Hamburg Institute of International Economics

# Impact of Welfare Sanctions on Employment and Benefit Receipt – Considering Top-Up Benefits and Indirect Sanctions

Ingrid Hohenleitner, Katja Hillmann

HWWI Research
Paper 189

Hamburg Institute of International Economics (HWWI) | 2019
ISSN 1861-504X

Corresponding author:
Ingrid Hohenleitner
University of Hamburg | Department of Economics
i.hohenleitner@gmx.de

HWWI Research Paper
Hamburg Institute of International Economics (HWWI)
Oberhafenstr. 1 | 20097 Hamburg, Germany
Telephone: +49 (0)40 340576-0 | Fax: +49 (0)40 340576-150
info@hwwi.org | www.hwwi.org
ISSN 1861-504X

# Impact of Welfare Sanctions on Employment and Benefit Receipt — Considering Top-Up Benefits and Indirect Sanctions *

Ingrid Hohenleitner[†] Katja Hillmann[‡]

January 8, 2019

**Abstract**

This comprehensive study on UB-II-sanctions in Germany, applying PSM, presents the ex-post effects of welfare sanctions on several employment states for diverse (sub-)groups of employable welfare recipients. Besides unemployed, we also regard employed, and indirectly affected household members. The monthly updated ATT show the development of the sanction effect over two years.

We find sanction effects as highly volatile over time and strongly dependent on individual factors and on circumstances like the timing of the sanction. In total, we suppose tendentially positive effects on the probabilities to enter employment and to exit welfare, at least in the short run. The positive effects tend to work stronger in the short run, and the negative effects tend to work stronger in the medium and long run. Hence, the shorter the time horizons of studies on welfare sanctions are, the more the positive effects are overrated systematically. Especially the frequently occurring cases with strongly negative slopes of cumulated ATT indicate that the early positive effects, mainly driven by people with good labor market perspectives, are at the cost of people with strongly detrimental sanction effects, even in the long run.

JEL classification: I38, J48, J64, J65, J68

Keywords: benefit sanctions, sanction effects, unemployment duration, welfare duration, long-term effects, unemployment benefits, unemployment policy, welfare policy, stratification

*We are grateful to Michael Funke for his valuable comments and suggestions to an early draft of this paper and to Thomas Straubhaar for his valuable comments and suggestions to the latest versions of this paper.

[†]University of Hamburg, Department of Economics, i.hohenleitner@gmx.de (corresponding author)

[‡]German Aerospace Center, Institute of Transport Research, katja.hillmann@dlr.de

# 1 Introduction

Most of the European countries have restructured their social security system towards shorter periods of eligibility in the unemployment insurance system. As a result, more unemployed people and their families have to rely on means-tested tax-based welfare payments for low-income earners and needy job-seekers. Despite the fact that an increasing number of people in Europe are either directly affected by the new structured welfare system at some point in their lives, or are at least indirectly affected by the side-effects on the labor market, scientific literature about the effects of the monitoring and sanction systems in the welfare policy of European countries is still scarce and rather selective. With this comprehensive study on the ex-post effects of sanctions against UB-II-recipients in Germany we want to contribute an important step towards filling this gap.

In Germany, the restructuring of the unemployment benefit and welfare system reached its peak in January 2005 with the implementation of the 'unemployment benefits II' (UB II, colloquially known as the 'Hartz IV' laws) which established the means-tested welfare payments for needy employable people and their related household members, and which brought a huge number of people from unemployment insurance receipt to welfare receipt — during the implementation of 'Hartz IV' and thereafter. And as the means tested welfare payments are defined to merely cover the minimum existence level, sanctions in the form of punitive benefit cuts, mostly lasting several months, have been criticized more and more often, or at least critically questioned for various reasons and regarding different aspects even, to some extent, in the economic literature and scientific policy advice.[1]

Yet despite the increasing public awareness of the potentially adverse effects of welfare sanctions, the majority of European studies on unemployment benefit sanctions still focus on unemployment insurance (UI) recipients who are usually closer to the labor market, and thus are more likely to take up unsubsidized employment, than employable welfare recipients. The latter are often either long term unemployed, low-income workers or have other restrictions which prevent them from taking up the kind of employment that would bring them out of benefit receipt, for example, caring for children, or for elderly and sick family members. Others are job starters who have just finished school or university; still others may want to re-enter the labor market after a longer period of exclusive

[1] See for example Ames (2009), Kumpmann (2009), Götz et al. (2010), Ehrentraut et al. (2014), Wolff (2014), van den Berg et al. (2015), and van den Berg et al. (2017).

family work or a long-lasting disease. So, even if a non-negligible part of them are well-educated, such as recent university graduates, the group of welfare recipients is much more heterogeneous than UI recipients are; and the majority of welfare recipients face stronger obstacles to attaining employment well-paid enough to cover their household's minimum subsistence level than recipients of unemployment insurance benefits (UIB) do, who are predominantly unemployed for less than one year, and usually have worked in regular employment for a longer period beforehand. Therefore it is more than questionable, whether the findings on UIB sanctions, predominantly revealing positive effects on entering employment, are transferable to people receiving welfare benefits.

Meanwhile, a couple of studies on welfare sanctions in Europe have been conducted, often restricted to either small regional entities or specific subgroups, like the Dutch studies on the municipality of Rotterdam (van den Berg et al. (2004), van der Klaauw and van Ours (2013)), and the studies on young male welfare recipients in Western Germany (van den Berg et al. (2014, 2015)). Similar to the literature on UI sanctions, most of the studies on welfare sanctions are focused merely on the transition from unemployment to employment (Boockmann et al. (2014), van den Berg et al. (2014)), although others also consider the option to leave the labor market, namely the non-employment option (Busk (2014) for Finland, Hillmann and Hohenleitner (2015) for Germany, and van den Berg et al. (2015) for young men in Western Germany).

