Labour Market Responses to Social Security Contributions and Social Benefits

Empirical Evidence from Germany

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft
des Fachbereichs Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Diplom-Volkswirt Michael Neumann
geboren in Leonberg

Berlin, September 2016

Gedruckt mit der Genehmigung des Fachbereichs Wirtschaftwissenschaft der Freien Universität Berlin.

Dekan:
Prof. Dr. Dr. Andreas Löffler

Erstgutachter:
Prof. Dr. Peter Haan, Freie Universität Berlin

Zweitgutachter:
Prof. Dr. Antoine Bozio, Paris School of Economics

Datum der Disputation:
8. November 2016

Erklärung über Zusammenarbeit mit Koautoren und Vorveröffentlichungen

Kapitel 1

Eine frühere Fassung des Papiers wurde veröffentlicht als:

Neumann, Michael, "Earnings Responses to Social Security Contributions", DIW Discussion Paper 1489, 2015.

Revise and Re-submit at *Labour Economics*.

Kapitel 2

Veröffentlicht als:

Müller, Kai-Uwe and Neumann, Michael, "The Economic Incidence of Social Security Contributions: A Discontinuity Approach with Linked Employer-Employee Data", DIW Discussion Paper 1578, 2016. (Eigenanteil: 50%)

Kapitel 3

In Zusammenarbeit mit Kai-Uwe Müller und Katharina Wrohlich (Eigenanteil: 33.3%). Bislang unveröffentlicht.

Kapitel 4

In Zusammenarbeit mit Luke Haywood (Eigenanteil: 50%). Bislang unveröffentlicht.

Acknowledgments

This doctoral thesis was prepared while I was a member of the Graduate Center of the German Institute for Economic Research (DIW Berlin) and a research associate at DIW's Public Economics department. I want to thank all the great people who work here and make it such an cooperative and enjoyable research environment.

I am especially indebted to Peter Haan, my first supervisor, who supported me constantly with a lot of time and patience as well as invaluable advice. I would also like to thank Antoine Bozio, my second supervisor, for the many fruitful discussions during the countless project workshops (the latter is also true for the other members of the ORA group).

I am not less grateful to my co-authors Luke Haywood and Katharina Wrohlich for the good and enjoyable collaborations. This is especially true for Kai-Uwe Müller. This dissertation would not have been possible without our endless (sorry Agnes) discussions, his tireless commitment and experienced advice (aside from music).

Much of the empirical work was conducted in the research data center of the Statistical Office for Berlin-Brandenburg. I would like to thank Ramona Voshage and her team for the always friendly and entertaining cooperation. This is also true for the staff at IAB Nuremberg who were always ready to give advise on their great data.

I also want to thank my colleagues at the department and my GC cohort for the great time: Felix Arnold, Hermann Buslei, Sascha Drahs, Stefan Etgeton, Ronny Freier, Johannes Geyer, Verena Grass, Adrian Hille, Daniel Huppmann, Friederike Lenel, Daniel Kemptner, Christian Krekel, Richard Ochmann, Maximilian Podstawski, Philipp Richter, Ulrich Schneider, Songül Tolan and Clara Welteke. I am also indebted to Nicole Haase and Yun Cao for their kind administrative support as well as Adam Lederer for his language advise.

I want to thank Rainer Siegers and Stefan Seifert (and all the others) for countless Foosball matches (except that one game) and Alexander Zerrahn for sticking by me.

In particular I want to thank my office mates Patricia Gallego Granados, Holger Lüthen, Julia Schmieder and Andreas Thiemann for the fun time, their research advise and, most importantly, their friendship. This is also true for Sarah Dahmann and Christian Zankiewicz whose frequent and entertaining visits make them feel as office mates.

Finally, and most important of all, I want to thank Gabi for her endless support and love. You made this possible.

Table of Contents

List of Tables

List of Figures

General Introduction

Motivation

The tax benefit system of a country has multiple objectives. This includes for example raising funds for public expenditures or the redistribution of income in an economy. As the main income source for private households, labour market earnings thereby play a crucial role. They constitute a major part of the tax base for income taxation and social security contributions (SSC) and determine eligibility for many social benefits. Taxes and transfers thus impact work incentives unfolding the potential to importantly impact labour market outcomes. This might be intended for conditional social benefits whose objective often is to induce a steering effect. The widespread *Making Work Pay* policies, for example, introduce work incentives for low-paid workers (Blundell, 2000). Labour market responses might however also represent unintended distortions as for income taxation or SSC. The latter for example are claimed to be detrimental to employment (see for example OECD (1994) and Prescott (2004)). In both cases policy makers need to take labour market responses into account when they aim at making informed decisions on the design of the tax benefit system.

This dissertation focuses on earnings responses to SSC and social benefits (in the following also jointly referred to as social policy) which amounted to on average 10.3% and 13.6 % of GDP, respectively, in the OECD countries in 2014. Their large magnitude makes it essential to understand and quantify their impact on labour market outcomes. The four chapters of this dissertation cover the two crucial channels of labour market reactions encompassed in earnings responses: hours of work and compensation per hour. The first two chapters analyse economic incidence of and employment responses to SSC by different reduced-form approaches exploiting both, longitudinal as well as cross-sectional variation. Chapter three ex-ante evaluates a specific type of in-work benefit for parents promoting full-time or close to full-time employment for low-paid workers. For this purpose a structural labour supply model is augmented by constraints on the labour market. Chapter four highlights the firms' role in responses to tax policy. An equilibrium job search model is structually estimated based on the German low-pay sector which is strongly influenced by non-linearities in the tax schedule. The model is applied to analyse counterfactual policy reforms that aim at removing these tax non-linearities.

SSC as well as most social benefits directly impact private households' net earnings associated with different working time arrangements. This might change their relative attractiveness inducing responses of hours of work or labour market participation. An increase in SSC, for example, might make households want to decrease their labour supply as leisure becomes relatively more attractive. Labour supply responses to social policies are thus naturally in the center of focus in the majority of corresponding empirical studies. That is, they either interpret identified changes in, e.g., hours of work as labour supply reactions or they estimate structural labour supply models.

Labour supply decisions of private households, however, are only one side of the labour market. Firms might also respond to social policy. This is particularly obvious for SSC. In most OECD countries, including Germany, SSC are partly paid by employers and thus (ceteris paribus) directly impact the cost of labour. Firms might respond to this as other input factors become relatively cheaper. The same is also true for some active labour market programs subsidizing firms for employing certain groups of employees. Integration allowances for employers in Germany or SSC subsidies for low-paid jobs in France are examples.

Eventually, earnings responses to social policy are formed in equilibrium. When social policies change the basis of decision-making for some economic agents a new equilibrium might emerge, applying to all market participants. Changes in the labour market quantities might, for example, induce reactions of the gross wage rate. When labour demand is reduced as a response to an increase in employers' SSC, the equilibrium price of labour, the wage rate, comes under pressure. This in turn induces labour supply reactions and so on. Through this channel, employees take part in bearing the burden of employers' SSC and vice versa. Economic might thus deviate from statutory incidence (Fullerton and Metcalf, 2002). The first two chapters of the dissertation at hand analyse both, changes in hours of work as a response to SSC as well as economic incidence of SSC. The interrelations between both margins are thoroughly discussed.

While incidence effects are similarly relevant for social benefits (Andrew, 2010), they are only one channel how effects of social policies unfold in equilibrium. Displacement effects might for example penalise groups not eligible for a large-scale in-work benefit program (Kolm and Tonin, 2011). Chapter four focuses on an alternative channel which is rarely discussed in the literature (see Chetty et al. (2011) for an exception). It provides compelling evidence that an individual's choice set depends on other workers' preferences because firms cater their job offers to aggregate preferences in the market. This has important consequences for labour market responses to social policy. Labour supply responses could be reinforced (Best, 2014). Workers might be forced to react to a policy when firms align offered hours to the incentives generated by that policy. This may even concern workers who are not directly affected by the respective policy. Workers might thus be restricted in their choice of working hours - in particular when their preferences deviate from the majority.

It has been largely recognised that such and other restrictions on the labour market play an important role in determining responses to social policy (Chetty et al., 2011). In standard labour supply models used for policy evaluations, however, actual hours of work have been interpreted in terms of revealed preferences implicitly assuming that choices are not constrained by regulations, institutions, or insufficient labour demand. Chapter three of this dissertation accounts for participation and working hours constraints when evaluating an in-work benefit for parents where eligibility is conditional on exceeding a certain number of working hours. Accounting for hours restrictions turns out to be important as restrictions are found to hamper workers drawing on the benefit. The predictions based on the standard model are shown to be biased in such a setting.

The main joint feature of all chapters in this dissertation thus is that they recognise the importance of looking beyond pure labour supply reactions when evaluating social policy. A second joint feature is the focus on Germany. While the research questions analysed are of general interest, the empirical applications are based on German data. There are several reasons why the German case is interesting for analyzing earnings responses to social policy. SSC account for a large part of total taxation in Germany amounting to about 14% of GDP. Their contribution to government funding in Germany is one of the highest among industrialised countries (OECD, 2016). Nevertheless, there is virtually no previous evidence of earnings responses to SSC based on German data. This is all the more a shortfall as economic incidence is robustly found to depend on wage setting mechanisms (Alesina and Perotti, 1997; Daveri and Tabellini, 2000). Findings of studies based on US or Scandinavian data which account for the majority of the literature are thus not easily transferable. Further, I exploit some interesting non-linearities of the German tax benefit system. Besides earnings caps for SSC which exist in many countries, the SSC and tax exemption for low-paid employments provides a unique setting. It generates a very large discontinuity which only applies for a fraction of workers competing in the same market.

The chapters at hand can, though, be differentiated with respect to several dimensions. By focusing on SSC the first two chapters study the revenue side of the social security system.

The remaining chapters analyse conditional social benefits standing in the tradition of *Making Work Pay* (MWP) policies like the Earned Income Tax Credit in the US or the Working Family Tax Credit in the UK. As argued above, earnings responses to SSC usually represent unintended distortions. By contrast, the tax exemption for low earnings analysed in chapter four has the objective to facilitate the entrance into the labour market. Behavioural responses are thus explicitly intended. On the other hand it is a good example that social benefits are not immune against unintended consequences. The withdrawal of the subsidy for increasing earnings generates high implicit marginal tax rates. This implies an incentive to reduce hours of work for already employed individuals which is typical for MWP policies (Bargain et al., 2010). Chapter three therefore evaluates an in-work benefit for parents whose design aims at minimizing negative externalities. Eligibility is conditional on working hours exceeding a certain amount. The objective is to stimulate employment with high hours by making them more attractive. As the size of the benefit depends on the individual wage rate instead of individual or household earnings negative incentives on hours of work are impeded.

Methodology also differs between the chapters of this dissertation. One potential differentiation mirrors the two competing paradigms for policy evaluation in economics (Chetty, 2009). I do not take sides and apply both, *reduced-form* and *structural* approaches. The first two chapters follow rather reduced-form strategies. They place priority on the identification of causal effects by exploiting quasi-experimental variation (Chetty, 2009). The identified parameters - labour demand and supply elasticities as well as economic incidence - however, are crucial for policy makers in general and are not specific to the investigated variation. Hours responses are analysed by a modification of the recently developed bunching approach (Saez, 2010). The estimation equation is directly derived from a behavioural model; it is thus a *reduced-form* approach in the literal sense. The cross-sectional variation used is induced by the earnings caps for SSC. The same variation is exploited in chapter two to identify economic incidence. The applied approach does not directly base on an underlying behavioural model, though (Alvaredo and Saez, 2007). In the first chapter economic incidence is identified by variation in SSC over time generated by a policy reform. A difference-in-differences approach is applied.

Chapters three and four follow more structural approaches. Their major advantage is to allow for ex-ante evaluations of a variety of (hypothetical) policies. Chapter three augments a standard discrete choice model of labour supply by hours and participation constraints. The model is applied to evaluate an in-work benefit for parents which has not been introduced so far. Chapter four structurally estimates an equilibrium job search model which allows for counterfactual simulations. It analyses the effects of removing the non-linearities in the tax schedule for low-paid employment in Germany.

Contributions

This dissertation focuses on earnings responses to SSC and social benefits. While going beyond purely focusing on labour supply is a contribution shared by all chapters, every one makes important contributions to its strand of literature.

A joint contribution of the first two chapters is to provide fresh evidence on earnings responses to SSC. The first chapter analyses both, responses of hours of work as well as economic incidence. The second chapter focusses on economic incidence only. Empirical evidence on hours responses to SSC is rare (Saez et al., 2012). Empirical evidence on economic incidence is more prevalent but inconclusive (Melguizo and Gonzalez-Paramo, 2013). On Germany there is hardly any evidence although economic incidence is robustly found to depend on the institutional setting (Alesina and Perotti, 1997; Daveri and Tabellini, 2000). Moreover, specific features of SSC like the formal burden sharing impede transferring the results from the related literature on income taxes.

Several features of my analysis makes this a particularly valuable contribution. First, the reform evaluated in chapter one, the increase of the earnings cap of SSC in East Germany, is very suitable for analysing economic incidence. For a subgroup of affected employees the reform increases average but not marginal SSC rates. Abstracting from income effects and effects on effort, a change in gross earnings can be ascribed to economic incidence. Second, chapter two is one of the first studies that formally applies a recently proposed cross-sectional approach to estimate economic incidence (Alvaredo and Saez, 2007). The approach exploits discontinuities in earnings distributions at kinks in the budget set which are informative about tax incidence. Contrary to most research on economic incidence of SSC it does not rely on policy reforms, panel data, or hours information. When the location of kinks does not change significantly, estimates are less affected by short-run adjustment frictions than results based on policy reforms. Similar arguments can also be brought forward for the cross-sectional analysis of hours responses in chapter one.

The third chapter turns away from SSC and ex-ante evaluates an in-work benefit for low-paid parents that particularly targets work incentives for secondary earners. As it is largely recognised that workers face important constraints in choosing labour supply (e.g. Stewart and Swaffield (1997a)) the utilised labour supply model accounts for participation and hours constraints. This might be particularly important here since eligibility of the program is conditional on hours of work exceeding a certain threshold. Restrictions might thus hamper individuals drawing on the benefit. A major challenge is to identify potential constraints separately from preferences (Beffy et al., 2016). Usually, both have to be inferred from the distribution of actual hours of work. The model set up in chapter three additionally uses survey data on desired hours of work which are exploited as revealed preferences. Constraints are identified by the deviation between desired and actual hours of work (see also Euwals and van Soest (1999) and Bloemen (2008)). This provides for a more robust identification of preferences and constraints. It further allows for explaining constraints by a variety of factors. Besides labour demand rationing the model consider working hours norms varying across occupations as well as restrictions which are not originated in the labour market. Insufficient formal childcare, for instance, might prevent parents from accepting jobs with many hours.

While chapter three credibly identifies restrictions on the labour market based on survey data restrictions are not modeled structurally. Job search models, by contrast, provide a way of explaining some of these constraints without resorting to ad-hoc restrictions on the set of labour supply choices available to individuals. Chapter four structurally estimates an equilibrium job search model based on data on the German low-pay sector. This part of the labour market is strongly influenced by nonlinearities in the tax schedule - even for individuals to whom this tax schedule does not apply. Compelling evidence is presented of what is in this dissertation designated as *population labour supply effects*. Because firms package their hours-wage bundles according to average preferences of job-seeking workers, the restricted set of any particular individual worker will depend on other workers' preferences. Chapter four shows that tax subsidises offered to certain workers in low-paid jobs end up affecting all workers in the low-pay sector.

The model is used to evaluate removing the non-linearities in the tax schedule and analyse their distorting effects. This is relevant beyond the specific German setting and applies to many tax subsidies for low-paid jobs. Chapter four thus contributes to the body of evidence on *Making Work Pay* (MWP) policies (e.g. Blundell (2000) for the US, Bargain and Orsini (2006) for Europe). It is shown that firms can reinforce the negative earnings responses of already employed workers which are found when evaluating MWP policies (Bargain et al., 2010). Chapter four particularly complements other work which discusses how unintended consequences of MWP policies unfold in equilibrium (Rothstein, 2010; Andrew, 2010; Kolm and Tonin, 2011; Tazhitdinova, 2015). While these studies focus on channels like incidence or displacement effects the population labour supply effect is added as additional channel. Moreover, discontinuous taxes are included into a standard equilibrium job search model. Whilst other authors include taxes in equilibrium job search models (e.g. Shephard (2012)) taxes there are assumed to be continuous. In line with the bunching literature (Saez, 2010), important additional welfare costs might be caused by the discontinuous nature of taxes. These might be further increased by firm responses (Tazhitdinova, 2015). In contrast to most equilibrium job search models, the model set up in chapter four is able to rationalise discontinuous bunching at tax thresholds in the interior of the earnings distribution. It is thus able to analyse counterfactual policy reforms that aim at removing the welfare costs induced by tax non-linearities.

Overview

Table 1: Overview of Chapters

	Chapter 1	Chapter 2	Chapter 3	Chapter 4
Title	Earnings Responses to Social Security Contributions	The economic incidence of social security contributions - A discontinuity approach with linked employer-employee data	An in-work benefit for parents - Evidence from a structural labour supply model with participation and hours constraints	Firms' responses to aggregate preferences of workers - Evidence from a German Making Work Pay policy
Main Findings	Economic=statutory incidence at the individual level; no hours responses; both imply low labour supply and demand elasticities	Economic=statutory incidence at the individual level in the long-run	Standard model over-estimates policy effect; financial incentives & removing restrictions both make mothers increase hours	Firms play a crucial role in workers' responses to taxes; smoothing Mini-job discontinuity increases workers' utility and tax revenues
Data	SIAB, FAST	GSES	FiD, GSOEP	SIAB, GSOEP
Method	Difference-in-differences, bunching	Alvaredo and Saez (2007)	Labour supply discrete choice model with participation and hours constraints	Equilibrium job search model allowing for discontinuous tax schedule
Co-author	–	Kai-Uwe Müller	Kai-Uwe Müller & Katharina Wrohlich	Luke Haywood
Publication Status	Revise and Resubmit at *Labour Economics*, DIW Discussion Paper 1489	DIW Discussion Paper 1578	–	–

Chapter 1

Earnings Responses to Social Security Contributions*

1.1 Introduction

In most industrialised countries social security contributions (SSC) represent a large share of total taxation. In 2010, SSC amounted to 9.1 % of GDP and 26.4 % of total tax revenues in OECD countries (OECD, 2015). Since they are (usually) nominally shared between employers and employees it is hardly surprising that SSC are often claimed to be detrimental to employment and economic growth (see for example OECD (1994) and Prescott (2004)). Some governments seem to have followed this reasoning by implementing reforms to decrease SSC or switch from SSC to other means of revenue generation (Melguizo and Gonzalez-Paramo, 2013). During 2011 and 2012, the US decreased the employees' share of payroll taxes by two percentage points (pp.) in order to boost consumption. Many countries including Germany, Denmark and France increased the value added tax in order to finance a decrease in SSC. Additionally, there seems to be a tendency to shift some legal burden of SSC from employers to employees. Germany, for example, abolished the equal sharing of SSC in 2005. To understand the effect of such reforms on disposable income, employment and welfare, the elasticities of labour supply and demand with respect to SSC as well as economic incidence are crucial.

Based on two large administrative data sets this study analyses these parameters for employees

*This chapter is based on Neumann (2015) which contributes to the project *The Impact of Social Security Contributions on Earnings: Evidence from administrative data in France, Germany, the Netherlands and the UK* sponsored by the *Deutsche Forschungsgemeinschaft.* I have been invited to revise and resubmit the paper for publication in *Labour Economics.*

in the upper part of the earnings distribution. Identification makes use of the fact that health insurance rates in Germany (as in many other countries) only apply up to certain thresholds of earnings, in the following called (earnings) caps. Firstly, economic incidence is estimated within a quasi-experimental approach as similarly done by Gruber (1994), Gruber (1997) and Saez et al. (2012). Secondly, behavioural responses are studied within a modified bunching approach (Saez (2010) and Chetty et al. (2011)).

More precisely, economic incidence is estimated by exploiting exogenous variation of SSC over time generated by a considerable increase of a regional earnings cap of health and long-term care insurance in 2001. I estimate difference-in-differences models using employees with earnings above the earnings cap as treatment group and employees with earnings somewhat below as control group. The amount of hours worked and respectively hourly wages are not observed in the administrative data used here. I, therefore, follow the literature (Anderson and Meyer (2000), Bennmarker et al. (2009) and Saez et al. (2012)) and identify economic incidence by the reaction of yearly gross earnings[1] under the assumption of no employment responses. In the present application, however, this usually restrictive assumption can be credibly defended based on two arguments. First, a group of employees is only exposed to a change of the average SSC rate. Second, the preceding analysis of behavioural responses to earnings caps lends additional support to the assumption.

This analysis of behavioural responses exploits that the discontinuous drop of the marginal SSC rate at an earnings cap generates a kink in the budget set. The modified bunching approach relates the size of a potential dip in the earnings distribution at such a kink to intensive margin employment responses to a change in SSC. Four earnings caps of health and long-term care insurance between 1998 and 2004 are analysed with the drop of the marginal SSC rate being approximately 16 pp.

I find employment responses at the intensive margin to be negligible. Further, according to my results economic and legal incidence coincide implying that employees and employers share the burden of SSC roughly in equal parts. Taken together, this suggests that labour supply and demand elasticities both are very low. Considering that earnings caps naturally affect rather high-skilled employees, this result is consistent with the previous literature. The findings on economic incidence and behavioural responses are consistent with a standard competitive labour market model where economic incidence is determined by the ratio of labour demand and supply elasticities.

[1] Throughout the text, gross earnings include employees' SSC but are net of employers' SSC. Net earnings are gross earnings minus employers' and employees' SSC and labour costs are gross earnings plus employers' SSC.

My main contribution to the literature is to provide fresh evidence on economic incidence of SSC. Previous empirical work found mixed results. One potential reason is that economic incidence of SSC seems to crucially depend on the institutional setting like the centralisation of the wage bargaining process (Alesina and Perotti (1997) and Daveri and Tabellini (2000)). Yet, hardly any evidence is available for Continental European countries which mostly feature an intermediate degree of centralisation of the wage bargaining process. This study fills that gap by providing evidence for high-skilled employees in Germany.

While the analysis of behavioural responses to SSC supports the credibility of the evidence on economic incidence, it is also interesting in itself as previous evidence is rather rare (Saez et al., 2012). This is astonishing because the specific features of SSC in comparison to income taxes require to analyse their effects separately. I translate the bunching method by Saez (2010) to 'upward' kinks[2] in the budget set. This avoids some of the major problems responsible for the insufficient evidence. First, it does not depend on the existence of an exploitable policy reform. Second, it does not require information on hours of work. I can, therefore, use large administrative data sets that frequently do not include the exact amount of working hours. Third, while most existing evidence is limited to short-term effects, I argue that in the present setting the bunching method allows for estimating long-term responses.

Further, the effect of increasing the earnings cap for SSC is interesting for policy-makers as it constitutes a feasible policy for many countries to increase revenues or to shift the burden to employees in the upper part of the earnings distribution. In the UK, for example, the earnings cap for employers' SSC was abolished in 1985 as were most earnings caps for SSC in France in the 1980s. This study contributes to the controversial discussion of welfare effects and efficiency of an increase of an earnings cap for SSC (as briefly summarised by Liebman and Saez (2006)).

The paper is organised as follows. Section 1.2 discusses the existing literature. Section 1.3 presents the German social security system. Section 1.4 reviews the evidence on behavioural responses based on cross-sectional variation in SSC. The evidence on economic incidence based on the increase of the earnings cap in 2001 is discussed in section 1.5. Section 1.6 discusses the findings of both approaches and section 1.7 concludes.

1.2 Literature

This study relates to the large strand of literature on the effects of taxes on labour market outcomes. Blundell and Macurdy (1999) as well as Meghir and Phillips (2010) survey the vast

[2] An upward kink in a gross-net earnings diagram is generated by a discontinuous drop in the marginal tax rate at a certain threshold. By contrast, Saez (2010) analyses downward kinks generated by an increase in the marginal tax rate.

evidence on income taxation. The specific characteristics of SSC have given rise to an own branch of empirical literature. There is a much larger body of research on economic incidence for SSC than for personal income tax, probably because labour taxes formally paid by employers are considered to be more prone to burden shifting than taxes paid by employees. In contrast, evidence on behavioural responses or employment effects to SSC is rather rare.

The early work on economic incidence was mainly based on cross-country national account data (see for example OECD (1990)) and usually found that labour taxes are completely shifted to workers. However, more recent multi-country studies draw a more differentiated picture and conclude that shifting to wages seems to be an inverse U-shaped function of the degree of centralisation of wage bargaining (Alesina and Perotti (1997) and Daveri and Tabellini (2000)) and increases with the link between contributions and benefits (Ooghe et al. (2003)).

Most empirical studies are based on individual data. With respect to methodology, these studies can roughly be classified into two groups. First, some studies exploit variation in SSC rates over time or cross-sectional units and frequently find that increases in payroll tax rates are just partly shifted to wages (see for example Holmlund (1983) for Sweden, Hamermesh (1979) and Neubig (1981) for the US).

Second, policy reforms are exploited as natural experiments in order to estimate both, economic incidence of SSC and effects on employment. In an influential paper, Gruber (1994) analyse the effects of the introduction of mandated maternity benefits in the US on gross hourly wage rates and employment with a difference-in-differences-in-differences approach. Based on survey data he finds substantial shifting to wages and no impact on overall labour input. Since then, a large amount of quasi-experimental studies have been conducted for many different countries. Identification is based on variation between, among others, firms (Anderson and Meyer (2000) for the US, Gruber (1997) for Chile), industries (Bell et al. (2002) for the UK), age (Skedinger (2014) for Sweden) and regions (Bennmarker et al. (2009) for Sweden, Korkeamki and Uusitalo (2009) for Finland, Baicker and Chandra (2006) for the US). The results are very mixed which is reflected in a meta-study based on 52 empirical papers (Melguizo and Gonzalez-Paramo, 2013). On average 66 % of labour taxes are estimated to be borne by employees with a standard deviation of 51 pp. Some of these studies also analyse the impact of SSC on employment but rarely find statistically and economically significant effects.

Only few studies exploit the discontinuity induced by an earnings cap for SSC. Similar to section 1.5 of this study, Lang (2003) and Liang et al. (2004) analyse significant increases of the earnings cap of the American Federal Insurance Contribution Act (FICA)[3] between 1968 and

[3]Similar to the caps analysed in this study, the marginal payroll tax rate drops to zero at the earnings cap of FICA.

1979. While Lang (2003) finds that earnings of treated individuals rose consistently stronger in years the cap increased, Liang et al. (2004) conclude that gross hourly wages were not significantly affected. Further, the latter study finds a small negative employment effect at the intensive margin. Saez et al. (2012) evaluate a Greek reform which created parallel regimes by increasing the earnings cap for SSC for all employees who started working on or after 1993. They, therefore, estimate credible long-term effects which most other studies could not provide and find that economic and legal incidence coincide as well as that labour supply effects are negligible.

A recent strand of literature initiated by Saez (2010) and Chetty et al. (2011) exploits cross-sectional variation in tax rates to analyse behavioural responses. This so-called bunching approach does not depend on exogenous policy reforms. Under some additional assumptions it facilitates the estimation of long-term responses. This is a major advantage over reform evaluations which analyse the response to a change in incentives and, therefore, usually estimate short- or medium-run effects. Identification of the aforementioned studies is based on discontinuous jumps in the marginal (income) tax rate. Discontinuous drops as they appear at earnings caps for SSC are not at all or only informally covered. In section 1.4 I, therefore, modify the bunching method to analyse behavioural responses to the German earnings caps of health insurance.

1.3 Institutions

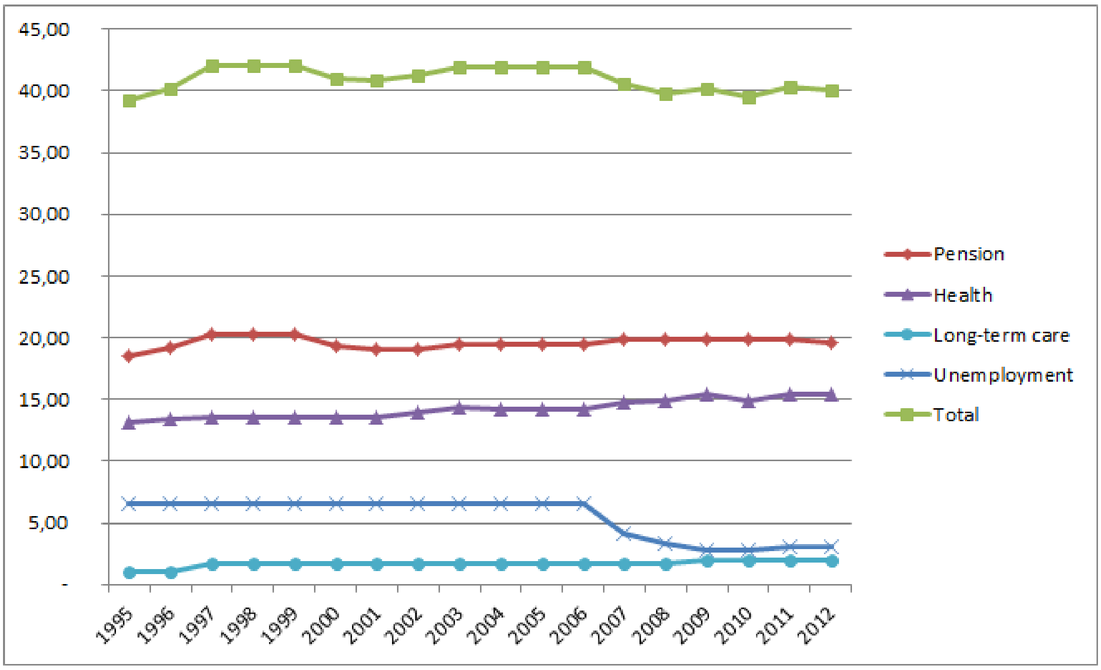
This study focuses on the impact of SSC on gross earnings in Germany. The German social security system consists of pension, health, unemployment, and long-term care insurance. In comparison to personal income taxes, SSC in Germany and many other countries have some specific features which are crucial for economic incidence and behavioural responses.

First, they are financed by flat contribution rates with daily gross earnings as tax base. Over the last twenty years the total SSC rate varied around 40 % of gross earnings with pension (around 20 %) and health insurance (around 14 %) as the most important branches (figure 1.1). The contribution rate to unemployment insurance decreased from 6.5 % before 2007 to 3 % in 2012. Apart from that the respective SSC rates did not change substantially.

Second, the burden is statutorily shared between employers and employees. Until 2004 this split was equal. Since 2005, 0.9 pp. (plus 0.25 pp. for childless employees) are exclusively paid by employees[4]. As of 2009, providers of public health care may charge so-called 'auxiliary contributions'. In 2011 the employers' SSC rate was fixed such that future increases are only born by employees. However, as these changes have been rather gradual, the present study

[4]There are specific rules for Saxony.

Figure 1.1: Development of SSC rates over time



Notes: The additional fee for childless employees, introduced in 2005, is omitted. The change of SSC which came into effect in July 2005 are considered as of 2006. Until 2006 SSC rates for health insurance varied between health insurance companies and the given numbers are averages. In 2001, for example, it varied between 11.0 % and 14.9 % (Grabka, 2004).
Source: http://www.statistik.baden-wuerttemberg.de; accessed 2015

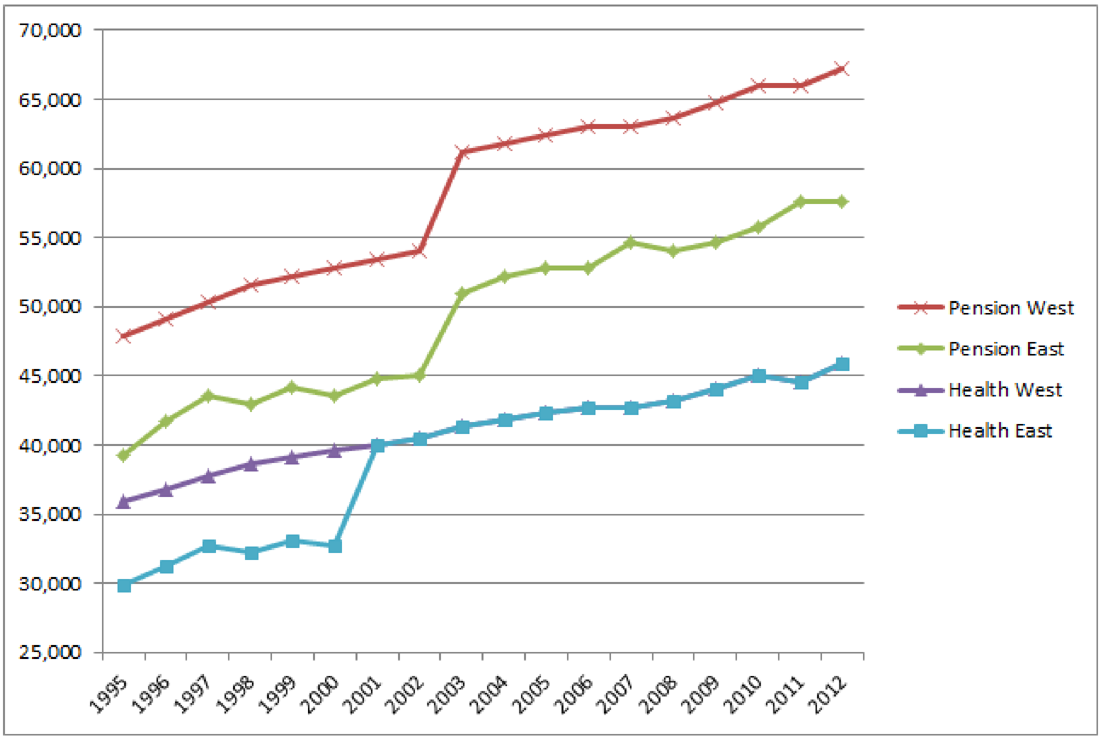
does not use them as a source of identification.

Third and most important for this study, marginal SSC rates only apply up to certain thresholds of earnings, in the following called (earnings) caps. There is a common cap for pension[5] and unemployment insurance that varies between East and West Germany (figure 1.2). A second, lower cap is defined for health and long-term care insurance[6]. Until 2000 it was also different for East and West Germany and until 2002 it was 3/4 of the cap of pension insurance. All caps are adjusted yearly according to the change in the average gross wage bill in the preceding year. However, there have only been very few substantial changes. There is a strong increase of the East German health earnings cap in 2001 when it is harmonised with the cap in West Germany. I exploit this jump to estimate economic incidence of SSC (section 1.5). Comparable jumps in the caps of pension insurance in 2003 cannot be utilised due to data limitations. In addition the cross-sectional variation in SSC generated by the earnings cap of health insurance is exploited to estimate behavioural responses to SSC (section 1.4).

[5]There is a special earnings cap for miners' pension insurance. However, as it is usually higher than the standard one, it does not pose a problem for the analysis.

[6]For the sake of readability I speak in the following of earnings caps of pension and health insurance instead of also mentioning unemployment and long-term care insurance.

Figure 1.2: Development of earnings caps over time



Notes: The vertical axis pertains to yearly gross earnings.
Source: http://www.statistik.baden-wuerttemberg.de; accessed 2015

Fourth, SSC differ from other taxes as they might be directly linked to benefits. While individual pension and unemployment benefits in Germany are a function of earnings, there is no direct link between benefits and the contributions or the cap of the here analysed health insurance[7].

1.4 Behavioural Responses

This section analyses behavioural responses to SSC by exploiting the cross-sectional variation in health insurance rates. To do that, I translate the bunching approach by Saez (2010) and Chetty et al. (2011) to upward kinks in the budget set where the marginal SSC rate drops discontinuously[8]. It is shown that a potential gap or dip in the earnings distribution at such a kink contains information about intensive margin employment responses to a change in SSC. However, analysing the German earnings caps of health insurance between 1998 and 2004, no gaps or dips are found (section 1.4.4). Assuming that economic incidence is entirely with employees, this implies that for employees in the upper part of the earnings distribution the

[7]Compensation in case of an illness lasting over 6 weeks is an exception.
[8]By contrast, Saez (2010) analyses downward kinks generated by an increase in the marginal tax rate.

compensated elasticity of labour supply is low. Although this assumption on economic incidence is common in the literature on labour supply, there are arguments against its validity. First, in contrast to income taxation, legal incidence of SSC is shared between employees and employers. Second, previous studies for other countries (section 1.2) and the findings of this study (section 1.5) lend support to the claim that employers and employees share the economic burden of SSC. While I maintain the standard assumption on economic incidence throughout this section, the implications of relaxing it are discussed in section 1.6.

Finally, the bunching approach is restricted to analyse employment responses at the intensive margin. This study does not present any evidence on participation effects. However, as the German earnings caps appear in the upper part of the earnings distribution, strong employment responses at the extensive margin are not likely.

1.4.1 Theory

Following Saez (2010), assume that preferences of employees can be represented by a quasi-concave utility function $u(c, y)$ which depends negatively on gross earnings y and positively on consumption c[9]. Let ability n be smoothly distributed over the population. In such a setting, downward kinks in the budget set (generated by a jump in the marginal tax rate) might create the incentive to locate at these kinks. Saez (2010) shows that the observed extent of bunching can be used to estimate the structural compensated labour supply elasticity. By contrast, an earnings cap at $\bar{y}$ where the marginal SSC rate drops from t_0 to t_1 generates the piecewise linear budget set (1.1) with an upward kink at $\bar{y}$ and R denoting non-wage income.

$$c = \begin{cases} (1-t_0)y + R & \forall y \leq \bar{y} \\ (1-t_0)\bar{y} + (1-t_1)(y-\bar{y}) + R & \forall y > \bar{y} \end{cases} \tag{1.1}$$

Building on the assumption of convex preferences it can be shown that there always exists an interval around an upward kink which does not contain a global maximum (Burtless and Hausman, 1978). If the labour supply elasticity is homogeneous this implies a gap in the earnings distribution (Saez, 2010). Intuitively, employees extend their working hours until the utility of an additional hour of work is offset by the hour of leisure lost. Therefore, employees who locate closely[10] to the right of the hypothetical kink in a linear tax system would extend their labour supply if an upward kink is introduced as the marginal utility of an additional

[9] Appendix A.1.1 specifies a quasi-linear and iso-elastic utility function to illustrate the general relationships derived in this section.

[10] Such that income effects are negligible.

hour of work increases discontinuously. The same is true for employees who locate close enough to the left of the hypothetical kink in a linear tax system such that the utility gain due to the additional lowly taxed hours beyond the kink at least offsets the utility loss due to the additional highly taxed hours up to the kink.

Assuming that $u(c, y)$ satisfies the single-crossing property there is an unique ability level $\tilde{n}$ above (below) which employees prefer locating above (below) the kink. This is shown exemplarily by means of a quasi-linear and iso-elastic utility function in Appendix A.1.1. Let $\tilde{y_0} < \bar{y}$ denote the optimal gross earnings of an individual with ability $\tilde{n}$ in the absence of a kink and $\tilde{y_1} > \bar{y}$ if there is an upward kink at $\bar{y}$. The elasticity of labour supply, e, is then defined as

$$e = \frac{\tilde{y_0} - \tilde{y_1}}{\tilde{y_0}} \frac{1 - t_0}{t_0 - t_1} = \frac{d\tilde{y}}{\tilde{y_0}} \frac{1 - t_0}{dt}$$

While $\frac{1-t_0}{dt}$ is known by the tax system, $\frac{d\tilde{y}}{\tilde{y_0}}$ has to be estimated in order to infer e. This can be done by measuring the position and extent of the arising gap in the earnings distribution around $\bar{y}$ which will cover all earnings between $\tilde{y_0}$ and $\tilde{y_1} = \tilde{y_0} + \Delta\tilde{y_0}$. As it is based on individuals who would locate close to the cap in the counterfactual situation this estimator of e is very local. However, if the elasticity is heterogeneous (and zero for at least some employees), an upward kink will result in a dip rather than a gap because $d\tilde{y}$ varies over individuals. Analogous to the bunching mass exploited in analyses of downward kinks (Saez, 2010), the missing mass MM relative to the density of counterfactual earnings, $h_0(.)$, can be used to approximate the average earnings response:

$$\begin{aligned} MM &= \int \int_{y_0}^{y_0 + \Delta y(e)} h_0(y, e) dy \ de \\ &\approx \int dy(e) \overline{h_0(y, e)} de \\ &= E_e[dy] \overline{h_0(y)} \end{aligned} \tag{1.2}$$

$\overline{h_0(.)}$ denotes the average counterfactual density in the area of the potential dip. Following most of the bunching literature (see for example Saez (2010)), equation (1.2) assumes that the tax change is small, so that $\overline{h_0(y, e)} \approx \overline{h_0(y)}$. Since this assumption also yields $\tilde{y_0} \approx \bar{y}$, the elasticity parameter e can, therefore, be estimated by

$$\hat{e} = \frac{\frac{\hat{MM}}{\hat{\overline{h_0(y)}}}}{\overline{y_0}} \frac{1 - t_0}{dt} \tag{1.3}$$

Appendix A.1.1 relaxes the assumption of a small tax change by specifying a quasi-linear and iso-elastic utility function. Equation (A.7), then, constitutes a direct relationship between MM, e and the parameters of the tax system.

1.4.2 Empirical Methodology

The empirical task is to detect a potential gap or dip in the gross earnings distribution around the earnings cap. To do that, the observed distribution is compared to an estimated distribution which approximates the counterfactual situation of a linear SSC schedule. Following Chetty et al. (2011) the counterfactual distribution is estimated by a polynomial of degree P fitted to the number of observations (C_j) in earnings bins. Bin j is defined by its midpoint Y_j € and its bin size k €. Earnings bins within the interval $(\bar{y} - Q, \bar{y} + Q)$ are assumed to be affected by a potential dip and are, therefore, excluded from the estimation of the counterfactual distribution.

$$C_j = \sum_{p=0}^{P} \beta_p Y_j^p + \sum_{i=-Q}^{Q} \gamma_i \mathbf{1}[Y_j = i] + \epsilon_j \tag{1.4}$$

The counterfactual earnings distribution is $\hat{C}_j = \hat{\beta}_p Y_j^p$ as estimated by equation (1.4)[11]. Missing mass MM, then, is the difference between the counterfactual and empirical earnings distributions within the interval $(\bar{y} - Q, \bar{y} + Q)$. In the following, the outlined approach is applied to the earnings caps of health insurance in West and East Germany.

1.4.3 Data

The estimation is based on the Wage and Income Tax Statistics (FAST)[12] which is a representative ten percent sample of households subject to German income tax (3.5 million observations). Due to this relatively large sample size FAST is particularly suitable for the semi-parametric bunching approach. Labour earnings are observed as uncapped yearly gross earnings without information about the amount of hours or months of work. This is sufficient for the purpose of this study because, first, the tax base of SSC is daily gross earnings and, second, earnings of jobs exempted from SSC are not included in this measure as the lower exemption level also applies to income tax. Self-employed and civil servants are excluded from the sample. The analysis is conducted separately for 1998 and 2004 as well as East and West Germany. Interpretation of the results needs to account for the peculiarities of both years and regions. While the earnings

[11] $\hat{C}_j$ is not an estimate of the true counterfactual earnings distribution in the absence of a kink. The reason is that all employees with counterfactual earnings above the cap have an incentive to respond to an introduction of the kink. Assuming that the substitution effect dominates, the observed density might, therefore, be less concentrated and shifted to the right.

[12] The German name is *Faktisch anonymisierten Daten aus der Lohn- und Einkommensteuerstatistik*.

cap of health insurance changed considerably in 2001 in East Germany, it was only subject to relatively small yearly adjustments[13] in the years preceding 2004 in West Germany and 1998 in both regions. As the earnings cap of health insurance and the threshold for compulsory public health insurance were decoupled in 2003, the analysis of 2004 are unaffected by the private health insurance system (see Appendix A.2.1).

1.4.4 Results

In table 1.1 $\hat{MM}$ denotes the estimated amount of missing employees normalised by the average counterfactual density in the excluded range. $\hat{e}$ is the elasticity estimate implied by equation (1.3). $\sigma_{\hat{MM}}$ and $\sigma_{\hat{e}}$ denote the corresponding standard errors, N the amount of individuals underlying the estimation. Graphs of the actual and counterfactual earnings distributions can be found in Appendix A.1.2. The basic specification is based on polynomials of degree seven and a bin size of 300 €/year. An earnings range of 1800 €/year below and above the kink is excluded. The latter this is chosen such that it covers the theoretical range of a gap assuming an elasticity of 0.5 (Appendix A.1.3). I do not find a statistically significant deviation between counterfactual and empirical earnings distribution at three of the four analysed caps (table 1.1). The only exception is West Germany in 2004 where I find a small but statistically significant excess mass. Nonetheless, this implies that there are neither gaps nor dips. Correspondingly, the point estimates of the elasticity are virtually zero ranging between -0.006 and 0.004 in the basic specification. The results turn out to be very robust to successively varying the degree of the polynomial, the bin size and the exclusion range (table 1.1). The only exceptions are specifications with a bin size of 100 € in 2004 resulting in even more negative estimates of up to -0.018. Summing up, these estimates suggest that the compensated labour supply elasticity with respect to SSC (of employees with earnings close to the respective caps) is zero or at least very low.

Opposed to some recent literature, the framework applied here does not allow for estimating (and accounting for) search or adjustment costs. It can, therefore, not be excluded that the observed response is attenuated by frictions (Chetty et al., 2011). Based on a quasi-linear and iso-elastic utility function I simulate an employee's instant utility gain (expressed in yearly net earnings) of responding optimally to the German earnings caps (Appendix A.1.3). For an individual who would locate exactly at the earnings cap in the absence of a kink the utility increases between approximately 170 € for an elasticity of one and the West German earnings cap of health insurance to 16 € for an elasticity of 0.1. Previous literature suggests that

[13] These regular yearly adjustments are based on changes in the average wage bill and fluctuated between -1.2 % and 2.9 % in the analysed period.

Table 1.1: Bunching analysis of employment responses

Specification	*1998*					*2004*				
	$\hat{MM}$	$\sigma_{\hat{MM}}$	$\hat{e}$	$\sigma_{\hat{e}}$	$\frac{N}{1000}$	$\hat{MM}$	$\sigma_{\hat{MM}}$	$\hat{e}$	$\sigma_{\hat{e}}$	$\frac{N}{1000}$
West										
Basic	−0.27	0.28	−0.003	0.003	353	−0.47	0.16	−0.004	0.001	490
Degree 9	−0.31	0.3	−0.003	0.003	353	−0.44	0.18	−0.004	0.002	490
Bin Size 100	−0.91	0.79	−0.009	0.008	356	−1.33	0.45	−0.012	0.004	495
Excl. 900	0.01	0.13	0.000	0.001	382	−0.19	0.08	−0.002	0.001	527
East										
Basic	0.32	0.43	0.004	0.005	323	−0.69	0.49	−0.006	0.004	248
Degree 9	0.42	0.47	0.005	0.006	323	−0.42	0.55	−0.003	0.005	248
Bin Size 100	0.1	1.43	0.001	0.017	326	−2.04	1.5	−0.018	0.013	251
Excl. 900	0.21	0.21	0.002	0.002	352	−0.33	0.23	−0.003	0.002	270

Notes: $\hat{MM}$ denotes the missing mass normalised by $\overline{\hat{h_0(y)}}$ as calculated by a modified version of the program by Chetty et al. (2011) which also includes a parametric bootstrap procedure to calculate standard errors $\sigma_{\hat{MM}}$. $\hat{MM}$ is positive (negative) if the empirical distribution comprises less (more) individuals than the counterfactual one. $\hat{e}$ denotes the implied elasticity estimate and $\sigma_{\hat{e}}$ its standard error. N denotes the number of observations used in the estimation. The basic specification estimates polynomials of degree 7 on bin sizes of 300 € excluding a range of 1800 €above and below a cap. The alternative specifications successively vary one of these parameters.
Source: FAST, own calculations.

adjustment costs might exceed these utility values[14]. Gelber et al. (2013), for example, estimate costs of adjusting earnings optimally with respect to a kink in the budget set to be $150 for the US. However, relaxing the assumption that economic incidence is fully with employees implies that employers also have the incentive to avoid earnings close to an earnings cap (section 1.6). The utility gains are, therefore, augmented by gains for the employer making the friction story look less likely.

Related to the discussion on adjustment costs is the question whether the results can be interpreted as long- or short-term effects. On the one hand, as earnings caps are relatively salient and employees and employers arguably have congruent incentives they can easily incorporate the earnings cap into wage negotiations. In contrast to reform evaluations which analyse the response to a change in incentives, the estimates of the bunching approach in the present setting might, therefore, be interpreted as long-term effects. On the other hand, the small yearly adjustments complicate optimization, especially because they are not announced before September in the preceding year. Appendix A.1.4, therefore, repeats the analysis based on the earnings cap of the preceding year. Still, no dips can be detected suggesting that the yearly adjustments cannot entirely explain the absence of behavioural responses.

1.5 Economic Incidence

Besides potential responses in working hours to a variation of SSC, changes in the wage rate are another margin of adjustment for gross earnings. As in most labour supply analyses, I

[14]Unfortunately, the FAST data set does not include information like industry sector or occupation which would allow analysing for example workers who we know can more easily adjust their labour supply.

assume economic incidence to be entirely with employees in the previous section. According to this assumption a raise of SSC implies that the hourly gross wage rate decreases because SSC are shared between employers and employees. The additional burden is, thus, shifted to employees. Alternatively, if economic coincided with legal incidence, the gross wage rate would be unaffected[15]. Should economic incidence fall entirely to employers, the gross wage rate would increase.

I exploit variation of SSC over time induced by a substantial increase of the East German earnings cap of health insurance to analyse economic incidence. Similar to Gruber (1994) and Saez et al. (2012) the reform is analysed as a quasi-experiment by a difference-in-differences approach.

1.5.1 The Reform in 2001

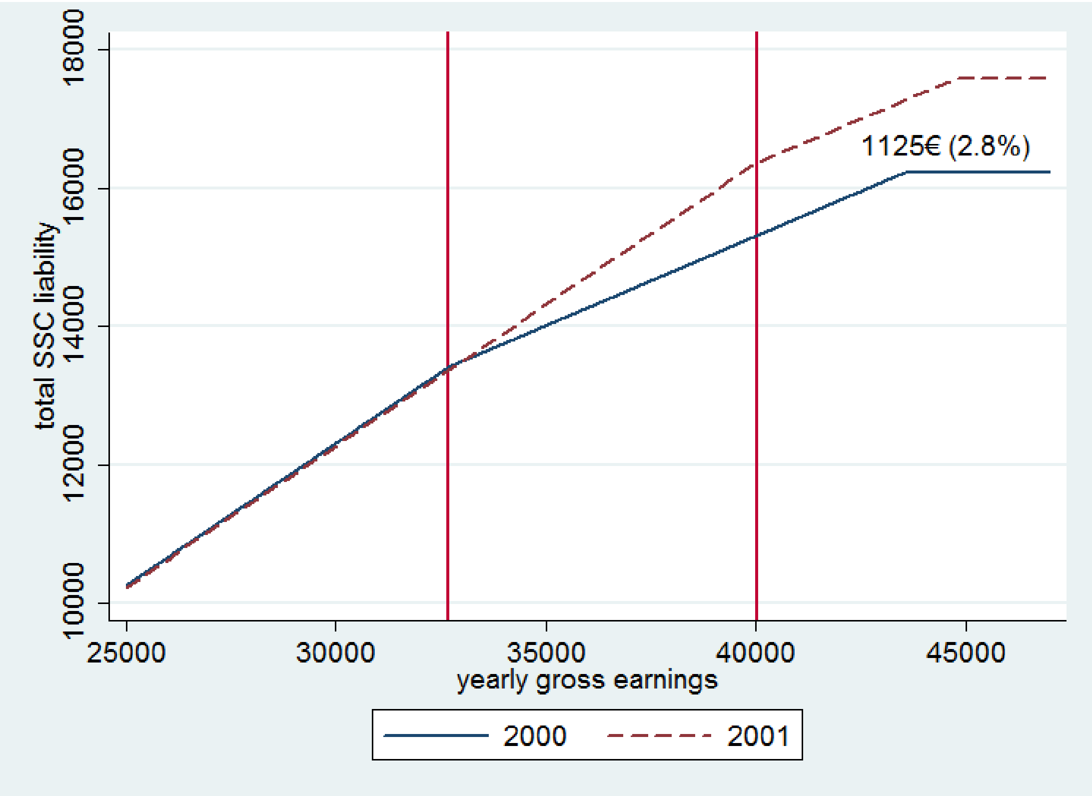
By January 2001[16], the East German earnings cap of health insurance was adjusted to the West German level (figure 1.2, section 1.3). This implied an increase from 32,672 € in 2000 to 40,034 € in 2001 or 22.5 %. Employees earning less than the pre-reform cap are not affected by the reform. For employees earning between the pre- and post-reform cap, the additional liability increased by their gross earnings (figure 1.3). Employees earning more than the post-reform cap are exposed to the maximum increase of 1125 €/year for employees and employers combined equaling 2.8 % of the earnings cap. Employees with earnings between the pre- and post-reform caps additionally experienced an increase of the marginal SSC rate by 15.3 pp. The reform did not affect the marginal SSC rate of those employees earning more than the post-reform cap.

At the time of the reform, the earnings cap for SSC coincided with the threshold above which most employees are allowed to leave mandatory public health insurance and opt for private health insurance. Therefore, employees affected by the reform might have been privately insured before the reform. However, as employers' SSC are capped at the same maximum value in private as well as public health insurance, their treatment due to the reform is comparable in most cases (see Appendix A.2.1 for a more detailed description of the private health insurance system in Germany and the implications for the reform evaluation). By contrast, the effect of the reform for privately insured employees (i.e. the actual change in their SSC share) cannot be quantified exactly.

Another potential confounding factor is an income tax reform which decreased the income tax rate in 2001. However, as this applied similarly for the control and treatment group it does not affect identification in the difference-in-differences approach.

[15]This abstracts from a potential effect on effort.
[16]The reform was announced in December 1999.

Figure 1.3: Total SSC by yearly gross earnings in 2000 and 2001



Notes: The solid (dashed) curve depicts total SSC by gross earnings in 2000 (2001), before (after) the reform. The difference between both curves is the treatment due to the reform which increased the earnings cap for SSC from the left reference line to the right.
Source: own calculations.

1.5.2 Data

The evaluation of the reform in 2001 is based on the Sample of Integrated Labour Market Biographies (SIAB)[17] which is a representative two percent sample of all individuals for whom an employer's record to the social security system exist. For the about 1.6 million employees in the sample information on their entire employment history is available, among others yearly gross earnings. The SIAB is particularly suited for the analysis of SSC variation over time for several reasons. First, the panel structure allows to calculate yearly earnings changes of individuals. Second, the information about earnings is very accurate due to the administrative character of the data. Third, as the earnings cap applies to the sum of all labour earnings[18], observing parallel employment spells is essential. Fourth, the sample size is comparatively large. Fifth, most factors important for wage dynamics are observed. These include age, sex,

[17]The German name is *Stichprobe der Integrierten Arbeitsmarktbiografien*. I used the weakly anonymised version of the data via on-site use at the research data center of the IAB in Berlin (*fdz*637). See vom Berge et al. (2013) for a detailed description.
[18]Plus earnings from most pensions which are not observed.

occupation and education at the individual level. Labour market experience and tenure within an establishment can be derived. Industry sector, establishment size and wage structure are especially noteworthy among establishment level characteristics. The regional unemployment rate can be matched based on the district of workplace.

The earnings information of the SIAB has two major drawbacks, though. First, working time is only differentiated between full- and part-time employment; as in many administrative data sets the exact amount of hours worked and hourly wages are not observed. Second, gross earnings are capped at the earnings cap of pension insurance. The imputation procedures proposed in the literature (see for example Büttner and Rässler (2008)) would not solve the problem here because observed individual earnings changes are crucial for identification. Therefore, I only focus on the increase of the earnings cap of health insurance.

The sample is restricted to full-time employees between 18 and 62 years. Spells with daily gross earnings below three euros and one-time payments are excluded. If an individual has two parallel full-time jobs or two parallel employment spells in the same firm, I exclude the employment spell with lower earnings. The quality of the education variable is inferior to most other variables because this information is not necessary for the administrative process (Dorner et al., 2011). I, therefore, use the first imputation procedure proposed by Fitzenberger et al. (2006) which is shown to perform best (Wichert and Wilke, 2012).

1.5.3 Methodology

The reform is evaluated within a difference-in-differences framework. Employees within a certain quantile of the earnings distribution above the earnings cap form the treatment and employees within a certain quantile below the control group. Since the wage rate is not observed, first differences in gross labour earnings serve as dependent variable as commonly done in the literature (Anderson and Meyer (2000), Bennmarker et al. (2009) and Saez et al. (2012)). Drawing conclusions about economic incidence, therefore, requires the assumption that the reform does not affect the amount of hours worked. Given that leisure is a normal good, the reform, however, creates incentives for both, employers and employees. In response to an increase in marginal SSC rates employees (employers) might want to decrease hours worked as leisure (other input factors) become(s) relatively more attractive (substitution effect). In response to an increase in average SSC rates, employees might want to increase their hours worked as the previous income can only be realised by working more (income effect). Two arguments support the assumption of no hours responses, however. First, employees who would have earned more than the post-reform cap in 2001 in the absence of the reform are only exposed to a change in the average SSC rate while their marginal SSC rate is constant. Restricting the treatment group to these employees requires only the exclusion of hours responses due to an income effect. This assumption is much weaker as income effects to taxation are often found to be very small (Gruber and Saez (2002)). Second, as shown with the cross-sectional bunching approach, the drop of the marginal SSC rate at the cap does not provoke any hours responses (section 1.4). A robustness check for the East German earnings cap of health insurance in 2000 (the year before the reform) based on the SIAB data set reaches the same conclusion (Appendix A.2.2). The finding of no behavioural responses to the earnings cap is consistent with the literature on taxation which usually finds that labour income of employees in the upper part of the earnings distribution is quite unresponsive to taxation (Blundell and Macurdy (1999), Saez (2003)).

The Treatment Group. All employees with potential earnings above the pre-reform earnings cap in 2001 in the absence of the reform are in theory directly affected by the reform. This is unobserved and the treatment group is, thus, approximated by employees earning more than the cap in the reform year 2001. The classification based on post-reform measures might lead to endogenous compositional changes of treatment and control group due to potential reform effects[19]. In the present setting, however, endogenous compositional changes are unlikely. For the treatment group consisting of all employees with earnings above the pre-reform cap these can only be driven by hours reactions which are shown to be negligible above[20]. Even full shifting to wages cannot explain a decrease of gross earnings below the lower group threshold, i.e. the pre-reform cap. For the treatment group restricted to employees with earnings above the post-reform cap the additional assumption of downward rigidity of the nominal wage rate is required[21]. Employees can only be analysed to a certain earnings level because the data is right-censored. Assuming a homogeneous treatment effect, this would not induce an endogeneity problem because the groups are defined by quantiles of the earnings distribution. When the treatment effect is allowed to be heterogeneous, however, a change in wage dynamics due to the reform might induce a downward bias. In a robustness check I, therefore, define the groups by earnings before the reform (Appendix A.2.6). This circumvents potential endogeneity problems, but comes at some cost. First, employees who would have decreased their earnings below the

[19]A similar endogeneity problem is prominently discussed by Blundell et al. (1998) who instrument the treatment indicator by cohort and education level. This yields very imprecise estimates in the present setting, however. Individual fixed-effects could be included to control for compositional changes of groups over time in mean regressions. However, as individuals naturally change groups over time by ascending or descending the wage distribution, one would have to carefully control for their earnings levels.

[20]In theory, responses at the extensive margin might have a similar effect. As the analysis is restricted to employees in the upper part of the earnings distribution, participation effects arguably are negligible as well.

[21]This assumption is supported by the fact that the average nominal earnings increase in the treatment group between 2000 and 2001 was almost 5 %. As the reform increased SSC by at most 2.8 % of gross earnings, even full shifting to employees could on average be achieved by a positive yearly earnings increase.

pre-reform cap in 2001 anyway are falsely allocated to the treatment group and employees who cross the cap in the reform year are falsely allocated to the control group. Second, classifying employees based on earnings before the reform eliminates the possibility to define a sub-group of employees exposed to a change in the average, but not the marginal SSC rate. Third, the right-censoring of earnings is much more problematic in that case (see below). Nevertheless, both approaches yield similar results.

The Control Group. The counterfactual earnings dynamics of the treatment group in the absence of a reform is estimated by East German employees earning less than the pre-reform earnings cap of health insurance. The reasoning is that both groups of employees are likely to be affected similarly by national, regional and sectoral shocks as well as inflation.

By contrast, as both groups by definition differ in earnings[22], rising income inequality or other changes in the income distribution not related to the reform might induce diverging trends. However, the analysis of the common trends assumption suggests that the deviation of the unconditional relative earnings changes between the two groups is constant over time (section 1.5.4). Nevertheless, a rich set of individual, regional and firm-level covariates is included in the estimation equation. Another threat to internal validity is that the reform generates incentives for firms to substitute between low- and high-earning workers. This might raise the wage rate for the control group (Lang, 2003) violating the stable unit treatment assumption of the difference-in-differences approach.

In order to maximise the comparability the control group is restricted by a lower threshold which is defined such that the sample sizes of treatment and control group are roughly equal in 2000.

Econometric Model. The causal effect of the reform on gross earnings is estimated by equation (1.5). Following Lang (2003), I employ median regressions in order to avoid the dependence on potential outliers[23]:

$$\Delta ln(Y_{ist}) = \alpha_s + \gamma_t + \beta D_{st} + \delta X_{ist} + \epsilon_{ist} \tag{1.5}$$

$\Delta ln(Y_{ist})$ states the yearly log earnings change of individual i in group s (i.e. treatment or control group) from year $t-1$ to t. α_s are fixed effects for the treatment and control group, γ_t are year fixed effects. D_{st} indicates observations in the treatment group in 2001. X_{ist} are potentially

[22] Using West German employees as control group would circumvent this issue. East and West German earnings dynamics are not comparable, however (Appendix A.2.3).

