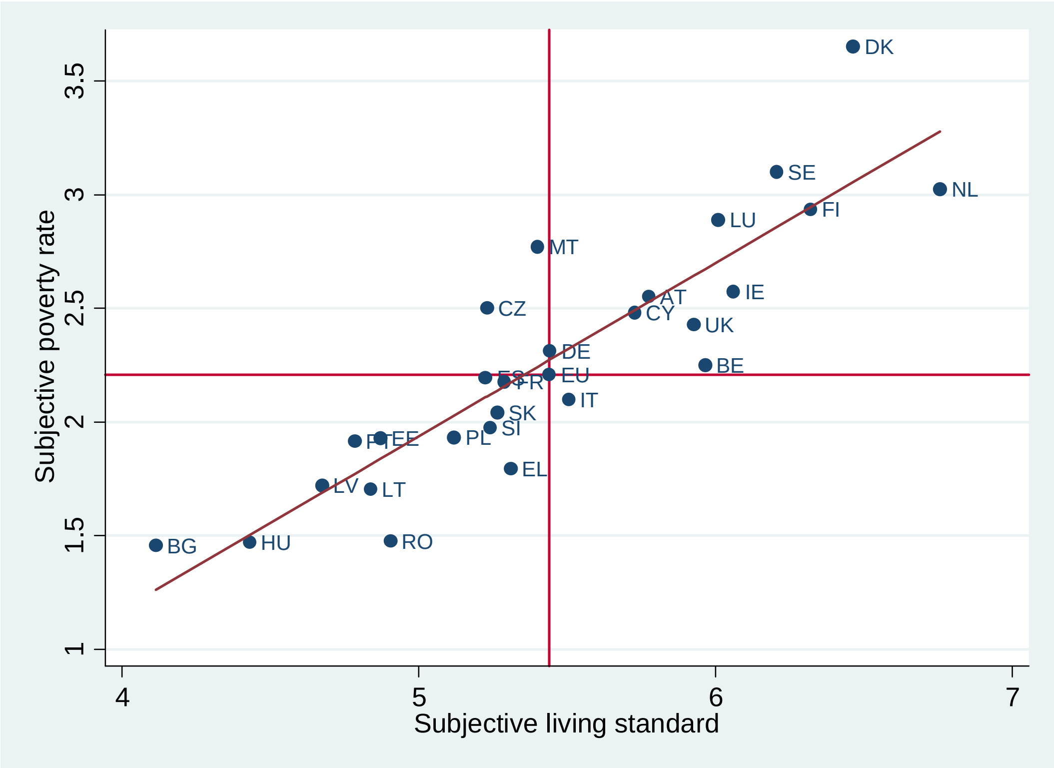


Table 5.9: Regression results and correlations (subjective poverty rate)

Dependent variable: Subjective poverty rate				
Explanatory variable	**Parameter estimate**	**|t|-statistic**	**Correlation**	**p-value**
Subjective wealth	0.76 ***	9.07	0.89	< 0.01
log(GDP)	0.98 ***	6.63	0.74	< 0.01

*, **, *** indicate significance at the 10%, 5%, 1% levels.

Figure 5.15 shows a scatter plot with the country-specific mean of the answer to the subjective poverty rate question on the y-axis, the country-specific mean of subjective living standard on the x-axis and the regression line (EU averages are indicated by the vertical and horizontal reference lines). Table 5.9 reports the coefficient for the regression line and the correlation coefficient of subjective poverty rate and subjective wealth. Additionally, and as a robustness check, table 5.9 also reports the results when GDP is used as the explanatory variable (graph not shown).

It can be seen that average subjective poverty is highly correlated with average subjective wealth (0.89 with a p-value smaller than 0.01) and that the slope of the regression line (0.76) is also highly significant (at the 1% level). Even though the evidence is less strong, the findings remain clearly evident when GDP is used as the income measure.