More extensive studies on welfare sanctions, each for whole countries, are provided by Schneider (2008, 2010) for Germany and Busk (2014) for Finland. As described, in more detail, in the following Section 2, Busk (2014) compares the effects of UI benefit and welfare sanctions in Finland, indeed finding differences in the effects; and Schneider (2008, 2010) provides a very early study on German welfare recipients using survey data, conducted shortly after the implementation of UB II. Although quite comprehensive, distinguishing exits to unsubsidized and subsidized work, and additionally regarding reservation wages and job search effort, Schneider (2008, 2010) finds only the effects on unsubsidized work to be partially positive significant. A reason for the mostly insignificant effect estimations of the remaining outcomes may be the fact that the data survey was conducted in the very early stages of implementing the new welfare system under 'Hartz IV', which took close to a full year to be nearly working as planned. Thus, the survey data was conducted during a time when the monitoring and sanction regime was still under construction and not fully effective.

Therefore, a more recent comprehensive analysis of the impact of welfare sanctions in Germany is still needed. With the present work, we intend to take a crucial step ahead in filling this gap. We use a data set specially designed for this research project by the Research Data Centre (FDZ) of the Institute for Employment Research (IAB), based on a 2% sample of administrative data of the German Federal Employment Agency (FEA), and comprising the years 2004 until 2010. In contrast to most other European studies on benefit and welfare sanctions who, by and large, apply the timing-of-events (ToE) approach, and similar to Schneider (2008, 2010) for welfare sanctions and Hofmann (2012) for UI benefit sanctions, we conduct propensity score matching (PSM) for our analyses. This non-parametric approach, needing no assumptions for a functional form like is necessary for ToE models, seems to be even more appropriate for the purpose of analyzing an extremely heterogeneous population group such as welfare recipients.

Our study is the only that conducts all analyses for the whole population of employable welfare recipients, as well as for various subgroups using distinct categories, e.g. for age, gender, region, and education, the latter indirectly comprised in a special variable for the individual 'labor market access' (LMA). Unique features, compared to other European studies on welfare sanctions, are that our analyses not only comprise unemployed but also employed welfare recipients, in Germany colloquially called *'Aufstocker'*, and additionally, employable people indirectly affected by sanctions against their family members, a scenario that we briefly refer to as 'indirect sanctions'.

Depending on the initial sample, we differentiate between multiple exit events: exit to mere employment versus employment with supplementary welfare receipt, which is defined similarly to unsubsidized versus subsidized employment like Schneider (2008, 2010) differentiates it, and the exit event of leaving welfare receipt, all for initially *unemployed* welfare recipients. The non-employment option of leaving the labor force is indirectly considered by interpreting the divergence between entering mere employment and exiting welfare. For *employed* people receiving top-up benefits — a group of welfare recipients almost neglected by scientific literature — we additionally consider the option of quitting employment for mere welfare receipt. And finally, our analyses of employable people who are *indirectly* affected by sanctions upon their family members also distinguish between initially unemployed and employed welfare recipients with the same corresponding exit events we analyze for the directly sanctioned.

In order to clearly draw the line between what we provide with our study and what are the limitations, the following aspects should be mentioned: sanctions in the form of temporary benefit cuts do not only affect the sanctioned individuals after the sanctions (ex-post effects), they also affect non-treated people receiving welfare or UI benefits, as they are threatened by potential sanctions, which may affect their behaviour already before the imposition of a sanction. These effects, still at the individual level, are referred to as ex-ante effects. Furthermore, the perception of the sanction regime of the welfare and benefit system by society as a whole may even cause ex-ante effects at the general level, like possible effects on the market wages of the labor market. The analysis of ex-ante effects on the market as well as on the individual level exceeds our research subject. We limit our research to the individual ex-post effects of welfare sanctions from an economic perspective, with a focus on labor market outcomes. We also refrain from considering other individual ex-post effects like effects on an individual's income, economic wealth, personal well-being, health state, and many other individual aspects possibly affected by welfare sanctions.

Moreover, we restrict our analyses to the first sanction an individual experiences and do not consider repeated sanctions. This is the common approach, applied by the overwhelming majority of studies on benefit and welfare sanctions; an exception is the study on sanctions against young welfare recipients in Germany by van den Berg et al. (2015) who explicitly disentangle ex-post effects of first and second sanctions. Also as is common in previous studies, we apply binary treatment variables and do not distinguish between different durations or extents of benefit cuts; an exception here is an earlier study on young welfare recipients in Germany by van den Berg et al. (2014) who distinguish between two categories of sanctions: mild and strong.

The remainder of the paper is organized as follows: Section 2 gives a brief overview of empirical literature on benefit sanctions in European UIB and welfare systems. Section 3 introduces the data sample and describes the treatment, outcome, and control variables. In Section 4 we explain the methodological approach. A detailed presentation of our numerical and graphical results is provided in Section 5. We critically discuss and assess these results against the background of previous studies considering methodological aspects in Section 6. And finally, we conclude our results in Section 7.

## 2 Literature

We provide a detailed overview of the well-known European empirical literature on sanctions against recipients of unemployment insurance benefits (UIB), *UI sanctions* for short, in our previous paper on benefit sanctions, Hillmann and Hohenleitner (2015). Thus, here, we just mention and briefly summarize these studies. The large majority of empirical studies merely analyze the ex-post effects of sanctions; only a hand full of studies also consider ex-ante effects of sanctions, like Lalive et al. (2005) and Arni et al. (2013), who both use data from several Swiss cantons which enable the authors to distinguish between the effects of warnings and imposed sanctions in order to disentangle ex-ante and ex-post effects of benefit sanctions. Hofmann (2012) and Arni et al. (2013) also give brief overviews of the European literature on benefit sanctions, where Hofmann (2012) explicitly mentions the quasi-experimental and laboratory experimental studies which also consider ex-ante effects. Other European studies on benefit sanctions include Abbring et al. (2005) and Svarer (2010) which use data sets from the Netherlands, and van den Berg and Vikström (2014) which uses Swedish data, all analyzing the effects of UI sanctions on the transition rate from unemployment to employment. The vast majority of these studies apply the timing-of-events (ToE) approach, mostly using mixed proportional hazard (MPH) models, except from Hofmann (2012) who applies propensity score matching (PSM).