[23] The basic specifications are re-estimated by mean regressions (Appendix A.2.7). As expected, the estimates are less robust although the qualitative results are similar.

time-varying covariates on the individual, firm or regional level (Appendix A.2.5). ϵ_{ist} is an error term. The treatment effect is estimated by β which measures the part of the difference in earnings changes between treatment and control group in 2001, which cannot be explained by the average[24] difference in other years (conditional on X_{ist}). The main specification includes observations from 1997 to 2001. This can be interpreted as a difference-in-differences approach in first differences with three periods before the reform as control periods. The choice of the period is mainly motivated by the common trends analysis below (section 1.5.4). I conduct robustness checks by extending the control years to six pre- and six post-reform periods and get similar results (Appendix A.2.6).

The focus on individual first differences controls for unobserved heterogeneity in earnings levels. It generates two additional problems, though. First, mean reversion might be an issue as employees in the treatment group are more likely to have a large transitory earnings shock in t compared to control observations. If mean reversion is assumed to be constant over time (exemplarily Gruber and Saez (2002)), it is controlled for by group fixed effects similar to other systematic time-constant differences between the control and treatment group.

The estimation sample is an unbalanced panel. The groups are defined separately for every year based on the position in the wage distribution instead of the actual earnings caps or absolute earnings levels. This prevents compositional changes for example induced by the significant increase of the cap of pension insurance in 2003 or by a potential treatment effect. The treatment (control) group contains employees between the 82.39 (71.23) % and 93.44 (82.38) % quantile and therefore roughly ten percent of the overall sample respectively. Employees above the 92.14 % quantile are handled as maximally treated. 6.56 % of my sample is considered to be censored. Firm and job position changes as well as employments which do not last the whole year are excluded in the basic specification. This isolates earnings changes within an existing contract circumventing some potential sources of differential trends. It also makes confounding labour supply responses less likely. A robustness check relaxing this restriction finds that I do not miss part of the story there (Appendix A.2.6).

1.5.4 Results

Descriptive Evidence. The treatment and control group are reasonably similar with respect to the mean values of some key characteristics (table 1.2). The existing differences are as expected considering the systematic difference in earnings. Treated employees are rather male, higher educated and slightly older. Whether these or other differences induce a differential trend

[24]Weighted by the number of observations in the respective years.

of the outcome variable, can be seen by depicting the evolution of average relative yearly changes in (nominal) gross earnings for the treatment and control group graphically (figure 1.4). From 1997, these vary around four percent every year with earnings changes being slightly higher in the treatment group. The trends, however, are almost parallel, both before and after the reform. Although it is not a formal test, this suggests that the control group provides a good approximation of the counterfactual outcome of the treatment group. Changes in earnings are considerably higher before 1997 which might still be driven by the reunification of Germany. The basic specification, therefore, only exploits information from 1997 to 2001. In the year of the reform the average earnings change increased similarly in both groups suggesting that earnings did not react to the increase of the earnings cap for SSC. There are also no hints for anticipation or lagged effects.

Table 1.2: Sample size and characteristics by treatment and control group in 2001

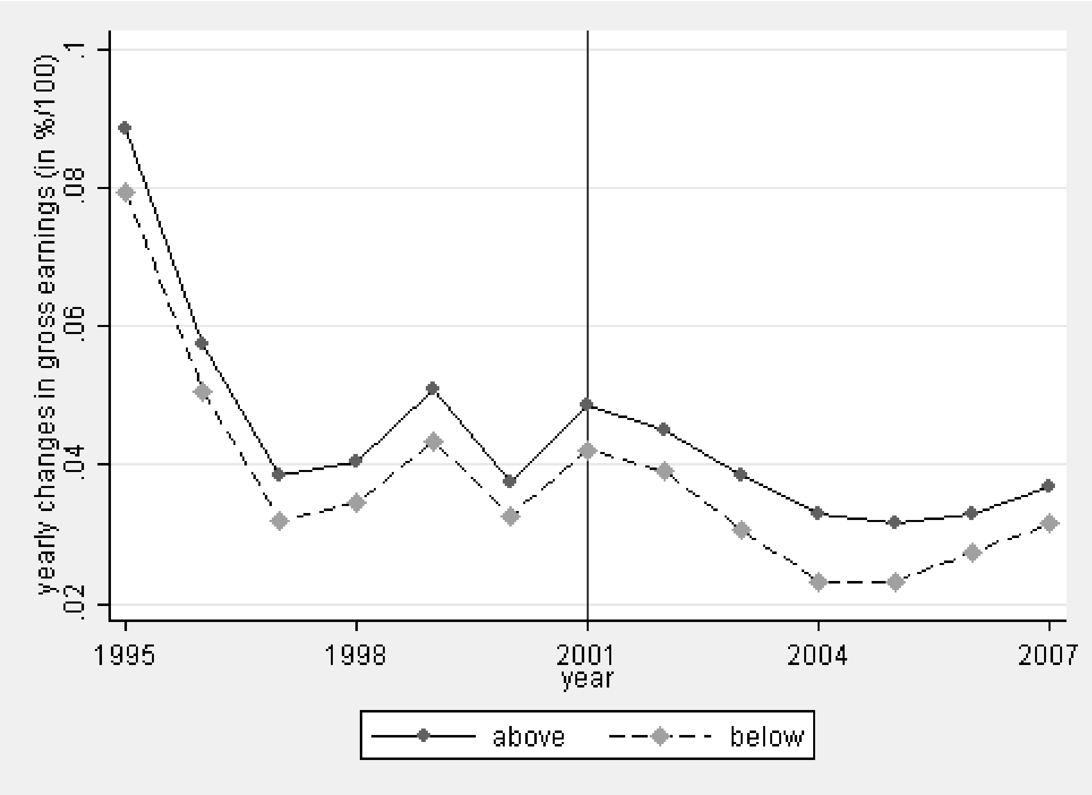
	treat	control			treat	control	
N	4,296	4,513			4,296	4,513	
variable	**mean**		p	**variable**	**mean**		p
Female	0.46	0.52	0.00	Unemp. Rate	13.90	13.86	0.59
Age	44.51	43.41	0.00	Yearly Earnings	37197	30630	0.00
Tenure	6.18	6.13	0.38	*Industry Sector*			
Experience	8.77	8.69	0.08	Agriculture	0.01	0.01	0.33
Education				Mining	0.01	0.01	0.09
Intermediate	0.00	0.01	0.00	Manufact.	0.17	0.18	0.98
Voc. training	0.51	0.73	0.00	Energy/Water	0.04	0.04	0.16
Grammar	0.07	0.07	0.53	Constr.	0.04	0.05	0.05
University	0.40	0.18	0.00	Commerce	0.06	0.06	0.63
Missing	0.01	0.01	0.03	Gastronomy	0.00	0.00	0.45
State				Transport	0.05	0.09	0.00
Brandenburg	0.17	0.18	0.11	Financial	0.06	0.05	0.00
Mecklenburg WP	0.13	0.13	0.64	Real Estate	0.09	0.07	0.00
Saxony	0.34	0.30	0.00	Public	0.18	0.18	0.58
Saxony Anhalt	0.20	0.20	0.99	Education	0.17	0.06	0.00
Thuringia	0.16	0.18	0.01	Health	0.06	0.17	0.00
Job Position				Misc. Services	0.05	0.04	0.09
Untr. worker	0.01	0.04	0.00	*Firm characteristics*			
Tr. worker	0.10	0.22	0.00	No. of Employees	688	672	0.53
Foreman	0.03	0.03	0.70	Median FT income	85.8	78.3	0.00
Clerk	0.86	0.71	0.00	High Union Cov.	0.42	0.30	0.00

Source: SIAB, own calculations

Difference-in-Differences Estimations. The upper panel of table 1.3 reports the results of estimating equation (1.5) by median regressions. The highlighted row contains the estimates of β, the average treatment effect on the treated. Coefficients of all included covariates (exemplarily for the specification of column (2)) are documented in Appendix A.2.5.

Columns (1) to (4) pertain to yearly earnings changes whereof columns (1) and (2) ((3) and (4)) report results based on a treatment group including employees with earnings above the pre-reform (post-reform) cap. The median change in gross earnings is lower for control observations than for observations in both treatment groups (see also figure 1.4). The difference even

Figure 1.4: Yearly earnings changes in treatment and control group



Notes: Yearly earnings changes in year t are defined as $ln(w_t/w_{t-1})$. These values are not comparable to average earnings increases of the whole population. First, I analyse a selective group of employees in the upper part of the earnings distribution and, second, the adjustments described in section 1.5.2 result in higher earnings increases.
Source: SIAB, own calculations.

increases when covariates are included (columns (1) and (3)). The estimate of the treatment effect based on all treated employees is not significantly different from zero without covariates (column (1)). Conditional on covariates (column (2)), it becomes significantly positive but stays very small (0.12 pp.). Taken at face value, the reform, which increased employees' SSC by 0.9 pp. on average, induced median earnings to increase by 0.12 pp. When the treatment group is restricted to employees earning more than the post-reform cap (columns (3) and (4)) the point estimate becomes slightly more positive if no covariates are included and decreases to virtually zero with covariates. This lack of a substantial reaction of gross earnings implies that (short-term) economic and legal incidence approximately coincide. This implies that neither employers nor employees are able to shift the additional burden. Recall that the unrestricted treatment group experiences a change in the marginal SSC rate while the restricted treatment group does not. Potential hours responses should induce a downward bias to the estimate of the former group (see section 1.5.3). The difference between those groups is, however, relatively small and not statistically significant. It also goes in the opposite direction. This lends support

Table 1.3: Reform effects on relative increase of gross earnings, labour costs and net earnings

	(1)	(2)	(3)	(4)	(5)	(6)
	TG 1		TG 2		TG 1	TG 2
Gross earnings						
2001	−0.0023***	−0.0008	−0.0020***	−0.0002	0.0226***	0.0215***
TG	0.0016***	0.0025***	0.0006	0.0044***	0.0070***	0.0123***
TG*2001	0.0009	0.0012**	0.0024**	0.0002	0.0010	−0.0021
CI l	−0.0008	0.0001	0.0001	−0.0020	−0.0011	−0.0092
CI u	0.0026	0.0023	0.0046	0.0024	0.0030	0.0050
Labour costs						
2001	−0.0032***	−0.0018***	−0.0029***	−0.0010	0.0149***	0.0141***
TG	0.0003	0.0010**	−0.0013*	0.0026***	0.0047***	0.0101***
TG*2001	0.0090***	0.0089***	0.0134***	0.0113***	0.0077***	0.0073***
CI l	0.0079	0.0077	0.0110	0.0087	0.0052	0.0017
CI u	0.0101	0.0102	0.0159	0.0139	0.0103	0.0128
Net earnings						
2001	−0.0008*	0.0009	−0.0009	0.0010*	0.0340***	0.0324***
TG	0.0037***	0.0047***	0.0032***	0.0067***	0.0102***	0.0160***
TG*2001	−0.0083***	−0.0097***	−0.0139***	−0.0156***	−0.0091***	−0.0168***
CI l	−0.0097	−0.0110	−0.0163	−0.0182	−0.0115	−0.0234
CI u	−0.0068	−0.0084	−0.0115	−0.0130	−0.0068	−0.0103
N	37188	37169	20780	20771	22848	12660
TG pre	13838	13828	1311	1310	7858	683
TG post	4296	4294	415	414	3289	276
CG pre	14541	14535	14541	14535	8240	8240
CG post	4513	4512	4513	4512	3461	3461
covariates	-	✓	-	✓	✓	✓
Δt	1	1	1	1	2	2

Notes: ***/**/* denote significance on the 1 %, 5 % and 10 % level respectively. Specifications labelled TG 1 (2) include all employees with earnings above the pre-reform (post-reform) earnings cap in the treatment group. In columns (1)-(4) the outcome variable in t is defined as $ln(w_t/w_{t-1})$, in columns (5) and (6) as $ln(w_t/w_{t-2})$. Gross earnings are observed, labour costs and net earnings are calculated based on gross earnings. Self-employed, employment spells lasting less than a year as well as job-to-job changes are excluded. The analysed period lasts from 1997 to 2001. Standard errors are bootstrapped by Stata's *bsqreg* command.
Source: SIAB, own calculations.

to the assumption that employment effects at the intensive margin are negligible.

The bootstrapped 95 %-confidence interval of column (4) can be used to provide upper bounds of burden shifting implied by the uncertainty[25] of the estimation. The reform increased SSC by at most[26] 2.8 % of gross earnings or 1.4 % for employees and employers respectively. This increase in SSC is estimated to lead to a decrease (increase) of gross earnings of at most 0.20 % (0.24 %)[27]. Thus, the left limit of the confidence interval translates into an upper bound of 14 % of the additional SSC of employers which might be shifted to employees. This implies that substantial shifting to wages can be rejected. The right limit of the confidence interval would translate into an upper bound of 17 % of the additional SSC liability of employees which might be shifted to employers. Taking into account that roughly one quarter of employees in

[25] At least the uncertainty induced by measurement error and the finite sample. Uncertainty induced by the common trends assumption as for example discussed in Donald and Lang (2007) is not considered.

[26] Employees in the restricted treatment group are treated homogeneously with respect to the absolute additional liability. As the earnings span within the group is rather small, the relative treatment neither varies much. It can, thus, be directly compared to the impact on the median. By contrast, this calculation would be problematic for the unrestricted treatment group as treatment intensity varies.

[27] This ignores the small bias induced by using the logarithmic change as percentage change.

the treatment group are privately health insured (Appendix A.2.1) and assuming that health insurance status and earnings are independent (conditional on earnings being above the cap), the upper bound rises to 23 %. This implies that considerable shifting to employers can be rejected as well.

The additional burden actually becomes manifest in a statistically significant increase (decrease) in yearly labour costs (net earnings) changes (second and third panel of table 1.3). Both measures are unobserved and have to be mechanically derived from gross earnings (Appendix A.2.4). These results underline that the finding of no effect on gross earnings does have explanatory power and is not owed to an imprecise estimation.

So far, yearly earnings changes have been analysed. Since the reform was announced in December 1999[28], the basic specification captures effects after two years' time. However, reform effects might be lagged even more or take place bit by bit due to frictions (Hamermesh, 1993). In order to capture medium-run effects the change in gross earnings between 2000 and 2002 is compared to other non-overlapping two-year changes in the observation period (columns (5) and (6)). For the unrestricted treatment group the point estimate does not change while standard errors increase, arguably due to restricting the sample to employees who have worked in the same company during two successive years. By contrast, the point estimate based on the restricted treatment group decreases sharply but stays insignificant. I find a statistically significant increase (decrease) of labour costs (net earnings) implying that considerable shifting in either direction can still be rejected. The more negative point estimate could be explained by modest lagged behavioural responses. The medium-term results for the restricted treatment group are, however, based on a very small sample size and should not be over-interpreted.

I conduct several robustness checks including the extension of the observation period and allocating employees to treatment and control group based on a pre-reform period (Appendix A.2.6). I also allow for job-to-job changes and employment periods of less than a year. At last, the analysis is restricted to sectors with low union coverage to ensure that sector-wide collective agreements do not drive the results. The qualitative picture of the results is shown to be reasonably robust. Re-estimating the main specifications by mean regressions comes to similar results (Appendix A.2.7).

[28] Anticipation effects are not expected as possible wage changes can be negotiated in advance. This is supported by figure 1.4.

1.6 Economic Incidence and Behavioural Responses

I do not find any behavioural responses to earnings caps of health insurance (section 1.4). Based on the assumption of economic incidence being entirely with employees, I concluded that the labour supply elasticity with respect to SSC is zero or very low. The results of this study (section 1.5) confirming previous literature for other countries (section 1.2) suggest that employers and employees might share the burden of SSC. If this is the case, both, the budget set and the cost function feature a kink at an earnings cap. To see that, imagine an additive cost function of an employer with y_j being the gross earnings of employee j, t_{0r} (t_{1r}) the SSC rate actually borne by employers if earnings are below (above) the cap and $C(y_{-j}, K)$ the costs for other input factors like other workers (y_{-j}) and capital (K):

$$C = \begin{cases} y_j(1+t_{0r}) + C(y_i, K) & \forall y_j \leq \bar{y} \\ \bar{y}(1+t_{0r}) + (y_j - \bar{y})(1+t_{1r}) + C(y_{-j}, K) & \forall y_j > \bar{y} \end{cases} \tag{1.6}$$

Assuming that employers have a quasi-concave production function $F(y_j, y_{-j}, K)$, the cost function (1.6) generates the incentive for employers not to contract an employee with earnings close to the kink. The mathematical reasoning is congruent to the discussion of a kink in the budget set (section 1.4). Intuitively, introducing a kink incentivises employers to increase hours of work for employees with gross earnings higher than $\bar{y}$ in a linear tax system as the marginal costs of an additional hour of work discontinuously drop. Employers would also wish to increase hours of work for employees with gross earnings slightly below $\bar{y}$ in a linear tax system as long as the cost reduction due to the additional lowly taxed hours beyond the kink at least compensates for the additional highly taxed hours up to the kink. A full representation of the firm's problem is beyond the scope of this study. This demonstration, for example, abstracts from the substitutability of production factors. However, even if employees at different positions of the wage distribution are highly substitutable, a gap or dip should emerge since firms would then have the incentive to substitute highly taxed hours below $\bar{y}$ with lowly taxed hours above. A gap or dip in the earnings distribution around the earnings cap for SSC might, thus, emerge due to behavioural responses of employees, employers or both. As long as economic incidence is not entirely with one side of the market, the method outlined in section 1.4 cannot separately identify the elasticities of labour supply and demand. It identifies an earnings elasticity which is composed of labour supply and demand responses. This is also true for the usual bunching estimator (Saez, 2010) when economic incidence of a tax is shared.[29]

[29] Allowing economic incidence to be heterogeneous and related to the elasticity the bunching approach does

Taking this into account, finding no dips or gaps in the earnings distribution around an earnings cap implies that both, labour supply and labour demand elasticity, have to be zero or very low[30]. Considering the selective sample and the fact that labour demand elasticities are usually found to be lower for skilled employees (Hamermesh, 1987), this result is not surprising. An alternative explanation for such low elasticities would be optimization frictions (section 1.4).

Irrespective of the reason for the low observed elasticities of labour supply and demand, they are consistent with the finding that the economic burden of SSC is neither shifted in large part to employees nor to employers. In a standard competitive partial-equilibrium model of the labour market, economic incidence is determined by the relation of the elasticities of labour supply and demand[31] (Fullerton and Metcalf, 2002). If labour demand is considerably more elastic than labour supply the burden of payroll taxes is completely shifted to employees which has been the standard assumption in most analyses of labour supply (Fullerton and Metcalf, 2002). However, demand and supply are both found here to be inelastic implying that the burden is equally shared between employees and employers which is empirically confirmed above.

Equal sharing corresponds to legal incidence for the German case. It is, therefore, impossible to determine whether the roughly equal sharing of the burden is driven by the elasticities of labour demand and supply as suggested by the model above. Alternatively, economic could be driven by legal incidence which would disprove the invariance of incidence proposition. Wage rigidities could in the latter case prevent gross wages to adjust in the short run. My analysis of economic incidence is based on changes in SSC over time. A long-run analysis could possibly yield different results. A potential mechanism for the long run might be relative-wage norms which are based on gross wages. Therefore, wage negotiations might be based on gross wages instead of the economically relevant labour costs and net wages (Saez et al., 2012). Laboratory experiments by Weber and Schram (2013) and Kerschbamer and Kirchsteiger (2000) found that subjects have on average a higher after-tax pay-off, when their negotiation partners have to pay the tax. They propose that the legal burden of a tax might induce a moral obligation to bear it.

not yield estimates of the average parameters in the population around the cap (Appendix A.1.5).

[30]An unlikely alternative explanation is that the labour market consists of two equally frequent types of employee-employer matches where the first (second) is characterised by inelastic employees (employers) and very elastic employers (employees).

[31]The choice of the model is not very restrictive as this central result also follows from most alternative models of the labour market.

1.7 Conclusion

This study provides evidence on the impact of SSC on labour earnings of high-skilled employees by exploiting the discontinuities induced by earnings caps of health insurance. Two complementary strategies based on two German administrative data sets are used. First, a quasi-experimental approach exploits a significant increase of the East German cap of health insurance in 2001. According to difference-in-differences estimates economic and legal incidence coincide in the short-term. Equivalently, the burden of SSC is shared roughly in equal parts. The main specification rejects that employers are able to shift more than 20 % of the additional legal burden to their employees. Although the average increase of SSC for employees cannot be pinned down exactly, full shifting of their additional legal burden to employers can be rejected as well. As I do not observe hours of work, the reform evaluation does not provide evidence on behavioural responses. It is based on the assumption of no impact on hours of work. I, therefore, translate the cross-sectional bunching approach (Saez (2010) and Chetty et al. (2011)) to upward kinks. Using four earnings caps between 1998 and 2004 with the drop of the marginal SSC rate being approximately 16 pp., I find that behavioural responses are non-existent or very low. If economic incidence is indeed shared between employers and employees, this suggests that labour supply and demand elasticities both are very low. I argue that both results are consistent in a competitive labour market model where economic incidence is determined by the ratio of labour demand and supply elasticities. If both market sides are similarly inelastic, the economic burden is shared in equal parts.

The combination of non-standard economic incidence and the absence of behavioural responses is in line with previous research (Saez et al. (2012), Skedinger (2014), Bennmarker et al. (2009) and Korkeamki and Uusitalo (2009)). It contrasts with other studies which found that SSC are completely shifted to employees (see for example Gruber (1994) for the US and Gruber (1997) for Chile). The difference might be explained by the institutional settings in the analysed countries. The more centralised wage bargaining in Germany might prevent shifting from employers to employees (Alesina and Perotti, 1997).

As in Germany equal sharing corresponds to the legal incidence, it cannot be separated whether this result is driven by the elasticities of labour demand and supply or if legal determines economic incidence. Further, positive economic incidence with employers does not automatically imply that parts of the increase of SSC has a negative impact on firms' profits. Other possible channels like burden shifting to prices or to the entire labour force cannot be captured here.

The result on economic incidence is restricted to a certain year and to East Germany which raises the issue of external validity. The result on behavioural responses is based on two years as well as East and West Germany. The theoretical relationship suggests that the finding on economic incidence can be extrapolated. Moreover, all estimates presented here exploit earnings caps for SSC which only affect employees in the upper part of the earnings distribution. The results are, thus, not necessarily transferable to employees with lower earnings because, for example, the labour demand elasticity might be higher for low-skilled employees (Hamermesh, 1987).

The consequences of increasing the earnings cap for SSC is interesting in itself as it constitutes a feasible policy for many countries to increase revenue or to shift the burden to high earners. In the UK, for example, the earnings cap for employers' SSC was abolished in 1985 as were most earnings caps for SSC in France in the 1980s. The finding of no employment responses at the intensive margin implies that the efficiency loss of an increase of an earnings cap seems to be negligible. The results on incidence show that employees and employers would share the increased burden.

Chapter 2

The economic incidence of social security contributions - A discontinuity approach with linked employer-employee data*

2.1 Introduction

In many countries payroll taxes or social security contributions (SSC) account for a material part of total taxation. Their average contribution to government funding among OECD countries is similar to personal income taxation (around 9% of total tax revenues, OECD, 2016). With earnings from employment usually being the sole tax base, the potential of SSC to distort efficient allocations on the labour market is large. Who bears the burden of SSC is thus a crucial question for distributional or welfare analysis.

By contrast to income taxation SSC are formally shared between employers and employees. According to standard economic theory legal incidence does not matter for the effective sharing of the tax burden though (invariance of incidence proposition). The SSC burden is primarily borne by employees because labour demand is considered to be more elastic than supply (Fullerton and Metcalf, 2002). More recent empirical research challenges the standard view and

*This chapter is based on joint work with Kai-Uwe Müller (see Müller and Neumann (2016)) within the project *The Impact of Social Security Contributions on Earnings: Evidence from administrative data in France, Germany, the Netherlands and the UK* sponsored by the *Deutsche Forschungsgemeinschaft.*

finds economic incidence to be identical with formal incidence (Saez et al., 2012). Overall the evidence on the burden-sharing of SSC is inconclusive (Melguizo and Gonzalez-Paramo, 2013). This might be partly explained by the variety of estimation approaches and their underlying assumptions.

The majority of the empirical literature on the economic incidence of SSC relies on longitudinal variation. Most earlier observational regressions lack clean identification, though. More credible approaches exploit exogenous variation generated by policy reforms (Gruber, 1994, 1997; Bennmarker et al., 2009; Saez et al., 2012). Suitable SSC reforms are infrequent, however. Only one policy change occurred, for instance, in Germany which created sufficient institutional variation and provided a setting with a valid control group (Neumann, 2015). Except for Saez et al. (2012) estimates from these studies are based on a difference-in-differences framework. They rely on the common-trends assumption, depict short-term responses and are biased towards zero under optimization frictions. Data restrictions exacerbate these limitations. Should hours of work not be observed – which is common in administrative data sets – behavioural responses (of labour supply and demand) to SSC have to be assumed away to interpret estimates in terms of incidence. Given the incomplete information, this presumption cannot be tested.

In this paper we therefore follow a different approach suggested by Alvaredo and Saez (2007). The alternative framework is based on cross-sectional data and utilises earnings caps in a given SSC schedule. The discontinuous drop of the marginal SSC rate generates a downward kink in the average tax schedule. Depending on how the SSC burden is shared, a positive or negative discontinuity emerges in the distribution of gross earnings at the cap. Only when statutory equals economic incidence, the gross earnings distribution is smooth around the cap. This can be exploited to estimate economic incidence.

The approach circumvents some of the aforementioned methodological issues. It neither requires exogenous policy reforms nor panel data or information on hours of work. Estimates represent long-term incidence effects as long as earnings caps do not change markedly. Adjustment frictions are thus less of a problem here. The first contribution of this paper is to refine the analytical framework: We elaborate on the identifying assumptions and discuss the empirical implementation using parametric as well as non-parametric estimators.

The second contribution is the application of linked employer-employee data (the German Structure of Earnings Survey, GSES). We argue that looking at distributions of other wage-related measures like net earnings or labour costs in addition to gross earnings densities produces more reliable results. Discontinuity estimates from a single earnings distribution are hard to interpret under measurement error. A small discontinuity might represent a substantial effect or result from noisy data. Besides providing precisely measured gross earnings, the GSES includes the exact amount of SSC paid by employees. We are therefore for the first time able to check whether incidence estimates at earnings caps are consistent across two different empirical densities.[32] This broader body of evidence allows to distinguish between the two competing interpretations.

We find at large small and insignificant discontinuities in empirical gross earnings distributions. Together with substantial negative discontinuities in net earnings this provides consistent empirical evidence that legal and economic incidence of SSC coincide in Germany. We demonstrate that the findings are robust with respect to methodological decisions, different caps and sample periods.

This finding is in line with recent evidence. Saez et al. (2012) evaluate a unique policy reform in Greece where a different SSC regime was implemented for employees who started working from 1993 onwards. They do not find gross earnings of otherwise similar workers, i.e of those who entered the labour market shortly before and after the reference date, to differ systematically. This implies that legal and economic incidence coincide. Skedinger (2014) finds small effects of payroll tax cuts for young Swedish workers. In Finland a similar reform targeted at older low-wage workers did not trigger any wage effects (Huttunen et al., 2013). These results contradict earlier studies for non-European countries which found economic incidence to be entirely with workers (Gruber, 1994, 1997). Our study provides further evidence that employers are hardly able to shift their SSC burden to employees under a more or less centralised wage setting regime as it exists in Germany and other European countries.

The remainder of this paper is organised as follows. Section 2.2 presents the methodological framework, discusses the underlying assumptions for identification and the empirical implementation of the approach. Section 2.3 applies the approach to German linked employer-employee data. Section 2.4 discusses the results in the light of previous findings and concludes.

2.2 Methodology

In this section we, first, introduce a refined version of the analytical framework developed by Alvaredo and Saez (2007). Second, we spell out the identifying assumptions. Third, we present parametric and non-parametric estimators to implement the discontinuity approach empirically.

[32]Saez et al. (2012) also observe the exact amount of SSC in their data for Greece. They use a different framework based on a cohort-specific policy reform.

2.2.1 Analytical framework

We start with a narrow framework centered around a general notion of economic incidence captured by a change in the hourly gross wage rate. The model does not describe any incidence mechanisms explicitly. Underlying labour supply and demand elasticities which drive economic incidence in the standard model are, for example, not included. Employees' and employers' taxes besides SSC are also not covered. The latter does not (qualitatively) affect the analysis as long as additional taxes do not vary systematically around the SSC caps.

Let t denote the symmetric SSC rate which has to be paid by employers (r) and employees (e), i.e. $t = t_r = t_e$. Assume ability or preference for work n to follow a smooth cumulative (marginal) distribution function denoted by $P(n)$ ($p(n)$). For simplicity let realised gross earnings y equal ability, i.e. $y = n$, if either the SSC rate t is zero or legal and economic incidence coincide. Should legal and economic incidence differ, y is distorted such that

$$y = \frac{n}{1 + ts} \tag{2.1}$$

with s being a homogeneous shifting parameter which is positive (negative) if employers (employees) shift some burden to employees (employers). This framework can be generalised along several dimensions. Instead of using a symmetric SSC rate t, the rates might differ between employers and employees, i.e. $t_r \neq t_e$ (Appendix B.1.1). Likewise shifting s can be asymmetric between employers and employees (Appendix B.1.2), or s can be heterogeneous across different employees and employers (Appendix B.1.3). Finally, a change in the marginal SSC rate might induce behavioural reactions, for instance an adjustment of working hours. This is particularly relevant as underlying labour supply and demand elasticities might drive both behavioural responses as well as economic incidence (Appendix B.1.4). As shown in Appendix B.1 neither of those extensions eliminates a discontinuity at an earnings cap.

We therefore stick to the parsimonious specification in equation (2.1) which is not restrictive for SSC institutions in most countries, such as Germany. First, statutory SSC rates for employers and employees are often identical. This holds for the period under observation (1995-2010) in this paper, too. Second, earnings caps apply to employers as well as employees. Separate identification of s_r and s_e with varying SSC rates requires differential variation in employers' and employees' SSC rates. Third, strong labour supply reactions make identification more difficult. Yet, there is no evidence for behavioural adjustments at earnings caps in the literature (Liebman and Saez, 2006; Alvaredo and Saez, 2007; Saez, 2010). Neither does the observed distribution of gross earnings suggest any hours reactions in our application (Appendix B.1.4).

An earnings cap for SSC at $\bar{y}$ where t drops to 0 yields the following relationships between the distributions of ability and realised earnings $(F(y), f(y))$:

$$F(y) = P(n) = \begin{cases} P(y(1+ts)) & \forall y \leq \bar{y} \\ P(y+\bar{y}ts) & \forall y > \bar{y} \end{cases} \tag{2.2}$$

$$f(y) = \begin{cases} p(y(1+ts))(1+ts) & \forall y \leq \bar{y} \\ p(y+\bar{y}ts) & \forall y > \bar{y} \end{cases} \tag{2.3}$$

Considering the left and right limit of the density of realised gross earnings when y approaches $\bar{y}$:

$$f(\bar{y})_- = p(\bar{y}(1+ts))(1+ts) = p(\bar{y}+\bar{y}ts)(1+ts) = p(\bar{n})_-(1+ts) \tag{2.4}$$

$$f(\bar{y})_+ = p(\bar{y}+\bar{y}ts) = p(\bar{n})_+ \tag{2.5}$$

As $p(n)$ is assumed to be smooth, there will be some discontinuity in the density $f(y)$ as long as $s \neq 0$, i.e. economic differs from legal incidence. The density drops (jumps), if $s > 0$ ($s < 0$), that is if employers (employees) are able to shift some of their legal burden. Full shifting to employees (employers) implies $s = +(-)1$. The deviation between the actual and a counterfactual density without earnings cap is not uniform over the area above the cap. For a negative (positive) slope the gap diminishes (increases) slowly with the distance to $\bar{y}$ (see Appendix B.2 for details). The size of the discontinuity at the cap is equivalent to the share of the SSC rate shifted (ts) in relative terms to (i.e. in per cent of) the density directly above:

$$\frac{f(\bar{y})_- - f(\bar{y})_+}{f(\bar{y})_+} = ts \tag{2.6}$$

Therefore s can be expressed as a function of observed measures. To put it more intuitively: Depending on their bargaining power and other factors like wage rigidities, employers or employees will shift some of their SSC burden. This manifests itself in a negotiated wage which is influenced by the average SSC rate. At an earnings cap the previously flat average SSC schedule features a downward kink. From that point the SSC burden decreases relative to earnings. Under constant shifting the wage rate is adjusted to a lesser degree with increasing distance of earnings to the cap. This implies that the gross earnings distribution expands (compresses) to the right of the cap when incidence is rather with employees (employers). A negative (positive) discontinuity emerges at the cap.

In theory analyzing gross earnings densities suffices to determine economic incidence. Discontinuities can be less clear-cut and hard to identify in empirical practice, however. Measurement error or other distorting factors lead to noisy data. Small or insignificant discontinuities, for instance, could be a substantial finding or simply result from such measurement problems. We argue here that the additional investigation of the net earnings and/or labour cost distribution(s) enhances the power of our estimation approach under those circumstances (see the two following sub-sections). Getting consistent estimates from two earnings distributions provides more robust evidence for actual economic incidence. Saez et al. (2012) show this for a different identification strategy.

The behavior of gross earnings around a cap has implications for labour costs (what employers effectively pay, denoted by z) and net earnings (what employees effectively receive, denoted by c). Both quantities are deterministically related to gross earnings:

$$z = \begin{cases} y(1+t) = \frac{n}{1+ts}(1+t) & \forall y \leq \bar{y} \\ y + \bar{y}t = n - \bar{y}ts + \bar{y}t & \forall y > \bar{y} \end{cases} \tag{2.7}$$

$$c = \begin{cases} y(1-t) = \frac{n}{1-ts}(1-t) & \forall y \leq \bar{y} \\ y - \bar{y}t = n - \bar{y}ts - \bar{y}t & \forall y > \bar{y} \end{cases} \tag{2.8}$$

It holds that $z = n$ ($c = n$) if $s = 1$ ($s = -1$), that is if the burden is shifted entirely to employees (employers). By similar arguments as above this implies that for $s = 1$ ($s = -1$) the density of labour costs (net earnings) is smooth, whereas the densities of gross earnings and net earnings (labour costs) are discontinuous. Should economic and legal incidence coincide, the gross earnings density is continuous while the density functions of net earnings and labour costs are not.

2.2.2 Identification

The emergence of a discontinuity in the density of gross earnings (net earnings or labour costs) which is proportional to the drop in the marginal SSC rate at an earnings cap requires a number of conditions to be met. A closer look at those assumptions is helpful to sort out different identification problems when we apply the framework to actual data. Assuming smoothly distributed abilities or preferences for work is a technical requirement. This is not restrictive in applications an will thus not be discussed further.

A first set of assumptions concerns different *shifting mechanisms*. The burden of SSC is shifted at the individual level. An alternative channel of incidence might be shifting to the workforce as a whole. Economic incidence at the aggregate level would invalidate the approach as individual earnings would not be affected by a cap and no discontinuity would emerge. Moreover, labour earnings subject to SSC are the exclusive item in negotiations about the sharing of SSC. The burden of SSC must not be shifted through other margins of employees' compensation (e.g. premia, non-pecuniary benefits, or paid/unpaid overtime). Finally, shifting may be heterogeneous but independent from the earnings level. Should s vary over individuals, the discontinuity at an earnings cap identifies average economic incidence of individuals located around the cap. Allowing for s and earnings to be correlated renders the estimate selective. (Appendix B.1.3)

A second set of assumptions comprises the (absence of) intervening factors which might distort incidence-induced discontinuities at earnings caps. There are no behavioural reactions in terms of labour supply or labour demand. Any adjustments at those margins would influence the shape of the density around an earnings cap (Appendix B.1.4) and therefore affect the identification of an incidence-induced discontinuity. Moreover, SSC rates and caps are perfectly salient. If this is not the case, (some) employers and employees might either ignore SSC when bargaining over wages, or perceived rates and earnings caps vary over individuals. In both instances a discontinuity would be blurred or even eliminated. And lastly, there are no optimization frictions. Wages can be adjusted according to SSC liabilities. By definition this also holds for re-negotiations of wages, e.g. after an institutional change in the earnings cap or an employee's promotion which implies a pay increase above an earnings cap. In addition, heterogeneous frictions across the earnings distribution render a discontinuity at an earnings cap and inference on economic incidence derived from it selective.

These assumptions seem very restrictive. However, all microeconometric evidence on economic incidence relies on them. Certain assumptions are less critical for the discontinuity approach where identification does no rely on policy reforms. Take optimization frictions as an example: If internalising the reduced SSC liability at the cap is costly, it takes only place when benefits exceed those costs. The drop in the average SSC rate slightly above the cap and hence the benefits of a reaction are indeed relatively small. On the other hand, the framework captures long- and not merely short-term effects. Given that earnings caps are salient, wage negotiations – which for example lead to the crossing of the cap – can directly incorporate the drop of the marginal SSC rate. In that sense the discontinuity represents a long-run equilibrium which is hardly affected by optimization frictions in the short-term.

Further arguments can be brought forward in support of certain assumptions. The salience assumption is underpinned by the fact that earnings caps are relevant for employers as well as employees. Employers should be particularly aware of their SSC liabilities when calculating labour costs, even more so when they have several employees with earnings above the cap.

Salience is plausible for the German setting where earnings caps have been in existence for a long time and have rarely been changed significantly (sub-section 2.3.1). The persistence of SSC institutions also facilitates the no-frictions assumption, as SSC liabilities are incorporated into equilibrium wages. Caps are adjusted regularly according to the change in the average gross wage bill in the preceding year. These minor changes might make a complete internalisation of legal SSC liabilities impossible but carry minor weight as they go along with the overall shift in the earnings distribution.

2.2.3 Estimation

It is straightforward to bring this framework to the data: The size of a potential discontinuity has to be determined at the cap of a given earnings distribution. We follow two estimation strategies, a non-parametric sorting test proposed by McCrary (2008) and a parametric approach where a polynomial is fitted to the distribution. The magnitude of the shifting ts is calculated by normalising the discontinuity estimate by the density directly right of the cap.
In the first step of the McCrary test a histogram is estimated where no bin contains observations below and above a potential discontinuity. Then the density to the left and right of the cap is estimated by local linear regressions to avoid bias at the boundaries. The test statistic consists of the log difference in the prediction slightly left and right of the cap. The procedure is subject to the choice of the bin size for the underlying histogram estimates, the selection of the bandwidth for the local linear smoother as well as the earnings interval underlying estimation.
In an alternative parametric approach we also estimate initially a histogram of the earnings distribution. Then a polynomial of degree nine is fitted to the density values at the center of each histogram bin. The discontinuity is measured by an indicator variable differentiating between bins above and below the cap. We restrict the coefficients to be identical on both sides of the cap to increase the precision of the fit for values close to the cap. This is justifiable since the estimation window is rather small and our specification is flexible enough to fit the slight change of the curvature at the earnings cap:

$$d_j = \sum_{p=1}^{9} \beta_p y_j^p + \delta \mathbb{1}(y_j > \bar{y}) + \epsilon_j \tag{2.9}$$

Here d_j denotes the height of bin j, y_j are earnings in the middle of bin j, $\mathbb{1}$ is an indicator function, β and δ are parameters, and ϵ is an error term. While the non-parametric test is not subject to functional form assumptions about the earnings distribution, it might be more sensitive to outliers close to the discontinuity. By allocating the same weights to all bins in the estimation window the parametric approach is more robust in that respect. The fact that identification is less dependent on individuals very close to the cap also decreases the diluting effect of optimization frictions (sub-section 2.2.2). In the basic specification we use an estimation interval of 2000 € which is split into 100 bins on either side of the cap. This implies a bin size of 10 €.

We provide some Monte-Carlo evidence on the performance of these estimators (Appendix B.2). Gross earnings distributions $f(y)$ are simulated on the basis of equation (2.3) under the assumption of full shifting of employers' SSC to employees ($s = 1$). Given our assumptions the simulated earnings densities feature a negative discontinuity at $\bar{y}$ in the magnitude of the drop in the marginal SSC rate of roughly 7.6% (figure B.1 in the Appendix). The McCrary test and the parametric test identify a discontinuity of -7.2% and -7.6%, respectively, which matches the underlying parameters s and t (figure B.2 in the Appendix). The confidence interval of the McCrary estimate translates to an interval of $(-1.24, -0.64)$ for the shifting parameter s. The corresponding interval for s based on the parametric test is slightly narrower: $(-1.21, -0.79)$.

2.3 Application to German data

2.3.1 Institutions

The German social security system consists of pension, unemployment, health, and long-term care insurance. The contribution rates are flat with daily gross earnings as tax base. SSC have formally been shared equally between employees and employers until the end of 2004. Since then a share of 0.9 pp. are paid exclusively by employees.[33] Most SSC rates have been quite constant. The total SSC rate varied around 40% with pension (around 20%) and health insurance (around 14%) as the most important branches (figure B.3 in the Appendix). Marginal SSC rates only apply up to earnings caps. There is one threshold for pension and unemployment insurance, another one for health and long-term care insurance. For the sake of readability we do not separately refer to unemployment and long-term care insurance throughout the paper. Both caps differed between East and West Germany until 2001 when the health insurance cap in East Germany was adjusted to the level of West Germany (figure B.3 in the Appendix).

SSC rates as well as earnings caps are subject to yearly gradual changes which are difficult to exploit for the identification of economic incidence. Considerable discontinuous changes are rare with the strong increases of the earnings cap of health insurance in East Germany in 2001 and pension insurance in 2003 as notable exceptions. While the former is evaluated by Neumann

[33] This share increases by 0.25 pp. for childless employees.

(2015), the latter is difficult to analyse because suitable German panel data is right-censored at the pension earnings cap. The cross-sectional approach outlined above does not rely on policy reforms and panel data. We are therefore able to draw on uncensored earnings data here. The caps are at different positions in the respective earnings distributions. The threshold for health insurance in West Germany is around the 75th quantile and by far the lowest (table 2.1). The health cap in East Germany and the pension cap in West Germany come in second at around the 90th quantile. The pension cap in East Germany is up high in the earnings distribution above the 95th quantile.

Table 2.1: Quantiles of earnings caps

Wave	**West Germany**		**East Germany**	
	Pension / Unemployment	Health / Long-term care	Pension / Unemployment	Health / Long-term care
1995	.91	.76	.97	.89
2001	.90	.76	.96	.93
2006	.92	.75	.96	.91
2010	.90	.72	.95	.89

Source: GSES 1995, 2001, 2006, 2010; own calculations.

One peculiarity of the German social security system which could affect our empirical analysis is the possibility to substitute public for private health insurance. Eligibility for private health insurance depends on exceeding the income threshold for compulsory insurance (called *Versicherungspflichtgrenze*), except for the self-employed and civil servants who are excluded here. This threshold had been equal to the earnings cap of health insurance until both caps were decoupled in 2003. Since then the threshold for compulsory insurance lies between the earnings caps of health/care and pension/unemployment insurance. The employees' share of SSC does not depend on earnings but on personal characteristics under private insurance. Employers have to pay half of private contributions but only up to the maximum value of employers' SSC under public insurance.

Private health insurance interferes with our analysis of gross earnings when it affects the change in the marginal SSC rate at the analysed earnings cap. This is only the case for the earnings cap of health insurance before 2003.[34] Being slightly below or above this earnings cap might change the insurance system applicable to the employee. Although we do not observe whether an employee is privately health insured, it is not obvious how severely this distorts a potential discontinuity at that cap. First, the cap for employers' SSC is in most cases identical under public and private insurance. Their incentives below vs. above the threshold are comparable as long as private is at least as expensive as public health insurance which normally holds

[34] As contributions to private health insurance are included in the GSES' measure of employees' SSC, the impact on the analysis of observed net earnings is more severe. See sub-section 2.3.2 for a detailed discussion.

(Neumann, 2015). Second, the switch to private health insurance usually occurs earliest in the first year after earnings increased above the threshold. A large fraction of employees with earnings very closely above the earnings cap, i.e. a large fraction of our estimation sample, are still publicly health insured in a given year.

2.3.2 Data

We use the German Structure of Earnings Survey (GSES, *Verdienststrukturerhebung*) which consists of repeated cross-sections. The information is provided by employers and is part of the official labour cost statistics of the German Statistical Office. Firms are therefore obliged to cooperate and provide the information.

Measurement error

Similar to the literature on behavioural responses at kinks or notches of tax schedules (Kleven et al., 2011; Kleven and Waseem, 2013), identifying a discontinuity at an earnings cap is demanding in terms of data quality. In order to identify the precise location of the cap at $\bar{y}$, the exact amount of labour earnings subject to SSC has to be observed. Measurement errors in data on gross earnings y could arise as a result of imprecisely/erroneously reported information. Let ϵ_i be a random disturbance term for individual i, the observed gross earnings are $y_i^{obs} = y_i + \epsilon_i$. As we do not observe ϵ_i, the assumed location of the earnings cap is $\bar{y} - \epsilon_i$ which varies across individuals in an unknown way. As a result, a discontinuity estimate might be biased towards zero. This type of distortion is purely a data problem and independent of the identifying assumptions. The influence of ϵ_i is a function of the sample and decreases with the number of observations. It also depends on the data collection and is more problematic in self-reported earnings data.

Reliability of individual earnings information is better in administrative data like the GSES. Therefore measurement error arguably is not a big issue (although we cannot determine its magnitude) as documented in other applications (Fitzenberger et al., 2013; Antonczyk et al., 2010). We primarily use monthly earnings information that refers to the month of October. As the GSES separately includes all earnings components which are subject to SSC (like regular earnings and compensation for overtime hours) as well as some of those which are not (like tax-free premia for shift work, working on Saturdays/Sundays/holidays, or night employment), locating the earnings caps is neither an issue.

Multiple earnings measures

As argued multiple earnings measures are helpful in empirical applications with non-systematic measurement error. Using incidence estimates from different earnings distributions generates a more robust empirical foundation to draw policy conclusions. Consider the following example. The gross earnings distribution is found to be smooth around the cap. This could be explained by no shifting or measurement error. According to the first explanation the distributions of labour costs or net earnings should feature a discontinuity consistent with the drop of the marginal SSC rate. Under the alternative scenario of measurement error all three distributions would be found smooth around the cap. The results would be inconclusive as far as incidence is concerned.

To our knowledge this is the first paper using a discontinuity approach on SSC incidence that exploits separate measures of gross earnings and actually paid employees' SSC. The latter include an employee's contributions (also voluntary) to social insurance (i.e. the employee's share of contributions to the pension, unemployment, health and care insurance). This also entails contributions to private health insurance as well as occupational insurance schemes. Together with information on gross earnings y we can calculate actual individual net earnings c. There is no independent information on employers' SSC. Labour costs z are thus not considered separately here.

Explicit information on participation in public or private health insurance is not available. Identification of discontinuities in the distribution of net earnings could thus be affected: A privately health insured employee i with earnings between the health and pension earnings caps, i.e. $\bar{y}^h < y_i \leq \bar{y}^p$, has *observed* net earnings $c_{i,obs} = y_i - ssc_{i,obs} = y(1 - t_e^p) - ssc_{i,private}^h$. The location of the earnings cap in the net earnings distribution is $\bar{c}_i^h = \bar{y}(1 - t_e^p) - ssc_{i,private}^h$. However, as we do not observe, whether an employee is privately health insured, we approximate $\bar{c}_i^h$ by assuming public health insurance: $\hat{\bar{c}}^h = \bar{y}(1 - t_e^p) - t_e^h \bar{y}^h$. This might attenuate a potential discontinuity. The threshold for compulsory public insurance exceeds the earnings cap of health insurance since 2003, though (see section 2.3.2). The estimates based on the health insurance cap in 2006 and 2010 are therefore not distorted by private health insurance.

Sample

We use all four waves of data currently available to us, i.e. data for the years 1995, 2001, 2006, and 2010. For the wave 1995 more than 700,000 observations are available, about 640,000 for the wave 2001, more than 1.2 million for 2006, and more than 1 million employees for the wave 2010. We analyse the caps for *health/long-term care* and *pension/unemployment* insurance

$\bar{y}^h, \bar{y}^p$. In order to increase the power of the estimation East and West Germany are pooled throughout the analysis. Waves are either analysed pooled or separately. There is variation in caps $\bar{y}_{rt}$ over time t and by region r (i.e. between West and East Germany). We, therefore, index individual earnings y_{irt} in the following simple way: $y_{irt}^{ind} = y_{irt} - \bar{y}_{rt}$. The resulting y_{irt}^{ind} is measured in deviations from the threshold which allows us to pool data for different r and t. The indexation of net earnings c at the cap is a little more intricate since the SSC rates applied change due to the health insurance earnings cap $\bar{y}^h$.

We exclude civil servants and self-employed as they do not contribute to the general SSC scheme. Home workers are also not included in the sample. We also exclude part time and marginally employed persons to limit behavioural adjustments. The majority of these employees has earnings below the caps, anyway. Small firms (below ten employees) are exempted from the GSES because of the administrative burden. Few branches are also missing for certain waves. Neither of those omissions is systematically related to SSC and earnings caps and affecting our estimates.

2.3.3 Empirical results

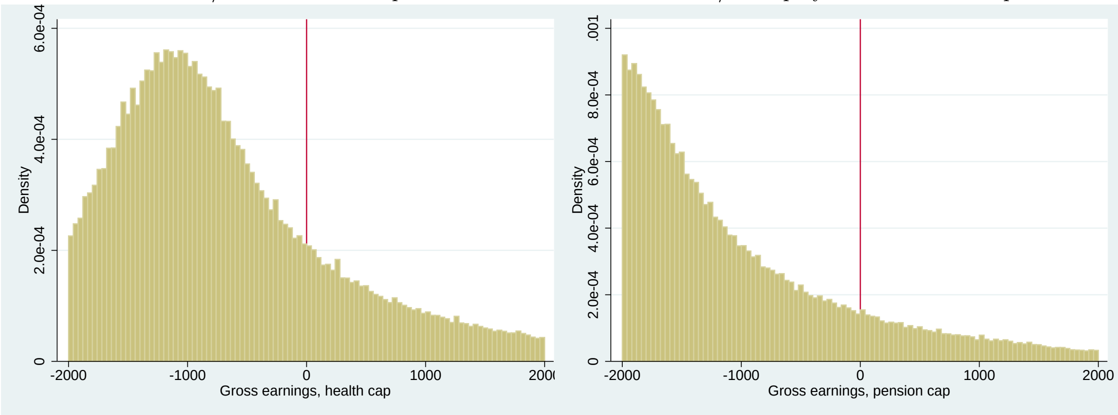
We first present results for the empirical gross earnings distribution. Results from gross earnings are in principle conclusive in terms of incidence. Yet, these estimates might be distorted by a number of factors as discussed above. The subsequent analysis of observed net earnings allows for plausibility and robustness checks.

Gross earnings

There is no clear-cut discontinuity in histograms of the monthly gross earnings distribution at the pension or the health cap based on pooled data for all years (figure 2.1). Although being indexed at earnings caps and based on a sizeable data set, the histograms are not completely smooth. There are some minor spikes at various points of the distribution. One of those visible spikes is located directly above the pension insurance cap. To a lesser degree this is also true for the health insurance cap which becomes more visible in histograms for single waves (figure B.5 in the Appendix).

The majority of those spikes can be explained by round number bunching as documented exemplarily in the non-normalised earnings distribution for West Germany in 2006 (figure B.4 in the Appendix). In similar fashion it is conceivable that contracted earnings are oriented towards prominent numbers like an earnings threshold. In addition, in some years the earnings threshold was a round number. We therefore exclude individuals with $y \in [\bar{y}, \bar{y} + 10]$ from the

Figure 2.1: Distribution of monthly gross earnings, all years pooled



Source: GSES 1995, 2001, 2006, 2010; own calculations.

estimation of discontinuities, in order to avoid bias driven by non-substantive factors, i.e. things not related to economic incentives. We adjust each histogram for the missing observations by moving the remaining distribution to the left to close the gap in the density.

Applying the non-/parametric estimators to the pooled sample we find very small negative and statistically insignificant discontinuities at both earnings caps (table 2.2, figure 2.2). We get the smallest and largest point estimates for the pension cap: $-.007$ with the parametric approach and $-.015$ with the non-parametric approach. In the pooled sample the average drop in the marginal SSC rate amounts to 8.1 pp. at the health and about 12.4 pp. at the pension insurance earnings cap for employees and employers, respectively (table 2.2). The shifting parameters implied by the discontinuity estimates, therefore, range between $s = -0.06$ for the pension cap to $s = -.17$ for the health cap. Both are estimated parametrically.

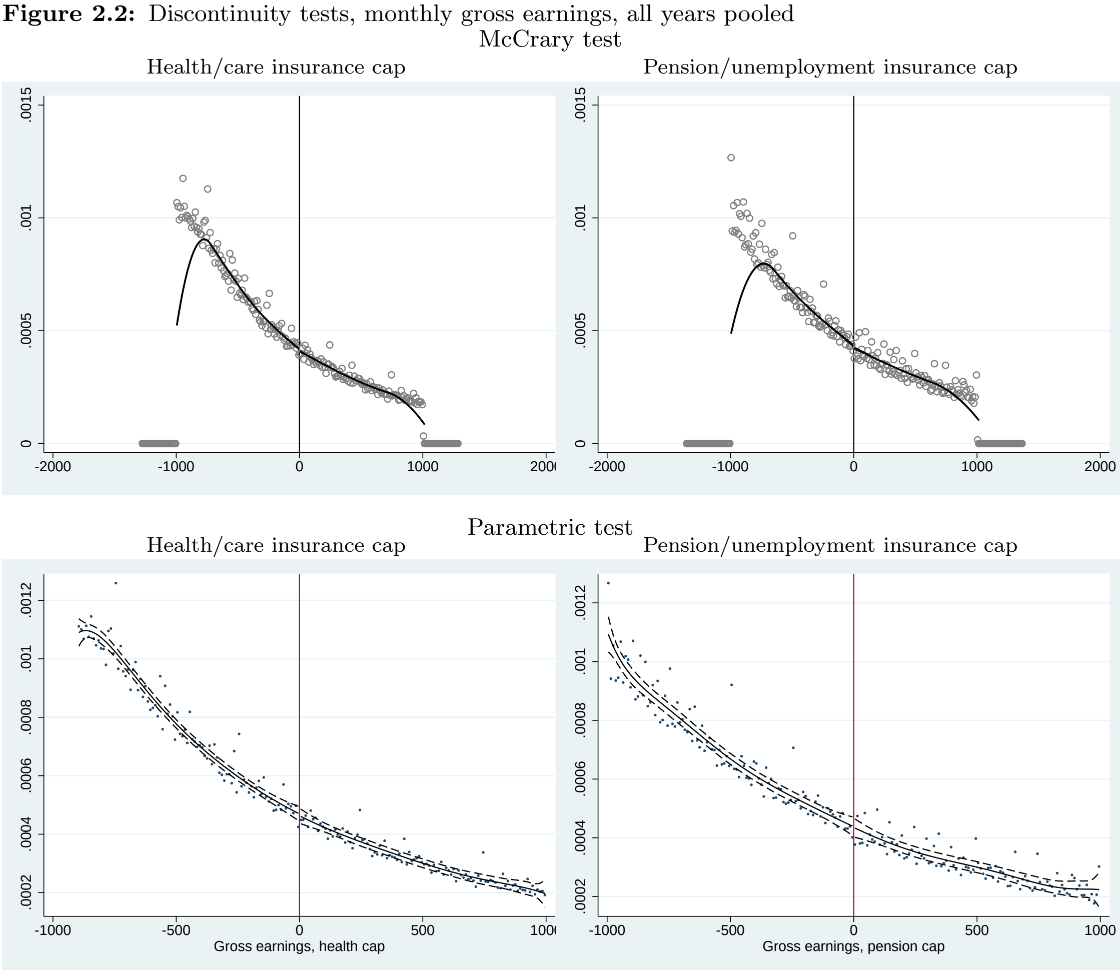
Table 2.2: Discontinuity estimates – gross earnings

	Pension/unemployment cap					**Health/care cap**				
	Δt	**McCrary**		**Polynomials**		Δt	**McCrary**		**Polynomials**	
		p.e.	t	p.e.	t		p.e.	t	p.e.	t
Pooled sample	-.124	-.015	-1.42	-.007	-.54	-.081	-.009	-1.16	-.014	-1.95
Single waves										
1995	-.126	.026	1.24	.026	1.1	-.071	.007	.46	-.005	-.35
2001	-.128	-.026	-1.18	-.044	-1.83	-.076	-.043	-2.92	-.066	-3.09
2006	-.130	.012	.59	.037	1.66	-.084	-.023	-2.11	-.016	-1.22
2010	-.114	-.053	-2.92	-.072	-2.48	-.088	.001	.1	-.001	-.07

Notes: Δt – average drop in marginal SSC rate, p.e. – point estimate, t – t-value (α=0.05).
Source: GSES 1995, 2001, 2006, 2010; own calculations.

The fact that point estimates are not significantly different from zero implies that there is no evidence that either employers or employees shift (some of) their SSC burden. It is neverthe-

Figure 2.2: Discontinuity tests, monthly gross earnings, all years pooled



Source: GSES 1995, 2001, 2006, 2010; own calculations.

less instructive to look at the confidence intervals for s to see the maximum amount of shifting supported by these estimates. The left (right) limit of the confidence interval for the discontinuities at the pension cap translates to $s = -0.15$ ($s = 0.26$). It includes thus scenarios where incidence is slightly more on either employers or employees. At the limits employees (employers) are able to shift 15% (26%) of their formal SSC burden. The confidence interval for the discontinuity estimate at the health cap implies an s between 0 and .35, i.e. economic is identical to statutory incidence or slightly more on employees. Employers shift at maximum 35% of their SSC burden to employees. Accordingly, substantial burden shifting can be rejected, even for the most distinct estimates. Formal and economic incidence of SSC to pension and health insurance (almost) coincide.

This is also supported by the estimates for the separate waves which by the majority turn out to be insignificant. Notable exemptions are significantly negative discontinuities for the pension cap in 2010 as well as the health cap in 2001 and (for the non-parametric approach) in 2006 (table 2.2). It would, however, be highly doubtful to interpret these estimates as substantive findings. Noise in the sub-samples of single waves is notably higher as illustrated by more spiky histograms (figure B.5 in the Appendix).

We conduct a number of sensitivity analyses related to choices made in the estimation. First, we vary the amount of bins: the estimation interval of 2000 €/month is split into between 10 and 1000 bins on either side of the cap. The baseline estimates are based on 100 bins implying a bin size of 10 €/month. Second, we vary the estimation interval between 1000 and 4000 €/month holding the amount of bins at the level of the baseline specification. At last, we double the interval of earnings above the caps which is excluded from the estimation. None of the alternative discontinuity estimates are qualitatively different from the baseline specification (table 2.3). The small and insignificant estimates are very robust in terms of estimation decisions comfirming the result that statutory equals economic incidence.

Table 2.3: Discontinuity estimates – gross earnings, sensitivity

	Pension/unemployment cap				**Health/care cap**			
	McCrary		**Polynomials**		**McCrary**		**Polynomials**	
	p.e.	t	p.e.	t	p.e.	t	p.e.	t
Baseline	-.015	-1.42	-.007	-.54	-.009	-1.16	-.014	-1.95
Number of bins								
10	.004	.32	.010	.64	-.004	-.45	-.019	-2.23
50	-.008	-.73	-.008	-.61	-.011	-1.48	-.015	-2.07
1000	-.013	-1.39	.004	.32	-.008	-1.36	-.013	-1.83
Estimation interval								
1000	-.010	-.59	.017	.92	-.006	-.55	.010	.96
3000	-.017	-1.77	-.009	-.94	-.011	-1.66	-.019	-3.09
4000	-.015	-1.79	-.001	-.10	-.008	-1.31	0.00	0.00
Excluded interval								
20	-.009	-.89	.002	.16	-.012	-1.62	-.018	-2.28

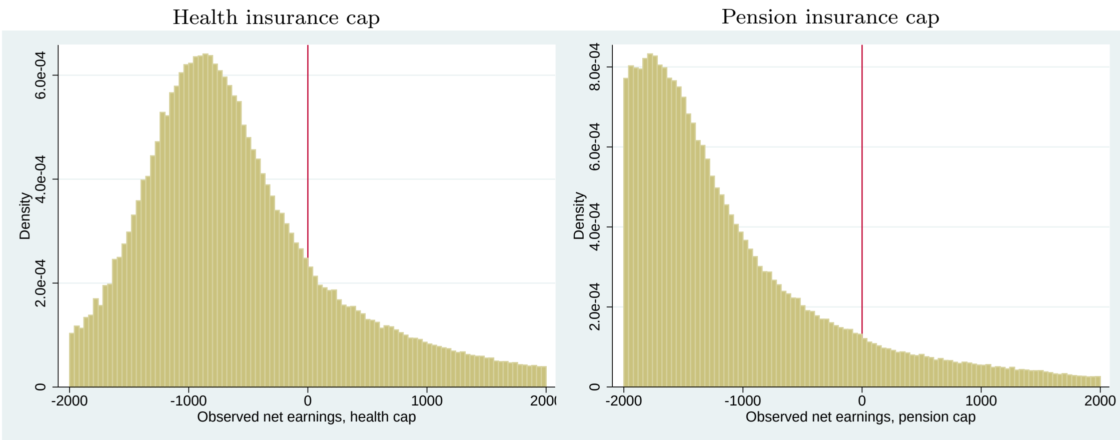
Notes: p.e. – point estimate, t – t-value (α=0.05).; robustness tested given the baseline parameter; baseline specification with 100 bins and estimation interval of 2000.
Source: GSES 1995, 2001, 2006, 2010; own calculations.