This is an interesting methodological point as we can observe a very high degree of consistency: If most people think that they are rich then most people think that there are few poor people in their country. This observation supports the point made by Deaton (2008) that people might indeed use the same comparison standard when asked to give answers on a point scale. In our case, this means that people know where they stand relative to other European countries and that people use this European standard to compare themselves to when they derive their answers. If people only used their own country as a comparison standard, figure 5.15 would look different. If someone is asked for his living standard and compares himself to the average inhabitant of his country, we would not see averages below “5” and above “6”. Following Sudman *et al*. (1996), the point scale represents the underlying distribution and as these numbers are positioned in the middle of the scale they therefore represent the “typical” person to the respondents (the EU average is roughly 5.5). Why would the average Bulgarian check “4” on the answer scale and the average Dutch “7” if they were comparing themselves to the country average? The same logic applies to the subjective poverty rate. First of all, it is hard for an individual to estimate the poverty rate as it is complicated to define poverty in absolute terms (e.g. in terms of minimum income), then guess how many people are affected and then calculate the actual poverty rate. Regressions at the country level (not shown) reveal that individuals who think of themselves as more wealthy tend to state a lower poverty rate. This may be wrong at the individual level but it is right on average because in a rich country (by European comparison standards) more people state a low poverty rate which is correct. The next part of the explanation is that people once again use a relative comparison standard. They do not know the poverty rate but they have an idea where their country lies relative to others and check the number that represents their idea of their country's position on the scale. In this case it is not even possible to pick a reference point other than a different country as the question is about the country itself and not about the individual. All this of course does not mean that the actual percentage rate is correct, but we have seen that the subjective poverty rate is also highly correlated with objective GDP which makes the relationship credible. Admittedly, this interpretation remains speculative but we think of it as support for the idea of an internationally standardized reference point for comparison.

5.6 Conclusions

The Easterlin paradox was discussed in the literature for a long time. We focused on its second part to show that well-being is related to a country's wealth (measured by GDP and subjective living standard) by using a large dataset with representative information for the entire EU which surveyed respondents over five consecutive waves during the financial crisis. We expanded the set of variables that is typically used in the literature (e.g. life satisfaction and happiness) by defining well-being in a broader sense, i.e. we used existential fears as determinants of (overall) well-being, measured by worries about income in old age, confidence in keeping the job and the self-rated probability of finding a job if one was to be laid off. Our results showed that these well-being measures are well explained by (are highly correlated with) wealth (measured by either GDP or subjective living standard) whereas the relationship is weak in the case of the self-rated probability of finding a job if laid off. Then we took a closer look at the EU as a whole and found that the cumulative distribution functions of the well-being measures differ by age but in an intuitive way: Compared to the youngest age group, people in the oldest group are slightly less worried about their income in old age and a bit more confident in keeping their job whereas they state a much lower probability of finding a job if they were to be laid off. Based on these findings, we analyzed the cross-country relationship of well-being and GDP again but differentiated the regressions by age. In the case of worries about income in old age we showed that the gradient increases with age and found the same for confidence in keeping the job while in the latter case the increase was not significant. When we analyzed the self-rated probability of finding a job if laid off, we did not find this pattern. As a next step, we ran cross-country regressions separately for the first and the last wave. We found that the relationship remains constant for the first two measures but it increases a bit for the last measure whereas the difference was not significant. In general, we saw a number of country patterns: For example northern countries were always better off than eastern European countries. Additionally, countries that were hit especially hard by the crisis like Greece and Portugal were typically among the underperforming countries whereas Ireland was often an exception. Finally, we showed that perceived poverty and average subjective living standard are closely linked. We interpreted this as a hint towards an internationally standardized reference point that is used in order to derive answers to point scale questions.