All these European studies on UI sanctions, referred to as benefit sanctions, find more or less positive effects of sanctions on taking up employment. Those who disentangle ex-ante and ex-post effects by analyzing warnings and imposed sanctions separately, such as Lalive et al. (2005) and Arni et al. (2013), find, in addition to positive effects of sanctions, positive effects of warnings upon the transition from unemployment to employment. If taking up regular employment or other employment, like subsidized work, is distinguished, as e.g. by Hofmann (2012), the positive effects on entering regular employment are stronger. And generally, earlier sanctions seem to be more effective than sanctions imposed later in the unemployment period (see for instance van den Berg et al. (2004) and Hofmann (2012)). Moreover, Hofmann (2012) analyzes diverse subgroups of UIB recipients and reveals that the positive sanction effects on taking up regular employment are mainly driven by younger unemployed. Nevertheless, as the findings of sanction effects on recipients of UIB are not necessarily transferable to welfare recipients, we mainly focus on literature about sanctions against welfare recipients.

Empirical literature about the effects of *welfare sanctions* in Europe is still scarce, although since the implementation of unemployment benefits II in 2005, the effects of welfare sanctions against UB-II-recipients in Germany have come more and more into the focus of policy advice and science. In our previous study on welfare sanctions, Hillmann and Hohenleitner (2015), we provide an overview of German and other European studies on welfare sanctions until mid-2015, which we thus mention here only briefly. Another overview, including qualitative surveys on German welfare sanctions, is provided by van den Berg et al. (2014, 2015), both focusing on young welfare recipients in Germany.

Two studies on welfare sanctions in the Netherlands use data about welfare recipients in the municipality of Rotterdam, and both find positive treatment effects on the transition from unemployment to work. The early Dutch study by van den Berg et al. (2004) additionally finds that sanctions at an early stage reduce the probability of long-term unemployment. The more recent Dutch study by van der Klaauw and van Ours (2013) additionally reveals re-employment bonuses to be an ineffective policy instrument. Both studies apply the ToE approach using mixed proportional hazard (MPH) models.

A more recent study from Finland by Busk (2014) analyzes the effects of unemployment insurance benefit (UIB) and welfare sanctions on employment, on participating in a program of the Active Labor Market Policy (ALMP), and on exiting from labor force. Busk (2014) estimates positive treatment effects of ongoing, i.e. currently executed, sanctions upon UIB and welfare recipients, and of completed sanctions upon welfare recipients on their probability to take up employment. The sanction effects on participating in a measure of the ALMP are slightly positive for welfare recipients, and not significant for UIP recipients. Finally, she finds the exit from labor force positively affected by both UIB and welfare sanctions.

The first study about welfare sanctions in Germany after the implementation of unemployment benefits II is provided by Schneider (2008, 2010) who uses cross-sectional survey data, conducted shortly after the implementation of UB II. Applying propensity score matching (PSM), she analyzes the effects of sanctions against unemployed UB-II-recipients on employment, reservation wages, and job search effort. She finds merely unsubsidized[2] employment as partially positive affected, while the

[2]Schneider (2008, 2010) defines ‘unsubsidized employment’ as jobs with an income that is high enough to leave UB II receipt; it may include also part-time employment. In contrast, ‘subsidized employment’ may include regular jobs with supplementary UB II receipt.

other outcomes turn out to be not significantly affected. She also finds earlier sanctions to be more effective in terms of unsubsidized employment than sanctions later in the unemployment spell.

More current German studies on welfare sanctions are provided by Boockmann et al. (2014), Hillmann and Hohenleitner (2015), and van den Berg et al. (2014, 2015). Boockmann et al. (2014) apply an instrumental variable (IV) regression in order to estimate the effect of welfare sanctions on the transition from unemployed UB-II-receipt to unsubsidized employment. Using a unique combined data set of German administrative and survey data, they find positive effects of welfare sanctions on taking up employment without supplementary welfare receipt. Another German study, provided by Hillmann and Hohenleitner (2015), uses a rich panel survey employing the timing-of-events (ToE) approach with mixed proportional hazard (MPH) models; it reveals positive effects of German welfare sanctions on employment entry as well as on leaving the labor force.

The two German studies by van den Berg et al. (2014, 2015) focus on the special situation of young UB-II-recipients who are sanctioned more severely and more frequently. Both investigations apply the ToE approach and use administrative data, analyzing the effects of sanctions against male unemployed, aged under 25 years, and living in Western Germany. The restriction to this sub-group of young welfare recipients is chosen in order to get a preferably homogeneous group for the analysis. While the study of van den Berg et al. (2014) is the first which distinguishes between mild and strong sanctions,[3] van den Berg et al. (2015) are the first who do not only analyze the effects of first sanctions, but also of second sanctions, considering only strong sanctions. Van den Berg et al. (2014) find a positive impact of mild and strong sanctions against young welfare recipients on their hazard rate to unsubsidized work,[4] whereby the effect is larger for strong sanctions. They further reveal that part of the sanction effect is caused by the expectation of intensified monitoring. In contrast to previous findings, mostly on UI sanctions, which identify earlier sanctions as more effective, the authors of this study do not find sanction effects dependent on the moment of imposition during the

[3]Mild sanctions are imposed for missing an appointment and amount to a 10% benefit cut for three months in the first instance; strong sanctions are imposed for all other breaches of duty and, for welfare recipients younger than 25 years, result in a 100% cut of the base benefit from the very first failure. More detailed information of the sanction regime applied to German UB-II-recipients are given, for example by Hillmann and Hohenleitner (2015) and van den Berg et al. (2014, 2015).

[4]Unlike Schneider (2008, 2010) who defines unsubsidized employment as jobs which pay enough to leave (supplementary) welfare receipt, in the study of van den Berg et al. (2014) ‘unsubsidized employment’ does not exclude receiving top-up benefits.

welfare spell.