Net earnings

Under measurement error the largely insignificant discontinuities for gross earnings might (in part) be resulting from a lack of identification at earnings caps and not represent actual incidence. Separate estimates based on distributions of gross and net earnings (or labour costs) provide a broader foundation for conclusions about economic incidence. The difference in the size of discontinuities across distributions is proportional to the drop in the SSC rate. Should s in fact be (close to) zero resulting in no (a small) discontinuity in gross earnings, a sizeable discontinuity would emerge in net earnings. In our case its magnitude would equal (be close to) the drop in marginal rates of $-.124$ for the pension and $-.081$ for the health cap which should be identifiable empirically. When, on the other hand, the small and insignificant discontinuities in gross earnings are driven by noisy data and imprecise estimates, we would expect to get insignificant results for net earnings as well.

A major advantage of the GSES is that actually paid employees' SSC are available. Note that using this information is different from calculating net earnings (labour costs) mechanically from observed gross earnings. The main systematic difference between such 'calculated' and 'observed' net earnings we use in this paper are contributions to private health insurance (and other voluntary contributions).[35] As a result individual locations of the caps in the observed net earnings distribution vary in an unknown way and estimates should be downward biased. Histograms for observed net earnings indeed seem to lack sharp discontinuities at the caps for pension and health insurance (figure 2.3, figure B.6 in the Appendix).

Figure 2.3: Distribution of observed monthly net earnings, all years pooled



Source: GSES 1995, 2001, 2006, 2010; own calculations.

Turning to our statistical tests, we find substantial and significantly negative point estimates for observed net earnings (table 2.4, figure 2.4 and figure B.7 in the Appendix). These discontinuities arise in the pooled sample and in almost all single waves. Both the non-parametric and the parametric test yield similar estimates. Discontinuities at the pension cap are larger in magnitude than those at the health cap. These patterns are consistent with the findings for gross earnings and expectations given the differences in SSC rates.

Closer examination of the magnitude reveals that the majority of these discontinuities is somewhat smaller than the drop in the marginal SSC rate. One exception for the pooled sample is the parametric test for the health cap yielding a discontinuity slightly larger in size than the actual change in SSC rates (table 2.4). At face value statutory and economic incidence do

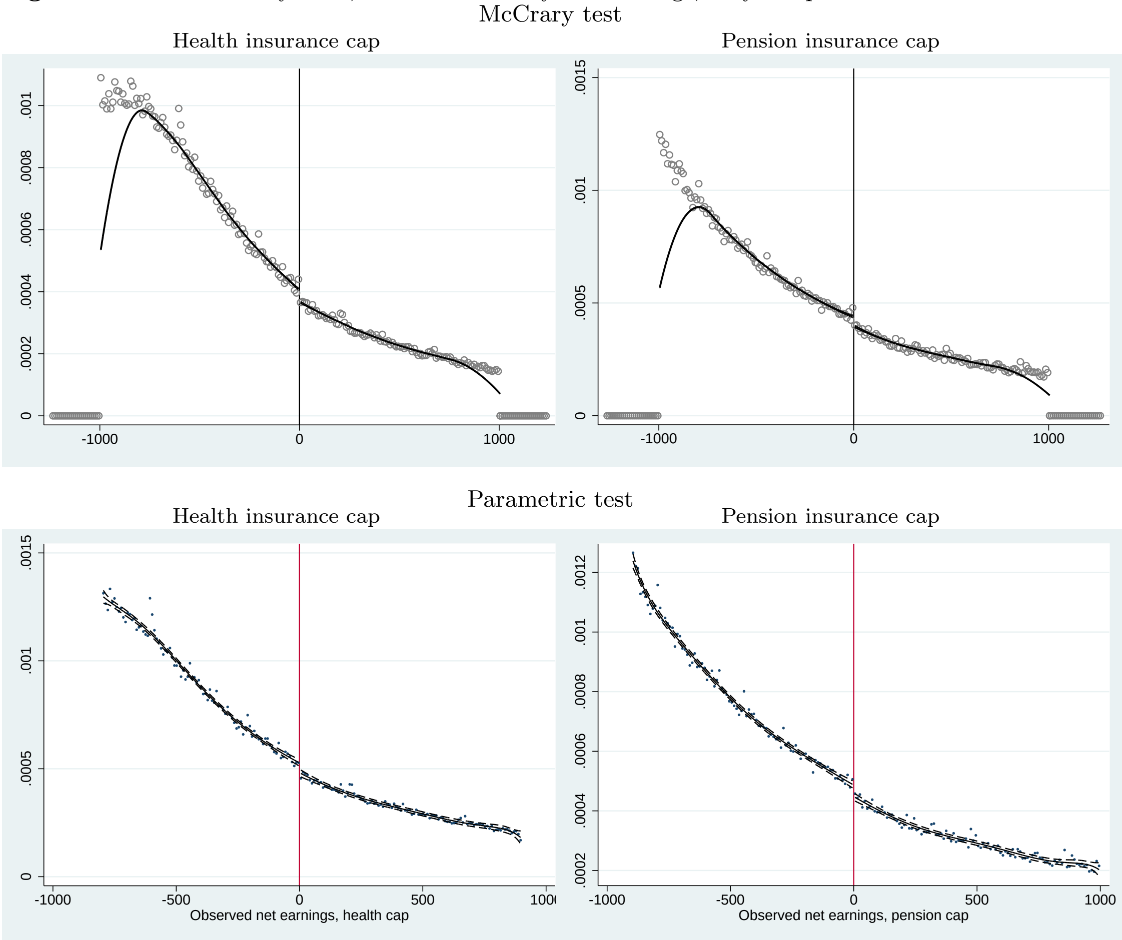
[35] For 'calculated' net earnings (see the consistency check below) public health insurance has to be assumed for everyone as the actual insurance status is not observed.

Table 2.4: Discontinuity estimates – observed net earnings

	Pension/unemployment cap					Health/care cap				
	Δt	**McCrary**		**Polynomials**		Δt	**McCrary**		**Polynomials**	
		p.e.	t	p.e.	t		p.e.	t	p.e.	t
Pooled sample	-.124	-.080	-6.86	-.087	-6.43	-.081	-.067	-8.38	-.096	-12.5
Single waves										
1995	-.126	-.044	-1.89	-.060	-2.58	-.071	-.029	-1.89	-.055	-3.84
2001	-.128	-.104	-4.92	-.105	-3.88	-.076	-.079	-6.01	-.075	-4.23
2006	-.130	-.024	-1.07	-.019	-.820	-.084	-.113	-8.32	-.112	-7.13
2010	-.114	-.128	-5.54	-.174	-6.82	-.088	-.087	-5.53	-.109	-6.35

Notes: Δt – average drop in marginal SSC rate, p.e. – point estimate, t – t-value (α=0.05).
Source: GSES 1995, 2001, 2006, 2010; own calculations.

Figure 2.4: Discontinuity tests, observed monthly net earnings, all years pooled



Source: GSES 1995, 2001, 2006, 2010; own calculations.

not seem to fully coincide. The pooled non-parametric estimate for the pension cap of $-.080$ implies a shifting parameter of $s = .35$ with a confidence interval ranging from .17 to .54. For the pension cap the non-parametric point estimate in the pooled sample is $-.067$ which means that the shifting parameter amounts to $s = .17$ with a confidence interval of $-.02$ to .37.

As argued above there is a reason for expecting discontinuity estimates to be somewhat downward biased. Given that we do not observe the insurance status there is measurement error in earnings caps for individuals with private health insurance. Therefore the discrepancy between the size of the discontinuity and the drop in marginal SSC rates should not be over-emphasised. In 2003 the threshold for compulsory health insurance was raised to a level between the health and pension earnings caps. We have thus 'cleaner' estimates at the health cap in 2006 and 2010 which may serve as an informal test for this interpretation.[36] Estimates from observed net earnings are indeed larger and close to the actual change in SSC rates (table 2.4). Smaller estimates in other years and at the pension cap might be mainly driven by measurement error for individuals with private health insurance. Altogether the evidence from gross earnings supports the interpretation that there is not much shifting going on for either employers or employees.

Given measurement error and round number bunching, it might be that earnings caps randomly coincide with a spike or gap in the density that is not related to incidence. We therefore conduct a sensitivity analysis and look at 'placebo thresholds' below the health and above the pension insurance cap for the pooled sample (table 2.5). The area between both earnings caps is not used for this robustness check to avoid potential interferences among them. We get significant placebo estimates slightly above the pension cap, in particular with the parametric approach. It is, however, clearly smaller in magnitude than the discontinuity at the threshold. Moreover, as indicated above there is a discrepancy between the counterfactual and observed density also above the cap, as both densities slowly converge (figure B.1 in the Appendix). It is not surprising that the parametric estimator picks up this difference above the cap. Consistent with theoretical expectations we do not find any discontinuities further up or down the respective densities. An exception is at $\bar{y}^h - 150$ where the point estimate is merely one third of the original estimate.

Table 2.5: Discontinuity estimates – observed net earnings, sensitivity w.r.t. breakpoint

	Pension/unemployment cap					**Health/care cap**			
	McCrary		**Polynomials**			**McCrary**		**Polynomials**	
Breakpoint	p.e.	t	p.e.	t	Breakpoint	p.e.	t	p.e.	t
$\bar{y}$	-.08	-6.86	-.087	-6.43	$\bar{y}$	-.067	-8.38	-.096	-12.5
$\bar{y}+50$	-.033	-2.75	-.063	-5.57	$\bar{y}-50$	-.013	-1.92	-.037	-4.92
$\bar{y}+100$	.020	1.53	-.006	-.70	$\bar{y}-100$	.012	1.70	.008	1.06
$\bar{y}+150$	.006	.47	0.00	0.00	$\bar{y}-150$	.010	1.41	.032	4.23

Notes: Δt – average drop in marginal SSC rate, p.e. – point estimate, t – t-value (α=0.05).
Source: GSES 1995, 2001, 2006, 2010; own calculations.

A final plausibility check concerns 'calculated' net earnings. There is a deterministic link

[36] The cap for pension insurance is higher up in the distribution and therefore still affected by this measurement error.

between gross earnings y and net earnings c (sub-section 2.2.1). By equation (2.7) almost smooth gross earnings distributions as in our case deterministically create discontinuities in net earnings about the size of the drop in the marginal SSC rate. Estimating discontinuities for calculated net earnings serves as consistency test. Discontinuities are clearly visible in the histograms for 'calculated' net earnings (figure B.7 in the Appendix). This is confirmed by statistically significant negative estimates for the pooled sample and for single waves (table B.3 in the Appendix). The point estimates are slightly above expectations given the changes in SSC rates under no shifting. This is, however, perfectly consistent with the small negative point estimates in gross earnings (table 2.2). We find the largest estimates for the pension cap in 2010 and the health cap in 2001 and 2006 where discontinuities in gross earnings were also bigger.

2.4 Discussion and conclusion

In this paper we analyse economic incidence of social security contributions (SSC) with cross-sectional estimators. The main advantage of this approach is that it does not rely on reform-induced, exogenous changes in SSC institutions over time. Panel data which often suffer from insufficient sample sizes or imprecise or missing information (e.g. hours of work) are therefore not needed. Moreover, identification does not depend on the validity of a control group and is not solely based on short-term responses.

We refine the original framework of Alvaredo and Saez (2007) and discuss the identifying assumptions more explicitly. An empirical implementation of the framework is sketched out and it is shown that non-parametric as well as parametric estimators capture the effects of economic incidence on different earnings densities at earnings caps of SSC.

The main contribution of this paper is the application of the framework to employer-employee data. The German Structure of Earnings Survey (GSES) used here satisfies the necessary requirements in terms of data quality: We pool four waves of data and generate a sample size sufficiently large to reduce random noise in earnings distributions. Being part of the official labour cost statistics the amount of earnings subject to SSC is measured with little error. More importantly, it contains a direct measure of employees' SSC which allows us to analyse discontinuities at earnings caps for gross as well as net earnings distributions. We, therefore, do not rely on a single cross-sectional distribution.

The overall picture of the empirical evidence is conclusive. We see no or very small significant discontinuities in the gross earnings distribution which means that legal and economic incidence coincide. This interpretation is confirmed by the identification of statistically significant negative discontinuities in the distribution of observed net earnings. The result is robust for the parametric and non-parametric estimator. Differences in magnitude of the point estimates are qualitatively consistent for varying SSC rates between the pension/unemployment and the health/long-term care insurance cap. Estimates from net earnings are slightly downward-biased because of measurement error for privately insured individuals. We thus do not over-emphasise the fact that discontinuities are somewhat below the drop in the marginal SSC rates at the cap leaving room for partial incidence on employees.

Our reading of the overall evidence is that economic and legal incidence are more or less identical for SSC in Germany. Employers and employees thus share the burden of SSC for health and pension insurance. Even in the extreme cases supported by our estimates neither employees nor employers shift a substantial share of their SSC burden to the respective other side of the market. Relying on precise data with gross earnings as well as actually paid employees' SSC we argue that measurement issues do not drive this result. Among the substantive explanatory factors salience or adjustment frictions are less convincing in our setting with long-standing earnings caps and our empirical framework which is based on equilibrium incidence. Our local estimates at earnings caps would not be informative about underlying incidence when the drop in average SSC rates above the cap was not or not yet internalised into bargained earnings during wage negotiations. In this scenario incidence would be systematically different below and above the cap which would bias the discontinuity estimates (as shown in Appendix B.1.3). The result of non-standard economic incidence is in line with recent studies for different countries which are mostly based on quasi-experimental identification (Saez et al., 2012; Skedinger, 2014; Bennmarker et al., 2009; Korkeamki and Uusitalo, 2009; Huttunen et al., 2013). It also corroborates previous evidence for Germany which exploits a reform of the earnings cap of the health insurance in East Germany (Neumann, 2015). There are, however, studies which provide evidence for complete shifting of SSC to employees (Gruber, 1994, 1997). A potential explanation could be the different institutional settings, in particular divergent wage setting mechanisms (Alesina and Perotti, 1997; Daveri and Tabellini, 2000).

Chapter 3

In-work benefits for parents - Evidence from a structural labour supply model with participation and hours constraints*

3.1 Introduction

It is well known that constraints on the labour market hinder individuals from freely choosing their labour supply Stewart and Swaffield (1997a). In recent years optimization frictions and restrictions on the labour market have regained interest (Chetty, 2012). This research has also influenced the literature on labour supply elasticities (Chetty et al., 2011), based on quasi-experimental variation (Kleven and Waseem, 2013) as well as structural frameworks (Beffy et al., 2016). Observed behavior on the labour market is not simply the result of utility optimization, but also reflects constraints and frictions prevalent on the labour market that might be related to regulations, institutions, or insufficient labour demand. The distinction is likewise relevant for economic policy analysis. Potential behavioural reactions to changing monetary incentives might be limited by various constraints (Stewart and Swaffield, 1997b). Empirical models of the labour market used for evaluation purposes need to capture these different mechanisms. We therefore augment a static discrete choice labour supply model by including different types of

*This chapter is based on unpublished joint work with Kai-Uwe Müller and Katharina Wrohlich.

constraints. Instead of relying on actual working hours as revealed preferences (van Soest, 1995) we draw on Euwals and van Soest (1999) and estimate labour supply preferences based on stated desired working hours. The model features restriction probabilities for each positive hours choice that represent participation and hours constraints. One can think of *participation constraints* as labour demand rationing (Laroque and Salanié, 2002; Bargain et al., 2010). Individuals may lack the productivity to attract job offers and become involuntarily unemployed. In addition, specific *hours restrictions* (Euwals and van Soest, 1999) may arise because demand for specific numbers of hours is lacking in certain occupations or labour market sectors: employers might for example either favour, or are only able to offer full-time jobs, less part-time jobs and hardly anything in between. On the other hand, jobs with specific working hours might also not be viable for employees because, e.g., adequate childcare is either not available or affordable.

We apply this empirical framework to analyse an in-work benefit for low-paid parents (IWBP) in the German institutional context which specifically (but not exclusively) targets secondary earners. The German tax and transfer system provides strong incentives for an uneven share of market work between mothers and fathers, in particular through joint taxation of married spouses. Therefore, mothers' employment rates and average working hours for those employed are still significantly lower compared to fathers. To tackle this issue eligibility for the IWBP is conditional on hours of work exceeding 25 hours per week. Previous in-work credits as the Earned Income Tax Credit (EITC) in the US (Meyer, 2010) or the Working Tax Credit (WTC) in the UK (Costa Dias et al., 2012) are usually means-tested at the household level implying additional adverse incentives for secondary earners. The amount of the IWBP therefore depends on the individual's gross hourly wage rate (Bargain and Orsini, 2006) and the number of children in the household.

The IWBP is designed to impact labour supply through financial incentives. To what degree this translates into actual employment adjustments depends, however, also on labour market restrictions. Modeling participation and hours constraints is thus crucial for the evaluation of such a policy proposal. Our model not only allows us to answer the research question of how mothers and fathers in couple households with young children would adjust their working behavior. Based on the model we can also disentangle the pure incentive effect on labour supply from the limiting impact of constraints. We can further analyse to what extent different types of labour market restrictions contribute to the low labour supply of secondary earners in households with young children. Effect heterogeneity is considered by distinguishing east and west Germany as well as different qualification levels.

The problem of hours restrictions has been acknowledged already by Moffitt (1982) who extends a Tobit model to account for institutional restrictions on part time work. Van Soest et al. (1990) augment a Hausman (1980) type labour supply model with hours constraints by letting individuals choose between a finite set of wage-hours packages[37]. In terms of participation constraints we draw on a similar tradition. Meyer and Wise (1983) took a first step towards this direction distinguishing involuntary unemployment from other sources of non-employment. Blundell et al. (1987) set up a continuous labour supply model complemented by a rationing risk equation. Laroque and Salanié (2002) estimate a static structural labour supply model and distinguish different types of non-employment.

The general challenge in this literature is to separately identify preferences and the job offer distribution from observed hours of work. Beffy et al. (2016) exploit situations when individuals face non-convex budget sets and are observed to work an irrational amount of hours. Bargain et al. (2010) use a discrete choice household labour supply model and specify a latent rationing equation for the extensive margin utilizing information on desired hours. Bloemen (2008) introduces stated desired hours of work into a job search model in order to separately identify preferences and the job offer distribution.

We follow here an approach by Euwals and van Soest (1999) and exploit survey information about the eligibility and search activities of individuals as well as actual and desired hours to identify preferences and restrictions within a labour supply model with participation and hours constraints. Our contribution to the literature is twofold. First, we extend this framework to couple households allowing for restrictions at the extensive and intensive margin for both partners. Second, we allow the constraints to vary across individuals by observed and unobserved characteristics. Our model accounts for a differential impact of restrictions for men and women, across regions, and for different education levels. We further link the explanatory factors to various underlying mechanisms: We consider labour demand rationing as well as working hours norms that vary across occupations. Restrictions which do not originate from the labour market are also covered. Insufficient formal childcare might, for instance, prevent parents from accepting jobs with many hours. Having explanatory variables representing different mechanisms allows us to simulate their effect by relaxing different kinds of constraints at a time. Preferences and restrictions are jointly estimated which facilitates correcting for selection by way of unobserved characteristics in the equations for the different hours restriction probabilities.

Focusing on a sample of couple households with young children we find that the IWBP for Germany would increase the mothers' participation rate by 0.7 percentage points and their av-

[37]Similar approaches are followed by Tummers and Woittiez (1991), Dickens and Lundberg (1993), Aaberge et al. (1995) and Bloemen (2000)

erage working hours by four percent. Behavioural effects of fathers are negligible. A standard labour supply model that is based on actual hours and that ignores labour market constraints would yield upward-biased estimates and too optimistic conclusions about the benefit's policy impact. A simulation of the effects of restrictions shows their substantial impact on employment. Working hours of mothers would increase by 50% and their participation rate by ten percent when constraints could be removed completely. This theoretical exercise also reveals the heterogeneous impact of restrictions along other dimensions: men and women are confronted with different labour market restrictions. Fathers face the largest constraints in part-time employment. Mothers are primarily constrained in categories with large working hours.

The remainder of the paper is in four parts. Section 3.2 describes our data set and presents some descriptive findings on the discrepancy of desired and actual working hours for couples in our sample. We discuss the econometric model in section 3.3. Empirical results are presented in section 3.4. After presenting parameter estimates, the selection correction in the estimation of restrictions and the model fit, we compare labour supply elasticities of our model with a standard labour supply model. Then we describe the design of the IWBP and apply our model to evaluate the employment effects. Section 3.5 discusses the findings and concludes.

3.2 Data and descriptive results

The empirical analysis is based on the survey "Families in Germany" (FiD, see Schröder et al., 2009). The FiD is an add-on of the German Socio-Economic Panel (SOEP), a representative survey of German households conducted since 1984 (Wagner et al., 2007). Like the SOEP the FiD contains information on different sources of income, working time, the previous labour market experience as well as detailed socio-demographic characteristics of the individual and the household.

We exploit the first four waves available covering the period 2010 to 2013. About 4,500 households are surveyed per wave that include about 8,000 children. This comprises four sub-samples which are representative for their respective populations: families with children born between 2007 and 2010, low income households, families with three or more children, and lone parents. We restrict the analysis to the sub-sample of families with children born between 2007 and 2010. All results presented in this study are thus representative for that group of households. While this prevents calculating effects for all parents in the German population, the sample size of households with young children for who we expect the largest effects of an in-work benefit for parents is comparably large[38].

[38]The FiD has also another advantage. In comparison to other household data sets (including the SOEP),

FiD respondents are not only asked about their (effective) actual working time, but also whether they are satisfied with these hours worked. The exact wording is: “If you could choose your own number of working hours, taking into account that your income would change according to the number of hours: Would you prefer to decrease, increase or maintain your number of working hours?”. If they prefer a change, they will be asked for their desired working hours.[39] There are two possible interpretations to this question (Callan et al., 2007): Respondents might choose their desired hours of work conditional on their partners’ actual working hours (constrained optimization). We deviate from this view and assume here that both spouses can freely choose their desired labour supply (unconstrained optimization of family utility).

As we are mainly interested in the effects of introducing a policy which aims at increasing work incentives for secondary earnings, our sample is restricted to couple households only. Moreover, we exclude observations with negative net income and couples with more than three children. Further, we only consider couples where both spouses are ‘flexible’ with respect to their labour supply, i.e. are neither in full-time education, on maternity leave, fully disabled, nor retired. Overall 3,604 choice situations of 1,369 households are used in the estimation. The descriptive statistics are given in table C.2 in the Appendix.

Cross-tabulating the distributions of desired and actual hours for men and women in our sample reveals some basic patterns (table 3.1). For men the by far most desired hours categories are full-time and overtime. Most of these men actually seem to be able to work their desired amount of hours. This changes significantly for men who desire to work 25 to 35 hours whereof almost 60% are restricted in the sense that desired and actual hours deviate. Virtually all of them are over-employed. Men do generally not desire to or actually work in the lower two hours categories. Approximately 8% of the male sample does not work, less than half of them voluntarily.

By contrast, non-participation is the dominating alternative for women. Almost half of the female sample prefers not to work. The great majority of employed women actually works and desires to work in one of the three part-time categories. 72% of women desire to work less than 25 weekly hours. Less than 10% prefer the full-time categories. The scope for increasing incentives for women to work more is thus fairly high.

Preferences and incentives are not the only reasons for women working few hours or not at all. Roughly 10% of non-working women are involuntarily unemployed. Under-employment

the FiD provides more detailed information on potential restrictions. FiD respondents are not only asked about their (effective) actual working time and their desired working hours but also about the reason for a potential deviation.

[39] For non-employed individuals the question on desired hours of work differentiates only between non-work, part-time and full-time. According to their preferences we assign them randomly a specific hours category, e.g. different lengths of part-time, with the probability weights given by the observed shares for employed people.

Table 3.1: Joint frequency distribution of desired and actual hours of work

Desired \ Actual	0	1-14	15-24	25-35	36-40	>40	Total
Men							
0	138	0	0	1	0	1	140
1-14	2	16	1	2	1	0	22
15-24	3	0	19	3	6	8	39
25-35	33	2	6	196	119	114	470
36-40	83	12	12	50	1293	203	1653
>40	32	1	0	3	77	1167	1280
Total	291	31	38	255	1496	1493	3604
Women							
0	1619	0	0	0	0	0	1619
1-14	14	314	8	1	0	0	337
15-24	41	110	472	33	11	1	668
25-35	72	29	74	434	61	25	695
36-40	18	7	7	33	143	7	215
>40	2	1	0	0	2	65	70
Total	1766	461	561	501	217	98	3604

Notes: Desired hours of involuntary unemployed discriminate only between part-time, full-time and *both*. Finer categories are allocated randomly and proportionally by considering the crude information.
Source: Own calculations based on FiD, wave 2010-2013.

also concerns employed women, though. While approximately half of them desire to work more than 25 hours, only about 40 % actually do. Supply side reasons like child care duties are stated to be similarly responsible for not being able to increase hours of work as labour demand rationing. This is however only true for women. For under-employed men supply side reasons are of minor importance.

There is evidence that the deviation between actual and desired hours as stated in a survey is informative about working hours constraints. Blundell et al. (2008), for example, show that desired hours have predictive power for next year's working hours. Euwals (2005) comes to the same conclusion based on SOEP data. Bryan (2007) finds that local labour market conditions only affect hours of work for individuals whose actual deviate from their desired hours.

The labour demand variables measured at the regional level are taken from a data set called "Indicators and Maps on the Spatial Development" ("Indikatoren und Karten zur Raumentwicklung", INKAR, see Helmcke, 2008). It includes indicators at different regional levels for Germany. We use data at the county level which can be matched based on regional identifiers in the FiD.

3.3 Econometric model

The model is based on a discrete choice specification (van Soest, 1995) of a standard labour supply framework (Blundell and Macurdy, 1999). Given their expected wage employees choose among job offers with different hours of work. Households maximise their aggregate utility which is determined by income and leisure. We go beyond a purely neoclassical conception of the labour market by considering participation and working hours constraints that individuals face when choosing their labour supply (Stewart and Swaffield, 1997b). Observed and unobserved heterogeneity at the household level is allowed in preferences and constraints which are estimated jointly to control for selection in the restriction part.

Following Euwals and van Soest (1999) the model consists of two main building blocks. First, *labour supply preferences* are modeled within a standard discrete-choice labour supply framework (van Soest, 1995). We consider cohabiting partners with young children who are assumed to decide collectively on job offers. Estimation of preferences is based on stated desired, rather than actual working hours (van Soest et al., 2002; van Soest and Das, 2001). Desired working hours are assumed not to be affected by actual hours or potential restrictions and thus indicate the undistorted labour supply of both partners.

Second, choice restrictions are modeled representing *participation and hours constraints*. Participation constraints accrue from labour demand rationing (Laroque and Salanié, 2002; Bargain et al., 2010) as certain individuals are not productive enough to receive any job offers and are involuntarily unemployed. Conditional on participation, hours restrictions might arise either when jobs with preferred working hours are not offered by employers (Euwals and van Soest, 1999). Working hours norms might, e.g., vary between occupations or sectors. Or jobs with specific hours might not be viable for certain employees because adequate childcare is not available or affordable. Note that hours constraints can also lead to observed involuntary unemployment: people might choose zero hours with their preferred hours choice(s) not being available. The constraints covered here are more comprehensive than adjustment costs for hours of work or informational frictions on the labour market that are commonly captured in job search frameworks (Rogerson et al., 2005). We, however, do not structurally model the restriction mechanisms. We further abstract from dynamic considerations in both model parts (Keane et al., 2011).

Constraints are specified as restriction probabilities on individual employment states with positive working hours. Estimation is based on survey information about stated desired and observed actual working hours. Assuming that constraints are not directly dependent on the spouse's restrictions, the deviation between desired and actual hours is informative about individual restrictions.

This approach comes with potential selection problems, though. Information on a given hours category is only available, if an individual desires to or actually works in this state. Since certain groups of people are less or more likely observed to desire or to work in specific categories, the identification of hours constraints is based on groups of individuals that are not representative for the whole population. It is conceivable that a couple's propensity for consumption is systematically related to its restriction probabilities. This is similar to the standard selection problem in labour economics (Heckman, 1979). In our model preferences and choice restrictions are therefore related through observed and unobserved factors and estimated jointly. Unobserved heterogeneity in constraints is modeled non-parametrically by two latent types for men and women. The combination of types within couples results in four latent household types differing in their preference for consumption.

In the remainder of this section we first discuss preferences (sub-section 3.3.1) and then the constraints part of the model in greater detail (sub-section 3.3.2). After presenting the likelihood function (sub-section section 3.3.3) we explain how we model unobserved heterogeneity and selection issues (sub-section section 3.3.4).

3.3.1 Preferences for work

Couples choose a labour supply arrangement from $j = 1, 2, \ldots, J$ different labour market states[40]. Households maximise the direct utility function $u_j(.)$ in the arguments net income y_j and leisure l_j^m, l_j^w for men and women, denoted by m and w respectively. Leisure time is defined as the difference of total time endowment $TE = 80$ and hours of work h_j^m or h_j^w. Similar to Euwals and van Soest (1999), van Soest et al. (2002), van Soest and Das (2001), or Callan et al. (2007) we take desired working hours h_j^d (as opposed to actual hours h_j^a) as revealed preferences. Consistent with the unitary labour supply model we assume that preferred hours are stated under the condition that both spouses choose freely without facing restrictions. Labour supply preferences can then be inferred directly from desired working hours as stated in our data.

Disposable income $y_j = y(h_j^m, h_j^w, w^m, w^f, y^{nl}, X^m, X^w, X)$ depends on both spouses' labour supply, their before tax wage rates (w^m, w^w), non-labour household income y^{nl} and individual or family characteristics (X^m, X^w, X) which determine taxes, contributions and transfers. We use a microsimulation model (Steiner et al., 2012) to compute y_j for all possible labour supply choices of each household. Wage rates are estimated outside of the labour supply model[41].

Following van Soest (1995) we assume that couples choose out of a finite number of mutually exclusive alternatives. A family's unrestricted choice set contains all pairwise combinations of the man's and the woman's hours categories. Alternative j corresponds to the combination of

[40] For the sake of readability, we do not specify a household index in this exposition.

[41] Wage rates are derived from reported gross monthly wage earnings and observed working hours for the employed. Hourly wages of non-employed persons are predicted on the basis of parameters from wage equations and then inserted into the labour supply model. The wage equations control for selectivity as proposed by Heckman (1979) and are estimated separately for men and women as well as East and West Germany. Hourly wages are assumed to be exogenous and constant for different hours categories throughout this analysis (Appendix C.1).

the man's and woman's working hours category h_j^m, h_j^w, and the resulting family income y_j. Hours categories for both spouses include non-employment, marginal employment (10 hours), part-time (20 hours), reduced full-time (30 hours), full-time (40 hours) and overtime (45 hours). The unrestricted choice set thus consists of $J = 36$ alternatives.

We use a linear-quadratic specification of the utility function and allow preferences for leisure and consumption in their linear terms to vary with observed individual or household characteristics. Adding alternative-specific error-terms e_j leads to the following random utility specification:

$$\begin{aligned} u_j \quad = \quad & l_j^m \beta^{l^m} + (l_j^m \boldsymbol{X}^{l^m})' \boldsymbol{\beta}^{l^m X^{l^m}} + (l_j^m l_j^m) \beta^{l^m l^m} + \\ & l_j^w \beta^{l^w} + (l_j^w \boldsymbol{X}^{l^w})' \boldsymbol{\beta}^{l^w X^{l^w}} + (l_j^w l_j^w) \beta^{l^w l^w} + (l_j^m l_j^w) \beta^{l^m l^w} + \\ & y_j \tilde{\beta}^y + (y_j \boldsymbol{X}^y)' \boldsymbol{\beta}^{y X^y} + (y_j y_j) \beta^{yy} + e_j \end{aligned} \tag{3.1}$$

where $\boldsymbol{X}^l$ represent observed individual characteristics including age, dummies for German nationality and East German residence, disability, and young children (aged below three years) in the household. $\boldsymbol{X}^y$ only includes an indicator for young children in the household. β denotes parameters. The coefficient of the linear consumption term can be decomposed into a fixed and a random part: $\tilde{\beta}^y = \beta^y + \kappa$. The random component κ varies between households (section 3.3.4 below). Bold letters indicate column vectors. Assuming e_j to be i.i.d. type I extreme-value results in the well-known closed form solution for the choice probabilities $P(j)$ (McFadden, 1974).

3.3.2 Choice restrictions

Instead of modeling restriction mechanisms structurally, we follow Euwals and van Soest (1999) and exploit information in our data to identify restriction probabilities for all choice categories with positive hours and each spouse. We utilise individuals' stated willingness to work and job search activities as well as their stated desired and actual working hours. This survey information loosely reflects the two substantive labour market mechanisms constituting choice restrictions: *participation and hours constraints.*

More concretely, the probability that a certain hours category is available depends, first, on the likelihood that an individual has the general ability to find a job. In order to distinguish employed persons from involuntarily unemployed and (voluntarily) inactive individuals, we exploit information in the data: All people that are observed to work positive hours are considered employed. Individuals that state to actively search for a job and to be available to the labour market, but are observed to work zero hours are regarded as involuntarily unemployed. Observed involuntary unemployment does not necessarily identify a participation constraint, though, but could also result from constraints in preferred hours categories. In the empirical specification we thus do not explicitly differentiate between both mechanisms. We use survey information on whether unemployed individuals seek a part-time and/or full-time position to infer which choices are restricted[42]. Voluntarily inactive persons who do not search and are not available for work do not contribute to the identification of hours restrictions.

Second, conditional on participation an individual might not have the opportunity and capacity to work the desired number of hours. Whether an individual is able to choose a given number of hours or whether this choice is restricted cannot be observed directly. We infer this information from the individual's stated hours preferences (h^d) and the observed actual working hours (h^a). This approach rests on the assumption that the mapping from h^d to h^a is exclusively determined by individual hours restrictions and is, e.g., not motivated by the partner's deviation from desired hours[43]. For $h^d = h^a$, we know that h^a is available and the individual only contributes to the identification of the restriction probability of h^a. Again, voluntarily inactive people ($h^d = h^a = 0$) do not contribute to identification as we do not have any information on their potential hours restrictions.

When actual and desired hours deviate, it is immediately obvious that h^d is not available but h^a is. Take the example that an individual who desires to work 20 hours, but actually works 40 hours per week. He or she would contribute to the estimation of restriction probabilities for the 20 hours (as being constrained) and the 40 hours category (as being unconstrained). Without making further assumptions it is not known which other choices are potentially available. We do not make any inferences about other choices for those individuals.[44] Restriction probabilities for each positive hours category are estimated by pooling the information from men and women in the sample.

The probability that a given number of working hours k cannot be chosen, is specified as a function of observed and unobserved characteristics and denoted by $\psi(k)$:

$$\psi(k) = D^w \gamma_k^w + \boldsymbol{X}^{emp\prime} \boldsymbol{\gamma}_k^{emp} + \boldsymbol{X}^{h\prime} \boldsymbol{\gamma}_k^h + \mu^m + \mu^w + \epsilon_k \qquad (3.2)$$

[42] We assume that involuntarily unemployed individuals seeking a full-time position would accept all jobs. They contribute to the identification of all hours categories as being constrained. Unemployed seeking a part-time position might simply have a large preference for leisure, though. They thus only contribute to the identification of restrictions in part-time categories. As a robustness check we similarly use unemployed individuals seeking a full-time position only to identify restrictions in full-time categories (Appendix C.4.1).

[43] Under the alternative interpretation that individuals answer the question about desired working hours given the constraints of their spouse a deviation between desired and actual hours would be a mixture of restrictions and optimization. As a robustness check we identify hours constraints only based on couples where at least one is working his/her desired hours (Appendix C.4.2).

[44] Euwals and van Soest (1999) assume, for example, that utility decreases with the absolute distance between actual and desired hours. In our example, the 10 and 30 hours categories would then be treated as not available as they are closer to the desired 20 hours than to the actual 40 hours. However, this assumption is not consistent with the IIA property underlying the preference part of the model which is why we do not follow this approach.

Explanatory variables are related to the aforementioned mechanisms of involuntary unemployment and specific hours constraints. The dummy for women D^w picks up overall differences in restriction probabilities between men and women. All covariates are assumed to have homogenous effects for men and women with the coverage rate for public childcare being the only exception. As women are usually responsible for the bulk of parental care we allow for its impact to vary by gender. Unobserved terms are also specified separately for men and women. The first set of observables $\boldsymbol{X}^{emp}$ contains variables related to representing an individual's productivity like age, education, health, German nationality, East German residence or the loss of human capital (depending on the time of un- or non-employment). Moreover, unemployment rates at the county level are included to reflect the performance of the regional labour market. These variables are primarily thought to affect the overall employment probability. As some of them are also related to specific hours restrictions, we do not constraint parameters $\boldsymbol{\gamma}_k^{emp}$ to be constant across hours categories.

In addition, $\boldsymbol{X}^h$ contains variables that influence constraints for specific hours choices. Related to the demand side of the labour market, we suppose restrictions to vary over different occupations. For certain occupations or types of jobs a full-time contract is the norm, whereas other areas feature more non-standard employment relationships (Eichhorst et al., 2013). To capture this variation, we use the "International Standard Classification of Occupations" (ISCO) of the current (first) employment for (non-)employed individuals[45]. Hours restriction might also rather emerge in smaller firms without the leeway to offer flexible contracts. On the other hand, hours constraints might also arise on the supply side of the market when, e.g., individuals are not capable of working their desired number of hours: Parents of young children who do not have access to public childcare (rationing on the childcare market) face this type of constraints. Thus, $\boldsymbol{X}^h$ includes full- and part-time coverage rates of public care for children aged below three collected at the county level and provided by the German Statistical Office.

The terms μ^m, μ^w vary between individuals and represent unobserved heterogeneity for men and women specified as random intercepts in all restriction equations (sub-section 3.3.4). Finally ϵ_k is an error term assumed to follow a logistic distribution. Thus, we get the closed-form representation of the Logit model for the (conditional) hours category-specific restriction probabilities $\Psi(k)$.

[45] As some individuals have never been employed, we orthogonalise the indicator variables of the different occupations.

3.3.3 Likelihood function

Conditional on unobserved characteristics, the probability for a household to be observed with a given combination of working hours j can be written down in terms of labour supply preferences and restriction probabilities. When both spouses are voluntarily inactive ($j = 0$) it amounts to:

$$P_j\left(h_j^{a,m} = h_j^{d,m} = 0, h_j^{a,w} = h_j^{d,w} = 0|\kappa\right) = \frac{exp\,(u_j|\kappa)}{\sum_r exp\,(u_r|\kappa)} \tag{3.3}$$

Since we do not have any information on their restriction probabilities, these households only contribute to the identification of labour supply preferences. Imagine, by contrast, a household where the male spouse desires to work 40 hours and is also observed to have a job in this category with the female spouse preferring 20 hours, but being observed to work 40 hours. This household's probability to be in this particular state conditional on unobservables is:

$$\begin{aligned} &P_j\left(h_j^{a,m} = h_j^{d,m} = 40, h_j^{a,w} = 40, h_j^{d,w} = 20|\kappa, \mu^m, \mu^w\right) \\ &= \frac{exp(u_j|\kappa)}{\sum_r exp(u_r|\kappa)}\left(1 - \frac{exp\big(\psi(h_j^m\!=\!40|\mu^m)\big)}{1+exp\big(\psi(h_j^m\!=\!40|\mu^m)\big)}\right) \times \\ &\frac{exp\big((\psi(h_j^w\!=\!20|\mu^w)\big)}{1+exp\big((\psi(h_j^w\!=\!20|\mu^w)\big)}\left(1 - \frac{exp\big((\psi(h_j^w\!=\!40|\mu^w)\big)}{1+exp\big((\psi(h_j^w\!=\!40|\mu^w)\big)}\right) \end{aligned} \tag{3.4}$$

We thus assume that conditional on observed and unobserved heterogeneity both partner's constraints are independent. They are related through shared household characteristics (e.g. similar labour market conditions and supply side restrictions), correlated individual attributes (e.g. level of schooling and qualification or labour market experience), and unobserved characteristics, though. Note further that restriction probabilities scale down labour supply choice probabilities. To calculated the actual expected state probabilities we numerically re-allocate the 'restricted probability mass' to all other categories according to their relative choice probabilities. This substitution pattern is directly implied by the IIA property underlying the preference part of the model (see Appendix C.3 for more details).

Having data on $n = 1, 2, \ldots, N$ couple households in potentially $t = 1, 2, 3, 4$ time periods, the individual conditional likelihood contribution of a household n at time t can be written as:

$$L_{nt}|\kappa, \mu^m, \mu^w = \prod_{j=0}^{J} P_{ntj}\left(h_{ntj}^{a,m}, h_{ntj}^{a,w}, h_{ntj}^{d,m}, h_{ntj}^{d,w}|\kappa, \mu^m, \mu^w\right)^{d_{ntj}} \tag{3.5}$$

where d_{ntj} is an indicator that is equal to one for the observed combination of actual and desired hours of both spouses in household n and time period t, and zero otherwise. In order to get the unconditional sample likelihood we have to integrate out the unobserved heterogeneity terms:

$$L = \prod_{n=0}^{N} \int f(\kappa, \mu^m, \mu^w) \prod_{t=1}^{T} (L_{nt} | \kappa, \mu^m, \mu^w) \tag{3.6}$$

The specification of the joint distribution of the unobserved heterogeneity terms ($f(\kappa, \mu^m, \mu^w)$) will be detailed in the following sub-section.

3.3.4 Unobserved heterogeneity and estimation

The unobserved components in the utility function κ and in the restriction probabilities μ^m, μ^w are assumed to follow discrete distributions (Heckman and Singer, 1984a,b). Mass points in these distributions are interpreted as unobserved household types in terms of preferences for consumption g^u and unobserved individual types with respect to hours restrictions of men g^h_m and women g^h_w. As described in (3.1) above, unobserved heterogeneity for the preference of consumption is specified as random coefficient in the linear term for consumption which now varies between unobserved types: $\tilde{\beta}^y_{g^u} = \beta^y + \kappa^{g^u}$. In the equations for hours restrictions (3.2) we include random intercepts which vary between men and women as well as unobserved types: $\mu^{g^h_m}, \mu^{g^h_w}$. We distinguish four unobserved types for consumption preferences and two for the restriction probabilities of men and women, respectively:

$$\begin{aligned} \kappa^{g^u} &= \mathbb{1}(g^u = 1)c^u_1 + \mathbb{1}(g^u = 2)c^u_2 + \mathbb{1}(g^u = 3)c^u_3 + \mathbb{1}(g^u = 4)c^u_4 \\ \mu^{g^h_m} &= \mathbb{1}(g^h_m = 1)c^h_{m,1} + \mathbb{1}(g^h_m = 2)c^h_{m,2} \\ \mu^{g^h_w} &= \mathbb{1}(g^h_w = 1)c^h_{w,1} + \mathbb{1}(g^h_w = 2)c^h_{w,2} \end{aligned} \tag{3.7}$$

where $\mathbb{1}(.)$ is an indicator function and c are parameters to estimate. Heterogeneity terms for hours constraints are normalised to be zero in expectation. Moreover, we allow the distributions of types g^u and g^h_m, g^h_w to be correlated by specifying a joint distribution (Haan and Uhlendorff, 2013). Let π denote probabilities, then we assume the following (non-parametric) joint distribution:

$$\begin{aligned} \pi_1 &= P(g^u = 1, g^h_m = 1, g^h_w = 1) \\ \pi_2 &= P(g^u = 2, g^h_m = 1, g^h_w = 2) \\ \pi_3 &= P(g^u = 3, g^h_m = 2, g^h_w = 1) \\ \pi_4 &= P(g^u = 4, g^h_m = 2, g^h_w = 2) \end{aligned} \tag{3.8}$$

The possible combinations of restriction types within a household add up to four household types which are assumed to differ in their consumption propensity. We thus assume a deterministic relationship between unobserved constraint types for men and women and household preferences for consumption[46]. Whether or not certain restriction types are associated with a higher or

[46] In a robustness check we allow two household types with respect to consumption preferences to be freely

lower preference for consumption is a priori not determined, but estimated.

We use an expectation-maximization (EM) algorithm (Train, 2009) for the estimation of the type probabilities π (Appendix C.2). This is crucial for correcting the described selection due to unobservable characteristics in the estimation of constraints as it allows for computing individual (not merely average) type probabilities. The distribution of unobserved types are thus allowed to vary across the sub-samples contributing to the identification of restriction probabilities in different hours categories. Estimating the parameters conditional on the distribution of unobserved types in the estimation sample prevents an omitted variable bias when explanatory variables are correlated with the probability of being a certain type. Individual type probabilities are further used to compute unconditional expected restriction probabilities for all individuals and for each hours category which do not suffer from selection bias (Appendix C.2 for more details). The unconditional sample likelihood has the following form:

$$L = \prod_{n=0}^{N} \sum_{g=1}^{4} \pi_g \prod_{t=1}^{T} \left(L_{ntg} | \kappa^{g^u}, \mu^{g^h_m}, \mu^{g^h_w} \right) \tag{3.9}$$

It is a weighted average over the four unobserved type combinations we have specified. Having longitudinal information, i.e. up to four time periods of data, in the sample bolsters identification of unobservables. The latent types are assumed to be time-constant whereas labour market states as well as the restriction status of some households may change over time. Note that the IIA property underlying the preference part of the model is partly be relaxed by unobserved heterogeneity (Appendix C.2).

3.4 Estimation results and policy simulations

In this section we first present estimation results related to our labour supply model with constraints. We discuss the theoretical consistency of key parameter estimates, the model's in-sample fit in terms of the distribution of observed actual hours of work, and illustrate the selection correction with respect to hours constraints (sub-section 3.4.1). Then we discuss labour supply elasticities which are simulated on the basis of the parameter estimates (sub-section 3.4.2). Second, our model is applied to a specific policy reform. We simulate the behavioural effects of an in-work benefit for parents which is specifically designed to improve incentives for secondary earners within the institutional context of the German tax and transfer system (sub-section 3.4.3). The model allows the overall employment effect induced by the benefit to be disentangled in the pure incentive effect and the impact of different hours restrictions. Throughout this section results from our labour supply model with hours constraints are com-

correlated with the restriction types (Appendix C.4.3).

pared with those from a standard labour supply model (based on actual working hours without restrictions) as a benchmark. Third, we simulate the effects of removing different kinds of constraints (sub-section 3.4.4).

3.4.1 Parameter estimates, selection correction, and model fit

The *parameter estimates* of the *labour supply equation* cannot be interpreted directly because of the non-linearities in the model. Comparing coefficients with estimates based on the standard model reveals that they are qualitatively similar (table C.3, Appendix C.5). Quantitative differences imply some discrepancies in elasticities (sub-section 3.4.2). There is a pattern in the relationship between unobserved types in terms of hours restrictions and consumption preferences: When the woman is a 'bad' restriction type, i.e. she has c.p. higher restriction probabilities, consumption is valued markedly lower than when she is a 'good' type. The consumption propensity is larger for households with 'good' type men only when their spouse is also a 'good' type. Less than half of households with the woman being a bad 'type' have a positive first derivative with respect to consumption. For the other two unobserved household types, first derivatives are positive for all households indicating that model estimates are consistent with the underlying economic theory. Households with both spouses being a 'good' type account for approximately half of the sample. The combination of a 'good' type man and a 'bad' type woman comes in second with a share of 26 %. The remaining two types represent slightly more than 10 % of the sample each.

Turning to hours constraints we first specified a group of variables rather affecting overall *participation constraints* through individual productivity or labour demand shortages. Living in a tight labour market as approximated by the regional unemployment rate is significantly related to a higher risk of constraints in all hours categories with the effect size increasing in hours of work (table C.4, Appendix C.5). Low qualified people have higher restriction probabilities in almost all hours categories. By contrast, the rationing risk is decreasing with age and also lower for employees with German nationality. As unobserved effects are normalized to be zero in expectation there is a 'good' and a 'bad' unobserved type for women and men.

Men have markedly lower restriction probabilities for full-time categories and are significantly more likely to be restricted in the three part-time categories. This points to a second mechanism: differential *hours constraints* conditional on participation. Patterns of occupation-specific heterogeneity in hours constraints confirm that different types of jobs are available depending on people's profiles and qualifications: Managers have lower constraints for overtime hours, but an above-average rationing risk when they desire a normal full-time job or want to work 25-35 hours (table C.4, Appendix C.5). Service workers have markedly higher restriction probabilities in all hours categories from 25 hours upwards. On the other hand, employees with elementary occupations[47] have particularly low restriction probabilities in jobs with low working hours. In terms of supply side constraints we find lower restriction probabilities for parents living in regions where more places of full-time childcare are provided. This effect is significant for normal part-time and normal full-time. As expected it is larger for women.

Looking at average observed and predicted restriction probabilities reveals that the *in-sample fit* for the *hours constraints* is very good (Table 3.2). We find very different patterns of hours restrictions for men and women in observed as well as predicted restriction probabilities: Women have significantly lower probabilities to be constrained in all hours categories below full-time jobs. In contrast, men face significantly lower restriction probabilities in full-time and overtime compared to their rationing risk in all other categories with lower hours and compared to women's rationing risk in the full-time and overtime choices.

Table 3.2: Observed and predicted restriction probabilities per category

	Men				**Women**			
Hours category	Observed	Predicted In-sample	Predicted Out-of-sample		Observed	Predicted In-sample	Predicted Out-of-sample	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-14	0.84	0.84	0.79	0.53	0.25	0.25	0.30	0.24
15-24	0.82	0.82	0.79	0.56	0.35	0.35	0.41	0.42
25-35	0.61	0.61	0.61	0.53	0.40	0.39	0.44	0.48
36-40	0.22	0.22	0.22	0.21	0.30	0.30	0.33	0.39
>40	0.13	0.13	0.15	0.12	0.31	0.31	0.32	0.35
Unobs. het.		✓	0	✓		✓	0	✓

Notes: Observed – Observed restriction probability, In-sample – predicted restriction probability for estimation samples of respective hours categories, Out-of-sample – predicted restriction probability for entire sample, Unobs. het. – unobserved heterogeneity.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

More importantly, we can also illustrate the degree of selectivity prevalent in the different sub-samples of each hours category on which the estimation of hours constraints is based. To that end in-sample and out-of sample predictions are compared. We can thereby disentangle the contribution of observed and unobserved characteristics to the sample selection by comparing model predictions without and with unobserved heterogeneity. For men the strong discrepancies in restriction probabilities between full-time and overtime hours on the one hand and all part-time categories on the other hand are mitigated when selection in category-specific sub-samples is corrected for - the difference stays substantial, though (Table 3.2). The selection due to unobservables is much higher than the contribution of observed variables. The restriction probability for men in normal part-time, e.g., goes down from 0.82 to 0.56 implying that over-

[47] This includes e.g. cleaners, refuse workers, caretakers as well as labourers in mining and agriculture.

proportionally many bad types contribute to its identification. Correcting for selection also decreases the restriction probability for the other part-time categories but hardly affects full-time categories. The overall restriction probability is thus over-estimated due to selection. The reason is rooted in involuntarily unemployed men being over-proportionally bad types. While employed individuals contribute to the identification of at most two hours categories, involuntarily unemployed may even contribute to all.

For women the restriction probabilities increase for all hours categories but small part-time when selection due to observed and unobserved heterogeneity is taken into account. The most prevalent changes are for large part-time and full-time where the restriction probability increases by almost ten percentage points. For women, selection thus results in under-estimating the overall restriction probability. The reason is that women out of work who do not contribute to the estimation have on average higher restriction probabilities than those who either work or desire positive hours.

Finally, we check the *in-sample fit of the model* for the men's and women's *distributions of actual working hours.* Expected choice probabilities from a standard labour supply model without constraints are taken as a benchmark. Our model fits the data reasonably well (Table 3.3). It performs slightly better than the standard model for men; both models perform equally for women.

Table 3.3: In-sample fit: observed and predicted hours distributions

	Men			**Women**		
Hours	Observed	Full	Standard	Observed	Full	Standard
0	0.081	0.013	0.037	0.491	0.466	0.415
1-14	0.009	0.014	0.043	0.128	0.235	0.259
15-24	0.011	0.035	0.055	0.156	0.125	0.141
25-35	0.071	0.094	0.097	0.139	0.078	0.084
36-40	0.415	0.340	0.258	0.060	0.053	0.054
>40	0.414	0.501	0.510	0.027	0.040	0.048

Notes: Observed – Observed actual hours, Standard – Discrete choice model based on actual hours, Full – Discrete choice model based on desired hours of work augmented by constraints (section 3.3).
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

For men, non-employment and full-time are clearly underestimated whereas part-time and over-time are over-predicted. For women, marginal employment is clearly over-estimated while both part-time choices with larger hours are under-predicted. That part-time categories tend to be over-predicted could be explained by (unobserved) fixed costs of working (Euwals and van Soest, 1999) that are not explicitly dealt with in the current specification. In our full model the high restriction probabilities for men in part-time categories work in that direction, though, which is why the fit is better than in the standard model.

3.4.2 Elasticities

Wage elasticities with respect to hours worked and participation are calculated numerically. Probabilities for choosing the different hours categories are calculated based on the parameter estimates and status quo incomes. Then, gross wage rates are increased by one percent for one spouse at a time, disposable incomes are re-simulated and choice probabilities re-estimated. Elasticities are inferred from the difference between probabilities in the counterfactual and the status quo. We consider changes in expected participation rates in percentage points (pp.) and in expected working hours in percent separately for women and men (table 3.4). Hours elasticities include the extensive and intensive margin.

Table 3.4: Own-wage elasticities of labour supply

	Men		**Women**	
Change working hours (%)				
Incentive	0.06	(0.04 ; 0.06)	0.34	(0.30 ; 0.37)
Standard model	0.17	(0.16 ; 0.19)	0.45	(0.42 ; 0.47)
Change participation rate (pp.)				
Incentive	0.00	(0.00 ; 0.00)	0.10	(0.09 ; 0.11)
Standard model	0.05	(0.05 ; 0.06)	0.13	(0.12 ; 0.14)

Notes: Incentive =Undistorted labour supply effects based on the preference part of our model (section 3.3.1), Standard=Discrete choice model based on actual hours, bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on FiD, wave 2010-2013.

The first row in each panel refers to the pure labour supply elasticities based on the preference part of our model with choice restrictions (sub-section 3.3.1). They are based on desired instead of actual working hours; restrictions are not relevant for the calculation of elasticities. Elasticities from the standard labour supply model based on actual working hours serve as a benchmark for comparison (second row in each panel, table 3.4).

Expectedly, elasticities of women are higher than those of men. This is true for both models. The level of elasticities is in the range of previous findings for Germany (Dearing et al., 2007; Steiner and Wrohlich, 2008; Müller and Wrohlich, 2015) given the focus of our sample on couple households with young children. The substantial difference between men and women is a well-established finding (Bargain et al., 2014; Bargain and Peichl, 2013). The difference can be explained by men's substantially higher employment rates and is mirrored in the changes in participation rates (which are virtually zero for men).

Elasticities from the labour supply model with hours constraints based on desired hours of work are smaller for both sexes compared to the standard model. This points to a moderate upward bias in the labour supply model that is estimated with actual working hours and confirms previous findings for Germany (Bargain et al., 2010; Haan and Uhlendorff, 2013).

3.4.3 Policy simulation: in-work benefit for parents

In this section we simulate the effects of introducing an in-work benefit for parents which aims at increasing incentives for full-time or large part-time work. We first present the design of the policy in detail and then discuss the simulation results.

3.4.3.1 Design of the in-work benefit

The German tax and transfer system provides strong incentives for an uneven share of market work between mothers and fathers, in particular through joint taxation of married spouses, subsidies to social security contributions for 'minijobs' (i.e. jobs with earnings up to 450 Euro per month) and free health insurance for non-working married spouses. Therefore, mothers' employment rates and average working hours for those employed are still significantly lower compared to fathers (section 3.2). This has far-reaching consequences for future career perspectives and old-age pensions of women. Furthermore, work incentives for parents with low wages and many children are generally low due to the social assistance scheme that is means-tested at the household level and withdrawn with increasing earnings, as well as child care costs.

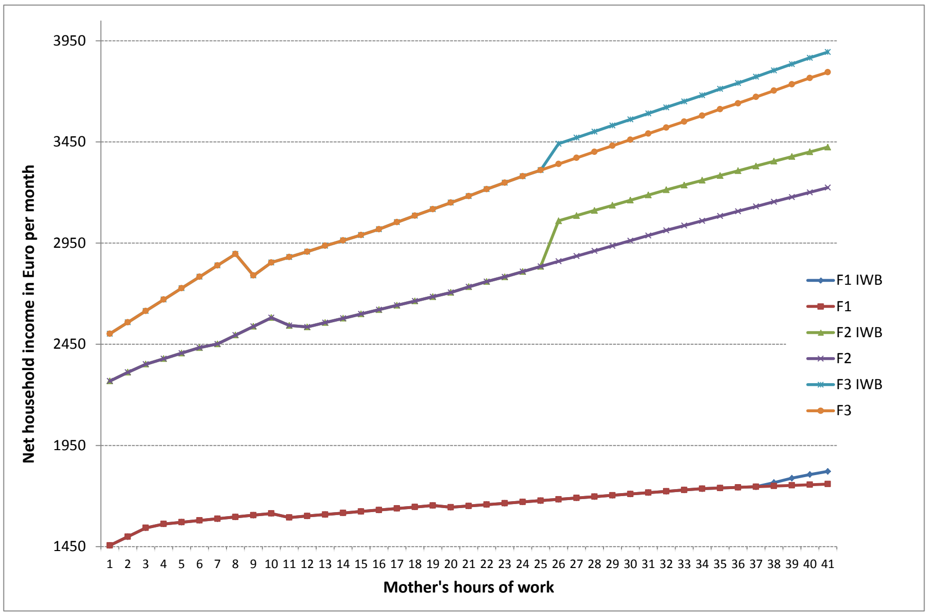
In this sub-section we present an in-work benefit for parents (IWBP) that is designed to tackle these two issues. It is conditional on the parents' hourly wage rate and on the number of children under 18 years. Each parent has an individual claim to this benefit, and it is not means tested at the household level. Eligibility requires at least 25 hours of work per week. The amount of the benefit is determined by the number of children and the individual's gross hourly wage rate. The basic amount is 1200 € per year and parent. This amount is increased by 600 € for every child. The benefit is withdrawn at a rate of 50% for each Euro the gross hourly wage rate exceeds a certain threshold that is 10 € for parents of one child and increases by 2 € for any further child. For a parent of one child this implies that the benefit is completely phased out for an hourly wage rate of 12 €[48].

Conditioning the in-work benefit on the individual hourly wage rate instead of the household income prevents adverse incentives, in particular for secondary earners with low wages. An increase in hours of work does not result in a decrease of the benefit and the partner's income

[48] For men in our sample, 10 € roughly refers to the tenth percentile of the wage distribution; 12 € refers to the 25th percentile. For women, 10 € is above the 25th percentile and 12 € is above the median of the distribution of hourly wages.

does not impact eligibility or the amount of the benefit. Although this has the disadvantage that some families might be subsidised who actually have a high income due to first earner's high income, it explicitly incentivises an increase of working hours for secondary earners with low wages[49]. The benefit is paid out tax-free but is deducted from the means-tested unemployment assistance (*Arbeitslosengeld II*).

Figure 3.1: Budget lines of different household types



Notes: F1 IWB: mother low wage (10 €), father not employed, IWB scenario; F1: mother low wage (10 €), father not employed, status quo scenario; F2 IWB: mother low wage (10 €), father full-time and low wage (14 €), IWB scenario; F2: mother low wage (10 €), father full-time and low wage (14 €), status quo scenario; F3 IWB: mother medium wage (13 €), father full-time and high wage (17 €), IWB scenario; F3: mother medium wage (13 €), father full-time and high wage (17 €), status quo.
Source: Own calculations.

Figure 3.1 plots budget lines for different exemplary households. Budget line F2 shows the example of a family with two children where the father (gross hourly wage 14 €) works full-time. The mother has an hourly wage of 10 € and is thus eligible for the program (whereas the father is not). The curve shows how family net income changes with working hours of the mother. Due to joint taxation, 40 hours of full-time work increase the family's income by only about 1,000 €. The in-work benefit for parents would make employment more attractive when the mother works at least 25 hours per week, as is shown by the curve F2_IWB. If the mother earns 13 € per hour, half of the benefit is withdrawn and the incentive to work 25 hours per week or more is considerably lower (see lines F3 and F3_IWB).

[49]For a discussion on the effects of individualised in-work credits versus family based in-work schemes, see Bargain and Orsini (2006).

Finally, F1 and F1_IWB show the case of a family (two adults and one child) with only one earner (mother or father). If the single earner has an hourly wage of 10 € per hour, the program has almost no effect on net earnings, since the in-work benefit for parents leads to a reduction of the means-tested unemployment assistance. Thus, for families where both parents have low hourly wages, this in-work benefit clearly provides incentives for both parents to work more than 25 hours instead of incentivizing the one-earner model.

3.4.3.2 Employment effects of introducing the in-work benefit

Table 3.5 shows the results of a policy simulation introducing the in-work benefit for parents based on our sample that is representative for families with children born between 2007 and 2010. Women's hours would increase by 4% and the participation rate would increase by 0.7 percentage points (Table 3.5, first column). The share of couples in which both spouses work more than 24 hours would increase by 1.4 percentage points. Starting from a base level of 16% in the status quo, this is an increase of almost 10%. For men, effects are negligible. This is not surprising as over 90% of men already work in one of the subsidised hours categories. That we do not observe negative effects for men can be ascribed to the program eligibility being based on the individual hourly wage rate impeding incentives to reduce hours of work.