Even though we contributed to a better understanding of variation in determinants of subjective well-being in the EU, it is still hard to derive clear-cut policy recommendations. As mentioned earlier, the Beyond GDP initiative aims at adding softer measures to accompany GDP as the standard developmental indicator but we have also seen that there is a correlation between GDP and measures of well-being. The question that is left open is how much of this relationship is attributable to causality. The more well-being is determined by GDP, the less we need other indicators. This said, we can think of different plausible measures we would want to add to GDP and they are probably all linked differently to GDP. So even if some part of well-being is directly caused by GDP, other parts may not be and in the end it depends on what the countries want to maximize in order to achieve the best possible life for their people. We therefore suggest that it is promising for future research to put even more effort into determining what policy should actually maximize and how to weight all these different subjective and objective measures of a nation's welfare. As we have shown here, this is complicated because these measures can depend on many circumstances (and we have shown that wealth and age are only some of them). Furthermore, future research should also concentrate on how subjective measures can be compared on an international level and how credible these comparisons are. If policy actions are to be based on subjective measures, we need to be sure that we are really measuring what we want to measure. The question whether people use an internationally equivalent reference point to answer questions on some scale is just one part of it. Lastly we want to encourage future research to find a way of measuring material well-being (Diener and

Biswas-Diener (2002)) in future surveys as this seems to be the most plausible way of measuring the individual financial situation and what people can make out of it whereas standard income measures are used for convenience and without being questioned very often.

5.7 Appendix

A5.1: Question wording as in the English version of the questionnaire

Worries about income in old age:

How worried are you, if at all, that your income in old age will not be adequate enough to enable you to live in dignity. Please express your opinion on a scale of 1 to 10, where 1 means "Not worried at all" and 10 means "Very worried".

Confidence in keeping the job:

How confident would you say you are in your ability to keep your job in the next 12 months?

- Very confident

- Fairly confident

- Not very confident

- Not at all confident

Subjective probability of finding a job if laid off:

If you were to be laid off, how would you rate on a scale from 1 to 10, the likelihood of you finding a job in the next six months? "1" means that it "would not at all be likely" and 10 means that "it would be very likely".

Fraction of poor people living in a country:

If you were to say how many poor people there are in (OUR COUNTRY), would you say that…?

- 1 person out of 3 – or about 30% – is poor in (OUR COUNTRY)

- 1 person out of 5 – or 20%

- 1 person out of 10 – or 10%

- 1 person out of 20 – or 5%

- Less than 5%

Table 5.10: Overview of country names and their abbreviations

Country name	Abbreviation
European Union	EU
Austria	AT
Belgium	BE
Bulgaria	BG
Cyprus	CY
Czech Republic	CZ
Denmark	DK
Estonia	EE
Finland	FI
France	FR
Germany	DE
Greece	EL
Hungary	HU
Ireland	IE
Italy	IT
Latvia	LV
Lithuania	LT
Luxembourg	LU
Malta	MT
Netherlands	NL
Poland	PL
Portugal	PT
Romania	RO
Slovakia	SK
Slovenia	SI
Spain	ES
Sweden	SE
United Kingdom	UK

List of figures

List of tables

Bibliography

Ai, C. and C. Norton (2003): Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123–129.

Andrews, F. M. and R. F. Inglehart (1979): The structure of subjective well-being in nine western societies. *Social Indicators Research*, 6(1), 73–90.

Anger, S. (2008): Overtime work as a signaling device. *Scottish Journal of Political Economy,* 55(2), 167–189.

Anger, S., D. A. Bowen, M. Engelmann, J. R. Frick, J. Goebel, M. M. Grabka, O. Groh-Samberg, H. Haas, B. Headey, E. Holst, P. Krause, M. Kroh, C. Kurka, H. Lohmann, R. Pischner, U. Rahmann, C. Schmitt, J.-P. Schräpler, J. Schupp, I. Sieber, T. Siedler, C. K. Spieß, M. Spieß, I. Tucci, and G. G. Wagner (2008): 25 Wellen Sozio-oekonomisches Panel. *Vierteljahrshefte zur Wirtschaftsforschung*, 77(3), 9–14.

Battistin, E., R. Miniaci, and G. Weber (2003): What do we learn from recall consumption data? *Journal of Human Resources,* 38(2), 354–385.