The later study of the authors, van den Berg et al. (2015), analyzing first and second sanctions against young male welfare recipients in Western Germany, additionally considers the non-employment option, namely the possibility of leaving the labor market. Furthermore separate models are estimated for people living alone and people living in multi-person households, as the latter ones may rely on other household member's income and thus might react less sensitive on sanctions. The authors find the employment effect of first sanctions most effective for single persons but still strongly effective for young men in multi-person households. Also second sanctions raise the exit rates into employment for young men in single households. For those living in multi-person households the second sanction was not significantly affecting the employment entry. Concerning the other exit-option, out of the labor force, van den Berg et al. (2015) find strong effects of the first and second sanction against young male unemployed living alone, but no significant effects on those living in multi-person households. Moreover, van den Berg et al. (2015) use the initial daily wage as an indicator of job match quality in order to identify possible adverse post-unemployment effects. And indeed, they find the positive employment effects accompanied by reduced wages. This implies that sanctions in the form of benefit cuts reduce the reservation wages of the treated.

# 3 Data

The data set we use for our analyses is based on an extract of the "Sample of Integrated Labour Market Biographies" (SIAB) supplemented by selected information from administrative data of the German Federal Employment Agency (FEA). The SIAB is a 2% random sample drawn from the "Integrated Employment Biographies" (IEB) of the IAB; the IEB comprises all individuals in Germany who are either employed or benefit recipients according to the German Social Code III or II (SC III since 1975, SC II since 2005) and who are officially registered as job-seekers with the German FEA or participants in programs of active labour market policies (ALMP) (in the data since 2000) at least once during the observation period. These data, which come from different sources, are merged in the IEB, where the labor market status is given on a daily base.[5]

[5]See vom Berge et al. (2013).

Our data set is assembled and prepared by the Research Data Centre (FDZ) of the Institute for Employment Research (IAB) at the German FEA especially for this research project. It is based on selected variables of the SIAB over the complete years from 2004 to 2010, supplemented by selected information about sanctions and household members obtained from process-produced data of the FEA's administrative sources. This combined data set, exclusively prepared and provided to us for this research project, comprises 978,459 observations in the form of 'spell data' (episodes of several employment status) for 223,725 individuals each having received Unemployment Benefits II (UB II) at least once in the observation period.

### 3.1 Samples and subsamples

For our analyses, we use two annual inflow cohorts — i.e. cohorts who come into UB II receipt in 2007 and 2008, each restricted to *employable* people in the age of 15 to 56 years. As we have a huge amount of results to present and as the findings of both years do not substantially differ, we present the results of the more current inflow cohort of 2008 in Section 5 and use the findings of the inflow sample of 2007 just as a kind of robustness check. The inflow samples are divided into two kinds of inflow status: *employed* people receiving supplementary UB II and *unemployed* UB-II-recipients. In doing so, the employment status on the day of entering welfare receipt is clearly defined. Moreover, we conduct all analyses separately for men and women.

In addition to the two main groups of unemployed and employed welfare recipients, we analyze several subgroups. Firstly, we differentiate between the following age groups: people aged between 15 and 25 years, because people younger than 25 (*under 25: u25*) face stricter monitoring and sanction conditions, and people aged 25 to 56 (*over 25: o25*). Secondly, we distinguish different places of residence: people from a federal state belonging to the former *Western Germany* (*WG*) or to the former *Eastern Germany* (*EG*). And finally, we analyze people with different levels of labor market access: *low*, *middle*, and *high*. This variable is based on the classification of the German FEA and the Jobcenters dividing their clients into so-called "market clients" (*Marktkunden*) with a *high* level of labor market access, "counseling clients" or "advisory clients" (*Beratungskunden*) with a *medium* level of access to the labor market, and "guided clients" (*Betreuungskunden*) with a *low* level of market access who are supposed to need support or even guidance in order to be able to take up employment.

As this variable of the clients' classification is often missing in the original data set, we estimate the missing values using further characteristics correlating with labor market access like (school and occupational) education and previous periods of (un-)employment.

### 3.2 Treatment variables

We carry out the whole analysis with two kinds of treatment variables: direct and indirect sanctions which indicate punitive cuts of unemployment benefits II that are imposed either directly against the UB-II-recipient or indirectly against a related household member. Such benefit reductions start at 10-%-cuts for minor failures (being late or missing an appointment) and 30-%-cuts for major failures (all other state of affairs causing a sanction), each calculated as percentage points of the base benefit, increasing for repeated failures of the same kind until 100-% of the UB II — including costs for accommodation and health insurance — is cut, with such cuts typically lasting for three months. In order to exclude statistical outliers with very short benefit cuts, and following Hofmann (2012), we ignore benefit cuts that last for seven days or less. As we neither have information about the amount of benefit cut nor about the reason for the sanction, we cannot distinguish between *minor* and *major* "breaches of duty" which cause different percentage points of benefit reductions.[6] Following the common practice of the vast majority of studies on the effects of benefit sanctions, we only consider the first sanction but not repeated sanctions.

We regard people as directly sanctioned from the beginning of the first punitive benefit cut that is imposed on them directly. We consider an individual to be indirectly sanctioned if they are not punished, themselves, but are indirectly affected by sanctions imposed upon one of their related household members. We regard them as indirectly sanctioned from the beginning of the first sanction against a household member on, and as long as the individual does not face a direct sanction. The moment that an indirectly sanctioned individual also faces a direct sanction, they are removed from the sample; he or she can neither be used in either of the two treatment groups nor, of course, in the control group. The reason for this is that we only use the first sanctions — be it either a direct or an indirect sanction — and thus we need people without previous sanctions for the treatment groups.

[6] Further details about different kinds of "breaches of duty" and the amount of benefit reduction they cause we provide in our previous paper about the effects of benefit sanctions against German UB-II-recipients; see Hillmann and Hohenleitner (2015).