Table 3.5: Employment effects: in-work benefit for parents

	Full	**Incentive**	**Standard**
Men			
Hours - change (%)	0.12	0.17	1.03
	(0.08 ; 0.20)	(0.12 ; 0.20)	(0.93 ; 1.16)
Part. - base (%/100)	0.99	0.99	0.96
	(0.98 ; 0.99)	(0.99 ; 0.99)	(0.96 ; 0.97)
Part. - change (pp.)	0.04	0.03	0.33
	(0.04 ; 0.05)	(0.02 ; 0.03)	(0.30 ; 0.37)
Women			
Hours - change (%)	4.09	3.66	4.74
	(3.63 ; 4.46)	(3.23 ; 4.05)	(4.42 ; 4.99)
Part. - base (%/100)	0.55	0.66	0.59
	(0.54 ; 0.56)	(0.65 ; 0.66)	(0.57 ; 0.59)
Part. - change (pp.)	0.70	0.73	1.02
	(0.62 ; 0.77)	(0.65 ; 0.81)	(0.95 ; 1.07)
Both spouses' hours > 24			
base (%/100)	0.16	0.19	0.16
	(0.15 ; 0.17)	(0.19 ; 0.20)	(0.15 ; 0.17)
change (pp.)	1.41	1.80	1.63
	(1.27 ; 1.53)	(1.66 ; 1.99)	(1.49 ; 1.73)

Notes: Full=Discrete choice model based on desired hours of work augmented by constraints (section 3.3), Incentive =Undistorted labour supply effects based on the preference part of our model, Standard=Discrete choice model based on actual hours, bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

The second column of Table 3.5 displays the pure labour supply effect based on the preference part of our model. In this case, the point estimate of the increase of the share of couples in which both spouses work more than 24 hours is significantly higher (1.8 pp.) than in the full model. This suggests that the pure labour supply response to the introduced financial incentives is reduced by approximately 10% due to restrictions hampering individuals from drawing on the benefit. Interestingly, hours responses of women are smaller when restrictions are not considered. This is rooted in a higher restriction probability for full-time work vis--vis large part-time. When complying women can choose freely, they rather prefer extending their labour supply to at most large part-time. When restrictions are taken into account this category is often not feasible, though. Our results suggest that in those cases many women accept taking up a less restricted full-time position in order to draw on the benefit.

Employing the standard labour supply model based on actual hours (third column) the estimated increase of the share of couples where both spouses work in the targeted hours categories lies in between the pure labour supply effect and the effect of the full model. The difference to the latter is only statistically significant at the ten percent level, though. It nevertheless suggests, first, that the standard model might over-estimate employment effects. Second, the standard model seems to implicitly incorporate some restrictions by using actual hours as revealed preferences.

Moreover, the standard model predicts much higher labour supply reactions for men, namely an increase in hours worked by 1% and an increase in the participation rate by 0.33 percentage points. For women, point estimates for the change in hours worked and in the participation rate are also higher based on the standard than on the full model.[50] This might among others be rooted in the higher elasticities estimated for the standard model (section 3.4.2).

Table 3.6 presents behavioural effects of the in-work benefit (based on the full model) for different subgroups. While we do not find differences in the behavioural reactions of fathers in east and west Germany, mothers in east Germany show stronger reactions: The participation rate of mothers in east Germany would increase by 1.3 percentage points, whereas this effect would only amount to 0.6 percentage points in west Germany, although the baseline participation rate in east Germany is much higher (69%) than in west Germany (53%). The same is true for working hours (5.3% vs. 3.9%). The share of couples where both spouses work more than 24 hours would increase by almost 3 percentage points in east, but only by 1.2 percentage points in west Germany. The main reason why we find so much stronger effects in east Germany is that wages are still lower in east than in west Germany, and thus the share of individuals eligible for the in-work benefit is higher.

Differences in the wage distribution obviously also explain varying effects across education groups. As expected, we find the highest behavioural effects for parents with low education. For mothers we also find considerable labour supply effects in the medium education group

[50] However, the difference in the change of working hours is not statistically significant, as the confidence intervals overlap.

Table 3.6: Effect heterogeneity: in-work benefit for parents

	Region			**Education**	
	West	**East**	**Low**	**Medium**	**High**
Men					
Hours - change (%)	0.12	0.11	0.46	0.10	0.01
	(0.07 ; 0.16)	(0.04 ; 0.19)	(0.27 ; 0.56)	(0.06 ; 0.14)	0.00 ; 0.03)
Part. - base (%/100)	0.99	0.99	0.98	0.99	0.99
	(0.98 ; 0.99)	(0.98 ; 0.99)	(0.97 ; 0.98)	(0.99 ; 0.99)	0.99 ; 0.99)
Part. - change (pp.)	0.04	0.08	0.14	0.04	0.01
	(0.02 ; 0.05)	(0.06 ; 0.10)	(0.09 ; 0.18)	(0.03 ; 0.05)	0.01 ; 0.01)
Women					
Hours - change (%)	3.91	5.33	8.06	4.54	1.77
	(3.53 ; 4.33)	(4.37 ; 5.97)	(6.94 ; 9.71)	(4.10 ; 5.05)	1.59 ; 1.97)
Part. - base (%/100)	0.53	0.69	0.41	0.55	0.60
	(0.52 ; 0.54)	(0.67 ; 0.70)	(0.39 ; 0.43)	(0.54 ; 0.56)	0.59 ; 0.61)
Part. - change (pp.)	0.60	1.31	1.17	0.79	0.36
	(0.54 ; 0.69)	(1.11 ; 1.50)	(1.02 ; 1.35)	(0.70 ; 0.88)	0.32 ; 0.39)
Both spouses' hours > 24					
base (%/100)	0.13	0.35	0.07	0.15	0.21
	(0.12 ; 0.14)	(0.33 ; 0.37)	(0.07 ; 0.08)	(0.14 ; 0.16)	0.20 ; 0.22)
change (pp.)	1.19	2.89	1.84	1.72	0.68
	(1.09 ; 1.32)	(2.64 ; 3.07)	(1.65 ; 2.09)	(1.60 ; 1.92)	0.61 ; 0.75)

Notes: All effects based on the full model, a discrete choice model based on desired hours of work augmented by constraints (section 3.3), Low education: ISCED level 0-2 (at most *Mittlere Reife*, no vocational training), Medium education: ISCED level 3-4 (A-levels or vocational training), High education: ISCED level 5-6 ((Applied) University degree and higher), bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

(Table 3.6).

3.4.4 The role of hours constraints

The full model that we estimate allows to distinguish the effects of financial incentives and constraints on labour supply behavior. In the sub-section above, we showed how financial incentives introduced by the IWBP affect labour supply behavior of parents and how restrictions impact the policy effects. In this sub-section, we simulate a removal of all constraints in order to show how labour supply would change in this case. Note, however, that this simulation cannot be interpreted as a policy reform since the reasons for the restrictions are numerous (for example, demand side restrictions such as unemployment, social norms regarding part-time employment of men, or insufficient child care). For the scenario in which all restrictions are removed, i.e. all individuals can actualise their desired working hours on the labour market, we find strong employment effects for women: The participation rate of women would increase by more than 10 percentage points and their average hours worked would increase by more than 50% (see Table 3.7, first column).

Note that our sample is only representative for families with children born between 2007 and 2010. The results thus reveal the strong labour market restrictions leading to substantial under-employment for this group and are not representative for the whole population. The effect on working hours for men in our sample goes in the opposite direction than for women: In case that all restrictions were removed, men would decrease their working hours by about 1%. This

Table 3.7: Employment effects: (partly) removing constraints

	All	**Unemployment**	**Childcare**
Men			
Hours - change (%)	-1.31	1.09	-0.25
	(-1.63 ; -0.58)	(0.78 ; 1.59)	(-4.57 ; 2.33)
Part. - base (%/100)	0.99	0.99	0.99
	(0.98 ; 0.99)	(0.98 ; 0.99)	(0.98 ; 0.99)
Part. - change (pp.)	0.44	0.18	0.16
	(0.36 ; 0.57)	(0.13 ; 0.24)	(-0.34 ; 0.28)
Women			
Hours - change (%)	51.98	9.76	27.87
	(46.25 ; 57.09)	(6.92 ; 13.11)	(17.83 ; 37.90)
Part. - base (%/100)	0.55	0.55	0.55
	(0.54 ; 0.56)	(0.54 ; 0.56)	(0.54 ; 0.56)
Part. - change (pp.)	10.54	1.88	2.25
	(9.84 ; 11.10)	(0.90 ; 2.69)	(-5.68 ; 8.35)
Both spouses' hours > 24			
base (%/100)	0.16	0.16	0.16
	(0.15 ; 0.17)	(0.15 ; 0.17)	(0.15 ; 0.17)
change (pp.)	3.51	0.95	1.74
	(2.95 ; 4.00)	(0.48 ; 1.27)	(-0.59 ; 6.35)

Notes: All: All restrictions are removed, *Unemployment*: Unemployment is reduced to the lowest observed value of all counties (2.1 %), *Childcare*: Full-time childcare coverage is increased to the highest observed value of all counties (84.5 %), All effects based on the full model, a discrete choice model based on desired hours of work augmented by constraints (section 3.3), bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

reflects the fact that men are more likely to be restricted with respect to part-time jobs than full-time jobs resulting in over-employment. However, the participation rate would increase by about 0.5 percentage points, reflecting a low predicted rate of involuntary unemployment for this group.[51]

Differentiating the behavioural effects between east and west Germany, we find that men are much more likely to be over-employed in west than in east Germany. If all restrictions were removed, average working hours of men in our sample would decrease by almost 1.5 percentage points in west, but only by 0.15 percentage points in east Germany (Table 3.8). For women, we find only small differences in hours and participation effects, albeit restrictions seem to deter west German more than east German women from working at least 24 hours. The difference is not statistically significant, though.

However, differentiating by education level, we find very strong differences: Mothers with low education seem to be particularly affected by under-employment. If all constraints were removed, mothers with low education would increase their working hours by almost 150%. Mothers with medium and high levels of education face considerably lower effects (42% and 35%, respectively). These effects are among others driven by changes in participation which shows a similar pattern. Similar results can be found for fathers. While low income fathers would increase average hours (by more than 5%) and their participation rate (by 1.3 percentage points), fathers with

[51] Our model (as well as the standard model) considerably underestimates non-employment for men (section 3.4.1). The predicted participation rate is almost 99 % (vs. 92 % in the observed data) leaving not much scope for participation effects.

Table 3.8: Effect heterogeneity: Remove all constraints

	Region		**Education**		
	West	**East**	**Low**	**Medium**	**High**
Men					
Hours - change (%)	-1.48	-0.15	5.37	-2.01	-3.02
	(-1.81 ; -0.59)	(-1.69 ; 1.12)	(3.28 ; 8.70)	(-2.26 ; -1.30)	(-3.22 ; -2.44)
Part. - base (%/100)	0.99	0.99	0.98	0.99	0.99
	(0.98 ; 0.99)	(0.98 ; 0.99)	(0.97 ; 0.98)	(0.99 ; 0.99)	(0.99 ; 0.99)
Part. - change (pp.)	0.44	0.48	1.34	0.33	0.24
	(0.36 ; 0.57)	(0.40 ; 0.66)	(1.02 ; 1.60)	(0.27 ; 0.45)	(0.19 ; 0.47)
Women					
Hours - change (%)	51.95	52.20	143.45	42.13	34.49
	(46.24 ; 57.53)	(39.92 ; 60.51)	(119.55 ; 172.29)	(37.58 ; 46.23)	(30.37 ; 37.50)
Part. - base (%/100)	0.53	0.69	0.41	0.55	0.60
	(0.52 ; 0.54)	(0.67 ; 0.70)	(0.39 ; 0.43)	(0.54 ; 0.56)	(0.59 ; 0.61)
Part. - change (pp.)	10.30	12.15	16.26	9.92	9.43
	(9.67 ; 11.00)	(10.90 ; 12.88)	(14.34 ; 17.75)	(9.10 ; 10.52)	(8.75 ; 9.95)
Both spouses' hours > 24					
base (%/100)	0.13	0.35	0.07	0.15	0.21
	(0.12 ; 0.14)	(0.33 ; 0.37)	(0.07 ; 0.08)	(0.14 ; 0.16)	(0.20 ; 0.22)
change (pp.)	3.67	2.38	4.09	3.70	2.94
	(3.19 ; 4.14)	(0.98 ; 3.63)	(3.33 ; 4.92)	(3.11 ; 4.24)	(2.42 ; 3.45)

Notes: All effects based on the full model, a discrete choice model based on desired hours of work augmented by constraints (section 3.3), Low education: ISCED level 0-2 (at most *Mittlere Reife*, no vocational training), Medium education: ISCED level 3-4 (A-levels or vocational training), High education: ISCED level 5-6 ((Applied) University degree and higher), bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

medium or high education would reduce average working hours by 2% and 3%, respectively. This hints at different restrictions fathers are facing on the labour market. While involuntary unemployment is mostly prevalent for fathers with low education, fathers with medium or high education levels rather suffer from over-employment.

Since the removal of all possible restrictions on the labour market is hard to interpret, we conduct two additional simulations that have a more straight-forward economic interpretation. First, we simulate a low level of unemployment in all regions of Germany. In this scenario, we set the unemployment rate for all counties to the value of the lowest local unemployment rate observed in our sample (2.1%). Note that the average local unemployment rate in our sample is 7.16%. Thus, the average "treatment" means a decrease of the local unemployment rate by more than 5 percentage points. The treatment, however, is asymmetrically distributed across Germany; the average simulated decrease of the unemployment rate in west Germany refers to 4.24, in east Germany to 9.69 percentage points. In this scenario, men would increase their participation rate by 0.2 and women by almost 2 percentage points. Average hours worked would increase by 1% for men and almost 10% for women (Table 3.7, second column).

As expected, behavioural effects in this scenario differ greatly between east and west Germany and by education groups. Fathers as well as mothers in east Germany would increase their labour supply to a much larger extent than in west Germany (Table 3.9). Moreover, unemployment affects workers with low education much more than workers with medium or high education levels. Fathers with low education would increase their average hours by almost 5%, whereas for fathers with medium or high education we find very low effects. Women with low education would increase average working hours by more than 20%, those with medium or high levels of education by almost 9 and 7%, respectively. The change in the participation rate, however, is similar for women from all three groups.

Table 3.9: Effect heterogeneity: Reduce unemployment

	Region		Education		
	West	East	Low	Medium	High
Men					
Hours - change (%)	0.82	2.95	4.83	0.72	0.11
	(0.54 ; 1.35)	(2.14 ; 3.64)	(3.63 ; 6.99)	(0.54 ; 1.03)	(-0.02 ; 0.34)
Part. - base (%/100)	0.99	0.99	0.98	0.99	0.99
	(0.98 ; 0.99)	(0.98 ; 0.99)	(0.97 ; 0.98)	(0.99 ; 0.99)	(0.99 ; 0.99)
Part. - change (pp.)	0.17	0.20	0.61	0.13	0.08
	(0.13 ; 0.23)	(0.15 ; 0.29)	(0.44 ; 0.78)	(0.10 ; 0.17)	(0.06 ; 0.15)
Women					
Hours - change (%)	8.48	18.40	21.89	8.67	7.06
	(6.08 ; 11.78)	(13.62 ; 22.06)	(15.38 ; 29.82)	(6.26 ; 11.33)	(4.86 ; 9.55)
Part. - base (%/100)	0.53	0.69	0.41	0.55	0.60
	(0.52 ; 0.54)	(0.67 ; 0.70)	(0.39 ; 0.43)	(0.54 ; 0.56)	(0.59 ; 0.61)
Part. - change (pp.)	1.67	3.32	1.70	1.86	1.99
	(0.74 ; 2.39)	(2.02 ; 4.64)	(0.21 ; 2.99)	(0.95 ; 2.63)	(1.10 ; 2.67)
Both spouses' hours > 24					
base (%/100)	0.13	0.35	0.07	0.15	0.21
	(0.12 ; 0.14)	(0.33 ; 0.37)	(0.07 ; 0.08)	(0.14 ; 0.16)	(0.20 ; 0.22)
change (pp.)	0.76	2.20	1.09	0.99	0.81
	(0.43 ; 1.06)	(0.79 ; 3.44)	(0.67 ; 1.46)	(0.55 ; 1.30)	(0.26 ; 1.24)

Notes: All effects based on the full model, a discrete choice model based on desired hours of work augmented by constraints (section 3.3), Treatment is reduction of unemployment to the lowest observed value of all counties (2.1 %), Low education: ISCED level 0-2 (at most *Mittlere Reife*, no vocational training), Medium education: ISCED level 3-4 (A-levels or vocational training), High education: ISCED level 5-6 ((Applied) University degree and higher), bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

Second, we simulate a scenario in which we dramatically decrease employment restrictions due to lack of child care. In this scenario we set the local availability of full time child care slots for children under three to the highest level observed in our sample, which is 84.5. Since the average rate amounts to 11%, the treatment simulated in this scenario is very large and – as in the scenario above – very unequally distributed throughout Germany. In west Germany, the average increase of full time child care coverage is much larger (+77.62 percentage points) than in east Germany (+50.19 percentage points).

As expected, employment effects of men are negligible. However, employment effects of women are much larger in this scenario than in the scenario that decreases the local unemployment rate. If child care was dramatically increased, women's participation rate would increase by 2.3 percentage points and total hours would increase by almost 28% (Table 3.7). The latter is roughly half the size of the hours response for removing all constraints; child care restrictions thus seem to play a crucial role for under-employment of employed women. For participation constraints child care rationing is less relevant. As can be seen from Table 3.10, we find similar behavioural effects for mothers in east and west Germany as far as average hours are concerned.

However, the participation rate would increase much stronger in east (4.5 percentage points) than in west Germany (almost 2 percentage points). Mothers with low education show the strongest reaction in this scenario: their average working hours would increase by 70% and their participation rate would increase by 3.4 percentage points. Mothers with medium or high education would increase average working hours by about 20% and the participation rate by 2 percentage points.

Table 3.10: Effect heterogeneity: Expand formal childcare

	Region		Education		
	West	**East**	**Low**	**Medium**	**High**
Men					
Hours - change (%)	-0.36	0.46	1.69	-0.42	-0.82
	(-4.81 ; 2.18)	(-2.14 ; 3.04)	(-6.08 ; 7.43)	(-4.62 ; 1.92)	(-3.77 ; 0.81)
Part. - base (%/100)	0.99	0.99	0.98	0.99	0.99
	(0.98 ; 0.99)	(0.98 ; 0.99)	(0.97 ; 0.98)	(0.99 ; 0.99)	(0.99 ; 0.99)
Part. - change (pp.)	0.17	0.06	0.63	0.09	0.07
	(-0.35 ; 0.30)	(-0.48 ; 0.21)	(-0.59 ; 0.99)	(-0.33 ; 0.19)	(-0.25 ; 0.17)
Women					
Hours - change (%)	28.09	26.39	69.25	23.60	19.63
	(17.97 ; 39.28)	(17.35 ; 37.88)	(37.20 ; 104.64)	(16.53 ; 32.62)	(12.71 ; 25.85)
Part. - base (%/100)	0.53	0.69	0.41	0.55	0.60
	(0.52 ; 0.54)	(0.67 ; 0.70)	(0.39 ; 0.43)	(0.54 ; 0.56)	(0.59 ; 0.61)
Part. - change (pp.)	1.91	4.54	3.44	2.10	2.05
	(-6.60 ; 8.56)	(0.52 ; 8.03)	(-5.48 ; 14.61)	(-6.46 ; 7.67)	(-5.02 ; 6.95)
Both spouses' hours > 24					
base (%/100)	0.13	0.35	0.07	0.15	0.21
	(0.12 ; 0.14)	(0.33 ; 0.37)	(0.07 ; 0.08)	(0.14 ; 0.16)	(0.20 ; 0.22)
change (pp.)	1.90	0.67	1.97	1.89	1.39
	(-0.14 ; 6.45)	(-3.96 ; 5.79)	(-0.09 ; 5.11)	(-0.36 ; 6.65)	(-1.30 ; 6.30)

Notes: All effects based on the full model, a discrete choice model based on desired hours of work augmented by constraints (section 3.3), Treatment is increase of full-time childcare coverage to the highest observed value of all counties (84.5 %), Low education: ISCED level 0-2 (at most *Mittlere Reife*, no vocational training), Medium education: ISCED level 3-4 (A-levels or vocational training), High education: ISCED level 5-6 ((Applied) University degree and higher), bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

From these simulations we conclude that restrictions on the labour markets such as involuntary unemployment or lack of childcare seem to play an important role, in particular for mothers in general and fathers with low education. Thereby, childcare rationing contributes crucially to under-employment of employed mothers and much less to participation constraints.

3.5 Discussion and conclusions

In this paper we have specified a static labour supply model based on desired hours of work. The model also includes participation and working hours constraints. We exploit sample information on search activities and actual as well as desired hours of work for the estimation. The framework is designed for the evaluation of policies where not only economic incentives but also constraints on the labour market determine employment outcomes. We illustrate this by applying the model to an in-work benefit for low-paid parents that targets work incentives for secondary earners. Eligibility is conditional on hours of work exceeding a certain threshold.

Restrictions might thus hamper individuals drawing on the benefit.

The in-work benefit increases with the number of children in the household and is based on an individual's hourly wage and conditioned on a minimum number of 25 hours of work per week. Simulation results based on our labour supply model with restrictions show that while behavioural effects of fathers are negligible, such a policy would affect labour supply of mothers: their participation rate would increase by 0.7 percentage points and their average working hours would increase by four percent. Mothers with low education would react most strongly, their average hours of work would increase by eight percent. A standard labour supply model on actual working hours without constraints predicts higher behavioural effects for all groups, in particular for fathers. The comparison shows that taking actual hours as revealed preferences and ignoring restrictions on the labour market may yield upward-biased estimates. This affects policy conclusions as the potential of improving labour supply incentives through an in-work benefit would clearly be overstated by the standard model in our application.

Based on our full model we are able to directly simulate the effects restrictions have on actual employment of individuals. In a first exercise all restrictions are removed completely to illustrate their impact on actual employment. Working hours of mothers would increase by 50% and their participation rate by 10%. Mothers with low education face the strongest restrictions on the labour market: their participation rate would increase by 16 percentage points and their working hours almost 150% when constraints could be removed completely. This exercise also reveals the heterogenous impact of restrictions along other dimensions: men and women are confronted with very different constraints on the labour market. Besides being overall restricted to a lesser degree fathers face the largest constraints in part-time employment whereas mothers are restricted in jobs of all hours categories, but in particular in full-time jobs.

A complete removal of all types of restrictions at the same time is hard to interpret economically. We therefore conducted two further simulations that either reduce labour demand restrictions (lowering the regional unemployment rate) or constraints on the childcare market (increasing childcare coverage). Under both scenarios participation and average hours worked of mothers would increase substantially. Fathers with lower productivity would also profit from an increase in labour demand. These simulations also illustrate, that different mechanisms on the demand and supply side of the labour market may prevent parents from realizing their desired employment levels. These different types of constraints are distributed unequally across regions and socio-demographic groups.

The application of our empirical framework to the evaluation of an in-work benefit underlines the importance of understanding different mechanisms on the labour market to gauge the potential and improve the design of labour market or family policies. In this paper we took another step towards identifying preferences and constraints in the household context based on rich survey information. We introduced observed and unobserved heterogeneity in preferences and constraints. This brought us closer to determine the impact of various types of constraints that affect different groups on the labour market in various ways. In terms of policy conclusions we showed that the incentive effect of an in-work benefit is reduced by restrictions employees face when making employment decisions. At the same time this result highlights other areas for intervention, e.g. alleviating constraints through regulatory or family policies.

Chapter 4

Firms' responses to aggregate preferences of workers - Evidence from a German Making Work Pay policy*

4.1 Introduction

It is largely recognised that workers face important constraints on the labour market in choosing labour supply (e.g. Stewart and Swaffield (1997a)). Employed workers cannot easily change hours, and unemployed workers might not be able to find their preferred job. This has important implications for tax policy design as constraints limit workers' ability to make optimal choices (Chetty et al. (2011)).

Many researches have, therefore, introduced hours constraints into labour supply models (e.g. Van Soest et al. (1990), Bloemen (2008), Beffy et al. (2016) and chapter three of this dissertation). Job search models provide a way of explaining some of these constraints without resorting to ad-hoc restrictions on the set of labour supply choices available to individuals. Constraints result from informational deficiencies - individuals are only aware of specific job offers and firms' vacancies do not reach all potential workers.

In equilibrium these frictions can result in additional hours constraints imposed by firms (Chetty et al., 2011): If firms cannot easily change the hours of a job offer it has been posted, they have an incentive to package their hours-wage bundles according to average preferences of job-seeking workers. The restricted choice set of any particular individual worker will depend on other workers' preferences. This generates mismatches on the labour market which might have

*This chapter is based on unpublished joint work with Luke Haywood within the project *The Impact of Social Security Contributions on Earnings: Evidence from administrative data in France, Germany, the Netherlands and the UK* sponsored by the *Deutsche Forschungsgemeinschaft*.

negative welfare consequences. We term this in the following *population labour supply effect.* Population labour supply effects represent a channel through which firms mediate responses to tax policies - even when labour costs are not directly affected. Thereby, individual responses to tax incentives are reinforced. Tax incentives for one group of workers might further have unintended effects to other groups of workers. (Chetty et al., 2011)

We present direct evidence of a population labour supply effect induced by non-linear tax reductions for specific workers in the German tax system. We use this specific institutional setting to estimate a simple equilibrium job search model which accounts for population labour supply effects and allows for a discontinuous tax schedule. Our structural model is then used for counterfactual policy simulations. We thereby focus on the indirect costs of (discontinuous) tax policies.

More specifically, wage earnings below a certain threshold are exempted from social security contributions (SSC) and income taxes. These so-called Minijobs stand in the tradition of other *Making Work Pay* policies like the Earned Income Tax Credit in the US or the Working Family Tax Credit in the UK - albeit Minijob eligibility is based on individual earnings and subsidies are less generous (Bargain and Orsini, 2006). Earnings exceeding the Minijob threshold are subject to full SSC as well as income taxes. The observed earnings distribution features huge bunching at the threshold. This illustrates the negative earnings responses of already employed workers which is a typical feature of MWP policies (Bargain et al., 2010). Recent research underlines the firms' role in responses to tax reforms in general (Kopczuk and Slemrod, 2006) and bunching in particular (Chetty et al., 2011; Best, 2014; Gudgeon and Trenkle, 2016). The population labour supply effect is one channel. Part of the bunching arises from firms posting discontinuously many job offers at the threshold in order to attract more potential workers. Individuals seeking to earn more than in Minijobs are then hampered by the reduced availability of higher-paying jobs.

This study complements work by Tazhitdinova (2015) who proposes lower wage rates and fringe benefits for Minijob workers as a channel for firms to reinforce bunching at the German Minijob threshold. Bunching of individuals who had additional earnings from another job, however, serves as compelling evidence in favour of a (supplemental) population labour supply effect. Before 2003, second job earnings had to be taxed jointly with first job earnings. There was thus no incentive for second job earnings to bunch. Further, Tazhitdinova (2015) does not bring her model to the data.

We estimate our model using a combination of German administrative and household survey data. This enables us to simulate counterfactual reforms which smooth the tax schedule for low earnings. We find that the tax and SSC exemption for low earnings provoke substantial distortions which can partly be prevented by introducing a smooth tax schedule.

We make two main contributions. First, we include discontinuous taxes into a standard equilibrium job search model. Whilst other authors include taxes in equilibrium job search models (e.g. Shephard (2012)) taxes there are assumed to be continuous. In line with the bunching literature (Saez, 2010), we believe important additional welfare costs are caused by the discontinuous nature of taxes. These might be further increased by firm responses (Tazhitdinova, 2015). In contrast to most equilibrium job search models, we are able to rationalise discontinuous bunching at tax thresholds in the interior of the earnings distribution. We are thus able to analyse counterfactual policy reforms that aim at removing the welfare costs induced by tax non-linearities.

Second, our unique institutional setting offers a compelling empirical setting to study what we term population labour supply effects. Tax subsidises offered to certain workers with low earnings end up affecting all workers in the labour market. We estimate our equilibrium job search model and analyse these distorting effects. This is relevant beyond the specific German setting and applies to many tax subsidies for low-paid jobs. We thus contribute to the body of evidence on *Making Work Pay* (MWP) policies (e.g. Blundell (2000) for the US, Bargain and Orsini (2006) for Europe). We show that firms can reinforce the negative earnings responses of already employed workers which are found when evaluating MWP policies (Bargain et al., 2010). We particularly complement other work which discusses how unintended consequences of MWP policies unfold in equilibrium (Rothstein, 2010; Andrew, 2010; Kolm and Tonin, 2011; Tazhitdinova, 2015). While these studies focus on channels like incidence or displacement effects we add the population labour supply as additional channel.

In section (4.2) we briefly discuss the market for low-paid employment in Germany. Section (4.3) presents our model and simulation results. In section (4.4) we present our data and some descriptive statistics supporting our basic modelling choices. Section (4.5) discusses identification and estimation. Section (4.6) presents estimation results and an application of the model. Section (4.7) concludes.

4.2 Minijobs in Germany

This study focuses on regular low-paid employment[52]. Since certain jobs in this sector are subject to special tax and social security treatment, it is important to clarify the institutional setting. In 1999 and 2003, two major reforms to the regulations of *low-paid* employment were introduced. This study and the following description of the institutional setting focuses on the time period between both reforms.

Employment contracts qualify for special tax and SSC treatment if yearly earnings do not exceed a threshold of 3900 €[53] and come with fewer than 15 hours per week. These employment relations are designated as Minijobs. The most important feature is that Minijob earnings are not subject to SSC for employees. This constitutes significant savings as employees' SSC amounted to approximately 21 % of earnings. Crossing the threshold results in a decrease of net earnings because earnings above the threshold are entirely subject to SSC. This implies that the budget set features a notch[54]. Minijob earnings do not entitle a worker to social security. Most individuals with low labour income are however either health-insured through their partners, the state or have access to cheap student contracts. This is not the case for pension and unemployment insurance. Every individual is eligible for social assistance, though. If Minijob earnings qualified for unemployment or pension insurance, contributions would not suffice to increase benefits above the social assistance level. We thus treat the SSC exemption as a tax exemption.

Employers' SSC are 10% (health) and 12% (pension) of Minijob earnings and thus approximately correspond to employers' SSC for other employment contracts. Minijobs are also subject to a specific treatment of income taxation. If there was no other individual income, it was possible to request a complete exemption from income taxation[55]. The size of the drop in net earnings is thus composed of SSC and income taxation[56]. Its size varies across individuals[57]. Income tax rates depend among others on household income including for example a partner's

[52]By contrast to the OECD's definition, we use *low-paid* employment to describe jobs with low earnings. In our sense this covers all employment relations that yield regular earnings below 800 € per month. We exclude short-term employment consisting of jobs that last for less than a certain number of days per year or calender year as they are subject to different rules.

[53]The threshold is actually defined in terms of monthly earnings. The monthly threshold can, however, be exceeded for two months within a year as long as yearly earnings are below the yearly threshold.

[54]In March 2002, a policy called *Mainzer Modell* aimed at improving incentives by balancing the additional SSC due to crossing the threshold by a subsidy which was gradually withdrawn for higher household earnings. The notch with respect to income taxation remained. We nevertheless limit the empirical analysis to the period before March 2002 (section 4.4).

[55]Spousal income is not relevant.

[56]For the sake of readability we speak of tax exemption instead of also mentioning SSC.

[57]Until 2006 SSC rates for health insurance varied between health insurance companies and thus also between individuals. In 2001, for example, it ranged from 11.0 % to 14.9 % Grabka (2004). Income taxes are, thus, not the only reason for the notch varying at the individual level.

labour income. Since the general income tax allowance was above the Minijob threshold, the income tax rate was zero for a significant fraction of workers (e.g. singles without other income). It could become very substantial for married individuals, though.

Important for this study, a side job in addition to a main job subject to SSC could not qualify as a Minijob and was, therefore, subject to full SSC and income taxation. There is, thus, no incentive to bunch at the threshold for side job earnings. Holding two Minijobs simultaneously was only possible when joint income did not exceed the threshold.

4.3 The Model

Population labour supply effects are constraints on the labour supply choice set imposed by firms in equilibrium. They arise when firms cannot easily adjust earnings attached to their job offers. Firms then have an incentive to tailor their job offers to aggregated preferences of potential workers. To model population labour supply effects an equilibrium search model with wage posting is the most appropriate. Job search models provide a way of explaining constraints without resorting to ad-hoc restrictions on the set of labour supply choices available to individuals. Constraints result from informational deficiencies - individuals are only aware of specific job offers and firms' vacancies do not reach all potential workers. By contrast to simple search models, equilibrium models endogenise the job offer distribution. If earnings were totally flexible (in a simple bargaining framework), we should still expect no bunching at thresholds that do not apply to individuals personally. The data suggests otherwise, leading us to assume that firms make take-it-or-leave-it offers along the lines of Burdett and Mortensen (1998) and Bontemps et al. (1999)[58]. The endogenous job offer distribution then reflects the preference distribution of potential workers.

As we are mainly interested in the responses to the special German tax treatment of Minijobs, we only model the market of low-paid jobs and assume that these jobs form a separate labour market. Section 4.4.3 shows that the industry sector distributions of Minijobs and full-time employments are very different and that the former are mainly concentrated in economic services and retail. The focus on the low-paid market implies that first jobs of second job holders are not modeled but treated as constant and exogenous.

In the remainder of this section we present our model in detail. After setting the stage we first present a stylised model based on the simplifying assumption of homogeneous hours. We rationalise analytically bunching in the interior of the earnings distribution (section 4.3.2). We

[58]An alternative could be to allow for negotiations for a fraction of workers along the lines of Chetty et al. (2011). In our setting focussing on low-paid employees the assumption of take-it-or-leave-it offers arguably is credible, though (Shephard, 2012)

then resort to numerical solutions to allow for heterogeneous hours and to relax assumptions on taxes (section 4.3.3). This more realistic version of the model is brought to the data in section 4.5.

4.3.1 A simple equilibrium job search model

The labour market is composed of a continuum of workers and firms. Some workers have another job but are nonetheless active in the market for low-paid jobs seeking a second job. We refer to these as *type s* workers, of which there are n^s. These workers are not qualified for the tax exemption, which is available for workers seeking a main job in the low-paid market, i.e. *type f* workers. Thus the budget set of type f workers exhibits a notch at gross earnings $z = z^*$ while it is smooth for type s workers.
Net earnings c for workers of typy j are calculated by[59]

$$c = \begin{cases} z & \forall z \leq z^* \ \& \ j = f \\ z(1 - t^{ssc})(1 - t^{inc}) & \forall z > z^* \ \& \ j = f \\ z(1 - t^{ssc})(1 - t^{inc}) & \forall j = s \end{cases} \tag{4.1}$$

where t^{ssc} denotes the total social security contribution rate and t^{inc} the income tax rate. We assume utility v be generated by $v = c^{\alpha}(\bar{h} - h)^{(1-\alpha)}$, with $\bar{h} - h$ denoting leisure and α the elasticities of v with respect to net earnings and leisure. $\bar{h}$ is total available hours which we set to 12 hours a day.
Workers search when unemployed and when employed with uniform search intensity by drawing from a known job offer distribution $F(.)$. The exogenous job offer arrival rate (λ) is allowed to differ between type f and type s workers but is independent of employment status. Workers lose their job at an exogenous rate δ which also differs between worker types. Workers seek to maximise the expected steady-state future utility v. The strategy of workers is to accept every job with utility exceeding homogeneous reservation utility v^r.
Firms seek to maximise profits $\pi = [p - w]hl(v)$ with p being turnover per working hour, $w = \frac{z}{h}$ the gross hourly wage rate and $l(v)$ the size of a firm's labour force which offers jobs with utility v (we assume that within firms job offers are identical). In the next sub-section we present a stylised version of the model which allows for analytical solutions. Thereafter, the full model is presented and brought to the data.

[59] We abstract from an individual index.

4.3.2 Stylised model

The objective of this section is to derive how a notch in the budget set of one group of workers affects the earnings distribution of all workers. We thereby rationalise analytically bunching in the interior of the earnings distribution. Two simplifying assumptions are introduced that allow for analytical solutions. Both are relaxed in the next section. First, hours are homogeneous. This implies that utility increases monotonously in net earnings. Second, tax and SSC rates sum to 100 % for type f and zero for type s workers. Type f workers will thus only accept job offers with earnings below or at the threshold. This group of workers generates the incentive for firms to over-proportionally offer such jobs. In this stylised setting earnings are a monotonous transformation from utility. For the sake of readability the model in this section is thus formulated in terms of earnings z. In the following, we use information about worker mobility to establish the firm size distribution $l(v)$, critical in determining firms' optimal choice of v in equilibrium.

4.3.2.1 Worker mobility

In equilibrium, the flows of workers of each type j who move from unemployment to employment and vice versa must balance (equation (4.2)).

$$\delta^j(n^j - u^j) = \begin{cases} \lambda^j u^j & \text{for } j = s \\ \lambda^f u^f F(z^*) & \text{for } j = f \end{cases} \tag{4.2}$$

where u^j denotes the number of unemployed type j workers and $\kappa^j = \frac{\lambda^j}{\delta^j}$. The flows differ between types as type f workers do not accept jobs with earnings $z > z^*$. The number of unemployed individuals thus are:

$$u^j = \begin{cases} \frac{n^j}{1+\kappa^j} & \text{for } j = s \\ \frac{n^j}{1+\kappa^j F(z^*)} & \text{for } j = f \end{cases} \tag{4.3}$$

Similarly, in steady-state the flow of workers of each type from unemployment to jobs with earnings no greater than z must equal the flow of workers of the same type leaving that group. The latter comprises employees moving from a job with value no greater than z to unemployment or to a better job (left-hand side of equation (4.4) with $G^j(.)$ denoting the cumulative density function of realised earnings for type j workers).

$$\lambda^j F(z) u^j = \begin{cases} [\delta^j + \lambda^j(1 - F(z))] G^j(z)(n^j - u^j) & \text{for } j = s \\ [\delta^j + \lambda^j(F(z^*) - F(z))] G^j(z)(n^j - u^j) & \text{for } j = f \ \& z \leq \ z^* \end{cases} \tag{4.4}$$

4.3.2.2 Firm size

In the steady-state the number of workers of type j which is employed at a firm offering jobs with earnings z can be expressed by equation (4.5) (Burdett and Mortensen, 1998).

$$l^j(z) = \lim_{\epsilon \to 0} \frac{(G^j(z) - G^j(z-\epsilon))(n^j - u^j)}{F(z) - F(z-\epsilon)} \quad \text{for } j \in (s, f) \tag{4.5}$$

Using (4.2), (4.4) and (4.5) Appendix D.1 shows that firm size is increasing in z both below and above z^*, but with a discontinuity at z^*, since type f workers do not accept jobs above this threshold value:

$$\begin{aligned} l(z) &= l^s(z) + l^f(z) \\ &= \begin{cases} \frac{n^s \kappa^s}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))} + \frac{n^f \kappa^f}{(1+\kappa^f(F(z^*)-F(z)))(1+\kappa^f(F(z^*)-F(z-\epsilon)))} & \forall z \leq z^* \\ \frac{n^s \kappa^s}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))} & \forall z > z^* \end{cases} \end{aligned} \tag{4.6}$$

4.3.2.3 Equilibrium job offer distribution

As firms are identical, it must hold that in equilibrium profits are equal for different z Burdett and Mortensen (1998). Firms offering jobs with low earnings achieve higher profits per worker but attract fewer workers than firms offering jobs with higher earnings. However, when earnings exceed z^*, n^f individuals drop out suspending this mechanism. The endogenous offer distribution might, therefore, include a mass point because firms not necessarily have an incentive to dissolve it by offering slightly higher earnings.

Proposition (I) *If we observe offers above z^*, there must be a mass point of job offers at z^*. The earnings offer distribution above z^* is continuous up to the highest earnings offers, $\bar{z}$.*
Proof: See Appendix (D.2).

The intuition of proposition (I) is that equal profits at and slightly above the threshold can only hold when the loss of type f workers is balanced by a discontinuously larger number of type s workers that can be attracted by exceeding the threshold. This can be provided for by a mass point at z^*. An earnings offer slightly above the threshold is able to poach all type s workers

at the mass point. As we observe positive mass above the threshold in the data (section 4.4), proposition (I) implies that there is a mass point at $z = z^*$ (i.e. that $f(z^*) > 0$).
A mass point in our setting implies implicates that any job offers with earnings just below the mass point ($z^* - \varepsilon$) will earn less profits, since margins per worker are only slightly higher, but firm size will be discontinuously lower since there is a mass of firms (offering z^*) that can poach a worker employed at earnings $z^* - \varepsilon$.

Proposition (II) *If there is a mass point at z^*, there will be a gap in the offer distribution just below the threshold.*
Proof: See Appendix (D.3)

Given the lack of offers just below the mass point, i.e. a gap in the offer distribution, the question arises whether any earnings lower than z^* are offered in equilibrium. This is the case when the increase in profit per worker is able to balance the lower firm size vis-à-vis locating at the mass point.

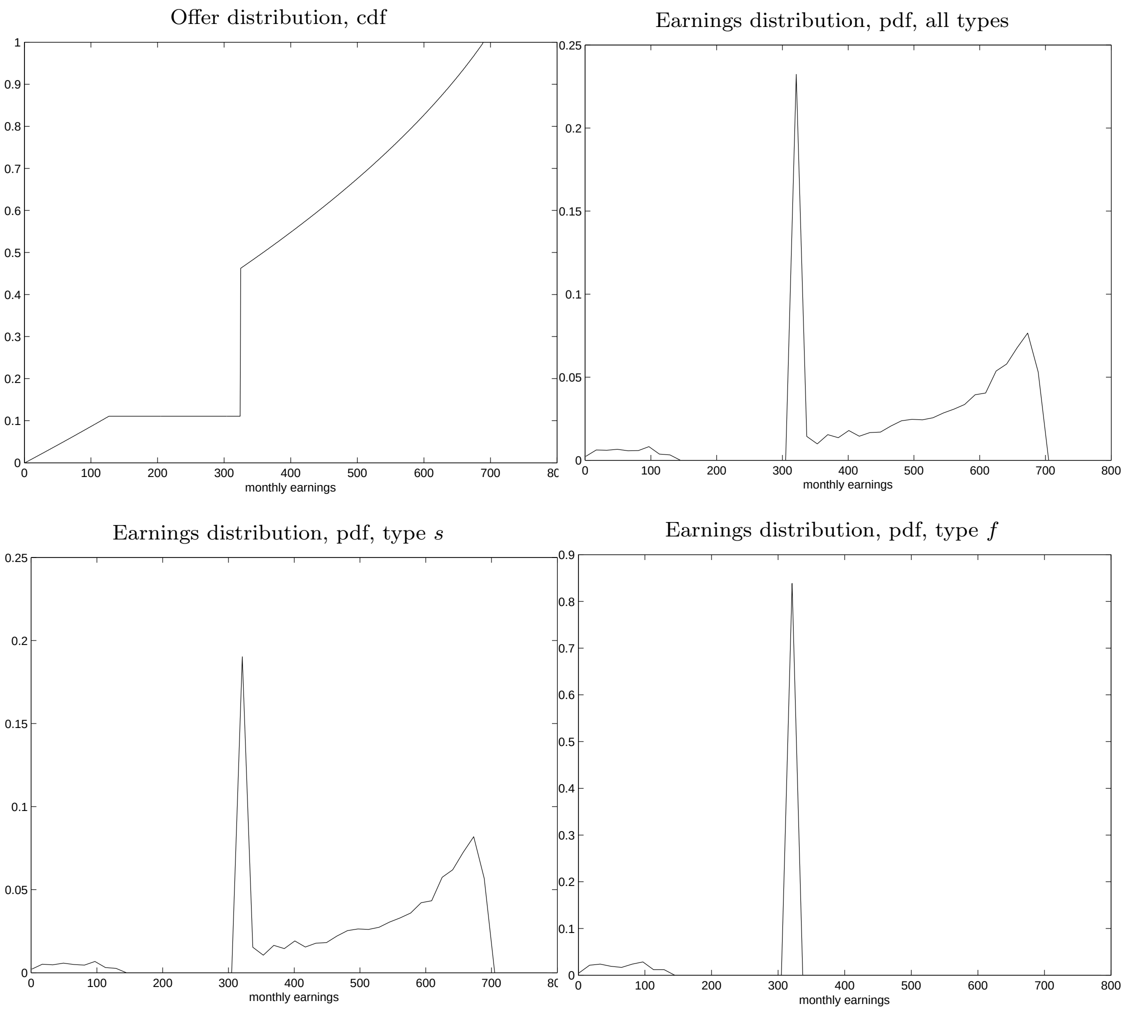
Proposition (III) *There may or may not exist earnings offers below the threshold z^* in equilibrium. The earnings offer distribution will then be continuous between the left limit of the gap, z'', and the homogenous reservation earnings, z^r.*
Proof: See Appendix (D.4)

4.3.2.4 Simulation

We now simulate the model outlined in this section assuming the following parameter values:
$p = 800$ €; $z^* = 325$ €; $\lambda^s = 0.2$; $\lambda^f = 0.2$; $\delta = 0.1$; $n^s = 1$; $n^f = 0.1$; $z^r = 0$

The resulting equilibrium cumulative offer distribution increases smoothly up to $z = z'' = 127$ and is constant in the interval $z \in [127, 325)$ (figure 4.1). Firms do not offer any earnings within that interval. The additional margin of reducing offered earnings compensates for the discontinuously lower firm size below the mass point not until a firm reduces its earnings to 127 €. The jump at $z = z^* = 325$ implies that there is a mass of offers at z^*. For $z > z^*$ the cumulative offer distribution again increases smoothly until $z = \bar{z} = 689$, the highest level of earnings offered. This shape of the offer distribution is reflected in the corresponding earnings distributions of the two types of workers (figure 4.1). Although type s workers do not have an incentive to bunch at z^*, the earnings distributions of both types exhibit a mass point at and a gap below this threshold. For type f workers the mass point is most distinct and – as assumed – there is no mass above. The upper right panel of figure 4.1 plots the resulting joint earnings distribution.

Figure 4.1: Offer and earnings distribution by types of workers

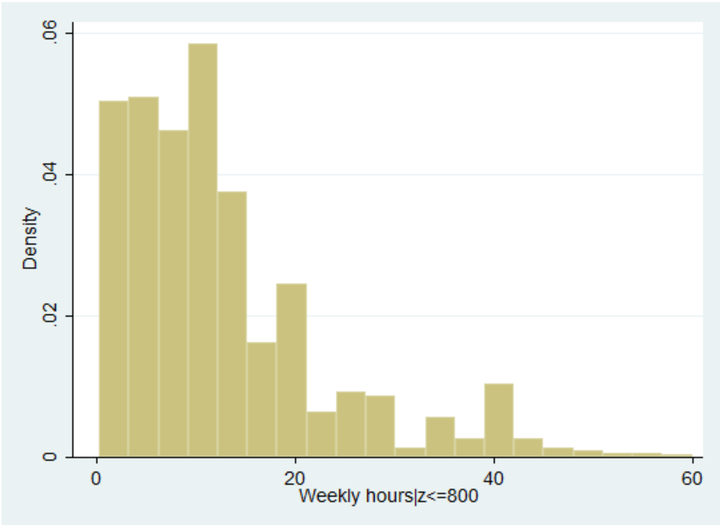


Notes: The graph of $f2$ workers uses a different scale.
Source: own calculations.

4.3.3 Heterogeneous hours, taxes and workers' heterogeneity

The model outlined in the previous section assumes homogeneous hours and a distribution of taxes such that type f workers do not accept any offers above the threshold. While these assumptions simplify an analytical solution, two important facts are neglected. First, working hours may differ. This is particularly relevant as the Minijob-threshold is based on monthly earnings, not hourly wages. Although there are clear spikes at round numbers, survey data

Figure 4.2: Hours distribution - less than 800 €/month



Source: GSOEP wave 2001

shows that there is some variation in hours for the jobs we consider (figure 4.2). This section thus considers how an equilibrium model can include workers and firm behaviour in the face of such variation. Second, although type f workers' net earnings decrease discontinuously at the threshold we certainly observe a positive mass of main jobs above the threshold. This section thus uses observed tax and SSC rates for t^{inc} and t^{ssc} in equation (4.1).

In the stylised model any mass point also requires a gap in the offer distribution below this mass point. This is not consistent with the data. Including variation in working hours in our model allows both for the existence of earnings offers above the threshold and a lack of a gap below the threshold if we include variation in working hours in our model.

4.3.3.1 Worker heterogeneity in preferences and productivity

The assumption of workers being homogeneous in productivity and preferences results in workers and firms agreeing to every match - even when the budget set exhibits a notch. This implies that the number of workers in employment does not vary with, for example, taxes. This is very restrictive not least because the tax exemption has the objective to increase employment. The mechanism could be that workers differ in productivity and preferences. Taxes and SSC might prevent workers with high opportunity costs, high preference for leisure and/or low productivity to find a suitable job.

Instead of attempting to precisely determine the set of different attitudes towards jobs above and below the threshold, we simply model two populations of type f job seekers. A fraction

θ of type f workers - in the following indexed by $f2$ - do not accept any offers with earnings above the threshold. Put differently, they only enter the market due to the tax exemption. The remaining type f workers continue to accept all offered jobs. This group of workers is indexed by $f1$. Their budget set exhibits a notch at the Minijob threshold, though, which varies between individuals (next sub-section). Both groups of type f workers thus generate the incentive for firms to offer earnings below or at the threshold.

4.3.3.2 SSC and Taxes

Net earnings are derived by equation (4.1). The SSC rate, t^{ssc}, is constant for all workers and set to the average legal value over the analysis period (20.69 %). The income tax rate, t^{inc}, depends among others on household characteristics and thus varies across individuals. As relevant variables are not observed we impute tax rates based on an auxiliary survey data set and a microsimulation model which accounts for the most important aspects of the German tax-transfer system (Appendix D.6). Tax rates ascribed to type s workers vary between roughly 8 % and just below 20 %. Approximately half of type f workers do not pay any taxes even when they exceed the Minijob threshold. Their earnings do not exceed the general tax allowance. Tax rates of the remaining typer f workers amount to up to 20 % (Appendix D.6).

4.3.3.3 Heterogeneous hours

We assume that hours vary across firms and workers care about hours and earnings. Following Chetty et al. (2011) we assume, however, that hours of work is not a choice variable of firms but predetermined by, for example, technology. The attractiveness of a job offer for firms depends crucially on how many job offers are viewed as superior in terms of utility and could poach a worker away from a firm. Consider a scenario where every firm has a different requirement of weekly hours and where the required hours of work vary continuously across firms[60]. Equal wage offers of different firms will correspond to different utility levels[61]. This will also mean that every firm will face a different wage threshold w^* corresponding to the wage level that generates monthly earnings of z^*. As the number of weekly hours is not a choice variable, the offered utility at z^* will similarly differ between firms. A mass point in the earnings distribution (for example at the threshold) does not result in a mass point in the utility distribution[62]. The

[60] Including discrete variation in hours is a non-trivial complication since there are now several thresholds, corresponding to the mini-job threshold with different working hours. The same strategic arguments as in the case of homogeneous hours will prevent firms from locating just below a mass point in the utility offer distribution. We demonstrate this in a version of the model with two hours sectors in Appendix D.5 and note that the model becomes intractable for many discretely different hours offers.

[61] This is true unless workers' marginal utility of leisure is zero, such that the utility function is independent of hours. This case is thus excluded here.

[62] If the distribution of hours of work includes mass points, a positive mass of firms share the same value of w^*. In that case, a mass point in the utility offer distribution is possible.

reasoning for a gap in the offer distribution below the threshold given homogeneous hours can, therefore, not be applied to a case with a continuous variation of hours.

4.3.3.4 Worker mobility and firm size

As in the case with homogeneous hours, workers either accept all jobs (type s and $f1$) or only jobs with earnings up to the threshold (type $f2$). Equation (4.2) is thus still valid albeit type $f1$ workers behave like type s workers. As before, in equilibrium the flow of workers into and out of jobs with utility no greater than v must coincide (equation 4.7).

$$\lambda^j F(v)u^j = \begin{cases} [\delta^j + \lambda^j(1 - F(v))]G^j(v)(n^j - u^j) & \text{for } j \in (s, f1) \\ [\delta + \lambda^f(1 - F(v|z \leq z^*))F(z^*)]G^{f2}(v)(n^{f2} - u^{f2}) = \lambda^f F(v)u^{f2} & \text{for } j = f2 \ \&\ z \leq z^* \end{cases} \tag{4.7}$$

where $(1 - F(v|z \leq z^*))F(z^*)$ describes the probability of another firm's offer exceeding v with earnings at or below the threshold.

For an individual firm with hours h, the strategic choice is then whether to offer utility $v(z, h)$ or utility $v(z', h)$ where strictly speaking the choice variable is the wage rate, but for fixed hours different wage rates monotonously translate to different earnings and utilities[63]. Profits are

$$\pi(w) = \begin{cases} (p - w)h(\frac{n^s\kappa^s}{(1+\kappa^s(1-F(v(w))))^2} + \frac{(1-\theta)n^f\kappa^f}{(1+\kappa^f(1-F(v(w))))^2} + \frac{\theta n^f\kappa^f}{(1+\kappa^f(1-F(v(w)|z\leq z^*)))^2}) & \forall z \leq z^* \\ (p - w)h(\frac{n^s\kappa^s}{(1+\kappa^s(1-F(v(w))))^2} + \frac{(1-\theta)n^f\kappa^f}{(1+\kappa^f(1-F(v(w))))^2} & \forall z > z^* \end{cases} \tag{4.8}$$

4.3.3.5 Characterization of Equilibrium

The model does not have a closed-form solution. The remainder of this section still considers how an equilibrium solution can look like with heterogeneous hours and realistic taxes.

The utility offer distribution The first and most important implication of heterogeneous hours is that the utility offer distribution will be smooth. That is, the utility distribution will not exhibit a mass point or a gap. If there was a mass point in the utility distribution, a firm could increase profits by slightly increasing the offered wage (and therefore utility). As hours of work are assumed to be distributed continuously, this is even possible at the threshold as $v(z^*)$ varies across firms.

We have a wage threshold function $w^*(h)$, where $w^*(h)\ h \equiv z^*\ \forall\ h$. The attractiveness of firms' offers will depend on whether or not a given wage offer corresponds to an earnings level that lies above or below the earnings threshold z^* - since for some workers ($j = f1$) utility decreases discontinuously at z^* and a positive number of workers ($j = f2$) do not accept any offers above.

[63]Note that this requires firms knowing the utility distribution in the population of workers as well as the distribution of tax rates.

Given a smooth utility offer distribution this implies that there will be a gap in the distribution of earnings above z^* because firm size drops discontinuously when the earnings offer exceeds the threshold.

There will be no gap below the threshold, though. The key difference to the case with homogeneous hours is that a mass point at z^* does not imply that firms offering $z^* - \varepsilon$ will attract discontinuously less workers than those offering z^*.

Hours and earnings Profits of heterogeneous firms need not to be equal in equilibrium. Strategies of firms with different hours requirements might differ. In particular, the decision between offering earnings below at or above the threshold might be assessed differently by different firms. As all firms want to make profits, two implications follow immediately.

First, firms with sufficiently small hours ($h \leq \bar{h} = \frac{z^*}{p}$) will offer earnings below the threshold. Second, when reservation utility is strictly positive, firms with sufficiently high hours ($h > \bar{\bar{h}} = v^{-1}(w^*, \underline{v})$) will offer earnings above the threshold[64].

For firms with hours between these two thresholds ($\bar{\bar{h}} > h > \bar{h}$) it is less obvious if they offer earnings less, at or above the threshold. It crucially depends on the usual trade-off between the marginal profit per worker (first term in equation (4.9) representing the first order condition of a firm with hours requirement h) and the firm size (second term in equation (4.9)). This is not trivial in our setting as firm size of a given firm is not monotonously increasing in the offered wage rate. Exceeding the threshold might even lead to a decrease in firm size as type f workers either suffer a utility drop at the threshold (type $f1$) or do not accept any jobs with $z > z^*$ (type $f2$).

$$\frac{\partial \pi(w)}{\partial w} = -h\left(\frac{n^s \kappa^s}{(1+\kappa^s(1-F(v(w))))^2} + \frac{(1-\theta)n^f \kappa^f}{(1+\kappa^f(1-F(v(w))))^2}\right) \tag{4.9}$$

$$+ (p-w)h\left(\frac{2n^s {\kappa^s}^2 \frac{\partial F(v(w))}{\partial v(w)} \frac{\partial v(w)}{\partial w}}{(1+\kappa^s(1-F(v(w))))^3} + \frac{2(1-\theta)n^f {\kappa^f}^2 \frac{\partial F(v(w))}{\partial v(w)} \frac{\partial v(w)}{\partial w}}{(1+\kappa^f(1-F(v(w))))^3}\right) \tag{4.10}$$

The change in firm size by increasing the wage rate crucially depends on $\frac{\partial v(w)}{\partial w}$ which varies between firms with different hours. Assuming a Cobb Douglas utility function, $v(w)$ is a monotonously increasing concave function in all points except the threshold where $v(w)$ drops discontinuously. Firms with low hours need to offer a high wage rate to reach threshold earnings. Due to $v(w)$ being concave high w implies $\frac{\partial v(w)}{\partial w}$ being small. Profit-maximizing earnings are likely to be small and below the threshold. Further, $v(w)$ varies between firms. There is an optimal number of hours h^* in the sense that for any given wage w, utility is largest for the firm with $h = h^*$. At a given w, $\frac{\partial v(w)}{\partial w}$ is larger for 'better' firms. Profit-maximizing earnings

[64] As $v(.)$ is monotonously increasing in w in all points except the threshold $v(w^*, h) < \underline{v}$ implies that a firm with h or more will not attract any workers by offering $w \leq w^*$.

of 'better' firms are thus more likely to be high and above the threshold. The popularity of a firm depends on the preference parameter α.

Whether a firm offers earnings above the threshold eventually depends on the equilibrium offer distribution $F(v)$, though. $\frac{\partial F(v(w))}{\partial v(w)}$ needs to be sufficiently high for $w > w^*$ such that enough additional type s and $f1$ workers can be attracted to balance the loss in marginal profit as well as of type $f2$ workers.

4.3.3.6 Simulation

The model outlined in this section is brought to the data in the following sections. The simulation presented here is based on the specification applied in the empirical part and already draws on the estimated parameters (table 4.5 in section 4.6.1).

The simulated cumulative earnings offer distribution exhibits a continuously decreasing slope up to the threshold. The jump at z^* implies a mass point in the density. It is flat above the threshold where the density exhibits a gap. From roughly 500 € it increases again up to approximately 850 € which is the highest earnings level in the simulated market.

This translates into a realised earnings distribution which slowly increases up to a discontinuous mass point at z^* where it sharply drops. This is consistent with the observed distribution (figure 4.4 in the next section). The increase in the predicted distribution is driven by type f workers. The density for type s workers up to the threshold is slightly decreasing. This is less consistent with the data although the observed density is indeed much flatter for type s workers (figure 4.4).

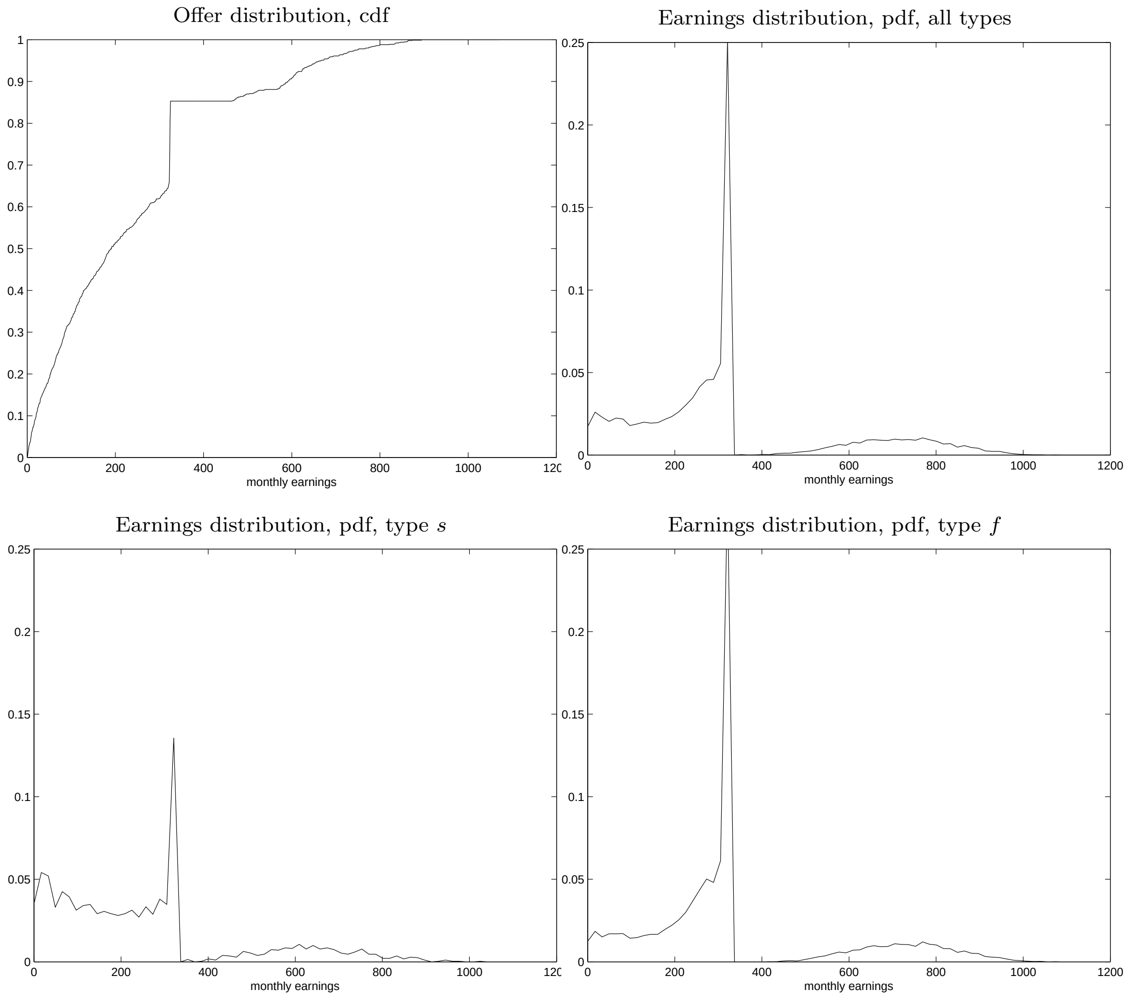
4.4 Data

We use administrative data collected by the German social security system. The Sample of Integrated Labour Market Biographies (SIAB) is a representative two percent sample of all individuals for whom an employer's report to the social security system exists[65]. For the present analysis, its specific characteristics makes the SIAB the most suited data set. First, exact total gross earnings of a period of an employment report are observed[66]. This information is very accurate due to the administrative character of the data. Second, with approximately 1.6 million sampled employees, the sample size is comparatively large and, therefore, includes a substantial number of individuals holding second jobs. It does not include civil servants and

[65] The German name is *Stichprobe der Integrierten Arbeitsmarktbiografien*. I used the weakly anonymised version of the data via on-site use at the research data center of the IAB in Berlin ($fdz637$). See vom Berge et al. (2013) for a detailed description.