Beckers, D. G. J., D. van der Linden, P. G. W. Smulders, M. A. J. Kompier, M. J. P. M. van Veldhoven, and N. W. van Yperen (2004): Working overtime hours: Relations with fatigue, work motivation, and the quality of work. *Journal of Occupational and Environmental Medicine*, 46(12), 1282–1289.

Bertrand, M. and S. Mullainathan (2001): Do people mean what they say? Implications for subjective survey data. *American Economic Review, Papers & Proceedings*, 91(2), 67–72.

Blanchflower, D. G. and A. J. Oswald (2004): Well-being over time in Britain and the USA. *Journal of Public Economics*, 88(7-8), 1359–1386.

Bound, J., C. Brown, and N. Mathiowetz (2001): Measurement error in survey data. In J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Volume 5, 3705–3843. Amsterdam: Elsevier.

Browning, M., T. F. Crossley, and G. Weber (2003): Asking consumption questions in general purpose surveys. *Economic Journal*, 113(491), F540–F567.

Cai, L., K. Mavromaras, and U. Oguzoglu (2008): The effects of health and health shocks on hours worked. *IZA Discussion Paper*, No. 3496.

Caruso, C. C., E. M. Hitchcock, R. B. Dick, J. M. Russo, and J. M. Schmit (2004): Overtime and extended work shifts: Recent findings on illnesses, injuries, and health behaviors. *DHHS (NIOSH) Publication*, No. 2004–143.

Christoph, B. and H.-H. Noll (2003): Subjective well-being in the European Union during the 90s. *Social Indicators Research*, 64(3), 521–546.

Clark, A. E. and A. J. Oswald (1996): Satisfaction and comparison income. *Journal of Public Economics*, 6(3), 359–381.

Cropper, M. L. (1977): Health, investment in health, and occupational choice. *Journal of Political Economy*, 85(6), 1273–1294.

Cunningham, M. R. (1979): Weather, mood, and helping behavior: Quasi experiments with the sunshine Samaritan. *Journal of Personality and Social Psychology*, 37(11), 1947–1956.

Dahlgren, A., G. Kecklund, and T. Åkerstedt (2006): Overtime work and its effects on sleep, sleepiness, cortisol and blood pressure in an experimental field study. *Scandinavian Journal of Work, Environment & Health*, 32(4), 318–327.

Deaton, A. (2008): Income, health, and well-being around the world: Evidence from the Gallup World Poll. *Journal of Economic Perspectives*, 22(2), 53–72.

Dembe, A. E., J. B. Erickson, R. G. Delbos, and S. M. Banks (2005): The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. *Occupational and Environmental Medicine*, 62(9), 588–597.

DGB – The Confederation of German Trade Unions (2010): Gesundheitsrisiko Arbeitslosigkeit – Wissensstand, Praxis und Anforderungen an eine arbeitsmarktintegrative Gesundheitsförderung. *Arbeitsmarkt Aktuell*, No. 9, Available online: http://www.dgb.de/themen/++co++44151b30-a92b-11df-6fd9-00188b4dc422
[Download: 11.10.2010].

Diederich, J. and T. Goeschl (2011): Giving in a large economy: Price vs. non-price effects in a field experiment. *University of Heidelberg Discussion Paper Series,* No. 514.

Diener, E. and R. Biswas-Diener (2002): Will money increase subjective well-being? *Social Indicators Research*, 57(2), 119–169.

Diener, E., E. Sandvik, L. Seidlitz, and M. Diener (1993): The relationship between income and subjective well-being: Relative or absolute? *Social Indicators Research*, 28(3), 195–223.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, 9(3), 522–550.

Duncan, G. J. (2002): The PSID and Me. In E. Phelps, F. F. Furstenberg, and A. Colby (eds.), *Looking at Lives: American Longitudinal Studies of the Twentieth Century*, 133–166. New York, NY: Russell Sage Foundation.

Easterlin, R. A. (1973): Does money buy happiness? *Public Interest,* 30, 3–10.