Nevertheless, we do not totally disentangle direct and indirect sanctions, as in the group of direct sanctioned there are concluded also later direct and indirect sanctions. This goes along with previous studies on benefit sanctions who also only consider the first sanction and define a person as sanctioned from the beginning of the first sanction on. But in contrast to the direct sanctioned, the group of indirect sanctioned contains only persons who are *exclusively* indirectly sanctioned.

As controls, self-evidently we can only use non-sanctioned people. Each individual that has been sanctioned as a UB-II-recipient since the implementation of unemployment benefits II in January 2005 is defined as (pre-)sanctioned until the end of the observation period, that is the end of December 2010, and hence cannot be used in the control group at all.

With regards to the stratification, as explained below in Section 4.4, the assignment to either a treatment or a control group works as follows: if a person is sanctioned within the stratum but is non-sanctioned before, the person enters the treatment group of this stratum. Hence, treated people are only placed into one treatment group, namely the treatment group of the current stratum when they face their first sanction; but they can neither be in the treatment group of a stratum before nor in the following strata. A disadvantage of the stratification is that the exact starting time of the sanction within the stratum is no longer considered. This loss of information resulting from stratification may cause a bias that we carefully discuss in Section 4.4.3.

### 3.3 Outcome variables

We use two different kinds of outcome variables for our analyses: continuous metric variables measuring the durations until the exit events ("duration outcomes"), and binary variables indicating the current (monthly) employment state ("probability outcomes"). Because of the necessity of stratification, which we explain in Section 4 and Subsection 4.4, we measure all outcome variables from the start of the stratum on, and not from the beginning of the treatment, namely from the imposition of the sanction on. This holds for duration outcomes as well as for probability outcomes. As mentioned above, the loss of information about the exact starting time of the treatment due to stratification can cause a bias that we discuss in detail in Section 4.4.3.

The duration outcomes are continuous variables measuring the duration from the start of the stratum until an exit event, i.e. employment entry, welfare exit, or exit into mere welfare receipt, the latter for the subgroup of employed people with supplementary welfare receipt. The binary outcomes, indicating the current employment state, concretely indicates whether the individual's initial employment state remains unchanged (value 0), or whether the employment state has changed (value 1) towards either employment entry, welfare exit, or employment exit, meaning welfare receipt alone as a possible outcome for people formerly employed with supplementary benefit receipt. These kinds of outcome variables show the shares of people with (value 1) or without (value 0) an exit event within the monthly prolonged observation periods lasting from the start of the stratum until the end of the consecutive final months. As the share of people with exit events in relation to the whole (sub-)sample reveals the probability of the exit event, we refer to these kind of outcome variables as "probability outcomes".

Furthermore, it must be stressed that we face two different samples for the two kinds of outcome variables: the samples for the metric duration outcomes excludes right-censored spells because durations can only be determined for people who experience an exit event until the end of the observation period, actually until the end of December 2010; in contrast, the samples for the monthly binary outcomes include right-censored spells because even if there is no exit event within the observation period, the outcome status can be defined, concretely as value zero indicating no change in labor market status. Hence, the analyses of probability outcomes comprise the full (sub-)samples, while the analyses of the duration outcomes are based on samples reduced by the right-censored spells.

### 3.4 Control variables

Applying Propensity Score Matching (PSM), as described in Section 4, it is essential to include as many covariates as available that are supposed to affect both the assignment to the treatment and the dependent variables.[7] This holds independently of whether or not the control variables *significantly* influence the treatment and the outcome. For our analyses of the effects of welfare sanctions, we use the explanatory variables presented in the following tables, distinguishing between binary (Table 1) and metric (Table 2) control variables; specifically, we use them for the propensity score estimation of the selection into the treatment of either direct or indirect sanctions.

When implementing propensity score matching (PSM) with dichotomous treatments applying either

[7] See, for example, Ho et al. (2007).

logit or probit estimation models, only binary or metric variables shall serve as controls; nominally or ordinally scaled variables ought to be transferred into binary variables.[8] Table 1 contains the correspondingly built dummies along with the binary variables used as controls.

The binary variables *child-u3* and *couple* indicate whether or not an individual has a child younger than 3 years of age, or a partner living in the same household. The dummy for professional education, *pquali*, denotes whether or not the individual has successfully completed a vocational training, including polytechnic or university degrees. The variable for school education, *squali*, is divided into three dummies: graduation from main school (*low*) (*Hauptschule*), secondary school (*middle*) (*Realschule*), and high school (*high*) (*Gymnasium*), where the reference category is having no graduation (*none*). The dummy variables for *age groups*, *nationality*, and *bula*, representing 15 of the 16 federal states (*German Bundesländer*), should be self-explanatory. The dummies for the quarterly inflow cohorts, (*qinfl*), indicate whether an individual entered welfare receipt in the 1st, 2nd, 3rd, or 4th quarter of the year 2008.[9] And finally, the dummies for the quarterly duration of the employment states, *employed* (*emp*), *unemployed* (*ue*), and employed with *supplementary* (*supp*) welfare receipt, indicate whether people have experienced either no (0), between zero and three, inclusive (3), between three and six, inclusive (6), up to nine (9), or up to twelve (12) months of the specific employment status during the

**Table 1:** Explanatory variables — binary controls

| Denotation | Dummy variables | Reference category |
|---|---|---|
| Age groups | age: 15-17, 18-24, 35-44, 45-56 | age: 25-34 |
| Child under 3 years | child-u3 | |
| Partner in the household | couple | |
| Nationality | German, non-EU-foreigner | EU-foreigner |
| School education | squali: low, middle, high | squali: none |
| Professional education | pquali | |
| Federal state | bula: 15 of 16 federal states | bula: Bavaria |
| Quarterly inflow cohort | qinfl: 0208, 0308, 0408 | qinfl: 0108 |
| Duration of previous emp | employed: 0, 3, 6, 9 months | emp: 12 mon. |
| Duration of previous ue | unemployed: 0, 3, 6, 9 months | ue: 12 mon. |
| Duration of previous supp | supplementary: 0, 3, 6, 9 months | supp: 12 mon. |

[8] An advantage of such a transformation is that several tests in order to assess the matching quality can be conducted more easily; see, for example, Müller (2012).