[66] A period lasts at most until the end of the calender year. Gross earnings are capped at the earnings cap of pension insurance which is not problematic here as we focus on low-paid jobs.

Figure 4.3: Offer and earnings distribution by types of workers



Notes: Underlying parameter values as estimated in section 4.6.1: $\lambda^s = 0.0329$, $\lambda^f = 0.2506$, $\delta^s = 0.0570$, $\delta^f = 0.0379$, $\theta = 0.3715$, $\alpha = 0.8648$. Optimization error is assumed as discussed in section 4.5.2 ($\sigma = 71.21$). Optimization error is the reason for realised earnings spanning the support of the offer distribution.
Source: own calculations.

self-employed. Third, the SIAB consists of complete employment biographies of the sampled individuals. We can, therefore, differentiate between first and second employments and observe, for example, the exact day a second employment starts. It further includes spells of reported job seeking. Fourth, we observe several characteristics important for labour market outcomes. On individual level this particularly includes age, sex, occupation and education. On firm level, industry sector, size and wage structure are especially noteworthy.

The SIAB has two limitations. First, the number of hours worked is not precisely measured. Hours information only exists in three broad categories. Second, we do not observe the household context like marriage status, the income of a potential spouse or other income sources which precludes the calculation of the exact income tax rates. We thus use household data to complement our administrative data. The German Socio Economic Panel (GSOEP)[67] is a yearly, nationally representative household survey which includes information on labour market status, earnings, hours and all tax-relevant individual and household information.

4.4.1 Sample definition

We restrict the sample to employees[68] between 17 and 65 years. We define as second job, the employment spell that has less earnings (any further jobs are dropped). We do not classify very short overlaps between jobs (less than 5 days) as parallel employment spells. Interruptions of less than one month of otherwise similar spells are ignored unless the interruption is filled by another spell[69]. We exclude individuals who receive benefits from unemployment insurance or social assistance. These benefits are withdrawn when labour earnings increase resulting in a high implicit marginal tax rate. In many cases it amounts to 100 % for net earnings exceeding 165 €. The Minijob threshold is thus not relevant[70].

All spells which ended after April 1999 or started before March 2002 are potentially included in the analysis[71]. Every spell is classified as type s or type f, with the same individuals sometimes seeking a main job, and sometimes a second job. We thus differentiate between four kind of spells representing type and employment status.

- Type f workers seek or have their main job in the market for low-paid jobs. Unemployed individuals who are observed to be job-seeking are classified as type f if they have held a job with earnings of less than 800 € at least once in the analysis period. The reasoning of the latter condition is to restrict the sample to workers actually active in the market of low-paid jobs[72]. Employed type f workers currently hold a single job with earnings of less than 800 €.

- Type s workers already have a main job, but are interested in or have a second job. Two conditions have to be fulfilled for a spell to be classified as type s. First, an individual

[67] See Wagner et al. (2007) for a detailed description.

[68] We exclude trainees, interns, employees in the military, individuals in old-age retirement, freelance home workers, disabled persons in approved workshops, short-term employees, seamen and artists.

[69] If an individual has two parallel full-time employments the employment with less earnings is excluded.

[70] They might still benefit from the tax exemption, though, since 165 € net earnings can c.p. be generated with less hours.

[71] As noted in section 4.2, the systems in place prior to 1999 and after March 2002 were slightly different.

[72] On the other hand we miss individuals who search for a low-paid job without registering formally. The data shows that a substantial number of new low-paid jobs are held by workers who had been out-of-labour force before. In a future version of this paper a robustness analysis will classify empty intervals between two jobs as unemployment.

needs to be employed[73]. Second, the individual must have had at least one second job with earnings up to 800 € in the analysis period. The latter condition is rooted in the lack of direct information on job-seeking activities for employed workers. Note that the exact definition has direct impact on the estimated frictional parameters[74].

The market of low-paid jobs is dominated by main jobs (type f). Second job spells (type s) account for around 10 % of all employment spells in our sample (table 4.1)[75]. While for type f workers most spells are employment spells (84 %), this is not the case for type s workers (38 %).

Whilst a high fraction of unemployment spells are right-censored (69.32 % for type s and 58.87 % for type f, table 4.1), by far the most frequent reason for this is the end of the analysis period. The probability of being right-censored is considerably lower for employment spells (33.31 % for type s and 39.35 % for type f). As Minijobs are only included in the data as of April 1999, all spells which started before April 1999 are treated as left-censored. This is the case for roughly 20 % of the sample. Job separation (60.90 % for type s and 51.91 % for type f) is much more frequent than job-to-job transitions (5.8 % for type s and 8.73 % for type f).

Table 4.1: Total number of spells by type and employment

Status \ Type	s	f
Unemployed	49,984	55,592
unemp. to job	16,290	19,281
right-censored	33,694	36,311
left-censored	11,176	6,485
Employed	30,121	287,255
job to job	1,690	23,803
job to unemp.	17,394	138,861
right-censored	11,037	124,591
left-censored	10,508	80,814

Notes: The by far most frequent reason for right-censoring is the end of the analysis period. For unemployment spells another reason is the start of a new employment with earnings higher than 800 €. Employment spells are treated as right-censored when their defining conditions lose validity before a transition occurs. This includes for example the increase of earnings beyond 800 € or for type s spells the end of the main job.
Source: SIAB; own calculations

4.4.2 Hours and taxes

While SSC rates are homogeneous in Germany, income tax rates depend on household characteristics. This implies that the size of the notch at the threshold varies between workers.

[73]We use all employment spells with earnings above 1000 €. We choose this figure to be sensibly above the maximum value for the market of low-paid jobs, to avoid individuals switching between markets when earnings vary.

[74]A future version of the paper will include robustness checks in that respect. A more restrictive definition would, for example, be to just include workers who will have a second job in the future.

[75]This is approximately also true for the number of employed individuals. In June 2001, for example, 96,789 individuals in our sample hold a main job, 9,441 a second job.

In particular, marital status and the potential partner's income is crucial. Since these are not available in the administrative data, we impute tax rates using detailed information on all tax-relevant characteristics provided by the GSOEP (Junge, n.d.). For the sake of calculation speed we allocate individuals to three tax groups instead of allowing for continuous variation. Type s workers face larger tax rates than type f workers (table 4.2) which is mainly driven by first job earnings. Appendix D.6 presents the imputation procedure in more details.

The model outlined in the previous section assumes hours of work varying continuously between firms. We do not have information on hours of work offered in the market, though. We do not observe individual hours of work in the data. We use the GSOEP to impute hours of work based on variables included in the SIAB: sex, gross earnings, sector and education. As the hours distribution resembles a log-normal distribution (figure 4.2) a generalised linear model with log link is used for imputation. We then predict individual hours of work in our sample and let firms draw from the smoothed predicted empirical hours distribution. Mean predicted hours per month are approximately 40 for type s workers translating into 10 weekly working hours (table 4.2). Low-paid main jobs come with on average 55 hours per month. Within-type variation in hours is substantial. Average implied wage rate is approximately 5.50 € for both worker types[76]. Appendix D.7 contains more details including the estimation results.

Table 4.2: Imputed hours, wage and tax rates

	type s		type f	
	mean	std	mean	std
weekly hours	8.95	4.61	12.84	7.66
wage rate	5.56	1.48	5.76	1.37
tax rate	32.61	3.27	28.03	6.02

Notes: Variation in tax rates is between groups. Tax rates include SSC (20.69 pp.).
Source: SIAB, GSOEP wave 2001, own calculations.

4.4.3 Descriptives

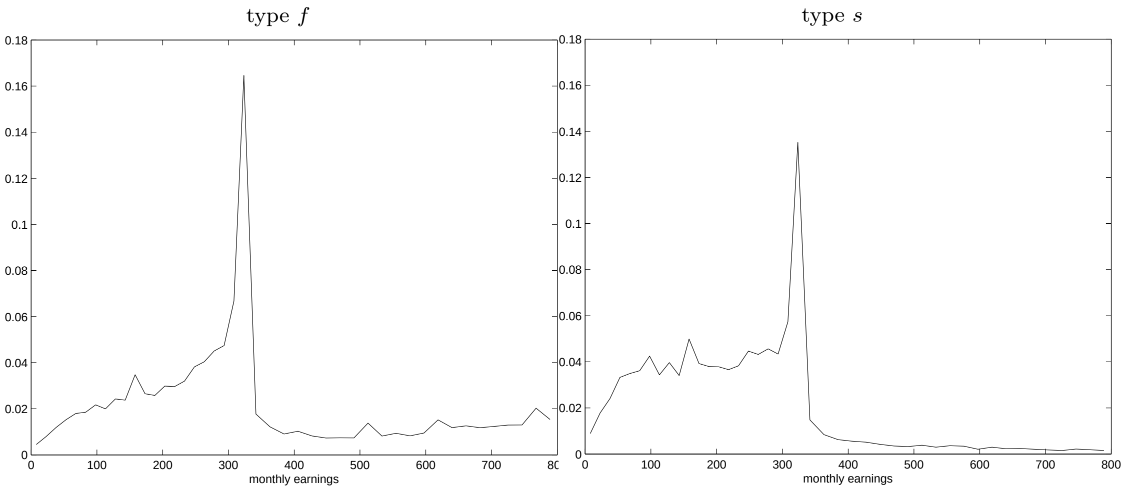
The earnings response to the tax exemption is dramatic. The observed earnings distribution of low-paid jobs as main jobs features a huge peak at the threshold for Minijobs (figure 4.4, left panel). This group of workers (type f) is subject to the Minijob threshold implying a strong incentive to locate below the 325 € threshold. This is not the case for second job earnings as the Minijob regulations do not apply. Nevertheless, the earnings distribution of type s workers features a clear peak at the Minijob threshold and hardly any mass beyond (right panel of

[76] This is quite low, in particular relative to the minimum wage of 8.50 € introduced a decade later. The median wage rate in the GSOEP sample used for the hours imputation is 7.2 €. The model fit is fairly good, though (Appendix D.7). The deviation is rooted in lower mean gross earnings in the SIAB sample (301.88 €/month) relative to GSOEP (374.79).

figure 4.4). We interpret this as compelling evidence for the existence of a population labour supply effect[77]. Firms seem to offer contracts based on employees looking for Minijobs as main jobs, consistent with our assumption of both types drawing from the same take-it-or-leave-it offer distribution. Furthermore, the earnings densities of both types increase gradually up to the mass point which is of similar relative size. The fraction of employees below the threshold is higher for type s ($\approx$ 90 %) than for type f workers ($\approx$ 75 %), though. We therefore allow the offer arrival rate to vary between types in our model.

We further observe differences in spell durations. Unemployment spells are, for example, shorter for type f workers (table D.8 in Appendix D.12) which again implies a larger offer arrival rate for that group of workers. Type f workers are on average also longer in employment. By contrast, the average duration between two jobs are similar between both types.

Figure 4.4: Distribution of observed earnings by types of workers



Notes: The graph includes all spells classified as type s or f as outlined in section 4.4. The vertical lines represent the Minijob threshold.
Source: SIAB.

The distribution of sectors of main and second jobs in the low-paid market are fairly similar (first two columns in table 4.3). Main sectors in both distributions are economic services as well as wholesale/retail, albeit the former (latter) is more frequent in low-paid jobs as main (second) jobs. By contrast, manufacturing is the major sector where second job holders work in their first job (third column in table 4.3). This, again, supports the assumption that main and second jobs in the low-paid market are drawn from the same job offer distribution. It further

[77] As we can observe whether a job benefits from the tax exemption, misclassification can be suspended as alternative explanation (table D.7 in Appendix D.12). Interestingly, the small spike slightly below 200 € is also an indication of a population labour supply effect. Workers receiving unemployment benefits have an incentive to earn at most 165 €. Although our sample excludes these workers we see a small spike at 165 €.

suggests that the low-paid market indeed constitutes a separate market.

Table 4.3: Distribution of sectors, 2000-2002

Sector	Low-paid job Type f	Type s	1st job of Type s
D Manufacturing	11.67	10.27	30.02
G Wholesale & retail	22.22	13.74	13.53
H Hotels & restaurants	9.08	8.62	2.21
I Transport	4.77	8.70	5.87
K Economic Services	18.98	28.70	10.52
L Public admin	2.18	2.97	7.62
N Health	10.87	8.08	12.72
O Other services	7.74	9.18	4.21

Source: SIAB

In our sample most workers holding a low-paid job as main job are relatively young women (table 4.4). East Germans are slightly over-represented and over 90% have at most vocational training. Unemployed type f workers are relatively more likely male, East German and badly educated. This difference suggests that some individuals seeking for a low-paid job might not register formally - this seems to be particularly true for women.

Table 4.4: Socio-demographic characteristics by type, 2001

	type s unemployed	type s employed	type f unemployed	type f employed
Female	48.20	45.15	52.67	76.26
Age	36.51	39.73	40.06	38.77
East German	9.07	7.76	29.82	13.22
Education				
Intermediate	12.99	12.97	28.13	24.24
Voc. training	74.81	74.71	65.16	67.16
Grammar	6.72	6.15	3.71	4.84
University	5.48	6.17	3.00	3.76
First job				
Monthly earnings	2025	2131	.	.
Part-time	17.4 %	18.3 %	.	.

Notes: First jobs of second job holder are only included when earnings exceed 1000 €.
Source: SIAB; own calculations

Resorting to survey data further clarifies who actually is in the market sector we analyse. Individuals with Minijobs as main job can be differentiated between housewives/-men, students, pensioners and officially unemployed individuals with housewives (unemployed) being the largest (smallest) group (Körner et al., 2013). Only roughly one quarter say that they do not work more because they cannot find a suitable position. This is another observation which initiates that we treat the low-paid market as separate. Furthermore, Minijobs do not seem to serve as a stepping stone to larger employment relations (Freier and Steiner, 2008).

The gender ratio of workers holding or seeking side jobs is much more even with slightly more men (table 4.4). The share of East German workers is less and type s workers are slightly better educated. Age is roughly similar between types. The first jobs of type s workers are mainly full-time positions with on average slightly more than 2000 € earnings per month. This is similar for employed and unemployed type s workers.

4.5 Identification and Estimation

We seek to estimate seven parameters[78]: the offer arrival rates of both worker types (λ_s and λ_f), the job destruction rates of both worker types (δ_s and δ_f), the preference parameter of the utility function (α), the fraction of type f workers who do not accept jobs with earnings exceeding the threshold (θ) and the level of productivity per hour (p).

In this section we first discuss identification arguments for these parameters (section 4.5.1). We then present the two-step maximum likelihood (ML) procedure which is used for estimation (section 4.5.2). This includes a discussion of the likelihood contributions for the different workers types and employment states.

4.5.1 Identification

The frictional parameters are identified by observed durations in (un-)employment. Type s workers are assumed to accept all jobs. Their job offer arrival rate λ^s is thus identified by the observed duration in unemployment for type s workers.

Some unobserved fraction of type f workers also accepts all offers - we denote these $f1$. A second group only accepts the proportion of job offers below the threshold - we denote this group $f2$. For this group, note that the acceptance probability is only known when the job offer distribution is identified. We exploit that type f workers who are observed to accept jobs above the threshold reveal them being of type $f1$ who accept all jobs. Their duration in unemployment identifies the job offer arrival rate of type f workers, λ^f.

The duration in employment identifies δ_s and δ_f. The upper support of the wage distribution identifies the underlying productivity of jobs. In addition, the distribution of accepted earnings of unemployed type s workers who accept all jobs represent a non-parametric estimator for the job offer distribution.

The distribution of realised earnings is informative about both the preference parameters α and θ[79]. This is especially true for the probability of earnings below, at or above the threshold. A high θ implies a high fraction of workers who do not accept offers with earnings $z > z^*$. This increases the attractiveness for firms offering earnings at or below the threshold. The same

[78]The homogeneous reservation utility is set to zero as we observe jobs with very small utility.
[79]Note that θ is not necessarily only a preference parameter (section 4.3.3.1).

is true for a low α implying a low relative weight of earnings (vis-a-vis leisure) in the utility function.

Identification of θ resorts to differences in state durations across worker types. For unemployed type $f2$ workers the probability of a match in one period is $\lambda_f F(z^*)$. For other workers this probability is simply λ_f or λ_s, respectively. Conditional on λ_f and the offer distribution (see above) the difference in durations between type s and f is, thus, informative about θ. The same is true for the difference in durations between type f workers below and above the threshold.

In the next section we present the ML estimation procedure where groups of parameters are estimated jointly. We thus do not only rely on the non-parametric estimators or identification arguments brought forward here.

4.5.2 Estimation procedure

We use a two-step ML procedure. In the **first step**, the frictional parameters are estimated based on duration data. The preference parameters are then estimated in the **second step** based on both duration as well as earnings data. Separate estimation is possible because the frictional parameters can be identified without knowing the job offer distribution (section 4.5.1). Although less efficient, two-step estimation procedures are more robust. In our case, joint estimation is numerically extremely costly as well.

In the **first step**, the durations in different employment states are used to estimate the transition parameters. We use all relevant spells of a worker but treat them as independent. The duration in employment is used to estimate the job separation rates (equation (4.11)). We use all employment spells. The first part of equation (4.11) is the probability that an employment spell is drawn. The second part pertains to the duration of employment which is assumed to be exponentially distributed with transition rate δ (Bontemps et al., 1999).

$$L_{e1}^{j}(t_{emp}) = \left(1 - \frac{n^j \delta_j}{\delta_j - \lambda_j}\right) \delta_j^{1-d} exp[-\delta_j(t_{emp})] \quad \text{for } j \in (s, f) \tag{4.11}$$

where t_{emp} is the length of the employment spell. Subsequent job spells are treated as contiguous. d is a dummy for censored observations (left- or right-censored).

We estimate job offer arrival rates using the duration in unemployment (equation (4.12) where t is spell length). Type $f2$ workers, however, do not accept jobs with $z > z^*$. Their duration in unemployment depends on the offer distribution. We, therefore, only use type f workers who accept earnings above the threshold. In doing so they reveal them being of type $f1$.

Dependence on the offer distribution also prevents exploiting the duration until a job-to-job transition occurs.

$$L^j_{u1}(t) = \frac{n^j \delta_j}{\delta_j - \lambda_j} \lambda_j^{1-d} exp[-\lambda_j(t)] \quad \text{for } j \in (s, f1) \tag{4.12}$$

Following the identification arguments, we use in the **second step** information on durations as well as earnings in order to estimate θ and α. Similar to Bontemps et al. (1999) we do not have an analytical expression for the endogenous job offer distribution $F(.)$ which only depends on structural parameters. Bontemps et al. (1999) apply a two-step procedure in which they first estimate the cdf and pdf of the realised wage distribution (G and g) non-parametrically by kernel density estimators. Using these estimates and their relation to the offer distribution (equation (4.7)), the likelihood contributions can be expressed in terms of the realised wage distribution and the model parameters. Once the transition parameters are estimated from workers' mobility patterns, these can be used to transform the observed distribution $G(.)$ to the offer distribution $F(.)$ which is the object about which theory provides us with predictions concerning the likelihood.

As this strategy only exploits equations with respect to workers' behaviour, the offer distribution is basically treated as exogenous. Firms' behaviour leading to the endogenous offer distribution is not exploited. In our setting, however, the restrictions on the earnings distribution helps to separately identify α, the parameter of the utility function and θ, the unobserved fraction of type f workers who do not accept any jobs with $z > z^*$ (type $f2$). In contrast to the two-step process proposed by Bontemps et al. (1999) we therefore calculate $f(.)$ and $F(.)$ by numerically solving the firms' problem in every iteration (see Appendix D.8 for more details).

$$\begin{aligned} L^j_{u2}(t,z) =& \frac{n^j \hat{\delta}_j}{\hat{\delta}_j - \hat{\lambda}_j} (\hat{D^j_u})^{1-d} exp[-\hat{D^j_u} t] * \\ & \left(\int_{-\infty}^{\infty} \hat{f}_z(z+\eta) dh(\eta) \right)^{(1-d_r)} \quad \text{for } j \in (s, f1, f2) \end{aligned} \tag{4.13}$$

The likelihood contributions in this estimation step are composed of the probability to be (un-)employed, the spell duration and the realised wage of the (potentially new) employment. For unemployment spells the first two parts are similar as above (equation (4.13)). $\hat{D^j_u}$ denotes the estimated arrival rate of acceptable offers. For $j \in (s, f1)$ it holds that $\hat{D^j_u} = \hat{\lambda}_j$. For $j = f2$, $\hat{D^j_u} = \hat{\lambda}_j \hat{F}_z(z^*)$. The third part of equation (4.13) is the probability that the accepted job has true earnings $\tilde{z_u} = z_u + \eta$ where z denotes observed earnings and $\eta \sim N(0, \sigma^2)$ represents random optimization error. The latter is assumed to be normally distributed with zero mean and variance of σ^2. The reasoning is that firms might not always offer exactly desired earnings because, for example, weekly hours or the wage rate usually is a round number. Frictions are a standard assumption in the bunching literature where bunching is usually found as a hump around the threshold rather than an exact spike (Saez, 2010)[80]. The likelihood of the realised earnings drop out of equation (4.13) when the unemployment spell is right-censored (i.e. $d_r = 1$).

For an employed individual the likelihood contribution is

$$L^j_{e2}(t,h,w) = \left(1 - \frac{n^j\hat{\delta_j}}{\hat{\delta_j}}\right) - \hat{\lambda_j}(\hat{D^j_e}(h,w))^{1-d} exp[-\hat{D^j_e}(h,w)t] * \\ (\delta^{1-j2j}\hat{\lambda_j}(\hat{D^j_e}(h,w))^{j2j})^{(1-d_r)} * \\ \left(\int_{-\infty}^{\infty} \hat{g_z}(z_e+\eta)dh(\eta)\right)^{(1-d_r)} \quad \text{for } j \in (s, f1, f2) \tag{4.14}$$

where $j2j$ indicates whether an uncensored employment spell ends with a job-to-job transition ($j2j = 1$) or a transition to unemployment ($j2j = 0$). It holds that $\hat{D^j_e}(v) = \hat{\delta_j} + \hat{\lambda_j}(1 - \hat{F}(v))$ for $j \in (s, f1)$. For $j = f2$, $\hat{D^j_e}(v) = \hat{\delta_j} + \hat{\lambda_j}(1 - \hat{F}(v|z \leq z^*))\hat{F}(z^*)$. The third part is the probability of a transition to unemployment (δ) or to a better job ($\hat{D^j_e}(h_e, w_e)$). It vanishes if the employment spell is right-censored. The last part is the probability that the current job has true earnings $\tilde{z_e} = z_e + \eta$. The density of realised earnings, g_z, is derived from equation (4.7) and the current estimation of the offer distribution.

While we observe if an individual is of type s or f, we cannot distinguish between type $f1$ and $f2$ workers if observed wages are below the threshold or not observed. For $k \in (u, e)$ the likelihood for type f workers, therefore, is

$$L^f_k(t,z) = \begin{cases} \theta L^{f2}_{k2}(t,z) + (1-\theta)L^{f1}_{k2}(t,z) & \text{for } z \leq z^* \\ L^{f1}_{k2}(t,z) & \text{for } z > z^* \end{cases} \tag{4.15}$$

[80] We do not allow for the case that true earnings are below the threshold while observed earnings are above or vice versa: When $z > z^* > \tilde{z}$ or $z < z^* < \tilde{z}$ we set $z = \tilde{z}$. The reason is that the observed distribution of earnings drops very sharply at exactly the threshold while it decreases much more gradual to the left (figure 4.4). We argue that the cost of crossing the threshold is so high such that it exceeds potential adjustment costs or costs for exact declaration.

We then estimate α, θ and σ by numerically optimizing the likelihood contributions[81].

4.6 Results

In the following we discuss the estimation results, the model fit and two counterfactual simulations of policy reforms which smooth the discontinuity generated by the Minijob regulations.

4.6.1 Estimation results

The job separation rate is larger for second jobs (type s) than for main jobs (type f, table 4.5). A second job is terminated every 18 months, a low-paid main job every 26 months. The job arrival rate differs even more strongly. A type s worker gets an offer every 30 months while type f workers get three job offers a year. Second jobs are, thus, less stable and workers need more time to find one. Both makes sense as workers seeking for second jobs arguably have less time for search and might be less dependent on a match[82].

About 37 % of type f workers in the market do not accept offers with earnings exceeding the threshold. This group of workers represents roughly two million individuals who are only in the market due to the tax exemption. At first sight the Minijob policy thus seems to achieve its main objective to facilitate new (formal) employment relationships. Conversely, this also implies that more than 60 % of workers with a low-paid job as main job would also work if there weren't any subsidy (type $f1$ workers). From the government's point of view the increase in employment comes thus at the expense of taxes and SSC missed out for this group. Intensive margin effects of type $f1$ workers are discussed in the next section by means of simulating the removal of the tax exemption.

The elasticity of consumption (leisure) with respect to utility is 0.86 (0.14) implying that workers reward both, consumption as well as leisure time. The coefficient is not statistically different from unity though. The standard deviation of the optimization error is estimated to be 71 €/month. This implies, for example, that about 15 % of jobs with observed earnings of less than 250 € actually have true earnings at the threshold.

Estimation precision for the transition parameters is very high. Standard errors are larger for the remaining parameters (table 4.5). The reason is that the stated standard errors for the latter include the uncertainty rooted in the random elements of the estimation procedure (section 4.5)[83].

[81] As the offer distribution is only evaluated at some grid points we do not resort to gradient-based optimization algorithms. See Appendix D.8 for more details.

[82] As discussed above the estimates of offer arrival rates depend on the definition of unemployment.

[83] Precision can be increased by basing these random procedures on more repetitions. The calculation of the offer distribution is, for example, repeated ten times per parameter set. Increasing this number increases the

Table 4.5: Parameter estimates

Parameter	p.e.	s.e.
δ_s	0.0570	0.0005
δ_f	0.0379	0.0002
λ_s	0.0329	0.0002
λ_f	0.2506	0.0028
θ	0.3715	0.0775
α	0.8648	0.1121
σ	71.212	6.963

Notes: p.e. – mean point estimate of 50 bootstrap repetitions, s.e. – standard deviation of 50 bootstrap repetitions; length of period is one month
Source: SIAB; own calculations.

4.6.2 Model Fit

Based on the parameter estimates we simulate the low-paid market under status-quo regulations. The predicted realised earnings distribution reflects the main characteristics of its observed counterpart fairly well (figure 4.5). For type f workers it increases gradually up to a discontinuous mass point where it sharply drops. The mass beyond the threshold is relatively small. This is perfectly consistent with the observed distribution. The size of the mass point is over-estimated, though. For type s workers, by contrast, the size of the mass point in the predicted and observed distribution almost coincide. Our model also correctly predicts that the side job earnings distribution is fairly flat to the left of the threshold. The predicted distribution has too little mass below and too much above the threshold, though.

Two more deviations between predicted and observed earnings distributions concern both worker types. First, our model does not predict any mass slightly above the threshold although some workers are observed to be located there. As utility and profits could be increased by decreasing earnings to the threshold our model cannot rationalise this[84]. Second, the endogenous upper earnings limit in the simulated low-paid market is over 1000 €/month (represented by the jumps at the upper limit in figure 4.5) while we cut the observed earnings data at 800. One parameter which crucially drives the support of the earnings distribution is productivity. We use the 95th quantile of the observed wage distribution as an estimator for the homogeneous productivity p which seem to be an over-estimation for at least some firms [85].

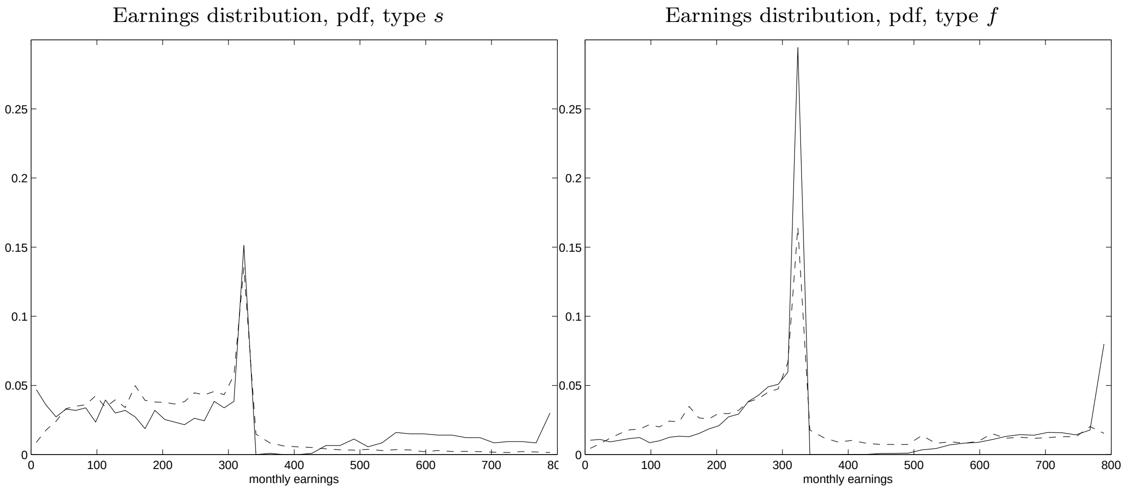
The employment probability is fit very well for both worker types (table 4.6). The model also correctly predicts that hours, wages and earnings are higher for type f than for type s

precision of the estimates. It is fairly costly in terms of computing time, though.

[84] Other studies interpret observed mass in a dominated region as evidence for optimization frictions (Kleven and Waseem, 2013). We indeed include an optimization error in the model but do not allow earnings to cross the threshold because this would smooth the discontinuity at the threshold (section 4.5.2).

[85] In the next version of this paper we plan to estimate p jointly with the other parameters. A disadvantage of that approach is that we then need to add a dimension to the grid search (section D.8) which is costly in terms of computing time.

Figure 4.5: Offer and earnings distribution by types of workers



Notes: The solid line represents the predicted distribution, the dashed line the observed one.
Source: own calculations.

workers. The fit of the earnings level is relatively good albeit on average slightly overestimated. The predicted underlying combination of hours and wages are not fit very well. Hours are over-estimated, the wage rate under-estimated.

Table 4.6: Fit of moments

	Type s Mean		Type f Mean	
Measure	obs	pred	obs	pred
$P(e)$	37.60	37.36	83.79	86.11
z	223.62	279.91	320.16	365.97
w	5.56	4.25	5.76	4.54
h	8.95	13.94	12.84	18.41

Notes: $\bar{w}$ *=mean wage,* $\bar{h}$ *=mean weekly hours,* $\bar{z}$ *=mean earnings,* $P(e)$ *=employment probability,* *unrest.* =all earnings, *unrest.* = $z \leq 800$
Source: SIAB; own calculations.

4.6.3 Smoothing the discontinuity

The main objective of MWP policies is to facilitate employment relationships which would not be formed without subsidy. This comes often at the expense of an incentive for already employed individuals to reduce hours of work (Bargain et al., 2010) as the withdrawal of the subsidy usually implies high implicit marginal tax rates. For German Minijobs distortions of the earnings distribution are particularly relevant since crossing the threshold implies a substantial increase in the average tax rate. The budget thus features a notch. This becomes obvious by the large bunching observed in the data (section 4.4.3). As shown above these potential distortions might be reinforced by firm responses which further generate externalities on workers who are not directly affected by policy.

In the following we therefore apply our model to evaluate two counterfactual policy reforms which remove the discontinuity induced by the Minijob tax exemption. We first simulate completely removing the tax exemption. Among others this helps understanding the externality on already employed individuals. In the second evaluation we replace the Minijob regulation by a smoother subsidy schedule which retains the objective to generate employment. It emphasises additional distortions induced by discontinuous tax policies like the German Minijob regulation. We first assume that frictional parameters are exogenous and constant. The number of vacancies in the economy and the intensity of firms searching for workers do thus not respond to a change in labour costs or labour supply. This labour demand restriction is ad-hoc[86].

We, thus, additionally calculate an scenario where we assume that the offer arrival rate increases for all remaining workers such that the total number of offers per month stays constant. How realistic both scenarios are depends on the substitutability of the dropped out type $f2$ workers. When the jobs are lost because the productivity of the activity is too small to allow for an acceptable wage rate, substitutability is arguably low. By contrast, low-productivity workers or workers with specific preferences might be substituted more easily. As we assume both reservation utility and productivity being homogeneous, these margins cannot be analysed in our simple model, though.

Removing the Minijob tax exemption When the tax exemption is removed all type $f2$ workers leave the market. This amounts to roughly one and a half million employment relationships (table 4.7). Naturally, these workers who only enter the market due to the tax exemption unambiguously lose utility when it is removed.

Extensive margin reactions of type s and $f1$ workers are excluded by assumption. In what way these worker types are affected by and whether they benefit from the tax exemption depends crucially on firms' responses[87]. As expected the earnings offer distribution becomes smooth when the tax exemption is removed (left panel of figure 4.6). Since jobs with earnings above the threshold become relatively more attractive for workers, firms offer more of these jobs. This reveals that firms are jointly responsible for negative effects of the Minijob regulation on

[86] Other assumptions like fixed and exogenous hours for a given firm are less restrictive. The relative size of firms with different hours might respond to a change in financial incentives. When tax rates increase, for example, leisure becomes relatively more attractive and so firms with fewer hours expand.

[87] Here, the assumption of constant frictional parameters becomes crucial. It implies that firms do not compensate for type $f2$ workers dropping out of the market. When these workers are (partly) substituted by workers of other types, offer arrival rates of the latter might increase. Appendix D.9 simulates an alternative scenario in that respect.

earnings of already employed (see below). The likelihood of offers below the threshold increases as well. The mass point at the threshold in the status-quo scenario is thus distributed over almost the entire support when the Minijob tax exemption is removed. The reason for the higher likelihood of low earnings offers is rooted in type $f2$ workers leaving the market. A higher share of type s workers in the market results in a decrease of the average job arrival rate because type s workers are estimated to receive fewer offers per time period (table 4.5). This reduces the competition on the labour market as the likelihood that another firm outbids a job offer decreases. Firms therefore offer less valuable jobs. This can also be interpreted as a population labour supply effect which thus goes far beyond the visible bunching reaction.

Abstracting from noise, the resulting realised earnings distribution of type s workers is monotonously decreasing over the entire earnings interval. The earnings distribution of type f workers, by contrast, increases up to approximately 500 € and decreases beyond. This difference mirrors the much higher offer arrival rate of type f workers (and potentially also the on average lower income tax rate).

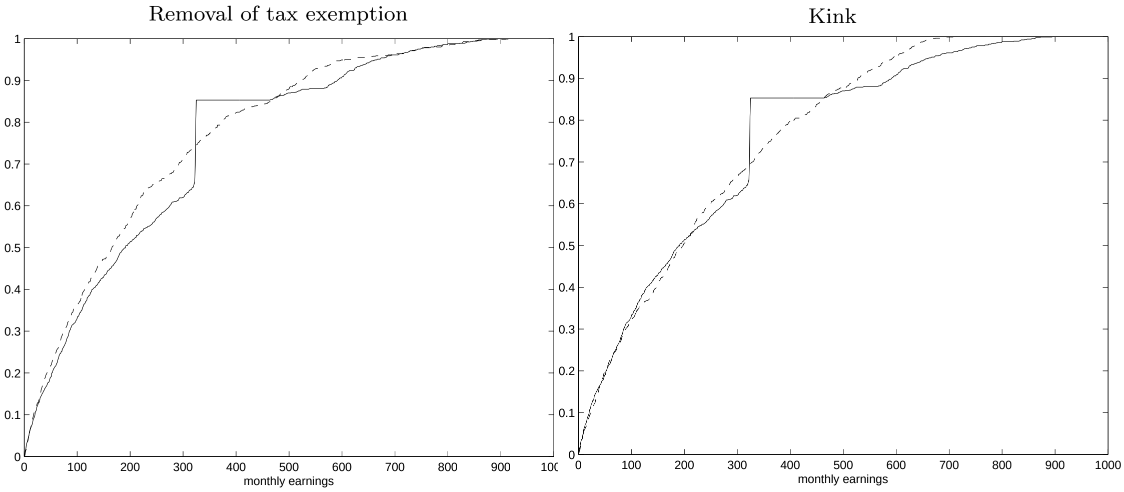
On average gross earnings of workers employed in the low-paid market increase (table 4.7). This partly represents a compositional effect, though, as type $f2$ workers had on average smaller earnings. For type s workers gross earnings even decrease slightly as an increase in hours is overcompensated by a decrease in the wage rate. Utility of type s workers also decreases. This implies that type s workers benefit more from the positive externality of an on average higher offer arrival rate than they lose due to the discontinuous tax schedule.

By contrast, gross earnings of type $f1$ workers strongly increase. This is driven by both, a positive effect on the wage rate and on hours of work. These workers are hampered by the Minijob threshold. Note that they still profit substantially from the tax exemption. Their special tax treatment in the Minijob-scenario outweighs the distortion of earnings.

Interestingly, wages and earnings of type s workers decrease while they increase for type $f1$ workers. This can be explained by the differential job offer arrival rates. Firms offer considerably more jobs with earnings below the threshold when the tax exemption is removed resulting in lower average starting earnings (and wages). The opportunities of advancement also increase, though, as firms offer more jobs with earnings above the threshold as well. The latter is mainly beneficial for type $f1$ workers, however, as they get considerably more offers.

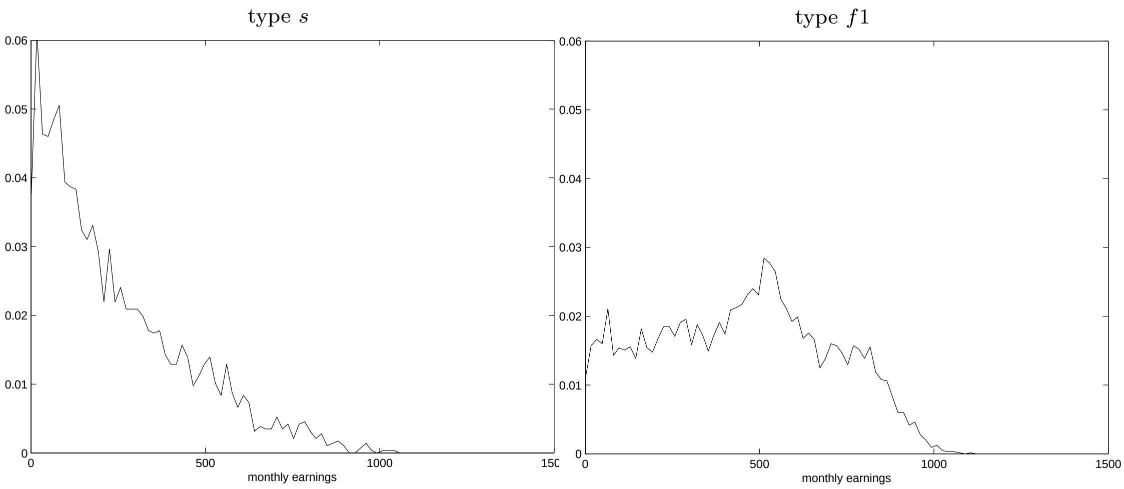
Total hours and gross earnings decrease by roughly 23 and 16 %, respectively (table 4.7). Tax revenues increase by 110 million Euro. This opens scope for compensation. When the additional tax revenue is equally redistributed to all individuals by tax-free lump-sum transfers, type s and type $f1$ workers' net earnings and utility would increase strongly (table 4.7). Both worker types would thus clearly benefit from removing the tax exemption. The reform is not Pareto improving in terms of utility, though, as type $f2$ workers lose albeit being compensated.

Figure 4.6: Cumulative earnings offer distribution - status quo vs. counterfactual



Notes: The solid line represents the status quo, the dashed line the counterfactual.
Source: SIAB, own calculations.

Figure 4.7: Distribution of earnings by types of workers - no Minijobs



Source: own calculations.

Smoothing the subsidy schedule In the following we simulate a reform which replaces the Minijob regulations by a smooth subsidy schedule. Average tax rates increase gradually in the interval between 325 € and 800 € up to their full values. The complete tax exemption below 325 € remains. The notch is thus replaced by a kink. The objective is to prevent additional distortions induced by a discontinuous tax schedule while positive employment effects

Table 4.7: Effects of smoothing the discontinuity

Change in	Removal of tax exemption Total	type s	type $f1$	type $f2$	Kink Total	type s	type $f1$	type $f2$
$\bar{w}$	0.23	-0.20	0.34	.	0.08	0.08	0.16	-0.04
$\bar{h}$	11.92	0.96	2.13	.	7.57	1.02	0.91	21.49
$\sum h$	-22.58	0.44	1.67	-100	10.04	1.69	0.98	38.87
$\bar{z}$	91.70	-4.56	52.81	.	37.97	5.30	8.99	99.97
$\sum z$	-15.62	-2.74	11.65	-100	10.94	1.89	2.10	40.03
$\bar{c}$	18.79	-2.66	-7.55	.	41.54	3.50	28.43	77.23
$\bar{v}$	6.34	-3.72	-6.47	-199.34	28.45	3.58	20.67	36.08
jobs	-1644	0	0	-1644	13.5	0	0	13.5
taxes	110.56	-2.39	199.17	-86.21	23.13	1.54	-50.08	71.67
After lump-sum transfers								
$\bar{c}$	35.51	14.06	9.16	.	45.04	7.00	31.92	80.73
$\bar{v}$	53.00	40.02	41.48	-174.02	71	36.56	64.64	78.32

Notes: $\bar{w}$, $\bar{z}$ and $\bar{c}$ in €; $\sum h$ and $\sum z$ in %; jobs in 1000s; taxes in million €; jobs and taxes extrapolated to population; $\bar{w}$, $\bar{h}$, $\bar{z}$, and $\bar{c}$ conditional on employment, $\bar{v}$ based on all individuals
Source: SIAB; own calculations.

are retained. Note that the withdrawal of the subsidy implies high implicit marginal tax rates in the interval between 325 and 800 €.

This hypothetical reform makes the tax schedule more generous. As discussed above, extensive margin responses in our model are represented by the reduced-form parameter θ which describes the group of individuals who only supply labour due to the tax exemption. We need to make two assumption in this respect. We assume, first, that no additional individuals enter the labour market due to tax subsidies beyond 325 €. We, second, need to specify how labour supply of type $f2$ workers varies with the average tax rate. In the status quo, type $f2$ workers do not accept any jobs with earnings above the threshold where the average tax rate jumps upwards. Now, the average tax rate increases gradually with earnings beyond the threshold. Type $f2$ workers will thus also accept some jobs above the threshold. We assume that $\theta(1 - H(t(z)))nf$ workers accept a job offer with earnings z where $H(.)$ is the cumulative uniform distribution (equation (4.16)). Intuitively, this implies that type $f2$ workers gradually leave the labour market when earnings – and therefore avarage taxes – increase.

$$H(t(z)) = \begin{cases} 0 \quad \text{for } z \in (0, 325] \\ \frac{t(z)-325}{800-325} \quad \text{for } z \in (325, 800] \\ 1 \quad \text{for } z > 800 \end{cases} \tag{4.16}$$

The resulting earnings offer distribution is smooth (figure 4.6). Most firms which offered jobs at the threshold in the status quo now offer jobs with higher earnings (right panel of figure 4.6). Some firms reduce their earnings offer, though. The likelihood of earnings farther below the threshold is not affected. The offer distribution in the reform scenario is steeper above the threshold with the highest earnings offer being considerably lower than in the status quo. The reason for that firm response is twofold. First, the high implicit marginal tax rate induced by the withdrawal of the subsidy comprises the incentive for workers to reduce earnings. Second, a marginal increase in offered earnings result in a marginal decrease of potential type $f2$ workers. Relative to the status-quo all groups of workers increase their average gross earnings (table 4.7, right panel). This is driven by both wage rate and hours effects. Gross earnings of type $f2$ workers increase the most. They now also accept jobs with earnings of more than 325 €. With nine Euros per month the average increase in gross earnings is fairly modest for type $f1$ worker. This is in particular true when compared to the increase resulting from the removal of the tax exemption (52.81 €). This again shows the negative impact of the high implicit marginal tax rate introduced by the reform which has often been found in the previous literature on making work pay policies (Bargain et al., 2010).

According to the effect on utility both groups of type f workers - those who only work due to tax subsidies and those who would work anyway - benefit substantially from the reform. Besides the modest increase in gross earnings type $f1$ workers mainly benefit from the additional tax subsidies beyond 325 € leading to substantially higher net earnings. Type $f2$ workers additionally benefit from a slight increase in employment because the likelihood of an acceptable offer increases for this worker group. As type s workers are not eligible their increase in utility is much smaller but still positive. The smoothing of the discontinuity is thus also beneficial for workers not directly affected.

Overall hours and gross earnings increase by approximately ten percent. Total tax revenue increases by 23 million €. From the government's point of view the positive behavioural responses are thus overcompensating the increase in generosity of the tax schedule.

A budget neutral version of the reform redistributes the additional tax revenues evenly to all individuals as lump-sum transfers. This clearly increases the positive effects on net earnings and utility for all groups of worker (table 4.7). With respect to utility replacing the notch by a kink is Pareto improving at the group level.

4.7 Conclusion

We present a simple equilibrium job search model in which firms tailor their offers to the aggregate preferences of job seeking candidates. As a result, a worker's labour supply choice set depends on other workers' preferences. This *population labour supply effect* represents a channel through which firms mediate responses to tax policies - even when labour costs are not directly affected.

We apply our model to a unique setting where we observe strong reactions to a *making work pay* policy for workers for who the policy is not directly relevant. In Germany earnings below a threshold are exempted from taxes and employees' SSC generating a substantial discontinuity in the budget set. These so-called Minijobs aim at facilitating new (formal) employment relationships. In the period of analysis this special tax treatment did not apply for second jobs. Yet, second job earnings feature strong bunching at the threshold. We interpret this as compelling evidence for the existence of a population labour supply effect.

Exploiting this specific institutional setting we set-up a model which allows for a discontinuous tax schedule. We are able to rationalise the discontinuous bunching in the earnings distributions of main and second jobs as firm responses. We structurally estimate our model to evaluate the Minijob tax exemption as well as counterfactual policy reforms which remove the discontinuity induced by the tax exemption.

We find that the tax exemption generates roughly 1.5 million jobs for workers who would otherwise be out of the labour force. It however has substantial negative effects on earnings of individuals who would also work without a subsidy. Removing the tax exemption increases earnings of the latter by on average over 50 €/month.

Replacing the Minijob tax exemption by a smooth subsidy schedule aims at preventing the distortions generated by a discontinuous tax schedule while retaining the positive employment effect. We show that replacing the notch by a kink prevents earnings bunching at the threshold. Total earnings in the low-paid market increase by over ten percent. Although the new tax schedule is more generous, total tax revenues increase as well. With roughly 10 €/month the positive effect on earnings of individuals who would also work without a subsidy turns out to be considerably smaller than when the tax exemption is removed, though. The reason is that the gradual withdrawal of the subsidy implies high implicit marginal tax rates. The effect being still positive implies that a discontinuous tax design can create additional distortions. Further, while removing the tax exemption generates a group of workers unambiguously losing utility, replacing the notch by a kink is Pareto improving at the group level.

Negative effects on earnings of already employed individuals is a usual side effect of MWP policies like the Earned Income Tax Credit in the US or the Working Family Tax Credit in the UK (Blundell, 2000; Bargain et al., 2010). We show that firms contribute crucially to these earnings effects. In both reform scenarios firms offer more jobs above the threshold than in the status quo. When subsidies are gradually withdrawn firms respond by reducing their earnings offers. Note that firms' taxes are not affected. Firms' responses represent reactions to changes in labour supply preferences or, put differently, population labour supply effects. We thus complement other work which discusses how unintended consequences of making work pay policies unfold in equilibrium (Rothstein, 2010; Andrew, 2010; Kolm and Tonin, 2011; Tazhitdinova, 2015). While these studies focus on channels like incidence or displacement effects we add the population labour supply as potential channel.

Firm responses to the Minijob regulations and the potential reforms also affect side job earnings. They are shown to increase when a smooth subsidy schedule is introduced translating into an increase in utility. Interestingly, workers with side jobs turn out to lose when the tax exemption is removed. The reason is that the group of workers who enter the market due to the tax exemption are estimated to have a higher job offer arrival rate inducing firms to post more valuable jobs. For workers with side jobs, this positive externality overcompensates the negative effect of the discontinuous tax schedule.

This shows that the population labour supply effect is pervasive and goes far beyond the visible bunching reaction. It has also implications beyond tax policy design. Less dominant groups in the labour market will find it difficult to get acceptable offers if their preferred hours differ from average preferences. Population labour supply effects might, for example, play a role in reinforcing low levels of labour supply of women when a substantial part of women desire working part-time.

A main limitation of our model is that it resorts to a reduced-form parameter to describe the extensive margin. Simulating a change in taxation thus requires additional assumptions. In a future version of this paper we will exploit the reform in 2003 to estimate the extensive margin as a function of the average tax rate. In 2003 the Minijob threshold was increased to 400 €/month. Gradually decreasing SSC subsidies beyond the threshold smoothed the discontinuity in the SSC schedule. Side jobs became eligible for the tax exemption as well (see Appendix D.11 for more details). The former two measures change the average tax rate at different levels of earnings. This can then be linked to the number of new low-paid jobs as main jobs.

At the same time our model offers a new perspective on the reform in 2003. One objective of the reform is to attenuate the distorting effect of the Minijob non-linearity by (partly) smoothing the drop in net earnings at the threshold. The extension of the tax exemption to side jobs antagonises this aim, though. The fraction of workers in the market eligible for the tax exemption increases raising again the incentive for firms to offer jobs below or at the threshold. This is in particular true because the discontinuity for side jobs is especially large. They, first, have on average a larger marginal income tax rate and, second, they are not eligible for the SSC exemptions beyond the threshold which partly smoothed the notch for low-paid main jobs. This contributes to explain that the discontinuity at the threshold in the distribution of earnings from main jobs rather increased contrary to the objective of the policy (figure D.3 in the Appendix).

General Conclusion

The analyses in chapter one and two consistently indicate that economic and legal incidence coincide for social security contributions (SSC) in Germany. Employers and employees thus seem to share the burden of SSC to health and pension insurance. Although both chapters exploit earnings caps of SSC for identification, the consistency of the results is significant as both analyses complement each other. Chapter one exploits an increase in the earnings cap of SSC to health insurance in East Germany as a natural experiment. It aims at measuring wage rate responses to an actual increase in the burden of SSC for given earnings. It thus rather focuses on short-term responses. When substantial adjustment frictions are present however short-term might deviate from long-term responses. Chapter two therefore aims at identifying long-term incidence. The empirical framework exploits the cross-sectional variation generated by earnings caps. Short-term adjustment frictions as explanatory factor are thus less convincing. Moreover, both analysis are based on different high-quality administrative data sets and different periods of analysis.

Several limitations have both approaches in common, though. Earnings caps of SSC are high up in the earnings distribution. The threshold for health insurance in West Germany for instance is approximately located at the 75th quantile which is by far the lowest of the four earnings caps analysed. The results are thus not necessarily transferable to employees with lower earnings. Their labour demand elasticity is for example found to be higher (Hamermesh, 1987) which might result in a higher share of the burden of SSC born by employees with low earnings. Further, both approaches analyse economic incidence as individual wage rate effects. Potential forward shifting to prices is analysed. Burden shifting at a more aggregated level like a firms' entire labour force is neither captured. When wage bargaining is conducted by unions burden shifting might even take place at the sectoral level. Chapter one therefore repeats the analysis restricted to sectors characterised by low union coverage but results do not change.

From a theoretical perspective the finding of economic coinciding with legal incidence is surprising. In most theoretical frameworks the relative elasticities of labour demand and supply are the crucial factors determining economic incidence. As labour demand is usually considered to be more elastic (Fullerton and Metcalf, 2002), SSC burden is predicted to be primarily borne by employees. The analysis of employment responses in chapter one offers an explanation of the supposed deviation between empirical finding and theoretical prediction. By modifying the bunching approach (Saez, 2010) I find no employment responses to the earnings cap of health insurance. I argue that both results are consistent in a competitive labour market model where economic incidence is determined by the ratio of labour demand and supply elasticities. If both market sides are similarly inelastic, the economic burden is shared in equal parts. Moreover, the finding of non-standard economic incidence is in line with recent studies for different countries which are mostly based on quasi-experimental identification (Saez et al., 2012; Skedinger, 2014; Huttunen et al., 2013).

For policy makers increasing the earnings cap for SSC is interesting as it constitutes a feasible policy for many countries to increase revenue or to shift the burden to high earners. In the UK, for example, the earnings cap for employers' SSC was abolished in 1985 as were most earnings caps for SSC in France in the 1980s. The finding of no employment responses at the intensive margin implies that the efficiency loss of an increase of an earnings cap seems to be negligible. The results on incidence show that employees and employers would share the increased burden. Although it is argued that the approach in chapter two is hardly impacted by short-term frictions, restrictions on the labour market in general might yet impact earnings responses to SSC - even in the long-run. To that effect the findings of chapters three and four are relevant. Compelling evidence that the labour market is importantly influenced by restrictions is presented corroborating previous evidence (e.g. Chetty et al. (2011)). For a substantial fraction of employees actual deviate from desired hours of work (chapter three). This comprises both participation and hours constraints. Employed mothers are found to be rather under-employed while employed fathers would like to decrease their workload. These deviations are found to be rooted in both, supply and demand side factors - albeit the former is particularly true for mothers who are still mainly responsible for child care. Labour demand constraints not necessarily result in under-employment or involuntary unemployment. Patterns of hours constraints are found to vary across occupations suggesting that different working hours norms predominate.

One potential mechanism is brought forward in chapter four. It presents evidence that firms are not able to easily adjust their job offers. They therefore have an incentive to package their hours-wage bundles according to average preferences of job-seeking workers restricting the choice set of any particular individual worker. This could result in deviations between desired and actual hours of work in general and in differing patterns across occupations and sectors in particular.

An equilibrium job search model is structurally estimated to rationalise this channel in the German low-pay sector. Wage earnings below a certain threshold are exempted from SSC and income taxes. These so-called Minijobs stand in the tradition of other *Making Work Pay* policies like the Earned Income Tax Credit in the US or the Working Family Tax Credit in the UK. Earnings exceeding the Minijob threshold are subject to full SSC as well as income taxes. The observed earnings distribution features huge bunching at the threshold. This is also the case for individuals who had additional earnings from another job which serves as compelling evidence in favour of this mechanism. Before 2003, second job earnings had to be taxed jointly with first job earnings. There was thus no incentive for second job earnings to bunch.

Moreover, earnings bunching illustrates the negative earnings responses of already employed workers which is a typical feature of MWP policies (Bargain et al., 2010). Bunching is particularly pronounced in the case of German Minijobs as benefits are withdrawn entirely when earnings exceed a certain threshold. The model set up in chapter four is used to evaluate policies which aim at removing the non-linearities in the tax schedule for low-paid workers.

Entirely removing the tax exemption is found to cost about 1.5 million jobs. It however has substantial positive effects on earnings of individuals who stay in the labour market as they are then not hampered by reduced availability of jobs paying above the threshold. Replacing the Minijob tax exemption by a smooth tax schedule, by contrast, aims at preventing the distortions generated by a discontinuous tax schedule while retaining the positive employment effect. Total earnings in the low-paid market increase by over ten percent. Although the new tax schedule is more generous, total tax revenues increase as well. The positive effect on earnings of individuals who would also work without a subsidy turns out to be considerably smaller than when the tax exemption is removed, though. The reason is that gradually withdrawing the subsidy also induce negative work incentives. The effect being still positive implies that a discontinuous tax design creates additional distortions. Negative effects on earnings of already employed individuals is a usual side effect of MWP policies like the Earned Income Tax Credit in the US or the Working Family Tax Credit in the UK (Blundell, 2000; Bargain et al., 2010). Our findings imply that firms contribute crucially to these negative earnings effects supporting recent research underlining the firms' role in responses to tax reforms in general (Kopczuk and Slemrod, 2006) and bunching in particular (Chetty et al., 2011; Best, 2014; Gudgeon and Trenkle, 2016).

The design of the in-work benefit for parents evaluated in chapter three aims at minimizing negative incentives. Eligibility is conditional on working hours exceeding a certain amount. The size of the benefit depends on the individual wage rate instead of individual or household impeding an incentive to reduce hours of work. The objective of the program is to stimulate employment with high hours. Although both spouses in a couple household can draw on the subsidy, the in-work benefit for parents is mainly targeted at secondary earners. When evaluating its effects chapter three accounts for participation and hours constraints. It thus considers that not every individual is able to find a subsidised employment. We find...

Appendix A

Appendix to Chapter 1

A.1 Appendix to Section 1.4

A.1.1 Model with Specified Utility Function

By assuming a small tax change the theoretical implications in section 1.4 can be derived without specifying a utility function. Following Saez (2010), I relax this assumption by means of a quasi-linear and iso-elastic utility function:

$$U(c,y) = c - \frac{n}{1+1/e}\left(\frac{y}{n}\right)^{1+1/e} \tag{A.1}$$

n denotes an ability or preference parameter, c consumption, e the homogeneous compensated elasticity of labour supply with respect to the net-of-tax rate and y gross earnings. The quasi-linearity simplifies the analysis by excluding income effects which is unproblematic as the change in the average tax rate is negligible close to the kink. A cap at earnings level $\bar{y}$ above which the marginal tax rate drops from t_0 to t_1 generates the piecewise linear non-convex budget set (A.2) with an upward kink at $\bar{y}$ and R being non-wage income.

$$c = \begin{cases} (1-t_0)y + R & \forall y \leq \bar{y} \\ (1-t_0)\bar{y} + (1-t_1)(y-\bar{y}) + R & \forall y > \bar{y} \end{cases} \tag{A.2}$$

To find the global maximum subject to a non-convex budget set, Burtless and Hausman (1978) propose to first calculate the unconstrained maxima assuming a linear tax system with the tax rates t_0 and t_1 respectively. These are given by $y_0^* = n(1-t_0)^e$ and $y_1^* = n(1-t_1)^e$. Three cases can be differentiated. If $y_0^* > \bar{y}$ and $y_1^* > \bar{y}$ ($y_0^* < \bar{y}$ and $y_1^* < \bar{y}$), y_1^* (y_0^*) is the global maximum as the individual is always better off on the boundary of the budget set than in the interior. If $y_0^* < \bar{y}$ and $y_1^* > \bar{y}$, the global maximum is the earnings level with the higher absolute utility. The determination of absolute utility is necessary, since double tangencies are possible if budget sets are non-convex. This complicates the analysis in comparison to the discussion of downward kinks in Saez (2010). Equations (A.3) and (A.4) state the indirect utility functions for the budget segments below and above $\bar{y}$ using the optimal interior earnings levels and the relevant budget constraints.

$$
\begin{aligned}
V^{below}(n) &= (1-t_0)n(1-t_0)^e + R - \frac{n}{1+1/e}\left(\frac{n(1-t_0)^e}{n}\right)^{1+1/e} \\
&= n(1-t_0)^{e+1}\left(1-\frac{1}{1+1/e}\right) + R
\end{aligned}
\tag{A.3}
$$

$$
\begin{aligned}
V^{above}(n) &= n(1-t_1)^e - (t_1(n(1-t_1)^e - \bar{y}) + t_1\bar{y} + R) - \frac{n}{1+1/e}\left(\frac{n(1-t_1)^e}{n}\right)^{1+1/e} \\
&= n(1-t_1)^{e+1}\left(1-\frac{1}{1+1/e}\right) - R + \bar{y}(t_1-t_0)
\end{aligned}
\tag{A.4}
$$

The difference between these interior maxima is linear in ability n with a strictly negative slope:

$$
V^{below}(n) - V^{above}(n) = [(1-t_0)^{e+1} - (1-t_1)^{e+1}]n\left(1-\frac{1}{1/e}\right) - \bar{y}(t_1-t_0) \tag{A.5}
$$

Thus, there is a specific ability value $\tilde{n}$ for which it holds that employees are indifferent between an allocation below and above the kink. Employees with higher (lower) abilities prefer the interior solution above (below) the kink. As locating between the earnings level y_0^* and y_1^* never is a global optimum, the earnings distribution should feature a gap. Equating the indirect utility functions below and above the kink determines $\tilde{n}$, the ability of the marginal jumper, as $\bar{y}$ multiplied by the factor $a(t_0, t_1, \bar{y}, e)$:

$$
\begin{aligned}
& V^{below}(n) \stackrel{!}{=} V^{above}(n) \\
\Leftrightarrow\quad & \tilde{n} = \frac{\bar{y}(t_1-t_0)}{(1-1/(1+1/e))[(1-t_0)^{e+1} - (1-t_1)^{e+1}]} = \bar{y}a(t_0, t_1, \bar{y}, e)
\end{aligned}
\tag{A.6}
$$

If $e \to 0$, a converges to one and no gap arises. For $e > 0$ it holds that $0 < a < 1$ ensuring that $\tilde{n}$ lies to the left of the kink. The marginal jumper's earnings response $\Delta\tilde{y}$ to the decrease in the marginal tax rate at $\bar{y}$ is defined by the jump from the interior solution below the kink to the one above:

$$
\tilde{y}_0^* + \Delta\tilde{y} = \tilde{n}(1-t_0)^e + \Delta\tilde{y} = \tilde{y}_1^* = \tilde{n}(1-t_1)^e
$$

Inserting equation (A.6) elasticity e can be inferred:

$$
\frac{\bar{y}(t_1-t_0)}{(1-1/(1+1/e))[(1-t_0)^{e+1} - (1-t_1)^{e+1}]}(1-t_0)^e + \Delta\tilde{y} =
$$

$$\frac{\bar{y}(t_1 - t_0)}{(1 - 1/(1 + 1/e))[(1 - t_0)^{e+1} - (1 - t_1)^{e+1}]}(1 - t_1)^e \tag{A.7}$$

While t_0, t_1 and $\bar{y}$ are known by the tax system, the earnings response $\Delta\tilde{y}$ can be estimated by the extent of the observed gap.