Easterlin, R. A. (1974): Does economic growth improve the human lot? Some empirical evidence. In P. A. David and M. W. Reder (eds.), *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz,* 89–125. New York, NY: Academic Press.

Easterlin, R. A. (1995): Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior and Organization*, 27(1), 35–47.

Frechette, G. R. (2001a): sg158: Random-effects ordered probit. *Stata Technical Bulletin,* 59, 23–27.

Frechette, G. R. (2001b): sg158.1: Update to random-effects ordered probit. *Stata Technical Bulletin*, 61, 12.

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

Frijters, P., J. P. Haisken-DeNew, and M. A. Shields (2004): Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification. *American Economic Review*, 94(3), 730–740.

Golden, L. and B. Wiens-Tuers (2006): To your happiness? Extra hours of labor supply and worker well-being. *Journal of Socio-Economics*, 35(2), 382–397.

Grossman, M. (1972): On the concept of health capital and the demand for health. *Journal of Political Economy,* 80(2), 223–255.

Hamermesh, D. S. (2004): Subjective outcomes in economics. *Southern Economic Journal*, 71(1), 2–11.

Helliwell, J. F. (2003): How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling*, 20(2), 331–360.

Helliwell, J. F. and S. Wang (2011): Weekends and subjective well-being. *NBER working paper*, No. 17180.

Hirshleifer, D. and T. Shumway (2003): Good day sunshine: Stock returns and the weather. *Journal of Finance*, 58(3), 1009–1032.

Hoderlein, S. and J. Winter (2010): Structural measurement errors in nonseparable models. *Journal of Econometrics*, 157(2), 432–440.

Johnson, J. V. and J. Lipscomb (2006): Long working hours, occupational health and the changing nature of work organization. *American Journal of Industrial Medicine*, 49(11), 921–929.

Kawada, T. and M. Ooya (2005): Workload and health complaints in overtime workers: A survey. *Archives of Medical Research*, 36(5), 594–597.

Kawakami, N. and T. Haratani (1999): Epidemiology of job stress and health in Japan: Review of current evidence and future direction. *Industrial Health*, 37(2), 174–186.

Kliger, D. and O. Levy (2003): Mood-induced variation in risk preferences. *Journal of Economic Behavior and Organization*, 52(4), 573–584.

Krueger, A. B. and D. A. Schkade (2008): The reliability of subjective well-being measures. *Journal of Public Economics*, 92(8-9), 1833–1845.

Krupp, H.-J. (2008): Die Anfänge: Zur Entstehungsgeschichte des SOEP. *Vierteljahrshefte zur Wirtschaftsforschung*, 77(3), 15–26.

McFadden, D., A. Bemmaor, F. Caro, J. Dominitz, B. Jun, A. Lewbel, R. Matzkin, F. Molinari, N. Schwarz, R. Willis, and J. Winter (2005): Statistical analysis of choice experiments and surveys. *Marketing Letters*, 16(3–4), 183–196.

McGonagle, K. A. and R. F. Schoeni (2006): The Panel Study of Income Dynamics: Overview and summary of scientific contributions after nearly 40 Years. *PSID Technical Paper Series*, No. 06–01.

Nakamura, K., S. Shimai, S. Kikuchi, H. Takahashi, M. Tanaka, S. Nakano, Y. Motohashi, H. Nakadaira, and M. Yamamoto (1998): Increases in body mass index and waist circumference as outcomes of working overtime. *Occupational Medicine*, 48(3), 169–173.

Nowlis, S. M., B. E. Kahn, and R. Dhar (2002): Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. *Journal of Consumer Research*, 29(3), 319–334.

Redelmeier, D. A. and S. D. Baxter (2009): Rainy weather and medical school admission interviews. *Canadian Medical Association Journal*, 181(12), 933.

Riener, G. and C. Traxler (2011): Norms, moods, and free lunch: Longitudinal evidence on payments from a pay-what-you-want restaurant. *Journal of Socio-Economics,* Available online.