[9] For the inflow cohort of 2007, used as a sensitivity check, we apply corresponding dummies.

year previous to the welfare receipt.

The first four of the metric control variables listed in Table 2 are taken from the labor market statistics of the German Federal Employment Agency (FEA); the statistics are provided to the public, and are available via the FEA's website. We merged the monthly data, each depicted separately for the 16 federal states, of the following four rates, reflecting the labor market situation of the federal state in the current month: The sanction rate (*sancrate*) of the FEA may be a bit misleading as it only denotes the share of the *currently* sanctioned people of the total of UB-II-recipients, and does not depict a person as sanctioned after the sanction period anymore. Thus, the current sanction rates of the FEA are much lower than in our data, as we also consider a person to be sanctioned after the end of the period of punitive benefit cut. The unemployment rate (*uerate*), the vacancy rate (*vacrate*), and the share of employable UB-II-recipients (*elbrate*) (*ELB: "erwerbsfähige Leistungsberechtigte"*) in relation to the whole workforce in Germany are also publicly provided by the FEA.

Information about the wages of previous jobs and incomes gained from employment during the year previous to the welfare receipt was obtained from our main dataset based on the SIAB, described above at the beginning of Section 3. We differentiate between the wages and the yearly income of the main employment on the one hand, and the average wage and the sum of yearly incomes over all jobs, on the other hand. Finally, we use the metric variable with the information about the duration of the three mentioned employment states *employed* (*emp*), *unemployed* (*ue*), and employed with *supplementary* (*supp*) welfare receipt that reveal the summarized duration of these states during the year previous to entering the current period of welfare receipt.[10]

**Table 2:** Explanatory variables — metric controls

| Denotation | Metric variables | Varying by |
|---|---|---|
| Sanction rate | sancrate | month, federal state |
| Unemployment rate | uerate | month, federal state |
| Vacancy rate | vacrate | month, federal state |
| Employable UB-II-recipients | elbrate | month, federal state |
| Previous wage | daily wage | main job, all jobs |
| Previous income | yearly income | main job, all jobs |
| Previous empl. states | duration in days of previous year | status: emp, ue, supp |

[10]The 'current period of welfare receipt' refers to those periods of the inflow cohort from 2008 that were actually

# 4 Methodological approach

We use a dataset based on administrative data, and hence the assignment to the treatment is not random like in case of experimental data. Thus, we have to account for the selectivity of the treatment process.[11] And as the probability of being treated is most likely influenced by unobserved characteristics that also affect the outcome, we have to solve the problem of endogenous treatment and confounding factors[12].

Furthermore, we examine different exit events which can be treated as so-called "competing risks" (CR) if they are mutually exclusive, as transition into mere employment ($O$) versus into employment with supplementary welfare receipt ($S$), or as transition into employment versus into non-employment would be. But we also examine exit events that are not mutually exclusive but overlapping, such as exit into mere employment ($O$) versus exit from welfare ($ExWel$), as the latter one comprises exits into non-employment as well as into mere employment.[13]

## 4.1 Choice of method

A common method to identify the effect of an endogenous treatment on the probability of a subsequent exit event is the *timing-of-events* (ToE) approach. Originally designed for endogenous treatments on a *single risk*, it is common to combine ToE models with another type of popular multivariate duration models, namely *mixed proportional hazard* (MPH) models for *competing risks* (CR).[14] As a crucial presupposition is often violated in practice, specifically, that the requirement that the competing risks have to be independent conditional on the covariates and the treatment, Drepper and Effraimidis (2016) developed an advanced combination of MPH competing risks and ToE models that allow for multiple CR which can be "dependent by way of unobserved characteristics", but where the competing

analyzed.

[11] A crucial advantage of using random samples of administrative data, however, is that they constitute a representative selection of individuals faced with "real world" conditions, and thus the external validity of analysis based on "real world" data is much higher than experimental data would be.

[12] We give a short definition of confounders and further information on how to deal with possibly unobserved confounding factors in Section 5.4.2 in the context of sensitivity analyses for checking the robustness of our estimations.

[13] For more details about the diverse exit events we examine, see the beginning of Section 5.

[14] Such timing-of-event (ToE) models, whether joint or not with MPH models for competing risks (CR), have been used for many empirical studies to evaluate the effect of active labor market programs or benefit sanctions on the probability of unemployed to enter employment, like van den Berg et al. (2004), Abbring et al. (2005), Lalive et al. (2005), and Rosholm and Svarer (2008).

risks still have to be mutually exclusive.[15]

Nevertheless, when applying the ToE approach, the specification of the model is still crucial; in particular, if the restrictions imposed on the heterogeneity distribution are not justified, a significant bias can result.[16] Additionally, MPH specifications impose restrictions on the functional form of the outcome equations that can severely distort parameter estimates.[17] Such kinds of restrictions on the functional form can be avoided using matching techniques such as *propensity score matching* (PSM).[18] And "model dependence", as it usually occurs "in parametric causal inference", can be reduced by applying "non-parametric matching procedures" like PSM, which offer "causal inference with fewer assumptions".[19]

Although a parametric regression model is used to estimate the propensity score (PS), *propensity score analysis* (PSA) with its two-step procedure is considered non-parametric,[20] because nonparametric density estimators, such as kernel functions, are used. And it is valued as a powerful matching technique which, properly applied, is able to balance the covariates in such a way that "the causal effect inference from observational data" becomes "as reliable as possible".[21] For these reasons, non-parametric matching methods like PSM are being applied to an increasing number of empirical studies in several disciplines, such as epidemiology and medicine, as well as social sciences and economics.[22]

Moreover, non-parametric approaches to solving the selection problem are extraordinarily beneficial in cases of possibly heterogeneous treatment effects, because they lead to consistent results even if the treatment affects diverse subgroups or individuals within the surveyed population in different ways.[23] This is a strong argument for us to use PSM, as the ex-post effects of benefit sanctions can be expected to be quite heterogeneous between different individuals and groups of welfare recipients. And indeed, the results of our analyses, presented in Section 5, confirm this supposition.