A.1.2 Graphical Evidence

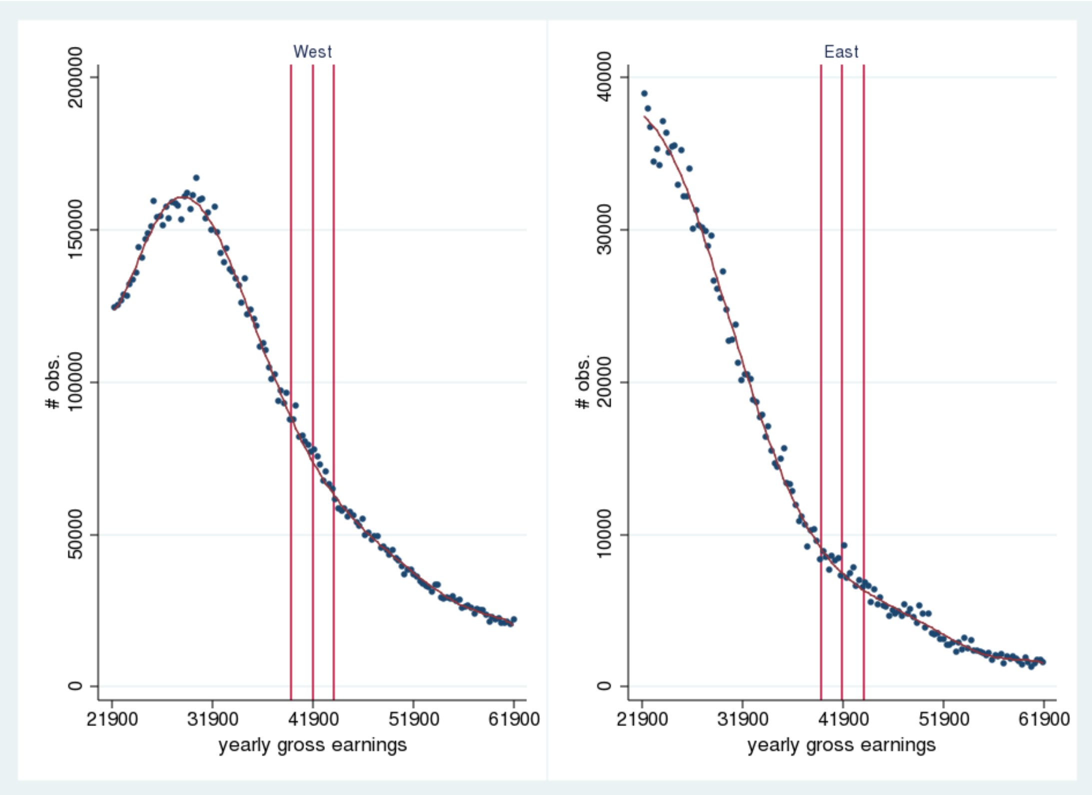
The empirical analysis of section 1.4 aims at identifying the existence and extent of a dip around the earnings cap of health insurance in the gross earnings distribution. Following Chetty et al. (2011) a counterfactual earnings distribution in the absence of an earnings cap is estimated by equation (1.4). Polynomials of degree seven are fit to the count data with earnings potentially affected by the dip being excluded from the estimation.

The corresponding plots to the estimated differences between the actual and counterfactual earnings distributions (section 1.4) are documented here (figures A.1 and A.2). Dots represent the observed, solid lines the counterfactual earnings distributions. The graphs show that the polynomials fit the data well and support the result that both distributions do not deviate.

A.1.3 Simulation of Utility Gain

Chetty et al. (2011) show that behavioural responses could be impeded by frictions as long as the related costs exceed the utility gains. In order to quantify the potential gains I simulate the change in absolute utility generated by moving from the optimal location in the counterfactual situation of a linear tax rate t_0 to the optimal location in the actual non-linear setting. Simulations are conducted for an individual with an ability such that it would be optimal to locate at the cap in the counterfactual situation. It, therefore, yields an upper bound of the utility gain. The gain for the employer is not considered. Tax rates and earnings cap are chosen according to the law applicable in Germany in 2004. The exercise is carried out for three compensated elasticities of labour supply: 0.1 is frequently estimated in static empirical studies; 0.5 and 1, alternatively, are quite high in the distribution of estimates (see Keane (2011) for a recent survey). The utility gain, expressed in yearly earnings, ranges from almost 170 € for $e = 1$ at the earnings cap of health insurance in West Germany to 16 € for $e = 0.1$ (table A.1). This simulation exercise also underlies the choice of the exclusion range which is set to 1800 € below and above the cap corresponding approximately to the earnings changes based on $e = 0.5$.

Figure A.1: Observed and counterfactual distribution around the earnings cap of health insurance in 2004



Notes: The dots represent the absolute frequency in earnings bins of size 300 €. The solid curve is a polynomial of degree seven fit to these frequencies excluding the bins between the first and third vertical reference line. The second reference line represents the earnings cap. This graph is generated by a slight modification of the program code provided by Chetty et al. (2011).
Source: FAST, own calculations.

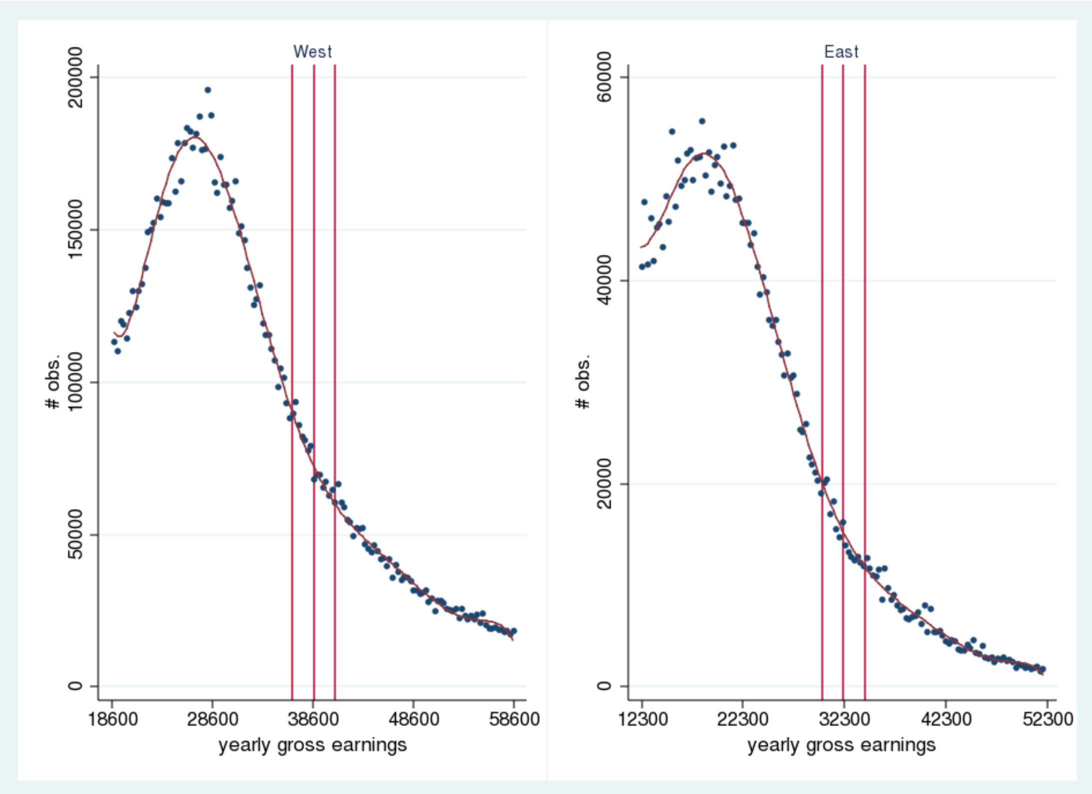
A.1.4 Analysis of earnings cap of preceding year

Table A.2 repeats the empirical analysis of section 1.4 but with respect to the earnings caps effective in the respective preceding years. Although yearly adjustments of the earnings caps are small, they might complicate optimization, especially because they are not announced before September in the preceding year. However, as I do not find any statistically significant dips at the earnings caps of the preceding year I conclude that the yearly adjustments cannot entirely explain the absence of behavioural responses.

A.1.5 Allowing for heterogeneous economic incidence

Relaxing the assumption that economic incidence falls entirely to employees and allowing economic incidence to be heterogeneous as well as to be correlated with the elasticity, the bunching

Figure A.2: Observed and counterfactual distribution around the earnings cap of health insurance in 1998



Notes: The dots represent the absolute frequency in earnings bins of size 100 €. The solid curve is a polynomial of degree seven fit to these frequencies excluding the bins between the first and third vertical reference line. The second reference line represents the earnings cap. This graph is generated by a slight modification of the program code provided by Chetty et al. (2011).
Source: FAST, own calculations.

estimator (equation (1.2) above) becomes:

$$\begin{aligned} MM &= \int \int \int_{y_0}^{y_0+\Delta y(e,s)} h_0(y,e,s)dy\ de\ ds \\ &\approx \int \int dy(e,s)\overline{h_0(y,e,s)}de\ ds \\ &= E_{e,s}[dy]\overline{h_0(y)} \end{aligned} \tag{A.8}$$

s is a parameter that denotes the extent and direction of burden shifting. Equation (A.8) implies that the missing or excess mass B does not represent the average earnings response in the population around the cap but an average earnings response weighted by the joint density of incidence and elasticity. However, since I do not find any missing mass (section 1.4 above) this is not relevant here.

Table A.1: Utility gain of employment response, in €

e	t_0	t_1	$\bar{y}$	y^*	ΔU
1.0	0.21	0.13	41850	46088	169.52
0.5	0.21	0.13	41850	43907	82.51
0.1	0.21	0.13	41850	42254	16.29

Notes: e is the assumed elasticity of labour supply. t_0 and t_1 are the SSC rates below and above the earnings cap $\bar{y}$. y^* is the optimal earnings level and ΔU states the utility gain from moving from $\bar{y}$ to y^*. The chosen values replicate the earnings cap of health insurance where $\Delta t = -7.96pp.$ in 2004.
Source: own calculations.

Table A.2: Bunching analysis of employment responses

	1998			***2004***		
Specification	$\hat{MM}$	$\sigma_{\hat{MM}}$	$\frac{N}{1000}$	$\hat{MM}$	$\sigma_{\hat{MM}}$	$\frac{N}{1000}$
West						
Basic	0.06	0.26	362	−0.52	0.15	494
Degree 9	−0.22	0.28	362	−0.52	0.17	494
Bin Size 100	0.33	0.69	364	−1.33	0.44	499
Excl. 900	−0.08	0.12	391	−0.12	0.08	531
East						
Basic	−0.04	0.44	317	−0.57	0.52	251
Degree 9	0.21	0.49	317	0.28	0.52	251
Bin Size 100	0.09	1.42	318	−1.41	1.54	255
Excl. 900	0.2	0.21	346	−0.44	0.24	273

Notes: $\hat{MM}$ denotes the missing mass normalised by $\overline{\hat{h_0(y)}}$ as calculated by a modified version of the program by Chetty et al. (2011) which also includes a parametric bootstrap procedure to calculate standard errors $\sigma_{\hat{MM}}$. $\hat{MM}$ is positive (negative) if the empirical distribution comprises less (more) individuals than the counterfactual one. N denotes the number of observations used in the estimation. The basic specification estimates polynomials of degree 7 on bin sizes of 300 € excluding a range of 1800 € above and below a cap. The alternative specifications successively vary one of these parameters.
Source: FAST, own calculations.

A.2 Appendix to Section 1.5

A.2.1 The German System of Private Health Insurance

The private health insurance system is a peculiarity within the German health insurance system which has important implications for the quasi-experimental analysis of section 1.5.

Employees with earnings above the income threshold for compulsory insurance, $\bar{\bar{y}}$, may leave mandatory public and opt for private health insurance. Usually this comes into effect the year after $\bar{\bar{y}}$ was exceeded for the first time and only if it can be anticipated that earnings are constantly above $\bar{\bar{y}}$. At the time of the reform, $\bar{\bar{y}}$ coincided with the earnings cap. Therefore, most treated employees were exempted from compulsory insurance and could have been privately insured before the reform. The employees' contributions to private health insurance do not depend on the wage but on individual characteristics. The cap and its increase are, thus, irrelevant[88]. The employers' share amounts to 50 % of total contributions up to the maximum value of employers' SSC in public health insurance. The reform affects the employers' share similarly in both regimes, only if the employers' share hits this maximum value before and after

[88] This is only valid with respect to health but not long-term care insurance. The contribution rate to the latter is relatively small, though.

the reform. This was the case for 72% of privately insured employees[89]. The 2001 reform also increased $\bar{\bar{y}}$ resulting in sudden compulsory insurance for employees earning between the pre- and post-reform caps. Although an exception rule allowed this group of employees to apply for a permanent exemption from compulsory insurance, most of them switched to public health insurance (table A.3). This might have increased or decreased an employee's share depending on the actual contribution rates; the employers' part increased unambiguously[90].

Privately and publicly insured employees cannot be differentiated within the SIAB data set. Another administrative data set, the employment panel of the Institute for Employment Research (BAP), includes the health insurance status, however, and allows to analyse a group similar to the treatment group in 2000 and 2001. Before the reform roughly one quarter of the treated employees was privately insured ("no contribution", table A.3). Another quarter was voluntarily publicly insured[91]. Thus, nearly half of the treated employees were compulsory publicly insured in 2000. They might have exceeded the threshold for the first time or just due to irregular payments. The lion's share of employees earning between the pre- and post reform cap did not seem to opt for staying in the private health insurance system such that over 90 % of the treated employees were publicly insured in 2001 (table A.3). This means that assuming that everyone is constantly publicly insured is justifiable for the treatment of employers but problematic for employees.

Table A.3: Status of health insurance, employees in treatment group

Status	2000	2001
no contribution	25.35	9.19
compulsory contribution	47.64	84.99
voluntary contribution	27.04	5.82

Notes: "no contribution": privately insured employees; "compulsory cont.": employees who are in the compulsory public health insurance system; "voluntary cont.": voluntarily publicly insured employees. The latter declaration is voluntary and it just includes employees whose employers transfer the contributions directly. If that does not apply employees are coded as "no contribution".
Source: BAP, own calculations

[89] This refers to my own calculations based on the socio-economic panel (SOEP). See Wagner et al. (2007) for a detailed description of the data set.

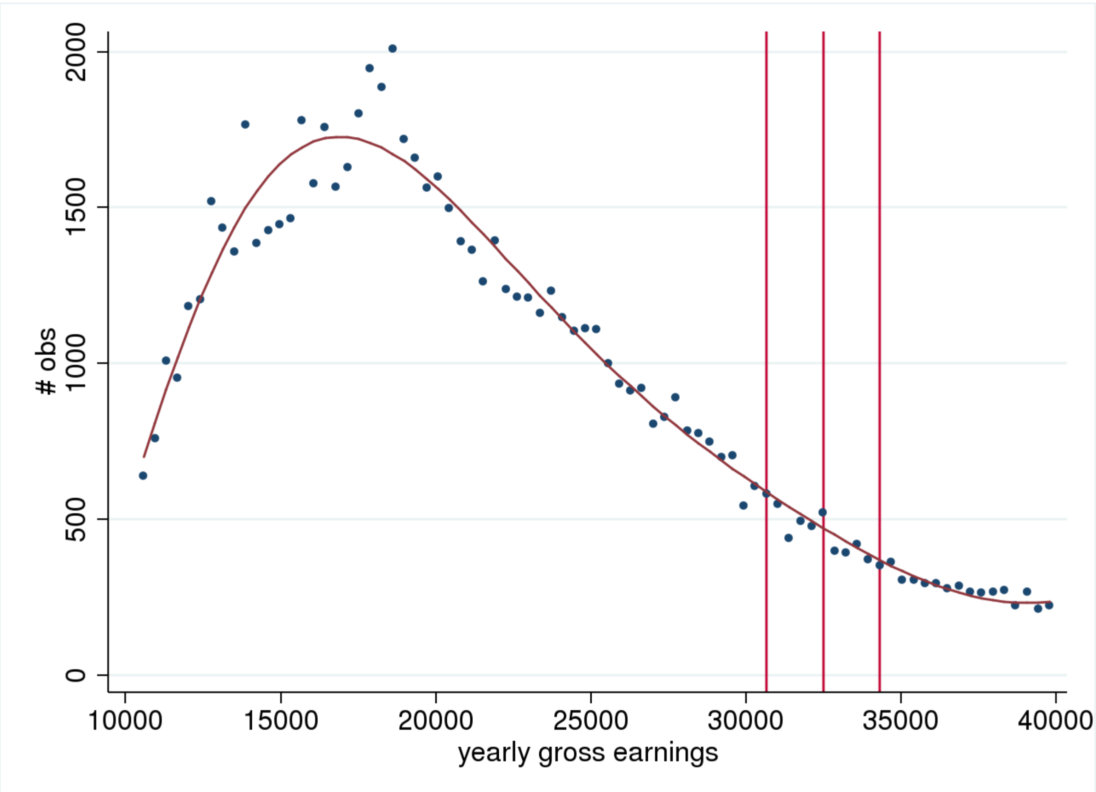
[90] Strictly speaking, a slight decrease is possible because the upper threshold is calculated based on the average contribution rates which differed between health insurance companies in 2001.

[91] However, this declaration is voluntary and it just includes voluntarily publicly insured employees whose employers transfer the contributions directly. If that does not apply employees are coded as privately insured. Hence, the share of privately insured (table A.3) represents an upper bound.

A.2.2 Bunching Analysis of the Earnings Distribution in 2000

As working hours and hourly wage are not observed in the SIAB, identification of economic incidence requires the assumption that the increase in the earnings cap of health insurance did not provoke hours reactions (Saez et al., 2012). In order to support this assumption, I repeat the bunching analysis of behavioural responses to the earnings cap (section 1.4) for 2000, the year before the evaluated reform. The analysis is based on the sample of employees included in the reform evaluation. There does not seem to be a deviation between the empirical and counterfactual earnings distribution around the cap (figure A.3); the point estimate is not statistically significant. This implies that the drop of the marginal SSC rate at the cap does not evoke hours reactions.

Figure A.3: Observed and counterfactual distribution around the earnings cap of health insurance in 2000



Notes: The dots represent the absolute frequency in daily earnings bins of size 1 €. The solid curve is a polynomial of degree 7 fit to these frequencies excluding the bins between the first and third vertical reference line. The second reference line represents the earnings cap.
Source: SIAB, own calculations.

A.2.3 Testing the Validity of an Alternative Control Group

The main disadvantage of the control group used in section 1.5 is that treated and control observations differ with respect to their position in the earnings distribution. This restriction could be relaxed by using a control group consisting of West German employees. They are not affected by the reform as their earnings cap increased only slightly by 1.2 % in 2001. On the other hand the cap is much higher in East Germany on relative terms. In 2000 it corresponded to 160 % of the median wage in my sample; the value for West Germany was just 135 %. Defining the control group based on the position in the earnings distribution is, thus, not feasible. The upper threshold would be too close to the cap of pension insurance above which earnings are censored. Therefore, I use West German employees with earnings above the West German cap as control observations.

I plot yearly changes in gross earnings over time for the treatment and the West German control group to compare their trends (figure A.4). The earnings change in t is defined as $ln(w_t/w_{t-1})$. The curves of both groups are far from parallel implying that the validity of the West German control group is dubious. I, therefore, do not use this control group for the analysis.

A.2.4 Labour Costs and Net Earnings

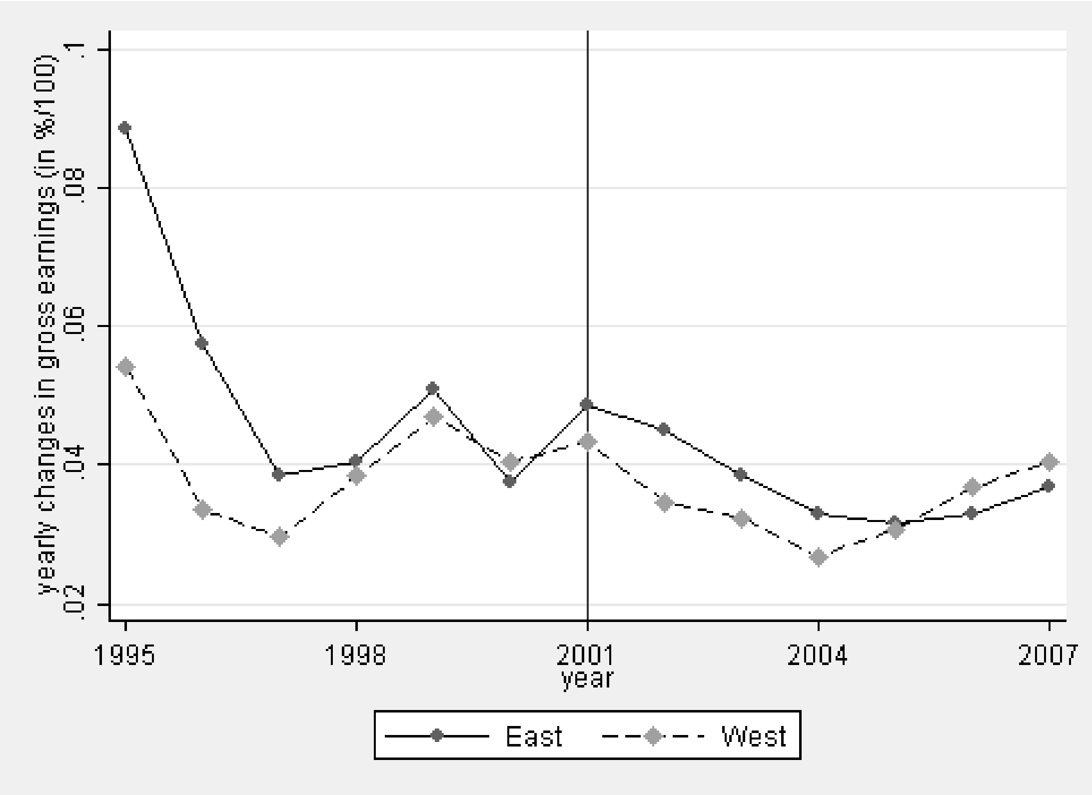
It is argued that if the reform had no impact on gross earnings there is a deterministic effect for labour costs (z) and net earnings (c; section 1.5.3). As I do not observe these quantities they have to be inferred from gross earnings as follows:

$$z = \begin{cases} y(1+t) = \frac{n}{1+t*s}(1+t) & \forall y \leq \bar{y} \\ y + \bar{y}t = n - \bar{y}t*s + \bar{y}t & \forall y > \bar{y} \end{cases} \tag{A.9}$$

$$c = \begin{cases} y(1-t) = \frac{n}{1-t*s}(1-t) & \forall y \leq \bar{y} \\ y - \bar{y}t = n - \bar{y}t*s - \bar{y}t & \forall y > \bar{y} \end{cases} \tag{A.10}$$

Doing so implicitly assumes that everyone is in public health insurance which is not true. This bias is much less important for labour costs than for net earnings (Appendix A.2.1). Therefore, the yearly changes are only plotted for labour costs (figure A.5). As for gross earnings the development of control and treatment group is almost parallel. One exception is the year 2001 when labour costs in the treatment group increased much stronger than in the control group suggesting a positive treatment effect. This eye test is supported by the estimation based on labour costs (section 1.5.4).

Figure A.4: Yearly earnings changes, East vs. West German employees



Notes: Yearly earnings changes in year t are defined as $ln(w_t/w_{t-1})$. These values are not comparable to average earnings increases of the whole population. First, I analyse a specific group of rather well earning employees and, second, the adjustments described in section 1.5.2 result in higher earnings increases.
Source: SIAB, own calculations.

A.2.5 Estimation Table

The estimation results (table A.4) of all covariates refer to the basic specification (section 1.5.4).

A.2.6 Robustness Checks

I present the results of alternative specifications which show the robustness of the main findings (section 1.5).

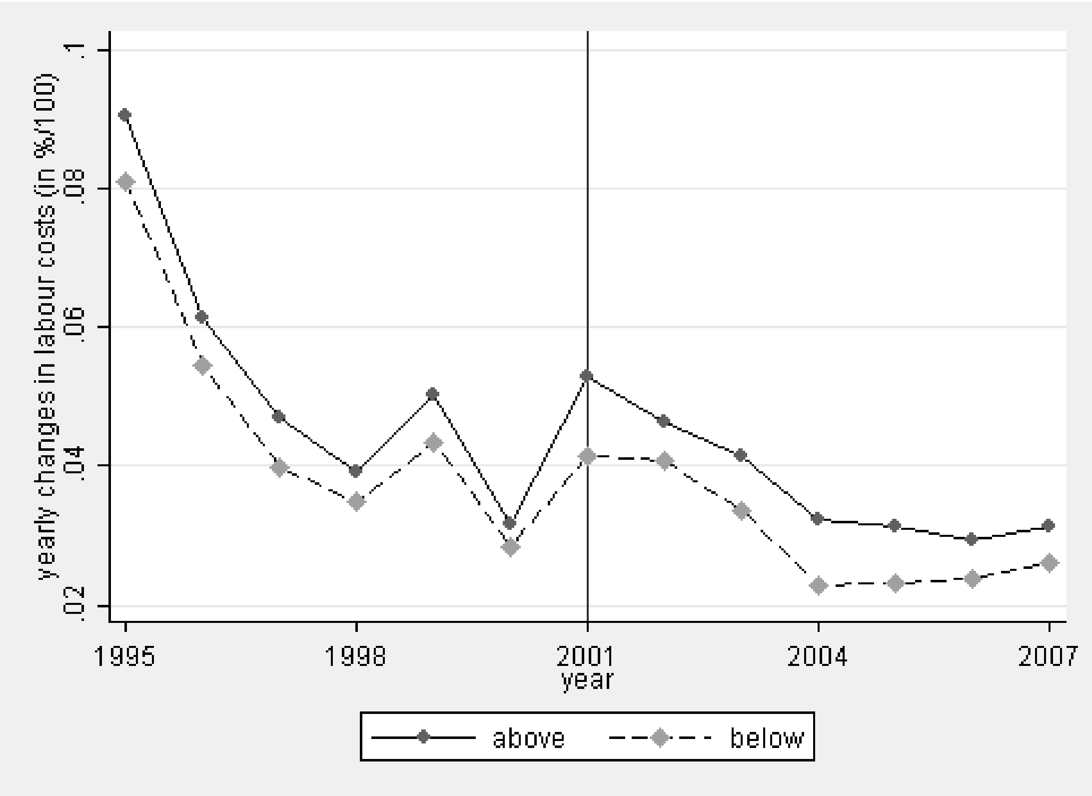
Alternative Classification of Treatment Group. An alternative treatment group consists of employees earning more than the earnings cap in the initial period of a pair of years (columns (1) and (2) of table A.5). This approach generates misspecification as, first, employees who would have earned less than the pre-reform maximum in 2001 anyway are falsely allocated to the treatment group. Second, employees who cross the cap in the reform year are falsely allocated to the control group. On the other hand, this approach is less prone to endogenous

Table A.4: Marginal Effects

	(1)
2001	−0.0008
TG	0.0025***
TG*2001	0.0012**
Male	0.0019***
Age	−0.0043***
Age^2	0.0000***
Tenure	−0.0003
$Tenure^2$	0.0000
Experience	−0.0003
$Experience^2$	0.0000**
Education (ref: Vocational training)	
Intermediate school degree or lower	−0.0025
Grammer school degree	0.0016***
University or college degree	−0.0009***
Missing	−0.0017
State (ref: Thuringia)	
Brandenburg	−0.0025***
Mecklenburg WP	−0.0022***
Saxony	−0.0015***
Saxony Anhalt	−0.0018***
Job position (ref: Trained worker)	
Untrained worker	0.0012
Foreman	−0.0058***
Clerk	−0.0046***
Industry sector WZ1993 (ref: Mining)	
Agriculture	0.0156***
Manufacturing	0.0178***
Energy/Water	0.0158***
Construction	−0.0025
Commerce	0.0136***
Gastronomy	0.0252***
Transportation	0.0235***
Financial	0.0110***
Real Estate	0.0117***
Public admin	0.0168***
Education	0.0129***
Health	0.0167***
Misc Services	0.0164***
Industry with high union coverage	0.0044***
Firm characteristics	
Number of employees	0.0000
Median daily income of fulltime employee	0.0001***
Regional unemployment rate	0.0000
1998	−0.0088***
2000	−0.0138***
Constant	0.1362***

Notes: ***/**/* denote significance on the 1 %, 5 % and 10 % level respectively. The estimation is based on the unrestricted treatment group and corresponds to column (2) of table 1.3. The outcome variable in t is defined as $ln(w_t/w_{t-1})$. Self-employed, employment spells lasting less than a year as well as job-to-job changes are excluded. The analysed period lasts from 1997 to 2001. Standard errors are bootstrapped by Stata's *bsqreg* command.
Source: SIAB, own calculations.

Figure A.5: Yearly labour cost changes, East German employees above and below the earnings cap for SSC



Notes: Yearly labour cost changes in year t are defined as $ln(z_t/z_{t-1})$. The additional fee for childless employees, introduced in 2005, is omitted. The adjustment of SSC which came into effect in July 2005, are considered as of 2006.
Source: SIAB, own calculations.

compositional changes. The result for yearly changes is almost unchanged. The estimate for two-year changes increases slightly. This suggests that the main specification does not suffer from the positive bias induced by the potential endogeneity discussed in section 1.5.3.

Period 1994-2007. The window of observation is extended to six pre-reform and six post-reform years as control observations (table A.5, columns (3) and (4)). The point estimates decrease and the significance of the medium-term estimate disappears. In the basic specification I refrain from using post-reform periods as control periods because the reform might have lagged effects. This would violate the stable unit treatment assumption of difference-in-differences estimators as some control observations would be affected by the reform.

Exclusion of sectors with high union coverage. Collective agreements might pose a problem for the identification strategy because the subject of wage negotiations, usually, is a basic wage which is multiplied by fixed factors in order to get wages for different groups.

Table A.5: Reform effects on relative increase of gross earnings, labour costs and net earnings - robustness checks 1

	(1)	(2)	(3)	(4)
	Alternative TG		All periods	
Gross earnings				
2001	−0.0015***	0.0083***	−0.0007	0.0205***
TG	−0.0013***	−0.0024**	0.0031***	0.0085***
TG*2001	0.0007	0.0022*	0.0005	−0.0006
CI l	−0.0005	−0.0003	−0.0007	−0.0029
CI u	0.0019	0.0047	0.0017	0.0018
Labour costs				
2001	−0.0025***	0.0136***	−0.0011**	0.0130***
TG	−0.0026***	−0.0048***	0.0026***	0.0072***
TG*2001	0.0076***	0.0096***	0.0066***	0.0059***
CI l	0.0063	0.0069	0.0053	0.0036
CI u	0.0088	0.0123	0.0078	0.0083
Net Earnings				
2001	0.0000	−0.0001	0.0002	0.0313***
TG	0.0006	0.0012	0.0039***	0.0102***
TG*2001	−0.0099***	−0.0090***	−0.0092***	−0.0092***
CI l	−0.0111	−0.0117	−0.0106	−0.0118
CI u	−0.0086	−0.0063	−0.0079	−0.0067
N	36842	14741	111877	42729
TG pre	13439	3862	50180	17581
TG post	4172	3313	4294	3289
CG pre	14670	4071	52891	18398
CG post	4561	3495	4512	3461
covariates	✓	✓	✓	✓
Δt	1	2	1	2

Notes: ***/**/* denote significance on the 1 %, 5 % and 10 % level respectively. In odd-numbered columns yearly changes are analysed, in even-numbered columns two-year changes. All specifications use the unrestricted treatment group. Gross earnings are observed, labour costs and net earnings are calculated based on gross earnings. Self-employed, employment spells lasting less than a year as well as job-to-job changes are excluded. Individuals are allocated based on the resulting year in columns (3) and (4) and on the initial year in columns (1) and (2). The analysed period lasts from 1997 to 2001 in columns (1) and (2) and from 1994 to 2007 in columns (3) and (4). Standard errors are bootstrapped by Stata's *bsqreg* command.
Source: SIAB, own calculations.

Apart from adjusting these factors it is not possible to differentiate relative wage increases by employees' characteristics. Therefore, employers subject to collective agreements might shift the additional burden of SSC to their entire workforce. This cannot be captured here as the control group would be affected as well. Coverage with respect to collective agreements is rather low in East Germany, however. In 2001, 44 % of all employees were covered by a sector-wide collective aggreement and an additional 12% by one on the firm-level (Hans-Boeckler-Stiftung, 2015). I do not observe, whether a single firm is subject to a collective agreement. Therefore, industry sectors with union coverage rates of more than 50% are excluded (columns (1) and (2) of table A.6). These account for roughly 40% of treated employees in 2001 whereof many work in the public sector. The point estimate of the reform effect becomes more negative, especially in the medium-term. This might suggest that employers not being subject to a collective agreement are more able to shift the additional burden of SSC. Yet, the estimate is still not statistically significant and the confidence interval still rejects full shifting.

Job-to-Job Tranistions and Employment Periods of less than a Year. Earnings changes of job-to-job transitions and employment periods of less than a year are excluded in the basic specification because these observations are more prone to measurement error and labour supply responses (section 1.5.3). The reform may affect these observations more strongly when it is easier for employees (or their employers) to adjust earnings. The robustness check, therefore, includes earnings changes of job-to-job transitions and employment periods of less than a year (table A.6, columns (3) and (4)). The results do not change qualitatively implying that I do not miss part of the story by excluding these observations.

Table A.6: Reform effects on relative increase of gross earnings, labour costs and net earnings - robustness checks 2

	(1)	(2)	(3)	(4)
	Small union coverage		Job-to-job	
Gross earnings				
2001	0.0006	0.0208***	0.0031***	0.0325***
TG	0.0045***	0.0119***	0.0052***	0.0107***
TG*2001	−0.0003	−0.0026	0.0007	0.0004
CI l	−0.0027	−0.0068	−0.0007	−0.0022
CI u	0.0022	0.0017	0.0021	0.0030
Labour costs				
2001	−0.0003	0.0130***	0.0018***	0.0247***
TG	0.0033***	0.0098***	0.0037***	0.0087***
TG*2001	0.0058***	0.0038*	0.0076***	0.0064***
CI l	0.0032	−0.0003	0.0063	0.0040
CI u	0.0084	0.0079	0.0090	0.0088
Net earnings				
2001	0.0019*	0.0325***	0.0049***	0.0444***
TG	0.0065***	0.0146***	0.0077***	0.0143***
TG*2001	−0.0102***	−0.0112***	−0.0105***	−0.0103***
CI l	−0.0128	−0.0152	−0.0119	−0.0128
CI u	−0.0076	−0.0073	−0.0091	−0.0077
N	23895	14509	53375	37300
TG pre	8034	4460	20419	12903
TG post	2506	1902	6410	5768
CG pre	10206	5739	20189	12783
CG post	3149	2408	6357	5846
covariates	✓	✓	✓	✓
Δt	1	2	1	2

Notes: ***/**/* denote significance on the 1 %, 5 % and 10 % level respectively. In odd-numbered columns yearly changes are analysed, in even-numbered columns two-year changes. All specifications use the unrestricted treatment group and allocate individuals based on the resulting year. The analysed period lasts from 1997 to 2001. Gross earnings are observed, labour costs and net earnings are calculated based on gross earnings. Self-employed are excluded. Columns (1) and (2) additionally excludes job-to-job changes and employment spells lasting less than a year. Columns (3) and (4) exclude all individuals employed in a sector with union coverage rates of more than 50 %. Standard errors are bootstrapped by Stata's *bsqreg* command.
Source: SIAB, own calculations.

A.2.7 Mean Regressions

Following Lang (2003), I employ median regressions to identify the treatment effect in the main specification (section 1.5). I re-estimate equation (1.5) by mean regression as an additional robustness check (table A.7). The highlighted row contains the estimates of β, the average treatment effect on the treated. The analysis is based on yearly earnings changes. The treatment group includes employees with earnings above the pre-reform (table A.7, columns (1) and (2)) and post-reform cap (table A.7, columns (3) and (4)). For none of the specifications a statistically significant treatment effect is found. The point estimate based on the unrestricted treatment group and covariates implies that the reform, which increased employees' SSC by 0.9 pp. on average, induced earnings to increase on average by 0.09 pp. (table A.7, column (2)). When the treatment group is restricted to employees earning more than the post-reform cap the point estimate becomes negative (table A.7, columns (3) and (4)). Although this difference might be driven by hours responses (section 1.5.3), it is rather small and not statistically significant.

Neither yearly labour costs nor net earnings changes are directly observed but have to be mechanically inferred from gross earnings (Appendix A.2.4). The significant treatment effects (second and third panel of table 1.3) imply that the finding of no effect on gross earnings does have explanatory power and is not owed to an imprecise estimation.

Table A.7: Reform effects on relative increase of gross earnings, labour costs and net earnings - mean regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	TG 1		TG 2		TG 1	TG 2
Gross earnings						
2001	−0.0024**	0.0042***	−0.0023*	0.0043***	0.0344***	0.0330***
TG	0.0059***	0.0107***	0.0131***	0.0218***	0.0176***	0.0326***
TG*2001	0.0007	0.0009	−0.0015	−0.0016	−0.0012	0.0006
CI l	−0.0024	−0.0022	−0.0088	−0.0086	−0.0056	−0.0104
CI u	0.0038	0.0039	0.0057	0.0055	0.0032	0.0116
Labour costs						
2001	−0.0033***	0.0033**	−0.0029**	0.0036**	0.0269***	0.0259***
TG	0.0045***	0.0092***	0.0106***	0.0192***	0.0157***	0.0289***
TG*2001	0.0072***	0.0074***	0.0093**	0.0093***	0.0051**	0.0102*
CI l	0.0042	0.0044	0.0021	0.0023	0.0008	−0.0008
CI u	0.0103	0.0104	0.0165	0.0163	0.0094	0.0211
Net earnings						
2001	−0.0011	0.0057***	−0.0013	0.0053***	0.0458***	0.0437***
TG	0.0079***	0.0129***	0.0169***	0.0257***	0.0205***	0.0381***
TG*2001	−0.0091***	−0.0089***	−0.0177***	−0.0178***	−0.0106***	−0.0136**
CI l	−0.0122	−0.0120	−0.0250	−0.0249	−0.0151	−0.0248
CI u	−0.0059	−0.0058	−0.0104	−0.0106	−0.0062	−0.0025
N	37188	37169	20780	20771	22848	12660
TG pre	13838	13828	1311	1310	7858	683
TG post	4296	4294	415	414	3289	276
CG pre	14541	14535	14541	14535	8240	8240
CG post	4513	4512	4513	4512	3461	3461
covariates	-	✓	-	✓	✓	✓
Δt	1	1	1	1	2	2

Notes: ***/**/* denote significance on the 1 %, 5 % and 10 % level respectively. Specifications labelled TG 1 (2) include all employees with earnings above the pre-reform (post-reform) earnings cap in the treatment group. In columns (1)-(4) the outcome variable in t is defined as $ln(w_t/w_{t-1})$, in columns (5) and (6) as $ln(w_t/w_{t-2})$. Gross earnings are observed, labour costs and net earnings are calculated based on gross earnings. Self-employed, employment spells lasting less than a year as well as job-to-job changes are excluded. The analysed period lasts from 1997 to 2001. Standard errors are bootstrapped by Stata's *bsqreg* command.
Source: SIAB, own calculations.

Appendix B

Appendix to Chapter 2

B.1 Model extensions

B.1.1 Varying SSC rates

We assume identical statutory SSC rates for employers (r) and employees (e) in our model. Relaxing this assumption yields

$$y = \frac{n}{1 + t_r max(0, s) + t_e min(s, 0)}. \tag{B.1}$$

As in our baseline model, $s > 0$ implies that employers are able to shift part of their burden to their employees. In that case, the distortion of n depends on the SSC rate of the employer, t_r. If $s < 0$, employees are able to shift part of their burden to their employers and the distortion of n depends on the SSC rate of the employee, t_e.

Considering the left and right limit of the density of realised gross earnings when y approaches $\bar{y}$:

$$\begin{aligned} f(\bar{y})_- &= p\left\{\bar{y}[1 + t_r max(0, s) + t_e min(s, 0)]\right\}[1 + t_r max(0, s) + t_e min(s, 0)] \\ &= p\left\{\bar{y} + \bar{y}t_r max(0, s) + \bar{y}t_e min(s, 0)\right\}[1 + t_r max(0, s) + t_e min(s, 0)] \\ &= p(\bar{n})[1 + t_r max(0, s) + t_e min(s, 0)] \end{aligned} \tag{B.2}$$

$$\begin{aligned} f(\bar{y})_+ &= p\left\{\bar{y} + \bar{y}t_r max(0, s) + \bar{y}t_e min(s, 0)\right\} \\ &= p(\bar{n}) \end{aligned} \tag{B.3}$$

The interpretation of the size of the discontinuity depends on its sign. If the distribution exhibits a drop (jump), full shifting is implied by a change of t_r% (t_e%).

B.1.2 Varying SSC rates with asymmetric incidence

When, in addition, employers' and employees' contributions are allowed to be shared independently, gross earnings in terms of ability become

$$y = \frac{n}{1 + t_r s_r - t_e s_e} \tag{B.4}$$

As before, the potential discontinuity in the earnings distribution at the threshold can be characterised as follows:

$$f(\bar{y})_- = p(\bar{n})\left[1 + t_r s_r - t_e s_e\right] \tag{B.5}$$

$$f(\bar{y})_+ = p(\bar{n}) \tag{B.6}$$

The discontinuity, thus, identifies an average of the SSC rates of employers and employees weighted by the corresponding shifting parameters. Separate identification of s_r and s_e would require differential changes of employer and employee SSC rates over time.

B.1.3 Heterogeneous shifting

The analysis in the main text is based on the assumption of s to be homogeneous. Allowing for heterogeneity in s raises the question whether a potential discontinuity in a given distribution at an earnings cap is informative about average economic incidence in the population. We thus use the simulation from the estimation section 2.2.3 (for details on the simulation see Appendix B.2) to check how representative this local estimate is. Homogeneity from the main simulation is relaxed and s is allowed to vary according to alternative distributional assumptions.

The first and second moments are held constant throughout this exercise: the shifting parameter s has a mean of 0.5 and a standard deviation of .239 for each distribution. This means that employers are on average able to shift half of their SSC burden to employees. For an assumed drop in the SSC rate of 7.6% at the cap, this yields a 'true' drop in the gross earnings density of .038 for the average individual. We start with a *uniform* distribution where s varies between 0 and 1. Alternatively we estimate discontinuities for underlying *normal* and *log-normal* distributions of s. Finally we allow for a correlation between the the shifting parameter and earnings. We use a *bivariate normal* distribution (with first and second moments for y and s as above) and a correlation coefficient of .5. This means that employers' ability to shift their SSC burden to employees is increasing in earnings.

For homogenous s we find a discontinuity of -0.037 with the non-parametric and -0.038 with the parametric approach (Tab. B.1). Both estimates match the 'true' discontinuity of $-.038$. Allowing for heterogeneity in s we see that both, the parametric and the non-parametric test employed in this paper yield discontinuity estimates of between -0.036 and -0.037. The estimates also closely resemble the true mean incidence in these scenarios regardless of the distribution of s with confidence intervals increasing slightly. As long as the distribution of shifting parameters is not related to earnings, the locally identified estimate is representative for mean incidence in the sample.

Table B.1: Discontinuity estimates – simulation of heterogeneous shifting parameter

	McCrary		**Polynomials**	
	p.e.	CI	p.e.	CI
Homogeneous s	-.037	[-.057, -.019]	-.038	[-.054, -.024]
Heterogeneous s				
Uniform	-.036	[-.063, -.010]	-.037	[-.052, -.021]
Normal	-.036	[-.058, -.014]	-.037	[-.052, -.022]
Log-normal	-.037	[-.060, -.013]	-.037	[-.052, -.023]
Heterogeneous s, correlation with y				
Bivariate normal	-.046	[-.065, -.027]	-.044	[-.058, -.030]

Notes: p.e. – point estimate, CI – 95% confidence intervall.
A mean of the shifting parameter of $s = 0.5$ and a drop in the marginal SSC rate of 7.6% yield a 'true' drop in density of .038 for the average in individual in the population.
Source: Own simulation.

The picture changes in a situation where the employers' ability of shifting their SSC burden is correlated with earnings. For a positive correlation coefficient of .5 we find a discontinuity estimate of $-.046$ (Tab. B.1). The upward bias in the estimate of mean incidence can be explained by selection: The cap is located in the upper part of the earnings distribution. Given the assumed positive combination with shifting this results in an above-average s for individuals around the cap. The local estimator is no longer able to capture mean incidence in the sample under these circumstances.

B.1.4 Behavioural responses

Alternative margins of adjustment, e.g. quantity adjustments on the labour market, are not modelled in our framework (sub-section 2.2.1). A change of the marginal SSC rate at an earnings cap, however, also entails labour supply and demand incentives. This is particularly relevant as underlying labour supply and demand elasticities might drive both behavioural responses as well as economic incidence. Alvaredo and Saez (2007) provide a model extension with variable labour supply. Assume individuals with skill (or preference for work) n to maximise quasi-linear utility $u^n(c, z)$ which is a function of net earnings c and labour costs z (with $c = z - T(z) = \frac{1-t}{1+t}z$ where T(z) is total tax liability). It holds that $\partial u/\partial c > 0$ and $\partial u/\partial z < 0$

(costs of labour supply). Quasi-linear preferences rule out income effects. This is not restrictive in our framework, though. As the discontinuity at the cap is driven by employees located close to the cap, the change in average tax rate is limited and so is the incentive for income effects. Under neoclassical labour supply a positive labour supply elasticity makes it suboptimal to locate close to the earnings cap (Saez, 2010). That is, if the labour supply elasticity is homogeneous and strictly positive, the distribution of earnings will feature a gap around the cap. Although a potential discontinuity generated by economic incidence would theoretically be present at the borders of the gap, it would be impossible to identify it empirically as the probability mass directly around the cap is zero.

If the labour supply elasticity is heterogeneous and zero for at least some individuals, the earnings distribution would feature a dip instead of a gap (Saez, 2010; Neumann, 2015). Depending on the size of the elasticities, it is theoretically still possible to identify a discontinuity in the density of gross earnings. The discontinuity is in principle not affected by the labour supply behavior. As discussed above a potential discontinuity is informative about economic incidence for individuals who locate close to the earnings cap. Those people have by definition small labour supply elasticities, otherwise they would not sit there. When labour supply behavior and economic incidence are correlated, the incidence estimate is highly selective.

In the case of a dip it is a practical empirical question whether an estimator picks up the curvature of such a dip and is still able to identify a discontinuity within this area. We conducted a number of Monte-Carlo simulations varying the size of the dip due to behavioural responses in order to study the sensitivity of the estimators in that regard (for details on the simulation see Appendix B.2). The non-parametric estimator performs in general better than the parametric estimator. It is able to pick up the curvature generated by behavioural responses and in most cases still identifies a discontinuity at the earnings cap (Tab. B.2). By contrast, estimates based on the parametric approach are completely out of place. The parametric estimator is not able to identify discontinuities in the presence of sizeable dips in earnings densities.

Table B.2: Discontinuity estimates – simulation of behavioural responses

Simulation parameters	**Actual discontinuity**	**McCrary** p.e.	CI	**Polynomials** p.e.	CI
$P(e = 0) = 1; s = 1$	-.075	-.072	[-.094;-.049]	-.076	[-.092;-.06]
$P(e \sim U(0, .7)) = .67$; $P(e = 0) = .33; s = 1$	-.075	-.051	[-.100;0]	.141	[.117;.165]
$P(e \sim U(0, .7)) = .67$; $P(e = 0) = .33; s = 1 - e$	-.054	-.048	[-.100;.003]	.089	[.065;.113]
$P(e \sim U(0, .7)) = .9$; $P(e = 0) = .1; s = 1 - e$	-.050	.033	[-.116; .196]	.251	[.219;.284]

Notes: p.e. – point estimate, CI – 95% confidence interval.
Source: Own simulations.

We start with a scenario where roughly two thirds of all employees have a positive labour supply elasticity e which is simulated to be uniformly distributed in the interval $(0, .7)$. Remaining employees are assigned an elasticity of zero. Economic incidence is assumed to be homogeneous, fully with employees ($s = 1$), and independent from e. The distribution thus features a dip around the earnings cap and, in addition, exhibits a discontinuity at $\bar{y}$. The non-parametric estimate of $-.051$ is smaller than the true value of $-.075$ but remains significant, although the behavioural distortion in general reduces precision. (Tab. B.2).

According to the neo-classical model incidence is determined by the labour supply elasticity ($s = 1 - e$): Highly elastic employees bear less of the SSC burden than individuals with lower elasticities. We model this relationship between e and s in the last two simulations. First, the elasticity is assumed to be uniformly distributed between zero and .7 for two thirds of employees while all others are inelastic as above. The incidence parameter is thus $s = .77$ on average resulting in a true discontinuity of $-.054$. The McCrary test almost exactly delivers that value (Tab. B.2). The better performance in comparison to the first simulation is a combination of a selective local estimate with the downward bias under behavioural adjustments.

Second, we increase the share of people with labour supply elasticities in the interval $(0, .7)$ to .9 which means that the the overall fraction with $e = 0$ is reduced to .1. The dip around the earnings cap is increased and the related average discontinuity decreases to -0.050. The non-parametric point estimate of .33 no longer represents the correct direction of incidence (Tab. B.2). It seems that the combination of a large dip, a small amount of probability mass at $\bar{y}$, and a relatively small actual discontinuity also disables identification based on the non-parametric approach. Taken altogether behavioural responses introduce bias to incidence estimates. When incidence is correlated with elasticities, the local estimator is selective. Under large behavioural responses identification of a discontinuity is lost even with a non-parametric estimator.

Empirical studies have not been able to identify a gap or a dip around earnings caps (Liebman and Saez, 2006; Neumann, 2015). We do not find evidence for behavioural responses either (section 2.3.3). People with high earnings may indeed have small labour supply elasticities. Alternatively adjustment costs could be too high to locate optimally given the relatively moderate increase in utility (Neumann, 2015).

B.2 Monte-Carlo evidence

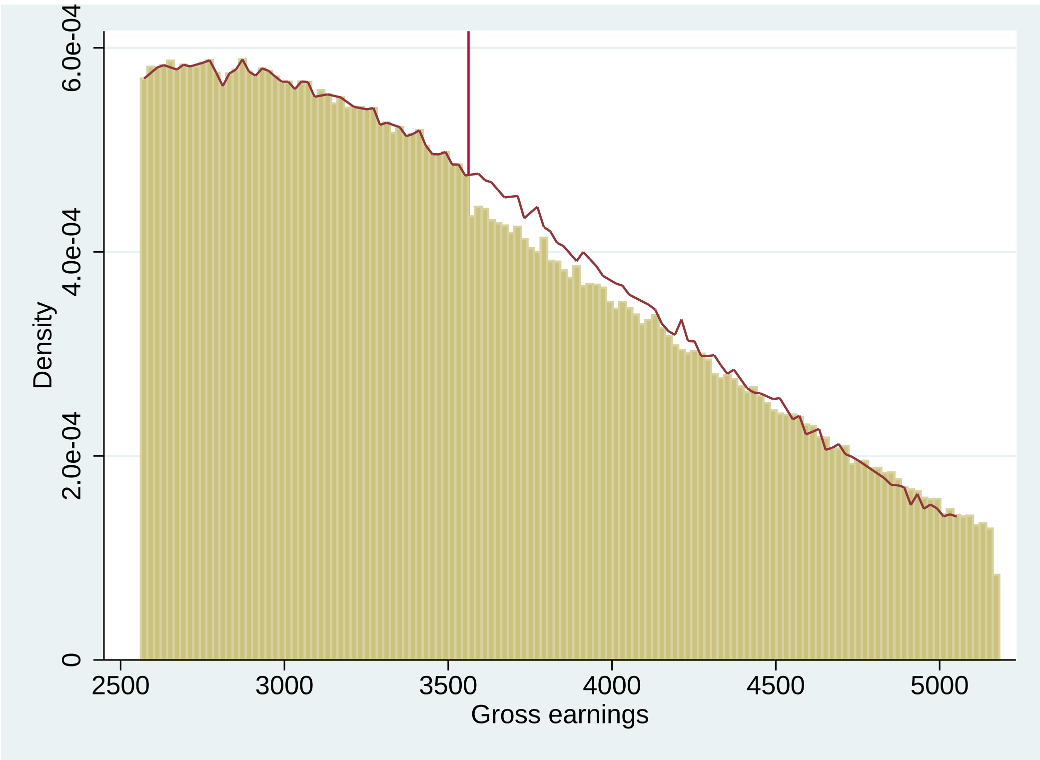
This section provides some Monte-Carlo evidence on the performance of the parametric and non-parametric estimator of a discontinuity in an earnings distribution. Gross earnings distributions $f(y)$ are simulated on the basis of equation (2.3) under the assumption of full shifting of employers' SSC to employees ($s = 1$). SSC rates for employees and employers are identical. We assume the underlying ability n to be normally distributed with $p(n) \sim N(2,913; 4,481)$ and take 1000 random draws of 2,000,000 observations from this distribution. We assume a drop in the marginal SSC rate of roughly 7.6% at 3,562 €. The choice of distributional and tax parameters draws on our empirical application to the health insurance cap in 2006. The resulting density features a clear negative discontinuity (Fig. B.1). The McCrary test and the parametric test identify a discontinuity of -7.2% and 7.6%, respectively, which matches the underlying parameters s and t (Fig. B.2). The confidence interval of the McCrary estimate translates to an interval of $(-1.24, -0.64)$ for the shifting parameter s. The corresponding interval for s based on the parametric test is slightly narrower: $(-1.21, -0.79)$.
The red curve added to Figure B.1 plots a counterfactual density assuming no earnings cap. The deviation from the histogram is not uniform over the whole area above the cap. To see the reason, consider both densities at the earnings level $\tilde{y} > \bar{y}$ (equations (B.7) and (B.8)).

$$f(\tilde{y})_{\text{Cap}} = p(\tilde{y} + \bar{y}ts) \tag{B.7}$$

$$f(\tilde{y})_{\text{No cap}} = p(\tilde{y}(1 + ts))(1 + ts) = p(\tilde{y} + \tilde{y}ts)(1 + ts) \tag{B.8}$$

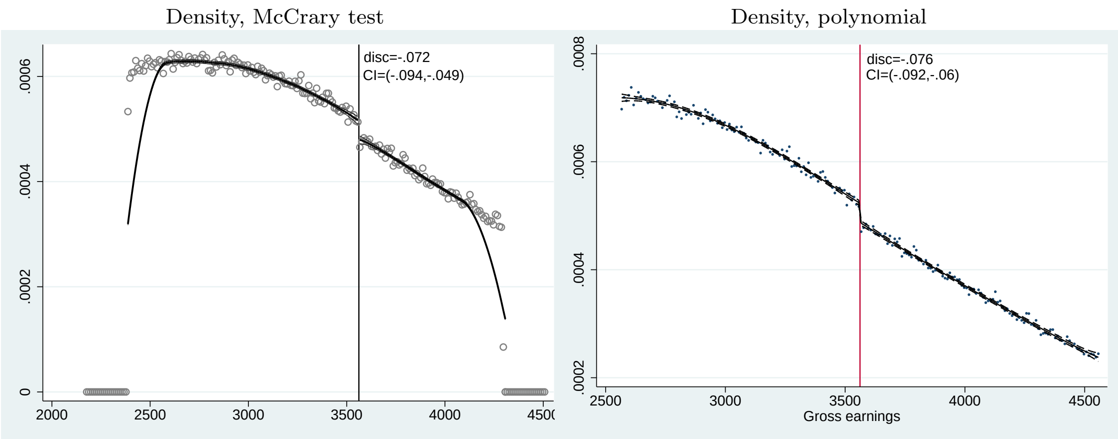
For a monotonously negative (positive) slope of the ability density above $n = \bar{y}$ it holds that $p(\tilde{y} + \bar{y}ts) > (<)\ p(\tilde{y} + \tilde{y}ts)$. The inequality becomes larger with the distance to $\bar{y}$. At the same time, for $s \neq 0$, the factor $(1 + ts)$ constantly increases the density in the case of no earnings cap. Assuming a negative slope, this implies that the deviation between both densities is at its maximum directly at the cap and diminishes slowly until it even might reverse (Fig. B.1). For a positive slope, the deviation increases with the distance to $\bar{y}$.

Figure B.1: Distributions of simulated gross earnings with negative discontinuity at $\bar{y}$



Source: Own simulation.

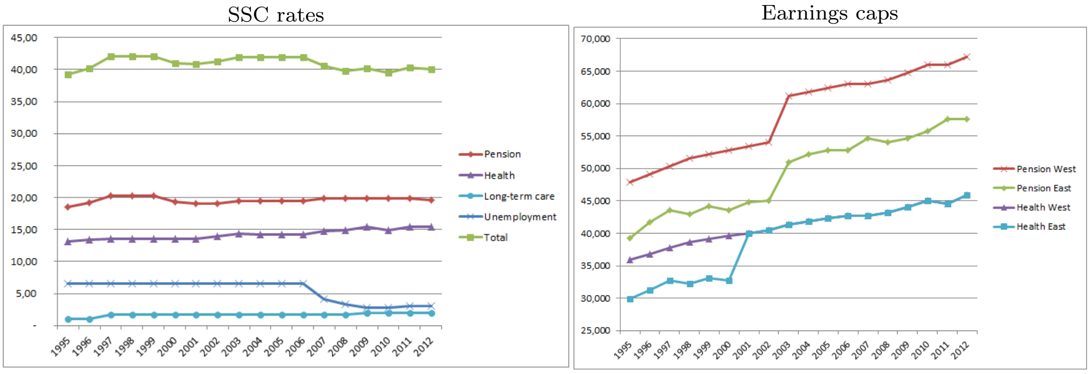
Figure B.2: Discontinuity tests, simulation



Notes: disc – p.e. of discontinuity, CI – confidence interval (α=0.05).
Source: Own simulation.

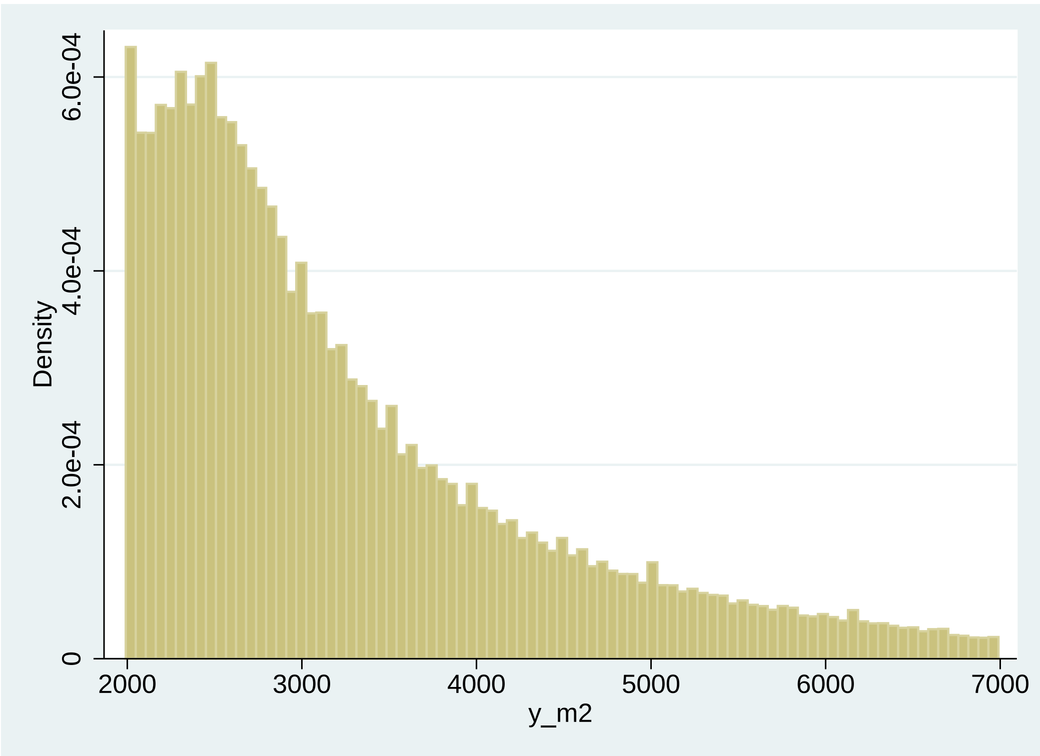
B.3 Additional figures

Figure B.3: Development of SSC rates and earnings caps over time



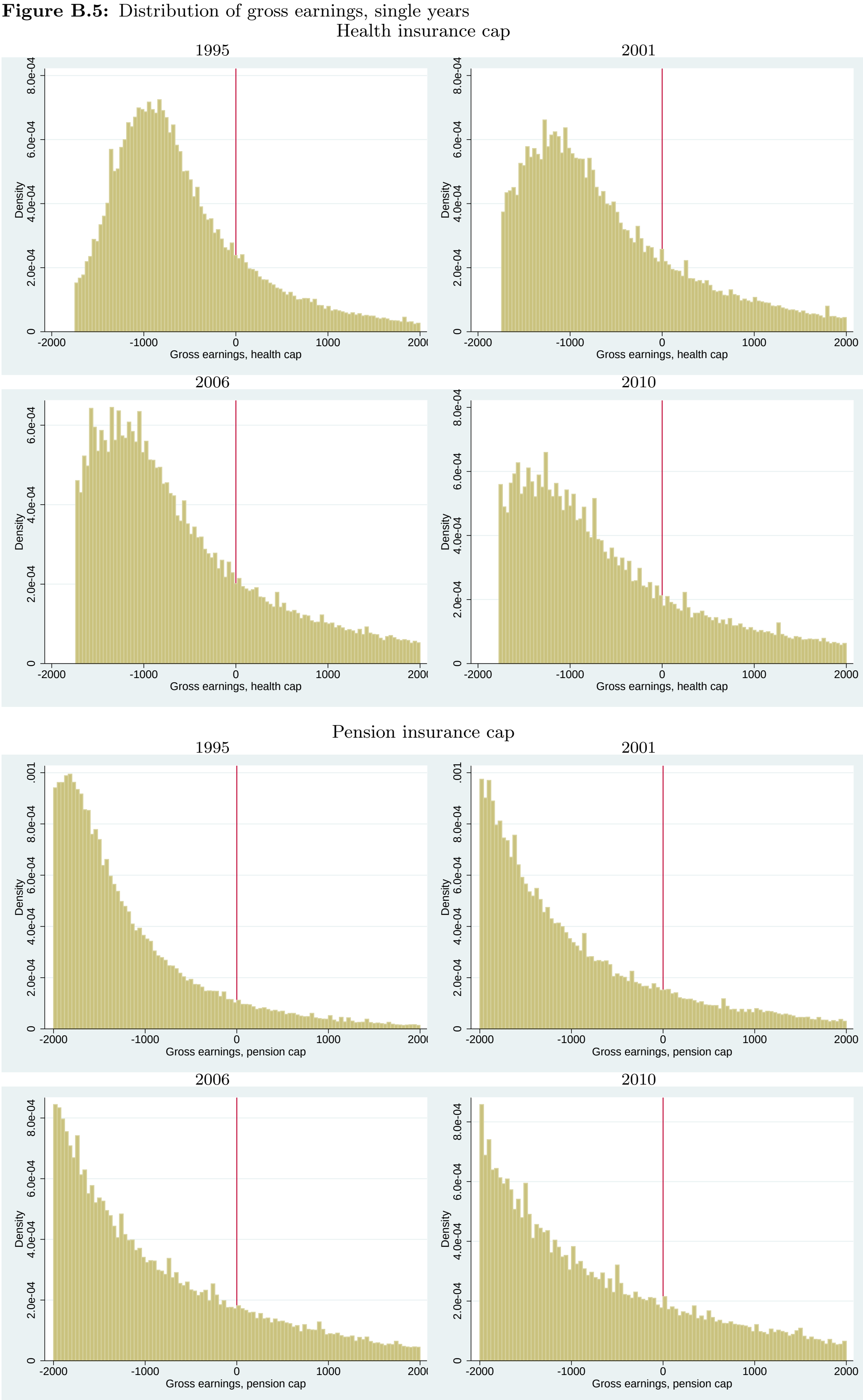
Notes: The additional fee for childless employees, introduced in 2005, is omitted; The change of SSCs which came into effect in July 2005, are considered as of 2006. Until 2006, SSC rates for health insurance varied between health insurance companies and the given numbers are averages. In 2001 for example, it varied between 11.0% and 14.9% Grabka (2004).
Source: German Statistical Office.