Ruder, M. and H. Bless (2003): Mood and the reliance on the ease of retrieval heuristic. *Journal of Personality and Social Psychology*, 85(1), 20–32.

Ruser, J. W. (1991): Workers' compensation and occupational injuries and illnesses. *Journal of Labor Economics,* 9(4), 325–350.

Ruud, P., D. Schunk, and J. Winter (2009): Uncertainty and rounding in survey responses: A laboratory experiment. *Unpublished manuscript*.

Saunders, E. M. (1993): Stock prices and Wall Street weather. *American Economic Review*, 83(5), 1337–1345.

Schwarz, N. and G. L. Clore (1983): Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45(3), 513–523.

Schwarz, N. and D. Oyserman (2001): Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22(2), 127–160.

Simonsohn, U. (2010): Weather to go to college. *Economic Journal*, 120(543), 270–280.

Sparks, K., C. Cooper, Y. Fried, and A. Shirom (1997): The effects of hours of work on health: A meta-analytic review. *Journal of Occupational and Organizational Psychology*, 70(4), 391–408.

Starrin, B., G. Larsson, S.-O. Brenner, L. Levi, and I.-L. Petterson (1990): Structural changes, ill health, and mortality in Sweden, 1963–1983: A macroaggregated study. *International Journal of Health Services*, 20(1), 27–42.

Stevenson, B. and J. Wolfers (2008): Economic growth and subjective well-being: Reassessing the Easterlin paradox. *Brookings Papers on Economic Activity*, Spring 2008, 1–87.

Sudman, S., N. M. Bradburn, and N. Schwarz (1996): *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass.

Tourangeau, R., L. J. Rips, and K. Rasinski (2000): *The Psychology of Survey Response*. New York, NY and Cambridge, UK: Cambridge University Press.

Trejo, S. J. (1991): The effects of overtime pay regulation on worker compensation. *American Economic Review*, 81(4), 719–740.

Veenhoven, R. (1991): Is happiness relative? *Social Indicators Research*, 24(1), 1–34.

Vogel, J. (1999): The European 'welfare mix': Institutional configuration and distributive outcome in Sweden and the European Union. A longitudinal and comparative perspective. *Social Indicators Research*, 48(3), 245–297.

Winter, J. (2002): Bracketing effects in categorized survey questions and the measurement of economic quantities. *Discussion Paper,* Sonderforschungsbereich 504, University of Mannheim, No. 02-35.

Winter, J. (2004): Response bias in survey−based measures of household consumption. *Economics Bulletin*, 3(9), 1–12.

Yardley, J. K. and R. W. Rice (1991): The relationship between mood and subjective well-being. *Social Indicators Research*, 24(1), 101–111.

Curriculum Vitae

Personal details:

Name:	Gregor Maximilian Tannhof
Date of birth:	13th March, 1981
Place of birth:	Nuremberg
Nationality:	German

Education:

10/2009 - 09/2012	**Ludwig-Maximilians-Universität (LMU), Munich** **Munich Graduate School of Economics (MGSE)** Ph.D. program in Economics
04/2004 - 07/2008	**Ludwig-Maximilians-Universität (LMU), Munich** - Studies of Economics (Dipl.-Volksw. / equiv. M.Sc. in Economics) - Erasmus exchange program (09/2005 - 02/2006) (Universität Zürich, Zurich - Switzerland)
10/2003 - 03/2004	**Philipps-Universität, Marburg** Studies of Psychology
06/2000	**Gymnasium Bad Aibling** Abitur

Other experience:

10/2009 - 09/2012	**Ludwig-Maximilians-Universität (LMU), Munich** Research and teaching assistant with Prof. Dr. Joachim Winter
08/2008 - 09/2009	**PricewaterhouseCoopers AG WPG, Munich** Consultant
09/2001 - 06/2003	**Deutsche Bank AG, Munich** Apprenticeship
07/2000 - 04/2001	**Bundeswehr, Brannenburg** Military service