Despite the advantages of using non-parametric matching techniques, the vast majority of previ-

[15] See Drepper and Effraimidis (2016).
[16] See Heckman et al. (1999), Gaure et al. (2007), and Hofmann (2012).
[17] See inter alia Hofmann (2012), Drepper and Effraimidis (2016), and Zhang (2017).
[18] See inter alia Hofmann (2012) and Zhang (2017).
[19] See Ho et al. (2007).
[20] See inter alia Gangl and DiPrete (2004), Reinkowski (2006), Ho et al. (2007) Urkaregi et al. (2014), and Zhang (2017)
[21] See Zhang (2017).
[22] See inter alia Stürmer et al. (2006), Ho et al. (2007), Caliendo and Kopeinig (2008), Urkaregi et al. (2014), and Zhang (2017).
[23] See Reinkowski (2006).

ous studies on the effect of benefit sanctions apply other methods, predominantly ToE and MPH approaches. One reason for this might be the complexity of propensity score analysis and the huge effort implementing PSA hence entails.[24] From our point of view, however, the advantages exceed the drawbacks, even more so as we analyze overlapping exit events and not just mutually exclusive ones, which would be necessary in order to apply MPH with competing risks.

Furthermore, a large and rich dataset is necessary, or at least conductive, when performing a reliable and robust PSA. Our dataset generally fulfills these requirements for the main part of our analyses. More precisely, the data set is sufficiently huge and extensive to apply PSM for our main topics and for most of the population groups under study.[25] Similar to Hofmann (2012), who uses a dynamic matching approach for her study on the ex-post effects of unemployment insurance (UI) sanctions in West Germany which takes the timing of the treatment into account, we also implement a *dynamic* approach of propensity score matching (PSM), applying stratification to deal with the flexible timing of the treatment and the missing start date for the untreated.

## 4.2 The matching approach

The general questions we want to answer with matching procedures are whether a specific treatment is causal for the outcome of the observed entities, and how distinct and strong such a possible impact on the outcome would be. The fundamental problem we face is that we can observe only the factual but not the counterfactual state for the same statistical unit under otherwise completely identical— i.e. apart from the treatment — conditions.[26] Instead of measuring, we have to estimate the counterfactual state. One possible way to put this into practice is to use comparable groups of *treated* ($D_i = 1$) and *untreated* ($D_i = 0$) individuals, where $D_i$ is a dichotomous treatment indicator for individual $i$, with value 1 for the status "treated" and 0 otherwise, and $i = 1, ..., N$, with $N$ denoting the number of observed individuals in the entire surveyed population.[27]

The outcomes then are defined by $Y_i(D_i)$ for each individual, $i$, and the individual treatment effect

[24]See Müller (2012).

[25]Nevertheless, because of the necessary stratification, there is a loss of observations per analyzed stratum which, in a couple of cases with small subgroups, can lead to convergence problems caused by too few exit events in the treatment group.

[26]See inter alia Roy (1951), Rubin (1974), Heckman and Smith (1995), Rubin (2004), and Gangl and DiPrete (2004).

[27]This applies to binary treatments; see Caliendo and Kopeinig (2008), Heinrich et al. (2010), and Müller (2012) who provide helpful practical guidance on implementing PSM.

(TE) is defined as

$$(individual\ TE) \qquad \tau_i = Y_i(1) - Y_i(0). \tag{1}$$

But the individual treatment effect, $\tau_i$, cannot be measured, as we can observe only one of the two outcomes per individual $i$, because a person can either be treated or not at one specific point of time. Therefore *individual* treatment effects cannot be estimated, and hence, analyses in general have to focus on estimating *average* treatment effects of the investigated group.[28]

The evaluation of treatment effects can generally be based on various measures like the *"average treatment effect"* (ATE), which refers to the impact of the treatment on the *entire group* of population under study, and the *"average treatment effect on the treated"* (ATT), referring to the *treated share* of the examined group. As the ATE would possibly also include individuals who are not targeted by the treatment, in the practice of evaluating treatment effects of political measures and other targeted interventions, the ATT is the predominantly used parameter.[29] For our study of the ex-post effects of benefit sanctions we follow this reasoning and the common approach to use an ATT estimator to analyze treatment effects.

The *true value* of the average treatment effect on the treated (ATT) is given by

$$(true\ ATT) \qquad \tau_{ATT} = E[\tau|D=1] = E[Y(1)|D=1] - E[Y(0)|D=1], \tag{2}$$

where $E[Y(1)|D = 1]$ is the expected value of the outcome $Y(1)$ for the treated $(D = 1)$ in case of being treated, $E[Y(0)|D = 1]$ is the expected value of the outcome $Y(0)$ for the treated in the hypothetical case of not being treated — referred to as the counterfactual state — and $E[\tau|D = 1]$ is the difference between the two, which is the expected value of the treatment effect on the treated. But as the average outcome in the counterfactual case, $E[Y(0)|D = 1]$, is not observable, we have to build an artificial simulacrum of the counterfactual state. A common approach to implement such

[28]See Caliendo and Kopeinig (2008). A potential way to determine the individual treatment effect, $(\tau_i)$, could be to measure the outcome before and after the treatment, where the outcome before the treatment would serve as the counterfactual state. This holds only if all other factors beyond the treatment which affect the outcome are stable between the two measurement points, but this is often not fulfilled in practice; see Reinkowski (2006) and Müller (2012). Furthermore not every kind of outcome can be measured before and after the treatment like it is the case with duration outcomes and exit events during an unemployment episode which ends when the event takes place.