Figure B.4: Distribution of non-normalised gross earnings, West Germany, 2006



Source: GSES 2006.

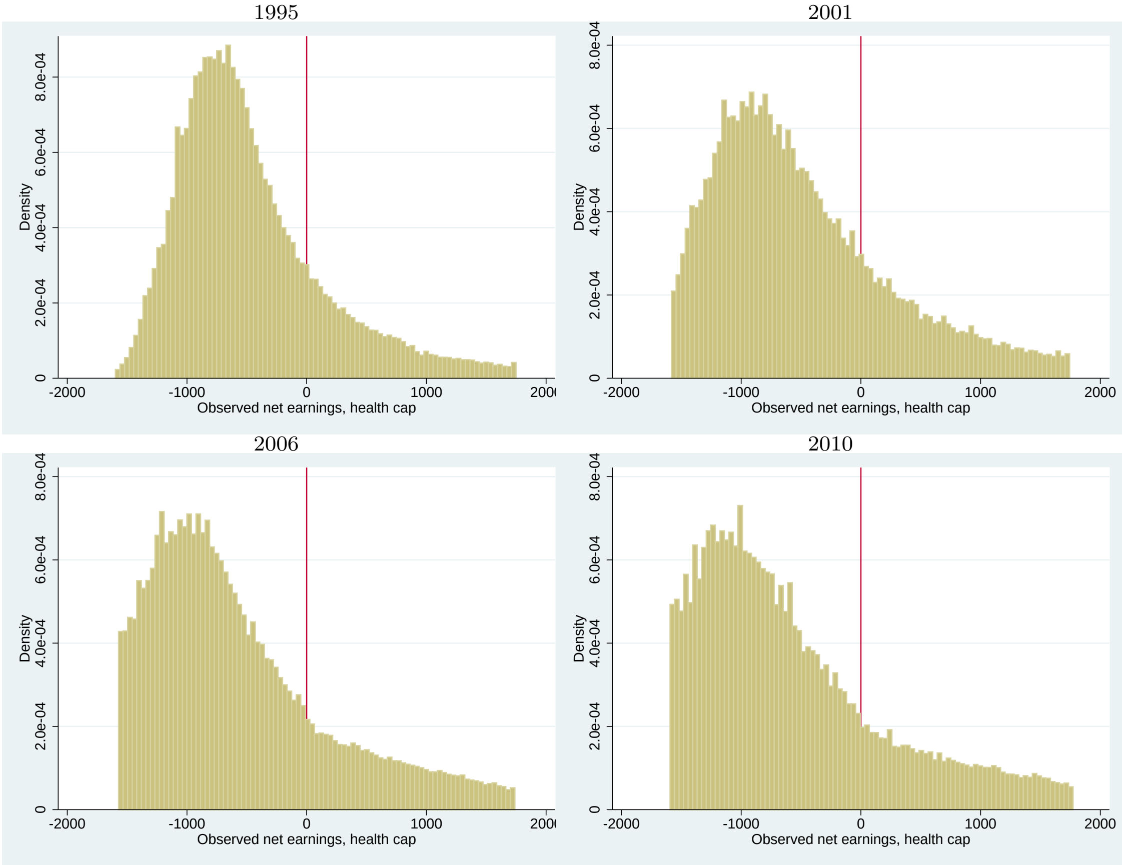
Figure B.5: Distribution of gross earnings, single years



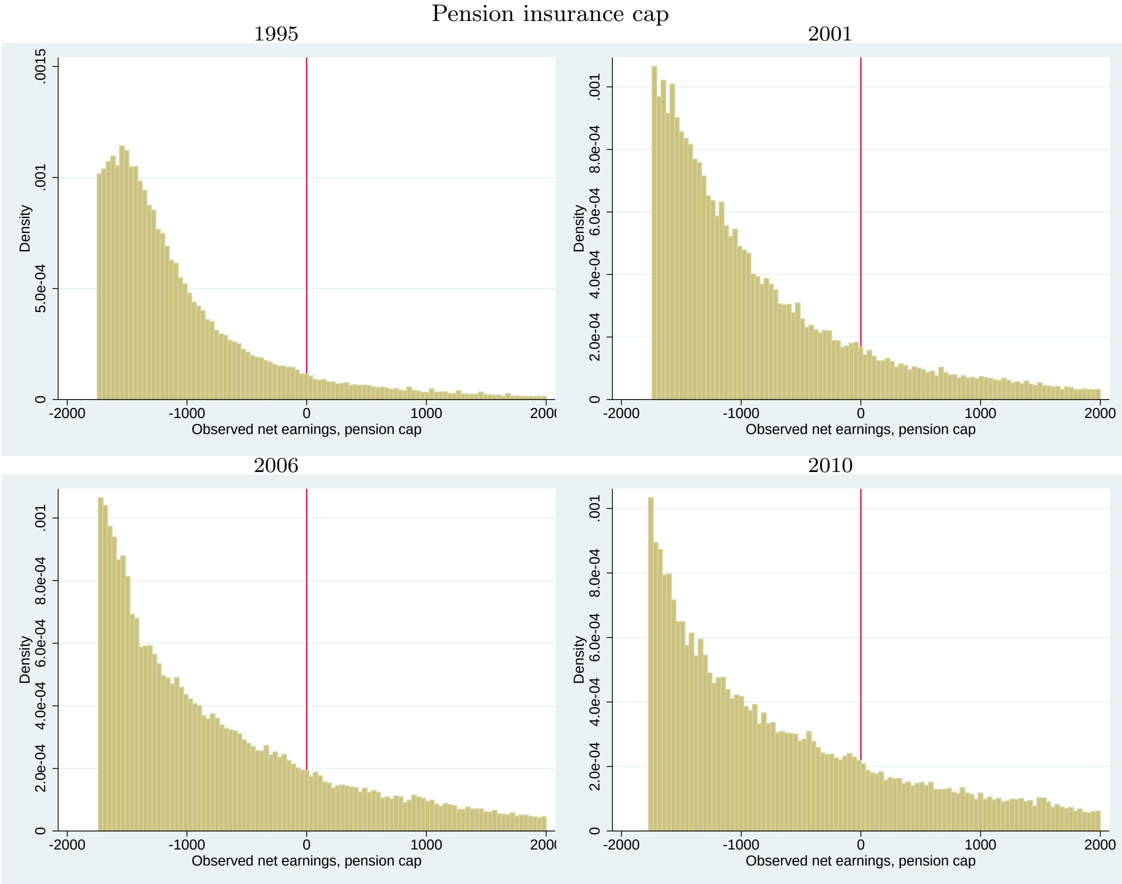
Source: GSES 1995, 2001, 2006, 2010; own calculations.

Figure B.6: Distribution of observed net earnings, single years

Health insurance cap

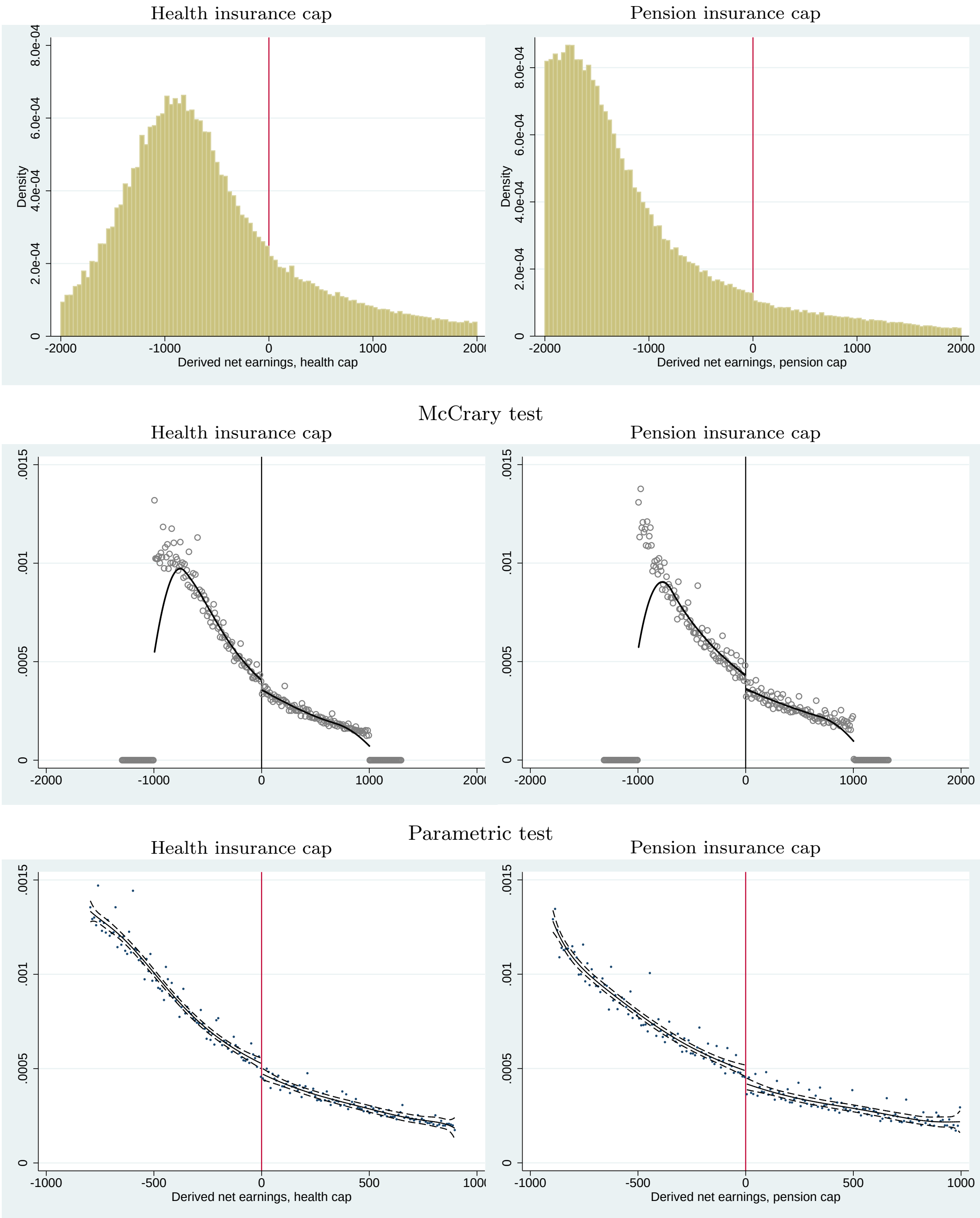


Pension insurance cap



Source: GSES 1995, 2001, 2006, 2010; own calculations.

Figure B.7: Distribution of calculated monthly net earnings, all years pooled



Source: GSES 1995, 2001, 2006, 2010; own calculations.

B.4 Additional tables

Table B.3: Discontinuity estimates – calculated net earnings

	Pension/unemployment cap					**Health/care cap**				
	Δt	**McCrary**		**Polynomials**		Δt	**McCrary**		**Polynomials**	
		p.e.	t	p.e.	t		p.e.	t	p.e.	t
Pooled sample	-.124	-.139	-12.38	-.163	-10.86	-.081	-.096	-13.28	-.113	-13.91
Single waves										
1995	-.126	-.103	-4.54	-.117	-4.39	-.071	-.096	-6.25	-.086	-5.96
2001	-.128	-.150	-7.11	-.195	-6.47	-.076	-.125	-8.62	-.165	-8.32
2006	-.130	-.119	-5.92	-.110	-4.13	-.084	-.114	-11.04	-.124	-8.87
2010	-.114	-.162	-9.27	-.246	-8.22	-.088	-.078	-5.79	-.072	-4.77

Notes: Δt – average drop in marginal SSC rate, p.e. – point estimate, t – t-value (α=0.05).
Source: GSES 1995, 2001, 2006, 2010; own calculations.

Appendix C

Appendix to Chapter 3

C.1 Wage estimation

Hourly wage rates in our model are treated as exogenous and constant for different hours categories. They are derived from reported gross monthly wage earnings and observed working hours including paid and unpaid overtime for the employed. Hourly wages of non-employed persons are predicted on the basis of parameters from wage equations. The estimation is conducted outside of the labour supply model and is based on GSOEP and FiD data pooled over the years 1999 to 2013. Individuals below 18 and above 65 are excluded. We estimate separately for men and women as well as East and West Germany. We control for selection into employment as proposed by Heckman (1979). Estimation results for the wage and the selection equations are presented in table C.5.

C.2 Unobserved heterogeneity: EM algorithm & IIA

C.2.1 EM algorithm

Information on the availability of a given hours category is only available, if an individual desires to or actually works in this state and thus crucially depends on preferences. If preferences and restrictions are not independent - i.e. if, for instance, households with a high propensity for consumption are more or less likely comprised of members with high restriction probabilities - the identification of hours constraints is based on groups of individuals that are not representative for the whole population. This is similar to the standard selection problem in labour economics (Heckman, 1979). In our model preferences and choice restrictions are therefore related through observed and unobserved factors and estimated jointly.

In order to correct the described selection due to unobservable characteristics in the estimation of constraints, we need to account for the distribution of unobserved types in the sub-samples underlying the identification of constraints in different hours categories. We therefore use an expectation-maximization (EM) algorithm (Train, 2009) for the estimation of the type probabilities π. It is appropriate for our case as the conditional likelihood contributions are weighted by the individual (not merely average) type probabilities in the estimation of the parameters.

The parameters are thus estimated conditional on the distributions of unobserved types in the respective sub-samples preventing an omitted variable bias when explanatory variables are correlated with the probability of being a certain type. Individual type probabilities further allow for calculating individual restriction probabilities which are essential to correctly predict policy effects, in particular when effect heterogeneity is analysed.
Methodologically, the EM algorithm we use comprises the following steps (Kabatek, 2013):

1. choose starting values for model parameters (θ^0) and sample group probabilities (π_g^0) for each type $g = 1, ..., 4$
2. calculate individual conditional likelihood contributions L_{ng} for each household n and type g based on current (iteration i) parameters θ^i
3. derive corresponding individual probabilities of types by $p_{ng} = \frac{\pi_g L_{ng}}{\sum_{g'=1}^{G} \pi_{g'} L_{ng'}}$
4. derive new sample group probabilities by $\pi_g^{i+1} = \frac{\sum_n p_{ng}(\theta^i)}{\sum_n \sum_{g'} p_{ng'}(\theta^i)}$
5. re-estimate parameters based on likelihood contributions weighted by $p_{ng} \rightarrow \theta^{i+1}$
6. repeat steps two to five until change in sample likelihood is sufficiently small

C.2.2 IIA property

The error term in the utility function is assumed to be i.i.d. type I extreme-value. This results in the IIA (independence of irrelevant alternatives) property of the choice probabilities $P(j)$ (McFadden, 1974). The IIA property indicates that the odds ratio of two alternatives is not affected by modifying a third alternative. We make use of this property to reallocate the probability mass of restricted alternatives (section C.3). When an hours category is excluded from the choice set due to constraints, probabilities of all other categories increase such that their odds ratios are not affected (Appendix C.3). This is an undesirable property when it is the case that (conditional on the shared characteristics) similar hours category are closer substitutes than categories with a very different amount of hours.
Modeling unobserved heterogeneity relaxes the IIA property. For our setting unobserved heterogeneity implies that if an hours category is excluded from the choice set due to constraints, probabilities of categories with a similar unobserved effect increases over-proportionally. As we specify unobserved heterogeneity as heterogeneous consumption propensity, hours categories with more similar net earnings are closer substitutes. The reason is that the IIA property sill holds conditional on being of a certain unobserved type. Imagine there are two unobserved types with one having a high and one a low propensity for consumption. In the case of the

former all hours categories with high hours have a higher likelihood to be chosen. When one of these categories is constraint, the other high hours categories benefit particularly much due to their high choice probabilities.

C.3 Computation of expected state probabilities and approximation of substitution pattern

In the standard model the probability for the actual state k ($P^a(k)$) is equal to the probability for the desired choice according to the household members' labour supply preferences ($P^d(k)$), i.e. $P^a(k) = P^d(k)$. Expected state probabilities $\hat{P}^a(k)$ are thus solely based on the parameters of the labour supply model.

In the model with labour demand and hours constraints the probability for the actual state ($P^a(k)$) depends on $P^d(k)$ as well as the rationing probability in state k, $\Psi(k)$ (and actually on $\Psi(.)$ and $P^d(.)$ of all other states). In order to calculate expected state probabilities $\hat{P}^a_j$, $\hat{P}^d_j$ needs to be adjusted by the expected probability of being able to work in the desired state $(1 - \hat{\Psi}(k))$. The probability mass of being constrained in category k, i.e. $\hat{P}^d(k)\hat{\Psi}(k)$, is re-allocated to all other states $l \neq k$ according to their relative choice probabilities. This procedure respects the IIA property of the underlying discrete choice labour supply model as odds ratio of states $l \neq k$ are not affected. At the same time alternative k receives probability mass from hours restrictions in all other states $l \neq k$ with $h^a_l > 0$ (there is by definition no hours restriction on non-employment). This probability mass is again subject to the rationing probability $\Psi(k)$. A fraction is thus again re-allocated to the other categories and so on. As analytical solutions for the expected choice probabilities get complex very quickly for increasing choice sets we implement the substitution pattern by a numerical algorithm which iterates until the probability mass which has to be re-allocated is sufficiently small (we use a threshold of 0.1% as convergence criterion).

As hours constraints are defined on the individual level, we perform the algorithm successively for men and women. One spouse's restricted probability mass for hours category k is re-allocated to all household categories in which this spouse does not work k hours. Note that the algorithm is run conditional on unobserved types. This ensures that the IIA property holds which provides us with the above described substitution pattern. The unconditional state probabilities are then calculated as the weighted average of the conditional state probabilities.

C.4 Robustness checks

In this section we examine whether the estimated effects of an introduction of the in-work benefit for parents are robust with respect to the most crucial specification decisions. Detailed estimation results, results for the other simulations as well as elasticities estimated are available on request. Overall the qualitative picture of the policy effects is fairly stable.

C.4.1 The role of involuntary unemployed in identifying hours constraints

Individuals that state to actively search for a job and to be available to the labour market, but are observed to work zero hours are regarded as involuntarily unemployed. This information contributes to the identification of hours constraints depending on what kind of position an individual states to seek. In the main text involuntarily unemployed individuals seeking a part-time position only contribute to the identification of restrictions in part-time categories. The reasoning is that these individuals might have a strong preference for leisure. Being involuntarily unemployed thus might not necessarily imply being restricted in full-time categories. Involuntarily unemployed individuals seeking a full-time position, by contrast, contribute to the identification of all hours categories. We assume that these individuals - although they prefer a full-time position - would also accept jobs with fewer hours. Inferring information about all hours categories additionally implies that these individuals are at least potentially confronted with all kind of job offers. It might be the case, though, that either unemployed individuals seeking a full-time position might decline job offers with fewer hours or that part-time job offers do not reach them as they only look for full-time job advertisements. In this robustness check unemployed individuals seeking a full-time position therefore only contribute to the identification of restrictions in full-time categories.

The estimated policy effects turn out to be very robust in that respect (column (1) in table C.1). The slightly larger effects might be rooted in the smaller average restriction probabilities.

C.4.2 Couple with one spouse working her desired hours

Desired hours of work is a crucial variable in our framework. The exact wording of the survey question underlying the information is: “If you could choose your own number of working hours, taking into account that your income would change according to the number of hours: Would you prefer to decrease, increase or maintain your number of working hours?”. If they prefer a change, they will be asked for their desired working hours. There are two possible interpretations to this question (Callan et al., 2007): Respondents might choose their desired hours of work conditional on their partners' actual working hours (constrained optimization). In the main text we deviate from this view and assume that both spouses can freely choose their desired labour supply (unconstrained optimization of family utility). A deviation between an individual's desired and actual hours then implies that she is not able to work her desired hours of work. Under the alternative interpretation this conclusion cannot be deducted. In this robustness check we therefore restrict the estimation sample to couples where at least one is working his/her desired hours. This ensures that a deviation between desired and actual hours of work is rooted in individual constraints. Simulation results are again based on the whole sample.

An in-work benefit for parents would increase the fraction of households where both spouses work more than 24 hours by 1 percentage point (column (2) in table C.1). This is significantly smaller than based on the main specification (≈ 1.4). The overall picture is preserved, though.

C.4.3 Free correlation of unobserved types

In the main text we impose a deterministic relationship between unobserved constraint types for men and women and household preferences for consumption. We choose a joint distribution such that the possible combinations of restriction types within a household add up to four household types which are assumed to differ in their consumption propensity. In this section we do not restrict the correlation between unobserved types in the two model parts. We assume two household types with respect to preferences and two individual types with respect to restrictions for men and women, respectively. Their joint distribution is estimated non-parametrically.

The estimated policy effects increase slightly (column (3) in table C.1). The difference is not statistically significant, though.

C.4.3.1 Unobserved heterogeneity of linear effect of leisure

In the main text we assume the coefficient of the linear consumption term to vary across households in an unobserved way. Alternatively, unobserved heterogeneity may be specified as a heterogeneous effect of leisure which is done in this robustness check. The policy effects almost halve relative to the basic specification (column (4) in table C.1). This is rooted in a significantly smaller estimate of female labour supply elasticities.

Table C.1: Employment effects: in-work benefit for parents, robustness

	Basic	**(1)**	**(2)**	**(3)**	**(4)**
Men					
Hours - change (%)	0.12	0.15	0.22	0.13	0.24
	(0.08 ; 0.20)	(0.10 ; 0.21)	(0.05 ; 0.38)	(0.05 ; 0.19)	(0.10 ; 0.35)
Part. - base (%/100)	0.99	0.99	0.98	0.99	0.98
	(0.98 ; 0.99)	(0.99 ; 0.99)	(0.97 ; 0.98)	(0.98 ; 0.99)	(0.97 ; 0.98)
Part. - change (pp.)	0.04	0.04	0.14	0.03	0.09
	(0.04 ; 0.05)	(0.03 ; 0.05)	(0.11 ; 0.16)	(0.03 ; 0.04)	(0.07 ; 0.11)
Women					
Hours - change (%)	4.09	4.13	3.80	4.66	2.07
	(3.63 ; 4.46)	(3.75 ; 4.53)	(3.17 ; 4.46)	(4.11 ; 5.28)	(1.63 ; 2.45)
Part. - base (%/100)	0.55	0.56	0.58	0.57	0.56
	(0.54 ; 0.56)	(0.55 ; 0.57)	(0.57 ; 0.58)	(0.56 ; 0.58)	(0.56 ; 0.58)
Part. - change (pp.)	0.70	0.72	0.45	0.77	0.34
	(0.62 ; 0.77)	(0.65 ; 0.81)	(0.41 ; 0.50)	(0.70 ; 0.84)	(0.24 ; 0.44)
Both spouses' hours > 24					
base (%/100)	0.16	0.16	0.14	0.16	0.18
	(0.15 ; 0.17)	(0.15 ; 0.16)	(0.14 ; 0.15)	(0.16 ; 0.17)	(0.17 ; 0.19)
change (pp.)	1.41	1.48	1.04	1.63	0.83
	(1.27 ; 1.53)	(1.33 ; 1.60)	(0.97 ; 1.12)	(1.50 ; 1.80)	(0.69 ; 0.94)

Notes: Basic: Specification as in the main text, (1): Involuntarily unemployed seeking a full-time position only contribute to the identification of full-time hours restrictions (section C.4.1), (2): Estimation sample is restricted to couples where at least one spouse is working in the desired hours category (section C.4.2), (3): Joint distribution of unobserved consumption and restriction types does not restrict correlation (section C.4.3), (4): Unobserved heterogeneity specified as heterogeneous propensity for leisure (section C.4.3.1), All effects based on the full model, a discrete choice model based on desired hours of work augmented by constraints (section 3.3), bootstrapped 95%-confidence bands in parentheses.
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

C.5 Additional tables

Table C.2: Summary statistics

Variable	Mean	Std. Dev.	N
Household characteristics			
Net income	3673.49	1788.15	3604
East	0.17	0.37	3604
Child aged $0-2$	0.64	0.48	3604
Child aged $3-6$	0.67	0.47	3604
Individual characteristics woman			
Hours of work	11.86	14.15	3604
Hourly wage	13.23	6.67	3604
Age	34.00	5.45	3604
German	0.85	0.36	3604
Handicapped	0.02	0.15	3604
Occ.: Managers	0.03	0.16	2784
Occ.: Professionals	0.17	0.38	2784
Occ.: Technicians	0.29	0.45	2784
Occ.: Clerical support workers	0.16	0.37	2784
Occ.: Service & sales workers	0.21	0.41	2784
Occ.: Craft & related trades	0.01	0.09	2784
Occ.: Agricultural workers	0.04	0.20	2784
Occ.: Plant & machine operators	0.01	0.12	2784
Occ.: Elementary	0.06	0.24	2784
Occ.: Armed Forces	0.00	0.04	2784
Edu.: Isced 0-2	0.12	0.32	3602
Edu.: Isced 3-4	0.59	0.49	3602
Edu.: Isced 5-6	0.29	0.45	3602
Small firm	0.09	0.29	1704
Mid-size firm	0.48	0.50	1704
Large firm	0.43	0.49	1704
Individual characteristics man			
Hours of work	38.17	13.37	3604
Hourly wage	19.38	10.36	3604
Age	37.16	6.24	3604
German	0.86	0.35	3604
Handicapped	0.02	0.14	3604
Occ.: Managers	0.08	0.27	3481
Occ.: Professionals	0.21	0.41	3481
Occ.: Technicians	0.17	0.37	3481
Occ.: Clerical support workers	0.06	0.23	3481
Occ.: Service & sales workers	0.05	0.22	3481
Occ.: Craft & related trades	0.02	0.13	3481
Occ.: Agricultural workers	0.23	0.42	3481
Occ.: Plant & machine operators	0.10	0.30	3481
Occ.: Elementary	0.07	0.26	3481
Occ.: Armed Forces	0.01	0.10	3481
Edu.: Isced 0-2	0.14	0.34	3569
Edu.: Isced 3-4	0.53	0.50	3569
Edu.: Isced 5-6	0.33	0.47	3569
Small firm	0.04	0.20	3230
Mid-size firm	0.40	0.49	3230
Large firm	0.56	0.50	3230
Regional characteristics			
Rate of unempl.	7.17	3.30	3604
CC quota PT	11.38	5.00	3604
CC quota FT	11.02	11.83	3604

Notes: Std. Dev.=Standard deviation, N=Amount of non-missing observations, East=Household lives in Eastern Germany, Occ.=occupation aggregated by ISCO code, edu.=Education aggregated by ISCED code, Small firm:< 5 employees, Mid-size firm:$5 - 199$ employees, , Large firm:> 199 employees, Rate of unempl.=Rate of unemployment on the county level, CC quota PT=Formal part-time child care slots for children aged below three relative to households with respective children (on the county level), CC quota FT=Available formal full-time child care slots for children aged below three relative to households with respective children (on the county level).
Source: Own calculations based on INKAR, waves 2010-2013 and FiD, waves 2010-2013.

Table C.3: Estimation results: preferences

	Full		**Standard**	
	coeff.	s.e.	coeff.	s.e.
Consumption				
c_1^u: bad-bad	0.93	0.25	2.49	0.13
c_2^u: bad-good	2.55	0.29	.	.
c_3^u: good-bad	0.90	0.19	.	.
c_4^u: good-good	3.63	0.22	.	.
x Age child $0-2$	-1.08	0.22	-0.97	0.16
Quadratic term	0.01	0.03	0.12	0.01
Leisure woman				
Linear term	1.58	0.10	1.99	0.08
x Age	-0.14	0.02	-0.17	0.02
x East	-0.77	0.06	-0.73	0.05
x German	-0.44	0.06	-0.48	0.06
x Handicapped	0.10	0.14	0.07	0.12
x Age child $0-2$	0.42	0.09	0.25	0.08
Squared term	-0.11	0.02	0.18	0.03
Leisure man				
Linear term	-0.55	0.10	0.24	0.08
x Age	0.05	0.02	-0.02	0.02
x East	0.17	0.07	0.14	0.06
x German	-0.21	0.06	-0.31	0.06
x Handicapped	0.60	0.12	0.54	0.11
x Age child $0-2$	-0.16	0.10	-0.29	0.08
Squared term	-0.44	0.03	0.25	0.03
Interactions				
Leisure woman x man	-0.27	0.03	0.01	0.02
Observations	3604.00	.	3604.00	.
Positive 1st Derivates (in %)				
U_{c1} (consumption)	0.39	.	1.00	.
U_{c2} (consumption)	1.00	.	.	.
U_{c3} (consumption)	0.39	.	.	.
U_{c4} (consumption)	1.00	.	.	.
U_{lm} (leisure woman)	0.98	.	0.90	.
U_{lf} (leisure man)	0.77	.	0.03	.

Notes: Full=Discrete choice model based on desired hours of work augmented by constraints (section 3.3), Standard=Discrete choice model based on actual hours of work, East=Household lives in Eastern Germany, *bad* (*good*) refers to the unobserved restriction types with c.p. a higher (lower) restriction probability, coeff.=regression coefficient, s.e.=standard errors.
Source: Own calculations based on FiD, waves 2010-2013.

Table C.4: Estimation results: hours constraints

	Hours categories				
	1-14	**15-24**	**25-35**	**36-40**	**>40**
Constant	-0.90	0.11	0.63	-0.26	-0.28
	(0.35)	(0.32)	(0.21)	(0.23)	(0.44)
Individual Characteristics					
Male	1.80	0.96	0.33	-0.97	-1.77
	(0.39)	(0.33)	(0.14)	(0.18)	(0.41)
Edu.: Isced 0-2	0.42	0.72	0.06	0.53	1.09
	(0.26)	(0.25)	(0.20)	(0.18)	(0.22)
Edu.: Isced 3-4	-0.14	-0.06	0.11	-0.08	0.28
	(0.11)	(0.09)	(0.07)	(0.07)	(0.12)
Age	-0.34	-0.30	-0.03	-0.16	-0.38
	(0.12)	(0.09)	(0.07)	(0.07)	(0.08)
German	-0.96	-0.97	-0.75	-0.74	-1.08
	(0.34)	(0.31)	(0.20)	(0.20)	(0.25)
East	0.96	1.92	-0.18	0.45	-0.55
	(0.81)	(0.65)	(0.33)	(0.40)	(0.58)
Handicapped	0.67	-0.20	-0.34	0.51	0.46
	(0.64)	(0.54)	(0.44)	(0.42)	(0.83)
Occupation (reference: professionals)					
Managers	-0.00	-0.30	0.58	0.56	-0.84
	(0.46)	(0.35)	(0.22)	(0.22)	(0.40)
Technicians	0.05	-0.01	-0.33	-0.28	-0.43
	(0.22)	(0.14)	(0.11)	(0.14)	(0.27)
Clerks	0.30	0.06	0.30	-0.02	0.72
	(0.26)	(0.22)	(0.20)	(0.25)	(0.34)
Service workers	-0.33	0.12	0.26	1.74	0.52
	(0.23)	(0.18)	(0.17)	(0.21)	(0.28)
Agriculture	2.36	0.54	-0.60	0.30	1.09
	(2.46)	(0.88)	(0.64)	(0.46)	(0.57)
Craft workers	0.75	0.44	-0.05	-0.26	0.52
	(0.34)	(0.35)	(0.16)	(0.13)	(0.20)
Plant	-1.47	-2.22	-0.47	-0.38	-0.52
& machine operators	(0.86)	(0.63)	(0.31)	(0.22)	(0.30)
Elementary	-0.79	-0.93	0.20	-0.48	0.61
	(0.44)	(0.34)	(0.26)	(0.27)	(0.33)
Firm size (reference: 5-199 employees)					
> 5 employees	-0.04	-0.15	-0.69	0.70	-1.04
	(0.40)	(0.30)	(0.33)	(0.32)	(0.69)
> 199 employees	0.11	0.03	0.07	-0.06	0.23
	(0.31)	(0.12)	(0.05)	(0.06)	(0.11)
Regional level					
Reg. rate of unempl.	0.23	0.26	0.25	0.35	0.65
	(0.15)	(0.13)	(0.08)	(0.10)	(0.12)
Childcare coverage 0-2					
pt m	-0.20	-0.21	0.01	-0.05	0.11
	(0.44)	(0.26)	(0.10)	(0.08)	(0.10)
ft m	-0.09	-0.83	0.29	-0.25	-0.03
	(0.65)	(0.37)	(0.15)	(0.13)	(0.23)
pt f	0.05	-0.14	-0.09	0.13	0.23
	(0.14)	(0.11)	(0.08)	(0.11)	(0.28)
ft f	0.18	-0.72	-0.18	-0.45	-0.10
	(0.35)	(0.24)	(0.12)	(0.19)	(0.38)
Observations	505.00	771.00	1186.00	2044.00	1675.00
u_m^1	2.01	2.01	2.01	2.01	2.01
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
u_m^2	-0.69	-0.69	-0.69	-0.69	-0.69
u_f^1	1.47	1.47	1.47	1.47	1.47
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
u_f^2	-0.89	-0.89	-0.89	-0.89	-0.89

Notes: East=Household lives in Eastern Germany, Occ.=occupation, Edu.=Education aggregated by ISCED code (reference Isced 5-6), standard errors is parantheses.
Source: Own calculations based on INKAR waves 2010-2013 and FiD, waves 2010-2013.

Table C.5: Estimation results: hourly wage

	Men, East		Women, East		Men, West		Women, West	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Age	0.014	0.004	0.039	0.003	0.013	0.002	0.027	0.001
squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Education, reference: primary								
Secondary 1st step	0.204	0.049	0.156	0.064	0.002	0.014	0.071	0.017
Secondary 2nd step	0.216	0.049	0.182	0.064	0.013	0.013	0.094	0.017
Post-secondary	0.226	0.052	0.265	0.066	0.045	0.015	0.140	0.018
Tertiary 1st step	0.230	0.051	0.243	0.066	0.055	0.015	0.147	0.019
Tertiary 2st step	0.287	0.052	0.359	0.066	0.155	0.015	0.272	0.019
Years education x German					0.007	0.001	0.010	0.001
Experience (in years)								
Full-time			0.001	0.002			0.010	0.002
squared			0.004	0.003			-0.011	0.006
x German							0.002	0.002
x German squared							-0.010	0.006
Part-time			0.003	0.002			0.002	0.003
squared			-0.006	0.007			-0.010	0.011
x German							-0.004	0.003
x German squared							0.019	0.012
Both	0.008	0.002			0.009	0.001		
squared	-0.035	0.004			-0.038	0.003		
x German					0.003	0.001		
x German squared					-0.001	0.003		
Tenure	0.011	0.001	0.019	0.001	0.026	0.002	0.029	0.002
squared	-0.017	0.003	-0.028	0.003	-0.045	0.005	-0.058	0.007
x German					-0.014	0.002	-0.014	0.002
x German squared					0.031	0.006	0.037	0.008
Loss human capital	-0.170	0.006	-0.080	0.005	-0.075	0.007	-0.014	0.005
x German					-0.063	0.008	-0.017	0.005
Firm size, reference: <5 employees								
$5-19$	-0.186	0.011	-0.198	0.009	-0.155	0.007	-0.157	0.005
$20-199$	-0.014	0.002	-0.002	0.003	-0.020	0.001	-0.004	0.002
> 200	0.109	0.006	0.102	0.006	0.060	0.002	0.089	0.004
Sector, reference: (Electrical) Machinery								
Energy	0.138	0.017	0.155	0.034	0.048	0.011	0.211	0.022
Wood/paper/chemicals	0.057	0.013	0.062	0.021	0.077	0.005	0.051	0.008
Construction	0.041	0.008	-0.016	0.020	-0.011	0.005	-0.029	0.014
Heavy industry	0.038	0.011	0.019	0.027	0.070	0.005	0.088	0.013
Textile/food	-0.062	0.048	-0.114	0.034	-0.067	0.019	-0.088	0.019
Whole sale/retail	-0.089	0.010	-0.070	0.008	-0.090	0.005	-0.071	0.004
Transport/communication	-0.006	0.010	0.035	0.016	-0.037	0.005	0.049	0.010
Public services	0.024	0.006	0.057	0.004	-0.039	0.004	0.020	0.002
Private services	-0.031	0.009	-0.046	0.008	0.001	0.005	-0.006	0.004
Other	-0.055	0.012	-0.091	0.011	-0.043	0.006	-0.061	0.007
Agriculture	-0.249	0.016	-0.262	0.026	-0.148	0.014	-0.164	0.024
Task, reference: Worker								
Skilled worker	-0.112	0.005	-0.147	0.011	-0.097	0.003	-0.105	0.011
Foreman	0.008	0.011	-0.012	0.038	-0.011	0.006	-0.060	0.025
Employee: no training	-0.237	0.018	-0.222	0.012	-0.311	0.010	-0.220	0.006
Employee: training	-0.108	0.012	-0.078	0.008	-0.160	0.008	-0.073	0.005
Employee: qualified tasks	0.044	0.008	0.057	0.004	0.030	0.003	0.095	0.002
Employee: management	0.329	0.008	0.323	0.009	0.287	0.003	0.300	0.006
Civil servant: middle grade	0.008	0.021	0.153	0.024	-0.098	0.009	0.129	0.014
Civil servant: upper grade	0.326	0.016	0.302	0.018	0.109	0.007	0.299	0.009
Hamburg					0.028	0.013	0.046	0.015
Lower Saxony					-0.007	0.009	-0.014	0.010
Bremen					-0.052	0.018	-0.030	0.019
Northrhine-Westphalia					0.017	0.008	0.006	0.009
Hesse					0.045	0.009	0.039	0.010
Rhineland-Palatinate					0.001	0.010	0.006	0.011
Baden-Wrttemberg					0.064	0.008	0.058	0.010
Bavaria					0.021	0.008	0.028	0.010
Saarland					0.012	0.014	-0.026	0.017
Brandenburg	-0.094	0.011	-0.113	0.011				
Mecklenburg WP	-0.096	0.013	-0.115	0.013				
Saxony	-0.156	0.010	-0.167	0.010				
Saxony-Anhalt	-0.146	0.011	-0.180	0.011				
Thuringia	-0.153	0.011	-0.164	0.011				

Table C.5: Estimation results: hourly wage (cont.)

	Men, East		Women, East		Men, West		Women, West	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
2000	0.024	0.015	0.039	0.016	0.007	0.008	0.015	0.010
2001	0.069	0.015	0.065	0.016	0.065	0.007	0.068	0.009
2002	0.096	0.015	0.081	0.016	0.100	0.008	0.087	0.010
2003	0.112	0.015	0.099	0.016	0.097	0.008	0.098	0.010
2004	0.107	0.016	0.080	0.017	0.091	0.008	0.088	0.010
2005	0.107	0.016	0.076	0.017	0.066	0.008	0.068	0.010
2006	0.091	0.015	0.064	0.016	0.072	0.008	0.078	0.010
2007	0.111	0.016	0.084	0.017	0.082	0.008	0.076	0.010
2008	0.112	0.015	0.109	0.016	0.082	0.008	0.103	0.010
2009	0.114	0.015	0.097	0.016	0.096	0.007	0.077	0.009
2010	0.129	0.015	0.108	0.015	0.110	0.007	0.089	0.009
2011	0.142	0.015	0.135	0.015	0.115	0.007	0.102	0.009
2012	0.144	0.016	0.182	0.017	0.091	0.007	0.110	0.009
Handicapped, degree	0.000	0.001	-0.001	0.001	0.001	0.000	0.002	0.000
squared	-0.001	0.001	0.001	0.001	-0.002	0.000	-0.003	0.001
Constant	1.806	0.083	1.259	0.086	1.991	0.036	1.447	0.036
Selection step								
Age	0.111	0.011	0.071	0.009	0.139	0.007	0.064	0.004
squared	-0.003	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
Education, reference: primary								
Secondary 1st step	0.804	0.103	0.551	0.108	0.367	0.042	0.148	0.039
Secondary 2nd step	1.184	0.097	0.953	0.104	0.676	0.040	0.424	0.037
Post-secondary	1.732	0.113	1.405	0.112	0.948	0.050	0.684	0.041
Tertiary 1st step	1.815	0.108	1.515	0.111	1.063	0.050	0.606	0.043
Tertiary 2st step	2.354	0.102	1.737	0.106	1.436	0.045	0.937	0.039
Experience (in years)								
Full-time			0.095	0.004		0.003	0.074	0.002
squared			0.030	0.011			0.007	0.006
Part-time			0.182	0.005			0.214	0.003
squared			-0.243	0.024			-0.459	0.010
Both	0.068	0.006			0.041			
squared	0.102	0.014			0.065	0.008		
Handicap degree	0.498	0.028	-0.004	0.003	0.302	0.020	0.001	0.001
Squared	-0.148	0.052	0.008	0.004	-0.105	0.035	-0.006	0.002
Current health, reference: very good								
Good	0.033	0.058	0.030	0.037	-0.059	0.037	0.005	0.018
Satisfactory	-0.058	0.033	-0.099	0.039	-0.087	0.021	-0.060	0.020
Bad	-0.068	0.044	-0.329	0.045	-0.069	0.028	-0.205	0.024
Very bad	-0.015	0.003	-0.705	0.073	-0.006	0.001	-0.711	0.041
Married	0.021	0.004	0.148	0.024	0.003	0.002	-0.232	0.015
Number of children aged								
< 3	0.097	0.042	-1.215	0.040	0.029	0.026	-1.608	0.023
$3 - 6$	-0.025	0.045	-0.686	0.043	-0.163	0.028	-1.100	0.023
$7 - 16$	-0.448	0.053	-0.299	0.029	-0.563	0.032	-0.656	0.016
> 16	-0.929	0.080	-0.142	0.039	-1.138	0.046	-0.262	0.021
Non-labour income	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hamburg					0.076	0.065	0.017	0.049
Lower Saxony					0.148	0.042	0.007	0.031
Bremen					-0.076	0.079	0.028	0.062
Northrhine-Westphalia					0.082	0.038	-0.056	0.029
Hesse					0.161	0.044	0.033	0.032
Rhineland-Palatinate					0.141	0.047	-0.107	0.035
Baden-Wrttemberg					0.370	0.041	0.008	0.030
Bavaria					0.236	0.040	0.044	0.030
Saarland					0.136	0.068	-0.044	0.050
Brandenburg	-0.352	0.042	-0.075	0.037				
Mecklenburg WP	-0.276	0.049	-0.101	0.042				
Saxony	-0.202	0.039	-0.078	0.033				
Saxony-Anhalt	-0.202	0.042	-0.172	0.036				
Thuringia	-0.104	0.043	-0.102	0.036				

Table C.5: Estimation results: hourly wage (cont.)

	Men, East		Women, East		Men, West		Women, West	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
2000	-0.089	0.055	-0.060	0.051	-0.140	0.041	-0.097	0.029
2001	0.008	0.055	0.096	0.051	-0.029	0.040	0.065	0.028
2002	-0.136	0.056	-0.043	0.051	-0.214	0.040	-0.065	0.029
2003	-0.033	0.056	0.073	0.051	-0.192	0.040	-0.042	0.029
2004	-0.068	0.057	-0.014	0.052	-0.229	0.041	-0.045	0.030
2005	-0.046	0.057	0.047	0.053	-0.230	0.040	-0.092	0.029
2006	0.146	0.059	0.187	0.054	-0.034	0.043	0.005	0.030
2007	0.233	0.062	0.221	0.055	-0.014	0.044	0.014	0.031
2008	0.388	0.061	0.399	0.055	0.011	0.042	0.130	0.031
2009	0.358	0.058	0.250	0.048	-0.159	0.038	-0.025	0.026
2010	0.398	0.057	0.351	0.048	-0.061	0.039	0.032	0.026
2011	0.450	0.060	0.382	0.051	-0.005	0.041	-0.103	0.029
2012	0.415	0.064	0.357	0.057	-0.197	0.041	-0.197	0.031
German					0.448	0.022	0.279	0.017
Constant	-1.600	0.224	-1.060	0.190	-2.028	0.133	-0.356	0.096
Mills λ	-0.092	0.020	0.030	0.019	-0.112	0.011	0.039	0.009
Observations								
Employed	17251		17802		58830		54442	
Non-employed	4120		6687		6644		24848	

Notes: Education is classified by ISCED. Reference categories for state: Schleswig-Holstein and West Berlin for West Germany, East Berlin for East Germany, coeff.=regression coefficient, s.e.=standard errors, x refers to an interaction effect.
Source: Own calculations based on GSOEP and FiD, waves 1999-2013.

Appendix D

Appendix to Chapter 4

D.1 Derivation of firm-size

In the following we derive the number of type s and f workers a firm can attract by offering earnings z. Thereby we make use of equations (4.2), (4.4) and (4.5). We start with type s workers:

$$
\begin{aligned}
l^s(z) &= \frac{\frac{F(z)}{1+\kappa^s(1-F(z))} - \frac{F(z-\epsilon)}{1+\kappa^s(1-F(z-\epsilon))}}{F(z)-F(z-\epsilon)}(n^s-u^s) \\
&= \frac{\frac{F(z)(1+\kappa^s(1-F(z-\epsilon)))-F(z-\epsilon)(1+\kappa^s(1-F(z)))}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))}}{F(z)-F(z-\epsilon)}(n^s-u^s) \\
&= \frac{\frac{F(z)+F(z)\kappa^s-F(z)\kappa^s F(z-\epsilon)-(F(z-\epsilon)+F(z-\epsilon)\kappa^s-F(z-\epsilon)\kappa^s F(z))}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))}}{F(z)-F(z-\epsilon)}(n^s-u^s) \\
&= \frac{\frac{F(z)+F(z)\kappa^s-(F(z-\epsilon)+F(z-\epsilon)\kappa^s)}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))}}{F(z)-F(z-\epsilon)}(n^s-u^s) \\
&= \frac{\frac{(F(z)-F(z-\epsilon))(1+\kappa^s)}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))}}{F(z)-F(z-\epsilon)}(n^s-u^s) \\
&= \frac{(1+\kappa^s)}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))}(n^s-u^s) \\
&= \frac{(1+\kappa^s)}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))}\frac{n^s\kappa}{1+\kappa^s} \\
&= \frac{n^s\kappa}{(1+\kappa^s(1-F(z)))(1+\kappa^s(1-F(z-\epsilon)))}
\end{aligned}
$$

Type f workers do not accept jobs with $z > z^*$. For $z \leq z^*$ we thus get

$$
\begin{aligned}
l^f(z) &= \frac{\frac{F(z)}{(1+\kappa^f(F(z^*)-F(z)))F(z^*)} - \frac{F(z-\epsilon)}{(1+\kappa^f(F(z^*)-F(z-\epsilon)))F(z^*)}}{F(z)-F(z-\epsilon)}(n^f-u^f) \\
&= \frac{\frac{F(z)((1+\kappa^f(F(z^*)-F(z-\epsilon)))F(z^*))-F(z-\epsilon)((1+\kappa^f(F(z^*)-F(z)))F(z^*))}{((1+\kappa^f(F(z^*)-F(z)))F(z^*))((1+\kappa^f(F(z^*)-F(z-\epsilon)))F(z^*))}}{F(z)-F(z-\epsilon)}(n^f-u^f) \\
&= \frac{\frac{(F(z)+F(z)\kappa^f F(z^*)-F(z)\kappa^f F(z-\epsilon)-(F(z-\epsilon)+F(z-\epsilon)\kappa^f F(z^*)-F(z-\epsilon)\kappa^f F(z)))F(z^*)}{F(z^*)^2(1+\kappa^f(F(z^*)-F(z)))(1+\kappa^f(F(z^*)-F(z-\epsilon)))}}{F(z)-F(z-\epsilon)}(n^f-u^f)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\frac{(F(z)+F(z)\kappa^f F(z^*)-(F(z-\epsilon)+F(z-\epsilon)\kappa^f F(z^*)))}{F(z^*)(1+\kappa^f(F(z^*)-F(z)))(1+\kappa^f(F(z^*)-F(z-\epsilon)))}}{F(z)-F(z-\epsilon)}(n^f-u^f) \\
&= \frac{\frac{(F(z)-F(z-\epsilon))(1+\kappa^f F(z^*))}{F(z^*)(1+\kappa^f(F(z^*)-F(z)))(1+\kappa^f(F(z^*)-F(z-\epsilon)))}}{F(z)-F(z-\epsilon)}(n^f-u^f) \\
&= \frac{(1+\kappa^f F(z^*))}{F(z^*)(1+\kappa^f(F(z^*)-F(z)))(1+\kappa^f(F(z^*)-F(z-\epsilon)))}(n^f-u^f) \\
&= \frac{(1+\kappa^f F(z^*))}{F(z^*)(1+\kappa^f(F(z^*)-F(z)))(1+\kappa^f(F(z^*)-F(z-\epsilon)))}\frac{n^f\kappa^f F(z^*)}{1+\kappa^f F(z^*)} \\
&= \frac{n^f\kappa^f}{(1+\kappa^f(F(z^*)-F(z)))(1+\kappa^f(F(z^*)-F(z-\epsilon)))}
\end{aligned}
$$

The total firm size, is then the sum of the number of workers of each type, which gives expression (4.6).

D.2 Sketch of Proof of Proposition (I)

Proposition (I) *If we observe offers above z^*, there must be a mass point of job offers at z^*. The wage offer distribution above z^* is continuous up to some $\bar{z}$.*

Assume there exists no mass point (i.e. $f(z^*) = 0$), the offer distribution for $z < z^*$ is then continuous and profits at the threshold are

$$\pi(z^*) = (p - z^*)\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*)))^2} + n^f\kappa^f. \tag{D.1}$$

The profit of a firm offering jobs with earnings slightly above the threshold (for $\epsilon \to 0$) is:

$$
\begin{aligned}
\pi(z^*+\epsilon) &= (p-(z^*+\epsilon))\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*+\epsilon)))(1+\kappa^s(1-F(z^*-\epsilon+\epsilon)))} \\
&= (p-z^*)\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*)))^2}
\end{aligned} \tag{D.2}
$$

Assuming $f(z^*) = 0$ it holds that $\pi(z^*+\epsilon) < \pi(z^*)$ implying that there is a gap to the right of the threshold.

Is there an earnings level $z' > z^* + \epsilon$ where the equal profit condition holds again? Equation (D.3) makes use of $F(z') = F(z^*)$ which holds if there is a gap in the interval $z \in (z^*, z')$.

$$
\begin{aligned}
\pi(z') &= (p-z')\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z')))(1+\kappa^s(1-F(z'-\epsilon)))} \\
&= (p-(z'))\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*)))^2}
\end{aligned} \tag{D.3}
$$

As $(p - z') < (p - z^*)$ it holds that $\pi(z') < \pi(z^*)$ implying that no job with earnings $z > z^*$ will be offered if there is no mass point at z^*.
Allowing for a mass point at z^*, $\frac{\partial\pi(z^*)}{\partial f(z^*)} < 0$ and $\frac{\partial\pi(z^*+\epsilon)}{\partial f(z^*)} = 0$ imply that there might be a value for $f(z^*)$ for which the equal profit condition between z^* and $z^* + \epsilon$ holds ($\pi(z^* + \epsilon) = \pi(z^*)$). For earnings $z' \in [z^* + \epsilon, \overline{z}]$, the usual trade-off between profit per workers and firm size ensures that the equilibrium offer distribution is continuous in that interval and determined by $\pi(z') = \pi(z^* + \epsilon)$.

D.3 Sketch of proof of Proposition (II)

Proposition (II) *If there is a mass point at z^*, there will be a gap in the offer distribution just below the threshold.*

We need to show that a mass point in the wage offer distribution is only consistent with equal profits if there is a gap in the wage offering distribution. We compare profits $\pi(z^*)$ with profits $\pi(z^* - \epsilon)$.
The profit of a firm offering jobs with earnings at the threshold is

$$\begin{aligned}
\pi(z^*) &= (p - z^*)l(z^*) \\
&= (p - z^*)\left(\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*)))(1+\kappa^s(1-F(z^*-\epsilon)))} + \frac{n^f\kappa^f}{(1+\kappa^f(F(z^*)-F(z^*)))(1+\kappa^f(F(z^*)-F(z^*-\epsilon)))}\right) \\
&= (p - z^*)\left(\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*)))(1+\kappa^s(1-F(z^*)+f(z^*)))} + \frac{n^f\kappa^f}{(1+\kappa^f f(z^*))}\right)
\end{aligned} \tag{D.4}$$

The profit slightly below the threshold is given by (for $\epsilon \to 0$):

$$\begin{aligned}
\pi(z^* - \epsilon) = (p - (z^* - \epsilon))(&\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*-\epsilon)))(1+\kappa^s(1-F(z^*-2\epsilon)))} + \\
&\frac{n^f\kappa^f}{(1+\kappa^f(F(z^*)-F(z^*-\epsilon)))(1+\kappa^f(F(z^*)-F(z^*-2\epsilon)))}) \\
&= (p - z^*)(\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*-\epsilon)))^2} + \frac{n^f\kappa^f}{(1+\kappa^f(f(z^*)))^2}) \\
&= (p - z^*)(\frac{n^s\kappa^s}{(1+\kappa^s(1-F(z^*)+f(z^*)))^2} + \frac{n^f\kappa^f}{(1+\kappa^f(f(z^*)))^2})
\end{aligned}$$

Given proposition (I), the data implies that there is a mass point at $z = z^*$ (i.e. that $f(z^*) > 0$). When $f(z^*) > 0$, it holds that $(1 + \kappa^j(1 - F(z^*) + f(z^*))) > (1 + \kappa^j(1 - F(z^*)))$ and $(1 + \kappa^f(f(z^*))) > 1$. Therefore, $\pi(z^*) > \pi(z^* - \epsilon)$ implying that there will be a gap to the left of the threshold.

D.4 Sketch of proof of proposition (III)

Proposition (III) *There may or may not exist wage offers offers below the threshold z^* in equilibrium. The wage offer distribution will then be continuous between $z \in [\underline{z}, z'']$for $z'' < z^*$.*

Define the highest wage offer below the threshold as z'', such that $F(z'') = F(z^* - \epsilon)$. Note that since there is a gap left of the threshold, if in equilibrium a z''-offer exists, it may be significantly below z^*. In equilibrium any z''-offer must make the same amount of profits as the threshold wage offer z^*.

$$\begin{aligned}\pi(z'') &= (p - z'')\frac{n^s \kappa^s}{(1 + \kappa^s(1 - F(z'')))(1 + \kappa^s(1 - F(z'' - \epsilon)))} + \\ &\frac{n^f \kappa^f}{(1 + \kappa^f(F(z^*) - F(z'')))(1 + \kappa^f(F(z^*) - F(z'' - \epsilon)))} \\ &= (p - z'')\frac{n^s \kappa^s}{(1 + \kappa^s(1 - F(z^*) + f(z^*)))^2} + \frac{n^f \kappa^f}{(1 + \kappa^f(f(z^*)))^2} \end{aligned} \tag{D.5}$$

Comparing equations (D.4) and (D.5) illustrates that $\pi(z'') = \pi(z^*)$ can hold as $\pi(z'')$ increases with decreasing z''. That is, there might be a certain size of the gap where $\pi(z'') = \pi(z^*)$ holds. Using that $F(\underline{z}) = 0$, we now determine the lowest wage offer $\underline{z}$ that will be made (if there are wage offers below z^*).

$$\begin{aligned}\pi(\underline{z}) &= (p - \underline{z})\frac{n^s \kappa^s}{(1 + \kappa^s(1 - F(\underline{z})))(1 + \kappa^s(1 - F(\underline{z} - \epsilon)))} + \\ &\frac{n^f \kappa^f}{(1 + \kappa^f(F(z^*) - F(\underline{z})))(1 + \kappa^f(F(z^*) - F(\underline{z} - \epsilon)))} \\ &= (p - \underline{z})\frac{n^s \kappa^s}{(1 + \kappa^s)^2} + \frac{n^f \kappa^f}{(1 + \kappa^f(F(z^*)))^2} \end{aligned} \tag{D.6}$$

If $z'' < z^*$ exists and $\underline{z} < z''$ there will then be a continuity of wage offers between these two values, generating equal profits with the standard trade-off between margins and firm-size.

D.5 Discrete variation in working hours

For simplification, we allow for two different numbers of hours h_k where $k \in [0, 1]$. In the market for low-paid jobs, this may correspond to 10 and 20 hours of working. We later consider how this model generalises to three or more hours sectors, and consider continuous variation in hours in the following section.

Firms set wage rates, w, workers care for wage rates as well as hours. In order to simplify notation, we follow Shephard (2012) and define $q_1(w) = w$ and $U(q_0(w), h_0) = U(q_1(w), h_1) = U(w, h_1)$, so $q_0(w)$ is a function that denotes the wage rate that makes individuals indifferent between working with few ($k = 0$) hours at $q_0(w)$ or working more ($k = 1$) hours at w. Depending on preferences, individuals may require a low-hours premium or accept a low-hours wage penalty.

D.5.1 Worker mobility

The flow from and into unemployment is equivalent to before (see equation (4.2)). The flow of workers of type $j \in (s, f1, f2)$ from and into jobs with hours h_k and wage rate w is

$$D^j(w)g^j_k(q_k(w))e^j_k = \lambda^j f_k(q_k(w))(u^j + G^j_0(q_0(w) - \epsilon)e^j_0 + G^j_1(w - \epsilon)e^j_1) \tag{D.7}$$

with $D^j(w) = [\delta + \lambda^j((1 - F_1(w)) + (1 - F_0(q_0(w))))]$ for $j \in (s, f1)$. Equation (D.8) states the corresponding definition for workers of type $f2$ who do not accept jobs with wage rates larger than w^*_k.

$$D^{f2}(w) = \begin{cases} [\delta + \lambda^j((F_1(w^*_1) - F_1(w)) + (F_0(w^*_0) - F_0(q_0(w))))] & \forall w \leq w^*_1 \text{ and } q_0(w) \leq w^*_0 \\ [\delta + \lambda^j(F_1(w^*_1) - F_1(w))] & \forall w \leq w^*_1 \text{ and } q_0(w) > w^*_0 \end{cases} \tag{D.8}$$

The LHS of equation (D.7) pertains to workers who leave a job in a sector k with wage $q_k(w)$. For $k = 1$ this group consists of workers who move from sector 1 to sector 0 ($\lambda^j(1 - F_0(q_0(w)))g^j_1(w)e^j_1$ for $j \in (s, f1)$), who move to a better paying job within sector 1 ($\lambda^j(1 - F_1(w))g^j_1(w)e^j_1$ for $j \in (s, f1)$) and who become unemployed ($\delta^j g^j_1(w)e^j_1$). The RHS pertains to workers who start a job in sector k with wage rate $q_k(w)$. For $k = 1$ this consists of workers who move from sector 0 to sector 1 ($\lambda^j f_1(w)G^j_0(q_0(w) - \epsilon)e^j_0$), who changes jobs within sector 1 ($\lambda^j f_1(w)G^j_1(w - \epsilon)e^j_1$) and who come from unemployment ($\lambda^j f_1(w)$). The overall flow (i.e. both sectors) between jobs with wage rate of no greater than w and unemployment is:

$$\begin{aligned} &(G^j_0(q_0(w))e^j_0 + G^j_1(w)e^j_1)D^j(w) = \lambda^j u^j F_0(q_0(w)) + \lambda^j u^j F_1(w) \\ &= \lambda^j u^j + \lambda^j u^j - \lambda^j u^j(1 - F_0(q_0(w))) - \lambda^j u^j(1 - F_1(w)) \end{aligned} \tag{D.9}$$

Using equation (4.2) gives

$$G^j_0(q_0(w))e^j_0 + G^j_1(w)e^j_1 = \frac{\delta n^j - u^j D^j(w)}{D^j(w)} \tag{D.10}$$

By combining equations (D.7) and (D.10) we obtain

$$g^j_k(q_k(w))e^j_k = \frac{\lambda^j f_k(q_k(w))(u^j + \frac{\delta n^j - u^j(D^j(w-\epsilon))}{D^j(w-\epsilon)}}{D^j(w)} \tag{D.11}$$

D.5.2 Firm size

The number of workers of type j in steady-state employed at a firm in sector k which offers wage rate $q_k(w)$ is

$$\begin{aligned} l_k^j(q_k(w)) &= \frac{g_k^j(q_k(w))e_k^j}{f_k(q_k(w))} \\ &= \frac{\lambda^j \delta n^j}{D^j(w)D^j(w-\epsilon)} \end{aligned} \tag{D.12}$$

The steady state firm size is then

$$\begin{aligned} l_k(q_k(w)) &= l_k^s(q_k(w)) + l_k^{f1}(q_k(w)) + l_k^{f2}(q_k(w)) \\ &= \begin{cases} \frac{\lambda^s \delta n^s}{D^s(w)D^s(w-\epsilon)} + \frac{\lambda^f \delta n^{f1}}{D^{f1}(w)D^{f1}(w-\epsilon)} + \frac{\lambda^f \delta n^{f2}}{D^{f2}(w)D^{f2}(w-\epsilon)} & \forall w \leq w^* \\ \frac{\lambda^s \delta n^s}{D^s(w)D^s(w-\epsilon)} + \frac{\lambda^f \delta n^{f1}}{D^{f1}(w)D^{f1}(w-\epsilon)} & \forall w > w^* \end{cases} \end{aligned} \tag{D.13}$$

Following the standard arguments of profit equalisation, we find the following (the reasoning is parallel to the case without hours variation):

Proposition (A1) *There can be (at most one) mass point in the wage offer distribution at the threshold in each sector, i.e. at wages* $w_k^* \equiv \frac{z^*}{h_k}$.

Sketch of Proof: The following argument closely mirrors the argument in the case of homogeneous hours. We compare profits at the threshold value with profits above. We find that if there are offers above, there must be a mass point at the threshold.

The profit of a sector k firm offering wage rate $q_k(w)$ can be expressed as $\pi_k(q_k(w)) = (ph_k - q_k(w)h_k)l_k(q_k(w))$. We first state the profits of a type-1-firm, assuming that $q_0(w_1^*) \leq w_0^*$.

$$\begin{aligned} \pi_k(w_1^*) &= \frac{\lambda^s \delta n^s}{D^s(w_k^*)D^s(w_k^*-\epsilon)} + \frac{\lambda^f \delta n^{f1}}{D^{f1}(w_k^*)D^{f1}(w_k^*-\epsilon)} + \frac{\lambda^f \delta n^{f2}}{D^{f2}(w_k^*)D^{f2}(w_k^*-\epsilon)} \\ &= \frac{\lambda^s \delta n^s}{[\delta + \lambda^s((1-F_1(w_1^*)) + (1-F_0(q_0(w_1^*))))][\delta + \lambda^s((1-F_1(w_1^*-\epsilon)) + (1-F_0(q_0(w_1^*)-\epsilon)))]} + \\ &+ \frac{\lambda^f \delta n^{f1}}{[\delta + \lambda^f((1-F_1(w_1^*)) + (1-F_0(q_0(w_1^*))))][\delta + \lambda^f((1-F_1(w_1^*-\epsilon)) + (1-F_0(q_0(w_1^*)-\epsilon)))]} + \\ &+ \frac{\lambda^f \delta n^{f2}}{[\delta + \lambda^j(F_0(w_0^*) - F_0(q_0(w_1^*)))][\delta + \lambda^j(f_1(w_1^*) + (F_0(w_0^*) - F_0(q_0(w_1^*)-\epsilon)))]} \end{aligned} \tag{D.14}$$

Evaluated slightly above the threshold, profits are

$$\begin{aligned} \pi_k(w_1^*+\epsilon) &= \frac{\lambda^s \delta n^s}{D^s(w_k^*+\epsilon)D^s(w_k^*)} + \frac{\lambda^f \delta n^{f1}}{D^{f1}(w_k^*+\epsilon)D^{f1}(w_k^*)} \\ &= \frac{\lambda^s \delta n^s}{[\delta + \lambda^s((1-F_1(w_1^*+\epsilon)) + (1-F_0(q_0(w_1^*+\epsilon))))][\delta + \lambda^s((1-F_1(w_1^*)) + (1-F_0(q_0(w_1^*))))]} + \\ &+ \frac{\lambda^f \delta n^{f1}}{[\delta + \lambda^f((1-F_1(w_1^*+\epsilon)) + (1-F_0(q_0(w_1^*+\epsilon))))][\delta + \lambda^f((1-F_1(w_1^*)) + (1-F_0(q_0(w_1^*))))]} \end{aligned} \tag{D.15}$$

Equations (D.14) and (D.15) shows that the equal profit condition can only hold if there is a mass point in the offer distribution of sector 1 at w_1^*. By symmetry, note that the same argument can be made with respect to a type-0 firm. However, if the utility of a threshold offer lies in the "gap area" due to a threshold in another sector, it may be the case that there is no mass point in that sector. This explains the restriction "at most one" in Proposition (IV) and completes this sketch of a proof.

We now consider the influence of thresholds in other hours sectors on the wage distribution. Consider a firm of type-1, i.e. seeking a worker to work for h_1 hours. The impact of a potential mass point in the offer distribution of sector 0 at w_0^* depends on the relation between w_1^*, $q_0(w_1^*)$ and w_0^*.

Proposition (A2) *There will be no wage offers at wage levels (and in a certain interval below this level) that offer the same utility as is available at threshold wages $w_{j\neq\ k}^*$ in other sectors.*

The intuition for Proposition (A2) is the following: It is a dominated strategy to offering a wage rate that is equal in utility to an offer made by several other firms. A slightly higher offer will attract all workers from these firms at only marginal cost. By Proposition (IV), wage offers at earnings thresholds generate mass points in the wage offer distributions. Thus for example a type-1 firm will offer a wage rate slightly larger than $\tilde{w}_1$ (where $U(\tilde{w}_1, h_1) = U(w_0^*, h_0)$.) in order to additionally attracts workers from this positive mass of sector 0 firms. This implies that there must be a gap in the wage offer distribution at $\tilde{w}_1$. How much below this utility value an offer can be sustained in equilibrium will depend on the parameters of the model in an analogous way to the potential existence of offers below the threshold offer in the homogeneous case.

Sketch of proof of Proposition (A2)
Let $\tilde{w}_1$ denote the wage rate which satisfies $U(\tilde{w}_1, h_1) = U(w_0^*, h_0)$. If $\tilde{w}_1 > w_1^*$ the profits of a sector 1 firm offering wage rate $\tilde{w}_1$ and slightly above are:

$$\pi_k(\tilde{w}_1) = \frac{\lambda^s \delta n^s}{D^s(\tilde{w}_k) D^s(\tilde{w}_k - \epsilon)} + \frac{\lambda^f \delta n^{f1}}{D^{f1}(\tilde{w}_k) D^{f1}(\tilde{w}_k - \epsilon)} + \frac{\lambda^f \delta n^{f2}}{D^{f2}(\tilde{w}_k) D^{f2}(\tilde{w}_k - \epsilon)}$$

$$= \frac{\lambda^s \delta n^s}{[\delta + \lambda^s((1 - F_1(\tilde{w}_1)) + (1 - F_0(w_0^*)))][\delta + \lambda^s((1 - F_1(\tilde{w}_1 - \epsilon)) + (1 - F_0(w_0^* - \epsilon)))]} +$$

$$+ \frac{\lambda^f \delta n^{f1}}{[\delta + \lambda^f((1 - F_1(\tilde{w}_1)) + (1 - F_0(w_0^*)))][\delta + \lambda^f((1 - F_1(\tilde{w}_1 - \epsilon)) + (1 - F_0(w_0^* - \epsilon)))]} +$$

$$+ \frac{\lambda^f \delta n^{f2}}{[\delta + \lambda^j((F_1(w_1^*) - F_1(\tilde{w}_1)) + (F_0(w_0^*) - F_1(w_0^*))))][\delta + \lambda^j((F_1(w_1^*) - F_1(\tilde{w}_1 - \epsilon)) + (F_0(w_0^*) - F_1(w_0^* - \epsilon)))]}$$

$$= \frac{\lambda^s \delta n^s}{[\delta + \lambda^s((1 - F_1(\tilde{w}_1)) + (1 - F_0(w_0^*)))][\delta + \lambda^s((1 - F_1(\tilde{w}_1)) + (1 - F_0(w_0^*) + f_0(w_0^*))))]} +$$

$$+ \frac{\lambda^f \delta n^{f1}}{[\delta + \lambda^f((1 - F_1(\tilde{w}_1)) + (1 - F_0(w_0^*)))][\delta + \lambda^f((1 - F_1(\tilde{w}_1)) + (1 - F_0(w_0^*) + f_0(w_0^*)))]} +$$

$$+\frac{\lambda^f\delta n^{f2}}{[\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1)))][\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1))+f_0(w_0^*))]} \quad \text{(D.16)}$$

$$\begin{aligned}
\pi_k(\tilde{w}_1-\epsilon) &= \frac{\lambda^s\delta n^s}{D^s(\tilde{w}_k-\epsilon)D^s(\tilde{w}_k-2\epsilon)}+\frac{\lambda^f\delta n^{f1}}{D^{f1}(\tilde{w}_k-\epsilon)D^{f1}(\tilde{w}_k-2\epsilon)}+\frac{\lambda^f\delta n^{f2}}{D^{f2}(\tilde{w}_k-\epsilon)D^{f2}(\tilde{w}_k-2\epsilon)}\\
&= \frac{\lambda^s\delta n^s}{[\delta+\lambda^s((1-F_1(\tilde{w}_1-\epsilon))+(1-F_0(w_0^*-\epsilon)))][\delta+\lambda^s((1-F_1(\tilde{w}_1-2\epsilon))+(1-F_0(w_0^*-2\epsilon)))]}+\\
&+\frac{\lambda^f\delta n^{f1}}{[\delta+\lambda^f((1-F_1(\tilde{w}_1-\epsilon))+(1-F_0(w_0^*-\epsilon)))][\delta+\lambda^f((1-F_1(\tilde{w}_1-2\epsilon))+(1-F_0(w_0^*-2\epsilon)))]}+\\
&+\frac{\lambda^f\delta n^{f2}}{[\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1-\epsilon))+(F_0(w_0^*)-F_1(w_0^*-\epsilon)))][\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1-2\epsilon))+(F_0(w_0^*)-F_1(w_0^*-2\epsilon)))]}\\
&= \frac{\lambda^s\delta n^s}{[\delta+\lambda^s((1-F_1(\tilde{w}_1))+(1-F_0(w_0^*)+f_0(w_0^*)))]^2}+\\
&+\frac{\lambda^f\delta n^{f1}}{[\delta+\lambda^f((1-F_1(\tilde{w}_1))+(1-F_0(w_0^*)+f_0(w_0^*)))]^2}+\\
&+\frac{\lambda^f\delta n^{f2}}{[\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1))+f_0(w_0^*))]^2}
\end{aligned} \quad \text{(D.17)}$$

$$\begin{aligned}
\pi_k(\tilde{w}_1+\epsilon) &= \frac{\lambda^s\delta n^s}{D^s(\tilde{w}_k+\epsilon)D^s(\tilde{w}_k)}+\frac{\lambda^f\delta n^{f1}}{D^{f1}(\tilde{w}_k+\epsilon)D^{f1}(\tilde{w}_k)}+\frac{\lambda^f\delta n^{f2}}{D^{f2}(\tilde{w}_k+\epsilon)D^{f2}(\tilde{w}_k)}\\
&= \frac{\lambda^s\delta n^s}{[\delta+\lambda^s((1-F_1(\tilde{w}_1+\epsilon))+(1-F_0(w_0^*+\epsilon)))][\delta+\lambda^s((1-F_1(\tilde{w}_1))+(1-F_0(w_0^*)))]}+\\
&+\frac{\lambda^f\delta n^{f1}}{[\delta+\lambda^f((1-F_1(\tilde{w}_1+\epsilon))+(1-F_0(w_0^*+\epsilon)))][\delta+\lambda^f((1-F_1(\tilde{w}_1))+(1-F_0(w_0^*)))]}+\\
&+\frac{\lambda^f\delta n^{f2}}{[\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1+\epsilon)))][\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1)))]}
\end{aligned} \quad \text{(D.18)}$$

$$\begin{aligned}
&= \frac{\lambda^s\delta n^s}{[\delta+\lambda^s((1-F_1(\tilde{w}_1))+(1-F_0(w_0^*)))]^2}+\\
&+\frac{\lambda^f\delta n^{f1}}{[\delta+\lambda^f((1-F_1(\tilde{w}_1))+(1-F_0(w_0^*)))]^2}+\\
&+\frac{\lambda^f\delta n^{f2}}{[\delta+\lambda^j((F_1(w_1^*)-F_1(\tilde{w}_1)))]^2}
\end{aligned} \quad \text{(D.19)}$$

As $f_0(w_0^*) > 0$ and $\epsilon \to 0$, it holds that $\pi_k(\tilde{w}_1-\epsilon) < \pi_k(\tilde{w}_1) < \pi_k(\tilde{w}_1+\epsilon)$. This implies that there will be no wage offers of value $\tilde{w}_1$. As $(ph-wh)$ increases with decreasing w, there might be a wage rate w' where it holds that $\pi_k(w') = \pi_k(\tilde{w}_1+\epsilon)$. This implies $f_1(.)$ exhibit a gap in the interval (w',$\tilde{w}_1$]. If $\tilde{w}_1 < w_1^*$, the terms of equations (D.16) and (D.18) referring to workers of type $f2$ drop out. Although this might reduce the extent of the gap, $\pi_k(\tilde{w}_1) < \pi_k(\tilde{w}_1+\epsilon)$ still holds. If $\tilde{w}_1 = w_1^*$ the necessary size of the mass point at w_1^* to balance the loss of type $f2$ workers decreases (in comparison to $\tilde{w}_1 \neq w_1^*$). How large the gap is, i.e. whether any offers will be made below $\tilde{w}_1$ will depend on the economic environment captured by the parameters of the model.

D.6 Predicting income tax rates

As in Germany income tax rates depend among others on household characteristics, the SIAB data set used in the analysis above does not include most income tax-relevant variables. We thus impute tax rates based on another data set, the GSOEP. The tax simulation accounts for the most important aspects of the German tax-transfer system (Junge, n.d.). Detailed information is available on request. For type f workers explaining tax rates by variables included in both data sets does not yield satisfying predictions. As we aim at predicting the marginal tax rate directly above the threshold, gross earnings, the most valuable predictor, do not vary. Other variables do hardly add explanatory power. (Junge, n.d.) We therefore replicate the SIAB sample in the GSOEP data and differentiate three tax groups based on the distribution of simulated tax rates. Observations in the estimation sample are then allocated randomly the mean tax rate of one of these groups. Since earnings at the threshold are less than the general tax allowance almost 50% of type f workers have an income tax rate of zero (table D.1). The imputed tax rate for the next 40 % is about eight per cent, for the highest decile it is approximately 20 %.

Tax rates for type s workers are explained by a tobit model with first job earnings and sex as explaining variables (table D.2) [92]. Based on the same information in the SIAB data we then predict the individual tax rates. Type s workers are also allocated to three groups (table D.1). Tax rates are higher for type s workers as they already have first job earnings. As our sample is restricted to observations with first job earnings of more than 1000 € (section 4.4.1), all observations exceed the general tax allowance and thus have a strictly positive marginal tax rate.

Table D.1: Income tax groups

	Type s			Type f		
Group	%	t	std	%	t	std
1	25	8.32	1.75	48.27	0	0
2	65	12.99	4.17	41.73	7.92	4.64
3	10	27.6	8.52	10	19.47	4.50

Notes: t= mean average income tax rate at 326 €; std=standard deviation; type s: aggregated predicted tax rates based on model in table D.2; type f: aggregated observed tax rates
Source: Own calculations based on GSOEP wave 1999-2002

D.7 Predicting hours of work

We do not have information on hours of work offered in the market. We, therefore, seek to impute actual working hours by means of survey data (GSOEP). To do that we model hours

[92] To increase precision we use all second jobs for the estimation of tax rates. We do not restrict the sample with respect to first job earnings as done for the main estimation.

Table D.2: Estimation results: tax rates type s workers

	coeff.	s.e.
Sex	-0.0422	0.003
Yearly first job earnings	4.27e-06	7.50e-08
Constant	0.0354	0.002

Notes: coeff.=regression coefficient, s.e.=standard errors.
Source: Own calculations based on GSOEP wave 1999-2002

based on variables included in both data sets: sex, gross earnings, sector and education. As the hours distribution resembles a log-normal distribution (figure 4.2) a generalised linear model with log link is used for imputation. As in the main analysis the estimation sample comprises two groups of employees. These are, first, employees with only one job and earnings of less than 800 €/month. We use their actual hours of work per week. The sample includes, second, employees holding a main job with earning above 1000 €/month and a side job with earnings below 800 €/month. Weekly hours of work in the side job are used as dependent variable. The model is estimated jointly for both groups (table D.3). The resulting hours distribution increases sharply up to a global peak at approximately 50 hours per month (figure D.1). The right-handy tail covers hours up to 200. The mean is fit very well. The variation is smaller in the predicted data, though (table D.4).

Table D.3: Estimation results: hours

	coeff.	s.e.
Gross Earnings	0.00	0.00
Sex	0.03	0.05
Sector		
Manufacturing	-0.38	0.11
Energy, Water	-0.28	0.20
Construction	-0.21	0.14
Wholesale and retail	-0.26	0.10
Hotels and restaurants	-0.23	0.15
Transport	-0.40	0.14
Finance	-0.43	0.17
Real Estate	-0.44	0.11
Public Admin	-0.51	0.14
Education	-0.48	0.13
Health	-0.57	0.11
Other Services	-0.28	0.12
Households	-0.35	0.14
Missing	-0.33	0.11
Education		
Missing	0.56	0.13
Basic	0.26	0.09
Middle Voc.	0.28	0.09
Higher Voc.	0.18	0.10
Constant	1.83	0.14

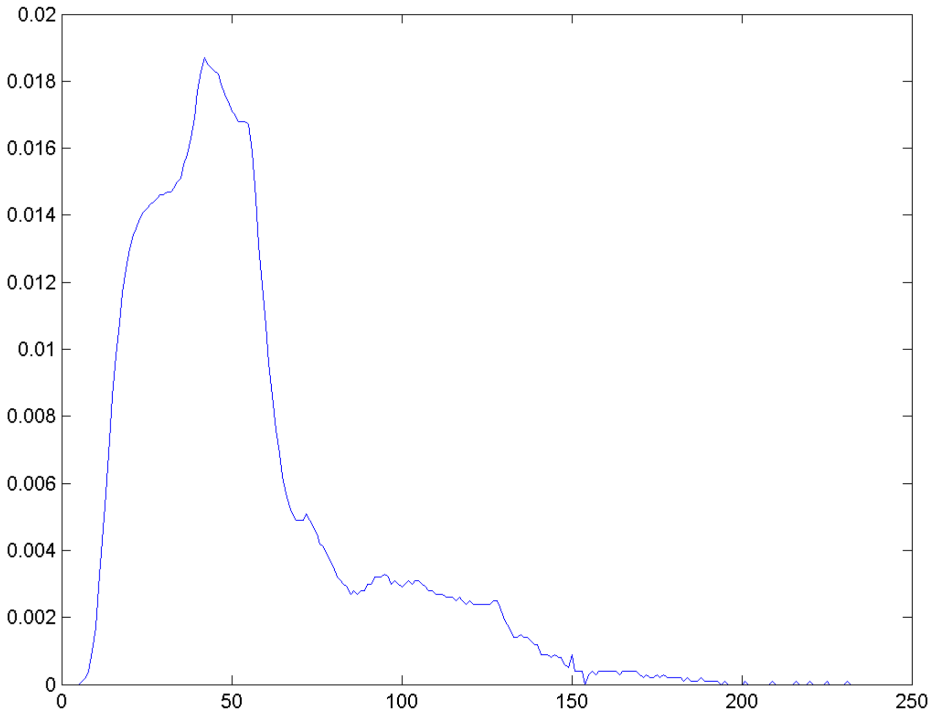
Notes: coeff.=regression coefficient, s.e.=standard errors., dependent variable: weekly hours
Source: Own calculations based on GSOEP wave 2001

Table D.4: Fit of hours distribution

	Observed	Predicted
mean	57.25	57.14
$p(25)$	23.11	34.45
$p(50)$	43.33	52.17
$p(75)$	86.67	76.91

Notes: $p(x)$: xth quantile
Source: SIAB; own calculations.

Figure D.1: Probability density function of monthly hours of work



Source: SOEP, SIAB, own calculations.

D.8 Numerical approaches

This section describes two numerical approaches applied in the estimation procedure. First, the numerical calculation of the job offer distribution $F(.)$ and, second, the numerical optimization of the likelihood.

D.8.1 Job offer distribution

In the case with homogeneous hours the offer distribution can be characterised by a system of equal profit conditions. Firms with different hours, however, may make different profits in equilibrium. The offer distribution in the case of heterogeneous hours is, thus, calculated by solving the maximization problem successively for all firms in the market. Firms take the current offer distribution as given which is then updated once a firm has decided on a wage offer. The order of firms deciding on an offer is determined by random sampling with replacement. We run the algorithm based on 100 firms which first draw once from the predicted hours distribution (see above). Firms can chose between 100 different wages. The algorithm is stopped after 1000 draws. After this, the distribution is fairly stable: The estimated fraction of offers with earnings below or at the threshold, for example, has a standard deviation of less than 0.02 when the algorithm is repeated ten times. To increase precision, we take the average offer distribution of ten repetitions.

D.8.2 Optimization of likelihood

The likelihood is optimised with respect to θ, α and σ by a two-step grid search procedure (Hansen, 2016). Gradient-based approaches are inappropriate because the costly calculation of the job offer distribution has to be repeated every iteration. It further contains random elements rendering the likelihood function unsmooth.
For θ and α we evaluate the likelihood in the first step at 11 equally spaced grid points in the interval $[0, 1]$. In the second step the grid is narrowed down to steps of 0.02. For σ we use the 50, 100 and 150 as grid points in the first step. The grid is narrowed down to steps of 20 in the second step.

D.9 Constant amount of vacancies

Extensive margin reactions in our model are exclusively driven by type $f2$ workers who leave the market when the tax exemption is disposed. Due to the assumption of exogenous and constant frictional parameters, the overall number of vacancies posted by firms decreases. Consider by contrast a scenario where type $f2$ workers are substituted by workers of other types. We model that by assuming that the number of vacancies per period posted by firms stay constant. Offer arrival rates for type s and $f1$ workers are thus increased. Keeping the ratio between both constant, this implies $\lambda_s = 0.0513$ and $\lambda_f = 0.3908$. This is comparable to introducing a matching function (Pissarides and Petrongolo, 2001) which accounts for on-the-job search and differing arrival rates across worker groups.

Both type s and $f1$ workers can substantially increase their net earnings and utility - albeit the latter forgoes the tax exemption. Total hours still decrease by about 12%. This is overcompensated by an increase in average wage rate, though, such that total gross earnings increase. Tax revenues increase accordingly.

The results of such a simulation can be interpreted as an upper bound. When economic incidence of the tax exemption, for example, is not entirely on employees, firms profit from lower wage rates for Minijobs relative to jobs subject to full SSC. Removing the tax exemption might then lead to decrease in vacancies. The substantial differences between both scenarios imply that labour demand responses are crucial for comprehensively evaluating the Minijob regulation. Our model indeed includes firms catering to the aggregate preferences in the market. It, however, abstracts from other important margins of firm responses.

Table D.5: Removing the tax and SSC exemption - upper bound scenario

Change in	Total	type s	type f	type $f1$
$\bar{w}$	0.78	0.09	0.92	.
$\bar{h}$	18.05	4.98	9.92	.
$\sum h$	-10.01	41.27	15.15	-100
$\bar{z}$	179.98	36.67	156.08	.
$\sum z$	9.72	47.45	41.38	-100
$\bar{c}$	83.96	24.98	70.31	.
$\bar{v}$	69.11	20.25	55.81	-199.34
jobs	-1360.5	163	120.5	-1644
taxes	341.04	39.64	387.61	-86.21

Notes: $\bar{w}$, $\bar{z}$ and $\bar{c}$ in €; $\sum h$ and $\sum z$ in %; jobs in 1000s; taxes in million €; jobs and taxes extrapolated to population; $\bar{w}$, $\bar{h}$, $\bar{z}$, and $\bar{c}$ conditional on employment, $\bar{v}$ based on all individuals
Source: SIAB; own calculations.

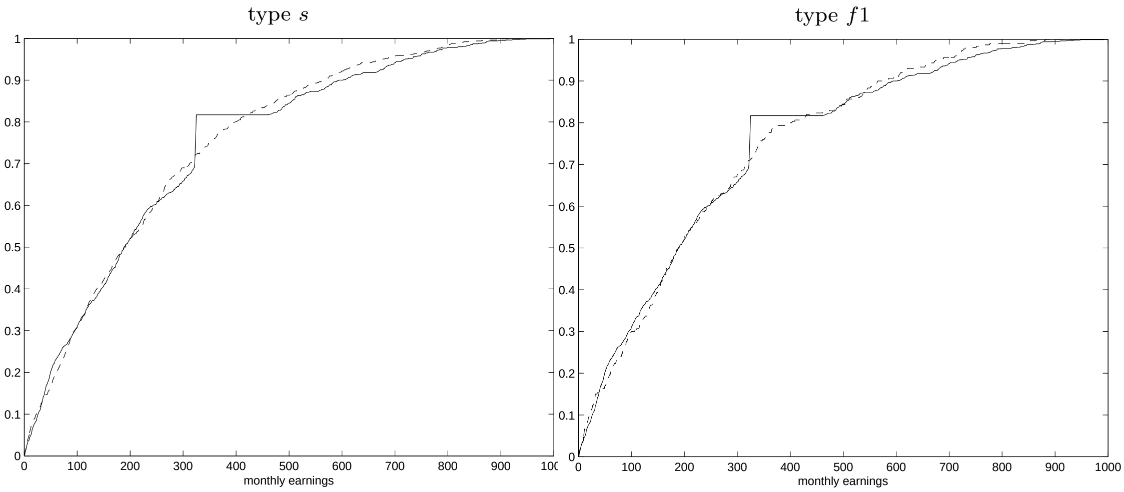
D.10 Low-paid market with one worker type

Imagine all workers in the low-paid market are of type $f1$. That is, all workers are eligible for a potential tax exemption but would work when subject to full taxes. Counterfactual simulations in such a market are insightful because they abstract from differential changes of incentives between worker types and compositional changes in the workforce.

When the tax exemption is removed, firms which offered jobs at the threshold in the status quo now mostly offer jobs with higher earnings (left panel of figure D.2). This is not surprising as jobs with high earnings are relatively more attractive when they are not subject to a disadvantaged tax treatment. Some firms reduce their earnings offer, though. The likelihood of earnings farther below the threshold is not affected. Removing the tax exemption has a positive effect on average (and total) gross earnings (table D.6). One reason is the higher likelihood of a high

earnings offer, another that workers more likely accept a high earnings offer.
Another way to smooth the Minijob discontinuity is a gradually increasing tax rate for earnings between 325 and 800 €. The full tax exemption below 325 € remains. Firms which offered jobs at the threshold in the status quo increase their earnings offers. Relative to the status quo the distribution above the threshold is compressed, though. The upper earnings limit decrease from x €in the status-quo to y €. Firms account for the change in workers' incentives due to the high implicit marginal tax rate introduced by the reform.

Figure D.2: Cumulative earnings offer distribution - status quo vs. reform - only f1 workers



Source: own calculations.

Table D.6: Effects of smoothing the notch - only f1 workers

Change in	Removal	Kink
$\bar{w}$	0.50	0.33
$\bar{h}$	-0.56	-2.60
$\sum h$	-0.58	-2.71
$\bar{z}$	40.73	6.93
$\sum z$	8.67	1.48
$\bar{c}$	-10.56	25.28
$\bar{v}$	-7.67	17.43
jobs	0.00	0.00
taxes	269.64	-76.41

Notes: Removal: Complete removal of tax exemption; Kink: Notch replaced by kink; $\bar{w}$, $\bar{z}$ and $\bar{c}$ in €; $\sum h$ and $\sum z$ in %; jobs in 1000s; taxes in million €; jobs and taxes extrapolated to population; $\bar{w}$, $\bar{h}$, $\bar{z}$, and $\bar{c}$ conditional on employment, $\bar{v}$ based on all individuals
Source: SIAB; own calculations.

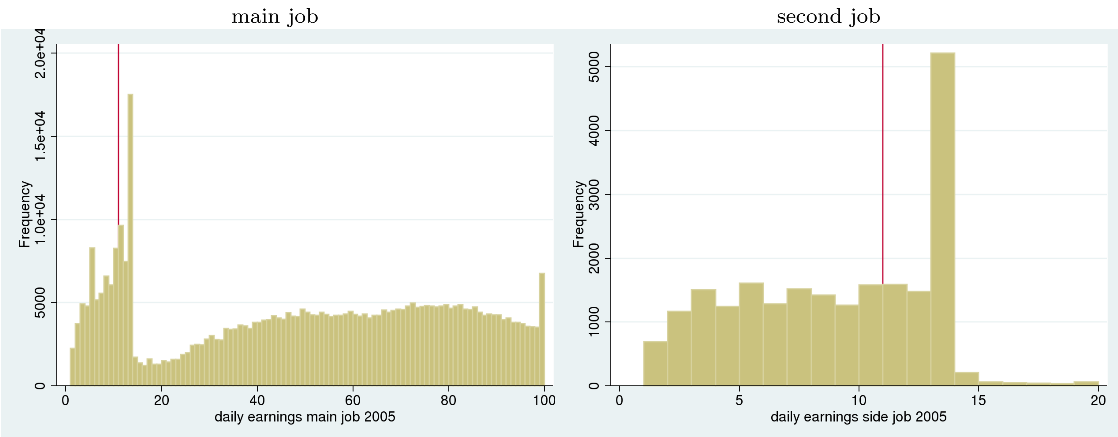
D.11 The 2003 Reform

Since 2003, earnings from at most one second job are subject to the advantageous tax treatment of Minijobs. Individuals who prior to the reform had only one job now have an incentive to take up a second job given the increase in marginal return to working an extra hour in form of a second job. Although the reform resulted in a sustained increase in number of second jobs, most individuals still just work in one job. We believe that set-up costs of a second job (bureaucracy etc.) prevent many individuals from doing this. Additional utility derived from additional earnings must thus at least cover the set-up costs. Subject to assumptions about the distribution of these costs we can use the earnings distribution of actual new second jobs to identify set-up costs which deter marginal second jobs.

The change in tax rules also created an incentive to reduce earnings in the existing job and take on a second job with earnings up to the maximum tax-free threshold (or increase earnings of an existing second job). However, adjusting hours is often not easy. The likelihood of observing such adjustments will depend on the benefits minus the costs - thus conditional on observing the benefits of this choice (taking into account leisure, consumption and job-seeking parameters), the distribution of the earnings adjustments in the first job then identifies the size of the adjustment costs.

Figure D.3 contains the distributions of main and second job earnings for 2005, two years after the reform. The red lines mark the old threshold. For both, main and second jobs, the peak moved to the new threshold. While for main jobs this seems to be the only noteworthy change, the number of second jobs increased tremendously.

Figure D.3: Distribution of earnings in 2005



Source: SIAB.

This version of the paper does not exploit this reform and focuses only on the period before 2003. In future versions, though, we plan to use the reform in order to either validate our model, relax restrictive assumptions concerning the comparability of workers looking for main and second jobs or to estimate additional structural parameters like adjustment or setting-up costs.

D.12 Additional tables

Table D.7: Job type

	type s		type f	
	$z \leq z^*$	$z > z^*$	$z \leq z^*$	$z > z^*$
Minijob	1,762	125	194,584	13,852
s.t. SSC	25,320	2,914	13,927	64,892

Notes: s.t. SSC: employment subject to social security contributions; $z \leq z^*$: earnings below or at the Minijob threshold; $z > z^*$: earnings above the Minijob threshold
Source: SIAB

Table D.8: Spell durations

Status \ Type	s	f
Unemployment spell		
uncensored	5.76	4.14
only right-censored	9.40	4.62
only left-censored	12.74	9.42
left+right-censored	10.87	6.2
Job spell		
uncensored	5.56	5.45
only right-censored	10.75	10.85
only left-censored	10.32	12.19
left+right-censored	28.94	28.09
Employment spell		
uncensored	6.10	8.49
only right-censored	11.76	14.77
only left-censored	11.94	17.72
left+right-censored	29.37	29.04

Source: SIAB, own calculations

Summary

This dissertation focuses on earnings responses to SSC and social benefits which amounted to on average 10.3% and 13.6 % of GDP, respectively, in the OECD countries in 2014. Their large magnitude makes it essential to understand and quantify their impact on labour market outcomes. The four chapters of this dissertation cover the two crucial channels of labour market reactions encompassed in earnings responses: hours of work and compensation per hour. The main joint feature of all chapters is that they look beyond pure labour supply reactions when evaluating social policy and recognise the importance of firms and equilibrium effects. While the research questions analysed are of general interest, the empirical applications are based on German data.

The first chapter exploits discontinuities induced by earnings caps for social security contributions (SSC) in Germany to analyse the effect of SSC on gross labour earnings. Empirical evidence is based on two complementary approaches utilising two administrative data sets. First, employment responses to SSC at the intensive margin are identified by a modified bunching approach that is applied to kinks in the budget set generated by the earnings caps. Second, I exploit an increase of a regional earnings cap of health and long-term care insurance as a natural experiment. In order to analyse economic incidence a difference-in-differences approach is used to estimate the effects on gross earnings. I find employment responses to be negligible and the burden of SSC to be shared equally between employers and employees. Both results turn out to be robust and are consistent with a competitive labour market model.

The second chapter estimates economic incidence of social security contributions (SSC) on the basis of cross-sectional earnings distributions. The approach exploits discontinuities in earnings distributions at kinks in the budget set which are informative about tax incidence. Contrary to most research on SSC incidence, it does not rely on policy reforms, panel data, or hours information. When the location of kinks does not change significantly, estimates represent equilibrium incidence and are less affected by short-run adjustment frictions than results based on policy reforms. The approach is applied to earnings caps of SSC in Germany

where the marginal SSC rate drops to zero. Substantial negative discontinuities are found at most earnings caps of SSC in the distribution of observed net earnings. Together with smooth gross earnings distributions around the caps this provides consistent empirical evidence that legal and economic incidence of SSC coincide.

The third chapter extends a static discrete-choice labor supply model by adding participation and hours constraints. Restrictions are identified by survey information providing for a more robust identification of preferences and constraints. Various restriction mechanisms are distinguished and their substantial impact on employment is shown: labour demand rationing, working hours norms varying across occupations, and insufficient public childcare on the supply side of the market. The empirical framework is applied to evaluate an in-work benefit for low-paid parents in the German institutional context. The benefit is supposed to increase work incentives for secondary earners. The structural model allows for disentangling behavioural reactions into the pure incentive effect and the limiting impact of constraints. The in-work benefit for parents is found to substantially increase working hours of mothers of young children, especially when they have a low education. A standard labour supply model that ignores labour market constraints, however, would yield upward-biased estimates and overstate the benefit's policy impact.

Labour supply in the market for low-paid jobs in Germany is strongly influenced by nonlinearities in the tax schedule - even for individuals to whom this tax schedule does not apply. Chapter four presents compelling evidence that an individual's choice set depends on other workers' preferences because firms cater their job offers to aggregate preferences in the market. An equilibrium job search model is estimated which rationalises earnings bunching at a tax discontinuity as firm responses using German administrative data. The model is used to simulate smoothing the tax schedule by reforming the tax and SSC exemption for low earnings in Germany. Results highlight the indirect costs of (discontinuous) tax policies which are shown to be reinforced by firm responses.

German Summary

Diese Dissertation untersucht den Effekt von Sozialversicherungsbeiträgen (SVB) und Sozialleistungen auf das Bruttoarbeitseinkommen. Mit durchschnittlich 10.3 % und 13.6 % des BIP der OECD-Länder in 2014 ist das Potential gro für Auswirkungen auf den Arbeitsmarkt. Diese zu verstehen und quantifizieren ist deshalb unbedingt nötig. Der in dieser Dissertation analysierte Effekt auf Bruttoarbeitseinkommen deckt dabei zwei entscheidende Mechanismen für Arbeitsmarktreaktionen ab: Arbeitstunden und Stundenlohn. Die vier Kapitel dieser Dissertation vereint, dass sie nicht nur auf reine Arbeitsangebotseffekte begrenzt sind, sondern auch die Bedeutung von Firmen und Gleichgewichtseffekten anerkennen. Auerdem basieren alle empirischen Anwendungen auf deutschen Daten, obwohl die untersuchten Forschungsfragen von generellem Interesse sind.

Das erste Kapitel untersucht den Effekt von SVB auf das Bruttoarbeitseinkommen. Dafür wird ausgenutzt, dass die Beitragsbemessungsgrenzen (BBG) der deutschen Sozialversicherung Diskontinuitäten in der Budget-Geraden erzeugen. Es werden zwei komplementäre Untersuchungsansätze angewendet, die auf zwei verschiedenen administrativen Datensätzen beruhen. Beschäftigungseffekte von SVB werden mit einer modifizierten Version des so genannten *bunching*-Ansatzes identifiziert. Dieser wird auf die Knickung der Budget-Gerade an den BBG angewendet. Die ökonomische Inzidenz von SVB wird mit einem quasi-experimentellen Ansatz geschätzt. Dabei wird eine starke Erhöhung der BBG der Krankenversicherung in Ostdeutschland als natürliches Experiment ausgenutzt. Ich kann keine nennenswerten Beschäftigungseffekte nachweisen. Die ökonomische Belastung der SVB weicht nicht von der gesetzlichen Aufteilung zwischen Arbeitgebern und Arbeitnehmern ab. Beide Ergebnisse sind robust und sind mit einem kompetitiven Arbeitsmarktmodell vereinbar.

Im zweiten Kapitel wird die ökonomische Inzidenz von SVB auf Basis von Arbeitseinkommensverteilungen im Querschnitt geschätzt. Dieser Ansatz nutzt aus, dass Diskontinuitäten in diesen Verteilungen an steuerinduzierten Knickungen der Budget-Gerade etwas über die Inzidenz einer Steuer aussagen. Im Gegensatz zu dem Groteil der empirischen Forschung über die

Inzidenz von SVB, basiert der hier gewählte Ansatz weder auf Reformen, Panel-Daten noch auf Arbeitszeitinformationen. Wenn sich die Position eines Knicks in der Budget-Geraden nicht signifikant verändert, können Gleichgewichtseffekte identifiziert werden. Diese sind weniger von kurzfristigen Friktionen beeinflusst als Ergebnisse, die auf Reformen basieren. Das Verfahren wird auf die Beitragsbemessungsgrenzen der deutschen Sozialversicherung angewendet. Übersteigt das Arbeitseinkommen die BBG, fällt der marginal Beitragssatz auf null. Es werden groe negative Diskontinuitäten an den meisten BBG in der Verteilung der Nettoarbeitseinkommen gefunden. Die Verteilung der Bruttoarbeitseinkommen dagegen verläuft kontinuierlich. Zusammen implizeren diese Ergebnisse, dass sich gesetzliche und ökonomische Inzidenz entsprechen.

Das dritte Kapitel erweitert ein statisches diskretes Auswahlmodell (*discrete choice model*) um Partizipations- und Stundenrestriktionen. Diese werden durch Befragungsdaten identifiziert, was eine robustere Identifikation von Präferenzen und Restriktionen ermöglicht. Es werden mehrere Restriktionsmechanismen unterschieden und ihren Einfluss auf die Beschäftigung simuliert: knappe Arbeitsnachfrage, über Berufe variierende Arbeitszeitnormen und unzureichende öffentliche Kinderbetreuung. Das Modell wird genutzt, um eine hypothetische staatliche Leistung für niedrigverdienende Eltern im deutschen Kontext zu evaluieren, die Arbeitsanreize für Zweitverdiener erhöht. Dabei ermöglicht das strukturelle Modell, Verhaltensreaktionen in reine Anreizeffekte und in den begrenzenden Einfluss der Restriktionen aufzuteilen. Es zeigt sich, dass die evaluierte Leistung die Arbeitsstunden für Mütter von kleinen Kindern substantiell erhöht - vor allem für Mütter mit schlechter Ausbildung. Ein konventionelles diskretes Auswahlmodell, welches Arbeitsmarktrestriktionen ignoriert, würde den Effekt der Leistung allerdings überschätzen.

Das Arbeitsangebot im Niedriglohnsektor in Deutschland ist sehr stark von Nicht-Linearitäten im Steuer-Transfersystem beeinflusst. Das gilt selbst für Personen, für die diese formell nicht gelten. Es wird gezeigt, dass die Wahlmöglichkeiten einer Person von den Präferenzen aller Arbeitnehmer abhängen. Der Grund ist, dass Firmen ihre Jobangebote auf die durchschnittlichen Präferenzen auf dem Arbeitsmarkt ausrichten. Unser Gleichgewichts-*job search*-Modell erklärt deshalb die Anhäufung von Arbeitseinkommen an bestimmten Punkten der Lohnverteilung, die Steuerdiskontinuitäten entsprechen, als Folge von Firmenreaktionen. Das Modell wird basierend auf deutschen administrativen Arbeitsmarktdaten geschätzt. Die Effekte von Reformen des Steuer-Transfersystems, die die Nicht-Linearitäten im Niedriglohnsektor beseitigen, werden simuliert. Es wird gezeigt, dass Steuern im Allgemeinen und nicht-lineare im Besonderen (indirekte) Kosten verursachen, die von Firmenreaktionen verstärkt werden können.

Bibliography

Aaberge, Rolf, John K. Dagsvik, and Steinar Strom, "Labor Supply Responses and Welfare Effects of Tax Reforms," *The Scandinavian Journal of Economics*, 1995, *97* (4), 635–659.

Alesina, Alberto and Roberto Perotti, "The Welfare State and Competitiveness," *The American Economic Review*, 1997, *87* (5), 921–939.

Alvaredo, Facundo and Emmanuel Saez, "The Effects of Payroll Taxes on Earnings: Evidence from Spanish Administrative Data," Discussion Paper 2007.

Anderson, Patricia M. and Bruce D. Meyer, "The effects of the unemployment insurance payroll tax on wages, employment, claims and denials," *Journal of Public Economics*, 2000, *78* (1-2), 81–106.

Andrew, Leigh, "Who Benefits from the Earned Income Tax Credit? Incidence among Recipients, Coworkers and Firms," *The B.E. Journal of Economic Analysis & Policy*, 2010, *10* (1), 1–43.

Antonczyk, Dirk, Bernd Fitzenberger, and Katrin Sommerfeld, "Rising wage inequality, the decline of collective bargaining, and the gender wage gap," *Labour Economics*, 2010, *17* (5), 835–847.

Baicker, Katherine and Amitabh Chandra, "The Labor Market Effects of Rising Health Insurance Premiums," *Journal of Labor Economics*, 2006, *24* (3), 609–634.

Bargain, Olivier and Andreas Peichl, "Steady-State Labor Supply Elasticities: A Survey," IZA Discussion Papers 7698, Institute for the Study of Labor (IZA) 2013.

_ **and Kristian Orsini**, "In-work policies in Europe: Killing two birds with one stone?," *Labour Economics*, 2006, *13* (6), 667 – 697.

__ , __ , **and Andreas Peichl**, "Comparing Labor Supply Elasticities in Europe and the United States: New Results," *Journal of Human Resources*, 2014, *49* (3), 723–838.

__ **, Marco Caliendo, Peter Haan, and Kristian Orsini**, "'Making Work Pay' in a Rationed Labour Market: The Mini-Job Reform in Germany," *Journal of Population Economics*, 2010, *23* (1), 323–351.

Beffy, Magali, Richard Blundell, Antoine Bozio, Guy Laroque, and Maxime To, "Labour supply and taxation with restricted choices," Discussion Paper 2016.

Bell, Brian, Jerry Jones, and Jonathan Thomas, "Estimating the impact of changes in employers National Insurance Contribution on wages, prices and employment," Bank of England Quarterly Bulletin 2002.

Bennmarker, Helge, Erik Mellander, and Bjrn ckert, "Do regional payroll tax reductions boost employment?," *Labour Economics*, 2009, *16* (5), 480 – 489.

Best, Michael Carlos, "The Role of Firms in Workers Earnings Responses to Taxes: Evidence From Pakistan," Discussion Paper, London School Of Economics 2014.

Bloemen, Hans G., "A model of labour supply with job offer restrictions," *Labour Ecnomics*, 2000, *7* (3), 297–312.

__ , "Job Search, Hours Restrictions, and Desired Hours of Work," *Journal of Labor Economics*, 2008, *26*, 137–179.

Blundell, Richard, "Work initiatives and in-work benefit reforms: a review," *Oxford Review of Economic Policy*, 2000, *16*, 2744.

__ **, Alan Duncan, and Costas Meghir**, "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 1998, *66* (4), 827–862.

__ **and Thomas Macurdy**, "Labor supply: A review of alternative approaches," in O. Ashenfelter and D. Card, eds., *O. Ashenfelter and D. Card, eds.*, Vol. 3 of *Handbook of Labor Economics* 1999, chapter 27, pp. 1559–1695.

__ **, John Ham, and Costas Meghir**, "Unemployment and Female Labour Supply," *The Economic Journal*, 1987, *97* (Supplement: Conference Papers (1987)), 44–64.

__ **, Mike Brewer, and Marco Francesconi**, "Job Changes and Hours Changes: Understanding the Path of Labor Supply Adjustment," *Journal of Labor Economics*, 2008, *26* (3), 421–453.

Bontemps, Christian, Jean-Marc Robin, and Gerard J. van den Berg, "An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms," *International Economic Review*, 1999, *40* (4), pp. 1039–1074.

Bryan, Mark L., "Free to choose? Differences in the hours determination of constrained and unconstrained workers," *Oxford Economic Papers*, 2007, *59* (2), 226–252.

Burdett, Kenneth and Dale T Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, *39* (2), 257–73.

Burtless, Gary and Jerry A Hausman, "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiments," *Journal of Political Economy*, 1978, *86* (6), 1103–30.

Büttner, Thomas and Susanne Rässler, "Multiple imputation of right-censored wages in the German IAB Employment Sample considering heteroscedasticity," IAB Discussion Paper 200844, Institute for Employment Research (IAB), Nuremberg 2008.

Callan, Tim, Arthur van Soest, and John R. Walsh, "Tax Structure and Female Labour Market Participation: Evidence from Ireland," IZA Discussion Papers 3090, Institute for the Study of Labor (IZA) 2007.

Chetty, Raj, "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 2009, *1* (1), 451–488.

_ , "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, *80* (3), 969–1018.

_ , John N. Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *The Quarterly Journal of Economics*, 2011, *126* (2), 749–804.

Daveri, Francesco and Guido Tabellini, "Unemployment, growth and taxation in industrial countries," *Economic Policy*, 2000, *15* (30), 47–104.

Dearing, Helene, Helmut Hofer, Christine Lietz, Rudolf Winter-Ebmer, and Katharina Wrohlich, "Why Are Mothers Working Longer Hours in Austria than in Germany? A Comparative Microsimulation Analysis," *Fiscal Studies*, 2007, *28* (4), 463–495.

Dias, Monica Costa, Jonathan Shaw, Costas Meghir, and Richard Blundell, "The long-term effects of in-work benefits in a life-cycle model for policy evaluation," 2012 Meeting Papers 93, Society for Economic Dynamics 2012.

Dickens, William T and Shelly J Lundberg, "Hours Restrictions and Labor Supply," *International Economic Review*, 1993, *34* (1), 169–92.

Donald, Stephen G. and Kevin Lang, "Inference with Difference-in-Differences and Other Panel Data," *The Review of Economics and Statistics*, 2007, *89* (2), 221–233.

Dorner, Matthias, Marion Knig, and Stefan Seth, "Stichprobe der Integrierten Arbeitsmarktbiografien. Regionalfile 1975-2008 (SIAB-R 7508)," Documentation on Labour Market Data, Institute for Employment Research (IAB), Nuremberg 2011.

Eichhorst, Werner, Paul Marx, and Verena Tobsch, "Non-Standard Employment across Occupations in Germany: The Role of Replaceability and Labour Market Flexibility," IZA Discussion Papers 7662, Institute for the Study of Labor (IZA) 2013.

Euwals, Rob, "The predictive value of subjective labour supply data: A partial-adjustment model with measurement error," *Empirical Economics*, 2005, *30* (2), 309–329.

_ **and Arthur van Soest**, "Desired and actual labour supply of unmarried men and women in the Netherlands," *Labour Economics*, 1999, *6* (1), 95–118.

Fitzenberger, Bernd, Aderonke Osikominu, and Robert Völter, "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," *Schmollers Jahrbuch: Journal of Applied Social Science Studies*, 2006, *126* (3), 405–436.

_ **, Karsten Kohn, and Alexander C. Lembcke**, "Union Density and Varieties of Coverage: The Anatomy of Union Wage Effects in Germany," *Industrial and Labor Relations Review*, 2013, *66* (1), 169–197.

Freier, Ronny and Viktor Steiner, "Marginal employment: Stepping stone or dead end? : evaluating the German experience," *Zeitschrift fr ArbeitsmarktForschung - Journal for Labour Market Research*, 2008, *41* (2/3), 223–243.

Fullerton, Don and Gilbert E. Metcalf, "Tax incidence," in A. J. Auerbach and M. Feldstein, eds., *A. J. Auerbach and M. Feldstein, eds.*, Vol. 4 of *Handbook of Public Economics*, Elsevier, 2002, chapter 26, pp. 1787–1872.

Gelber, Alexander M., Damon Jones, and Daniel W. Sacks, "Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test," NBER Working Papers 19491, National Bureau of Economic Research, Inc 2013.

Grabka, Markus, "Alternative Finanzierungsmodelle einer sozialen Krankenversicherung in Deutschland Methodische Grundlagen und exemplarische Durchfhrung einer Mikrosimulationsstudie.," Dissertation, Technical University Berlin 2004.

Gruber, Jon and Emmanuel Saez, "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 2002, *84* (1), 1–32.

Gruber, Jonathan, "The Incidence of Mandated Maternity Benefits," *American Economic Review*, 1994, *84* (3), 622–41.

_ , "The Incidence of Payroll Taxation: Evidence from Chile," *Journal of Labor Economics*, 1997, *15* (3), 72–101.

Gudgeon, Matthew and Simon Trenkle, "Frictions in Adjusting Earnings: Evidence From Notches in German Mini Jobs," Discussion Paper 2016.

Haan, Peter and Arne Uhlendorff, "Intertemporal labor supply and involuntary unemployment," *Empirical Economics*, 2013, *44* (2), 661–683.

Hamermesh, Daniel S., "New Estimates of the Incidence of the Payroll Tax," *Southern Economic Journal*, 1979, *45* (4), 1208–1219.

_ , "The demand for labor in the long run," in O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, 1 ed., Vol. 1, Elsevier, 1987, chapter 08, pp. 429–471.

_ , *Labor Demand*, Princeton University Press, Princeton, 1993.

Hans-Boeckler-Stiftung, "WSI-Tarifarchiv," 2015. http://www.boeckler.de/index_wsi_tarifarchiv.htm; accessed 2015.

Hansen, Bruce E., *Econometrics*, University of Wisconsin, 2016.

Hausman, Jerry A., "The Effect of Wages, Taxes, and Fixed Costs on Women's Labor Force Participation," *Journal of Public Economics*, 1980, *14* (2), 161–194.

Heckman, James J, "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–61.

Heckman, James J. and Burton Singer, "Econometric Duration Analysis," *Journal of Econometrics*, 1984, *24* (1), 63–132.

_ **and** _ , "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 1984, *52* (2), 271–320.

Helmcke, Thomas, "Regionalstatistik auf europäischer und nationaler Ebene," *Wirtschaft und Statistik*, 2008, *3*, 207–216.

Holmlund, Bertil, "Payroll Taxes and Wage Inflation: The Swedish Experience," *Scandinavian Journal of Economics*, 1983, *85* (1), 1–15.

Huttunen, Kristiina, Jukka Pirttilä, and Roope Uusitalo, "The employment effects of low-wage subsidies," *Journal of Public Economics*, 2013, *97*, 49 – 60.

Junge, Henrike, "From Gross to Net Wages in German Administrative Data Sets." Unpublished Manuscript.

Kabatek, J., "Iteration Capping For Discrete Choice Models Using the EM Algorithm," Discussion Paper 2013-019, Tilburg University, Center for Economic Research 2013.

Keane, Michael P., "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 2011, *49* (4), 961–1075.

_ , Petra E. Todd, and Kenneth I. Wolpin, "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications," in O. Ashenfelter and D. Card, eds., *O. Ashenfelter and D. Card, eds.*, Vol. 4 of *Handbook of Labor Economics*, Elsevier, 2011, chapter 4, pp. 331–461.

Kerschbamer, Rudolf and Georg Kirchsteiger, "Theoretically robust but empirically invalid? An experimental investigation into tax equivalence," *Economic Theory*, 2000, *16* (3), 719–734.

Kleven, Henrik J. and Mazhar Waseem, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *The Quarterly Journal of Economics*, 2013, *128* (2), 669–723.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Sren Pedersen, and Emmanuel Saez, "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark," *Econometrica*, 2011, *79* (3), 651–692.

Kolm, Ann-Sofie and Mirco Tonin, "In-work benefits and unemployment," *International Tax and Public Finance*, 2011, *18* (1), 74–92.

Kopczuk, Wojciech and Joel Slemrod, "Putting Firms into Optimal Tax Theory," *American Economic Review*, 2006, *96* (2), 130–134.

Korkeamki, Ossi and Roope Uusitalo, "Employment and wage effects of a payroll-tax cutevidence from a regional experiment," *International Tax and Public Finance*, 2009, *16* (6), 753–772.

Körner, Thomas, Holger Meinken, and Katharina Puch, "Wer sind die ausschlielich geringfgig Beschftigten? Eine Analyse nach sozialer Lebenslage," Excerpt from WISTA das Wissenschaftsmagazin, Statistisches Bundesamt 2013.

Lang, Kevin, "The Effect of the Payroll Tax on Earnings: A Test of Competing Models of Wage Determination," NBER Working Papers 9537, National Bureau of Economic Research, Inc 2003.

Laroque, Guy and Bernard Salanié, "Labour Market Institutions and Employment in France," *Journal of Applied Econometrics*, 2002, *17*, 25–48.

Liang, Xiaoli, Jeffrey Kubik, and Gary Engelhardt, "The Incidence of the Social Security Payroll Tax: Evidence from the 1977 Amendments to the Social Security Act," Discussion Paper 2004.

Liebman, Jeffrey and Emmanuel Saez, "Earnings Responses to Increases in Payroll Taxes," Discussion Paper 2006.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698 – 714.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press, 1974, pp. 105–142.

Meghir, Costas and David Phillips, "Labour Supply and Taxes," in "Dimensions of Tax Design: the Mirrlees Review," Oxford University Press, 2010.

Melguizo, Angel and Jose Gonzalez-Paramo, "Who bears labour taxes and social contributions? A meta-analysis approach," *SERIEs*, 2013, *4* (3), 247–271.

Meyer, Bruce D., "The Effects of the Earned Income Tax Credit and Recent Reforms," in "Tax Policy and the Economy, Volume 24" NBER Chapters, National Bureau of Economic Research, Inc, 2010, pp. 153–180.

Meyer, Robert H. and David A. Wise, "Discontinuous Distributions and Missing Persons: The Minimum Wage and Unemployed Youth," *Econometrica*, 1983, *51* (6), pp. 1677–1698.

Moffitt, Robert, "The Tobit Model, Hours of Work and Institutional Constraints," *The Review of Economics and Statistics*, 1982, *64* (3), 510–515.

Müller, Kai-Uwe and Katharina Wrohlich, "Two Steps Forward – One Step Back? Evaluating Contradicting Child Care Policies in Germany," *CESifo Economic Studies*, 2015, *forthcoming.*

_ **and Michael Neumann**, "The Economic Incidence of Social Security Contributions - A Discontinuity Approach with linked Employer-Employee Data," DIW Discussion Papers 1578, DIW Berlin 2016.

Neubig, Thomas, "The Social Security Payroll Tax Effect on Wage Growth," *Proceedings of the National Tax Association*, 1981, pp. 196–201.

Neumann, Michael, "Earnings Responses to Social Security Contributions: Evidence from two German Administrative Data Sets," DIW Discussion Paper No. 1489, German Institute for Economic Research (DIW Berlin) 2015.

OECD, "Employer Versus Employee Taxation: The Impact On Employment," Employment Output, OECD 1990.

_ , "The OECD Jobs Study Evidence and Explanations - Part II: The Adjustment Potential of the Labour Market," 1994.

_ , "OECD revenue statistics," 2015.
https://stats.oecd.org/Index.aspx?DataSetCode=REV; accessed 2015.

_ , "OECD revenue statistics," 2016.
https://stats.oecd.org/Index.aspx?DataSetCode=REV; accessed 2015.

Ooghe, Erwin, Erik Schokkaert, and Jef Flechet, "The Incidence of Social Security Contributions: An Empirical Analysis," *Empirica*, 2003, *30* (2), 81–106.

Pissarides, Christopher A. and Barbara Petrongolo, "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 2001, *39* (2), 390–431.

Prescott, Edward C., "Why do Americans work so much more than Europeans?," *Quarterly Review*, 2004, (Jul), 2–13.

Rogerson, Richard, Shimer, and Robert Randall Wright, "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 2005, *43* (4), 959–988.

Rothstein, Jesse, "Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence," *American Economic Journal: Economic Policy*, 2010, *2* (1), 177–208.

Saez, Emmanuel, "The effect of marginal tax rates on income: a panel study of 'bracket creep'," *Journal of Public Economics*, 2003, *87*, 1231 – 1258.

_ , "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010, *2* (3), 180–212.

_ , Manos Matsaganis, and Panos Tsakloglou, "Earnings Determination and Taxes: Evidence From a Cohort-Based Payroll Tax Reform in Greece," *The Quarterly Journal of Economics*, 2012, *127* (1), 493–533.

Schröder, Mathis, Rainer Siegers, and C. Katharina Spiess, ""Familien in Deutschland" – FiD. Enhancing Research on Families in Germany," Technical Report, SOEPpapers on Multidisciplinary Panel Data Research No. 556, German Institute for Economic Research (Deutsches Institut für Wirtschaftsforschung, DIW), Berlin 2009.

Shephard, Andrew, "Equilibrium Search and Tax Credit Reform," *mimeo*, 2012.

Skedinger, Per, "Effects of Payroll Tax Cuts for Young Workers," *Nordic Economic Policy Review*, 2014, *1*, 125–169.

Soest, Arthur Van, Isolde Woittiez, and Arie Kapteyn, "Labor Supply, Income Taxes, and Hours Restrictions in the Netherlands," *The Journal of Human Resources*, 1990, *25* (3), 517–558.

Steiner, Viktor and Katharina Wrohlich, "Introducing Family Tax Splitting in Germany: How Would It Affect the Income Distribution, Work Incentives, and Household Welfare?," *FinanzArchiv: Public Finance Analysis*, 2008, *64* (1), 115–142.

_ , _ , Peter Haan, and Johannes Geyer, "Documentation of the Tax-Benefit Microsimulation Model STSM. Version 2012," Technical Report, DIW Data Documentation No. 63, German Institute for Economic Research (DIW Berlin), Berlin 2012.

Stewart, Mark B and Joanna K Swaffield, "Constraints on the Desired Hours of Work of British Men," *Economic Journal*, 1997, *107* (441), 520–35.

Stewart, Mark B. and Joanna K. Swaffield, "Constraints on the Desired Hours of Work of British Men," *The Economic Journal*, 1997, *107* (441), pp. 520–535.

Tazhitdinova, Alisa, "Adjust Me if I Can't: The Effect of Firm Incentives on Labor Supply Responses to Taxes," Discussion Paper, University of California, Berkeley 2015.

Train, Kenneth E., *Discrete Choice Methods with Simulation* number 9780521766555. In 'Cambridge Books.', Cambridge University Press, 2009.

Tummers, Martijn P. and Isolde Woittiez, "A Simultaneous Wage and Labor Supply Model with Hours Restrictions," *The Journal of Human Resources*, 1991, *26* (3), 393–423.

van Soest, Arthur, "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 1995, *30* (1), 63–88.

_ **and Marcel Das**, "Family Labor Supply and Proposed Tax Reforms in the Netherlands," *De Economist*, 2001, *149* (2), 191–218.

_ **,** _ **, and Xiaodong Gong**, "A structural labour supply model with flexible preferences," *Journal of Econometrics*, 2002, *107* (1-2), 345–374.

vom Berge, Philipp, Anja Burghardt, and Simon Trenkle, "Stichprobe der Integrierten Arbeitsmarktbiografien. Regionalfile 1975-2010 (SIAB-R 7510)," Documentation on Labour Market Data, Institute for Employment Research (IAB), Nuremberg 2013.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp, "The German Socio-Economic Panel Study (SOEP) Scope, Evolution and Enhancements," *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift fr Wirtschafts- und Sozialwissenschaften*, 2007, *127* (1), 139–169.

Weber, Matthias and Arthur Schram, "The Non-Equivalence of Labor Market Taxes: A Real-Effort Experiment," Tinbergen Institute Discussion Papers 13-030/I, Tinbergen Institute 2013.

Wichert, Laura and Ralf A. Wilke, "Which factors safeguard employment?: an analysis with misclassified German register data," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2012, *175* (1), 135–151.

Declaration

Erklärung gem. 4 Abs. 2 der Promotionsordnung

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

Berlin, September 2016

Michael Neumann

Erklärung gem. 10 Abs. 3 der Promotionsordnung

Hiermit erkläre ich, dass ich für die Dissertation folgende Hilfsmittel und Hilfen verwendet habe: Stata, Matlab und MS Excel.

Berlin, September 2016

Michael Neumann