[29]See Heckman (1997) and Müller (2012).

an artificial counterfact is to use control groups of untreated individuals in order to compare their outcomes with the outcomes of the treated. The estimation of treatment effects is then conducted by measuring the average outcome differences between the groups — properly weighted if required, depending on the matching technique. *Naively*, the ATT could be estimated as

$$\widehat{\tau_{ATT}} = E[\tau|D=1] \approx E[Y(1)|D=1] - E[Y(0)|D=0]. \tag{3}$$

But as people differ in their properties, $E[Y(0)|D=0] \neq E[Y(0)|D=1]$ regularly holds, and thus the estimation of the counterfactual state with Equation (3) cannot be error-free. Specifically, if potential differences in the average properties of the treatment and control group also affect the outcome variable, the estimated treatment effects are distorted by so-called "confounding factors".[30] The estimation error caused by such confounders based on different characteristics between treatment and control group is referred to as *"selection bias"* (SB)[31] and can be formalized as

$$(\textit{selection bias}) \qquad\qquad SB = E[Y(0)|D=1] - E[Y(0)|D=0]. \tag{4}$$

Because of this potential selection bias of unknown quantity, Equation (3) must be supplemented by Equation (4) in order to get a *proper estimation parameter* of the ATT:

$$\widehat{\tau_{ATT}} = E[\tau|D=1] = E[Y(1)|D=1] - E[Y(0)|D=0] + SB. \tag{5}$$

It has to be stressed that differences in the average characteristics of the treatment and control groups are only problematic if they also affect the outcome. Otherwise they don't distort the estimation of the ATT.[32] The precondition which requires that the differences between the outcomes of the treatment and control groups must be independent of the selection process into the treatment, and thus are caused exclusively by the treatment itself, is called *"conditional independence assumption"*

[30]See Caliendo and Kopeinig (2008) and Müller (2012).
[31]See Heckman et al. (1998).
[32]See Müller (2012).

(CIA)[33] or *"unconfoundedness"*[34]. It can be formalized as

$$(CIA/unconfoundedness\ given\ X) \qquad\qquad Y(0), Y(1) \amalg D|X, \quad \forall \quad X, \tag{6}$$

where $X$ represents the vector of covariates and $\amalg$ stands for stochastical independence.[35] This strong assumption requires that all confounding factors must be eliminated or at least held constant, and hence they ought to be included in the estimation procedure as control variables.[36] If the CIA is not satisfied, the estimated ATT is distorted by unobserved confounders.[37] Therefore sensitivity analyses need to be conducted in order to check the robustness of the results against possibly unobserved confounding factors.[38] In Section 5.4.2, we give more detailed information about robustness checks and the kinds of sensitivity analyses we carried out, as well as about their results.

Another assumption that has to be satisfied in order to properly apply matching procedures is the so-called *"stable-unit-treatment-value-assumption"* (SUTVA). This precondition asserts that the treatment effect on one individual must not be influenced by the treatment of another entity, and thus the stability of the causal effect should be given.[39] In our analyses of benefit sanctions, we can presume that the SUTVA as satisfied. On the one hand, because of the huge number of people in the investigated group of UB-II-recipients in Germany, the sanction of people living in other households and families generally does not affect the labor market outcome of another sanctioned individual. And on the other hand, we disentangle the effect of sanctions against more than one member of a household by distinguishing between direct and indirect sanctions.[40]

If the preconditions are satisfied, and ideally a large and rich dataset is available with which to investigate the research question, matching techniques are a powerful and reliable method to solve the selection problem. As mentioned above, the treatment and control groups are not usually identical in their characteristics, and possible confounding factors must be considered and dealt with. The

[33]See Rosenbaum and Rubin (1983) and D'Orazio et al. (2006).
[34]See Lechner (1999).
[35]See Reinkowski (2006) and Caliendo and Kopeinig (2008), whereby Reinkowski (2006) uses the symbol $\perp$ for stochastical independence.
[36]See Müller (2012).
[37]See Reinkowski (2006) and Caliendo and Kopeinig (2008).
[38]See Caliendo and Kopeinig (2008), Heinrich et al. (2010), and Müller (2012).
[39]See Gangl and DiPrete (2004) and Müller (2012).
[40]For further details see Section 3.2.

matching approach then tries to solve the selection problem by constructing an artificial simulacrum of the counterfact of each treated individual using a properly generated sub-control-group of untreated.[41] In order to minimize the selection bias caused by the distorting effects of possible confounders, the matching procedure should lead to an artificially generated control group of untreated which is, on average, as similar as possible to the treatment group, for a properly defined set of covariates.[42] Even though there are several matching procedures available, because of its advantages we use the now common technique of *propensity score matching* (PSM) for our analyses.

### 4.3 Matching on propensity score

Matching on propensity score is an elegant and powerful technique for balancing the covariates of the treatment and control group. It simplifies the matching procedure by using a one-dimensional parameter instead of searching for so-called "statistical twins"[43].

With an increasing number of covariates, we get an exponentially growing number of possible matches. Hence, the more covariates are included into the vector $X$, the more severe the problem of emerging dimensionality will be if we try to find a statistical twin as a substitute for the counterfactual case.[44] In order to solve such kinds of dimensionality problems, Rosenbaum and Rubin (1983) suggest to apply a so-called *"balancing score"* (BS). Balancing scores are defined as functions $b(X)$ of the relevant covariates $X$ which holds

$$(\textit{balancing score}) \qquad X \amalg D | b(X), \quad \forall \quad X. \tag{7}$$

Formula 7 means that, given the balancing score $b(X)$, the *conditional distribution* of the set of control variables $X$ is independent of the treatment, and thus $X$ is, on average, identical for the treatment and the artificially built control group after the matching procedure.[45] It is crucial to distinguish between the initial group of the non-treated whose members are at available to serve as

[41]See Reinkowski (2006).

[42]See Müller (2012).

[43]"Statistical twins are cases that resemble their statistical siblings in selected variables", see Bacher (2002); strictly defined, statistical twins are meant to be identical for the whole set of selected variables.

[44]See Reinkowski (2006) and Caliendo and Kopeinig (2008).

[45]See Rosenbaum and Rubin (1983) and Reinkowski (2006); note that we use a notation similar to Caliendo and Kopeinig (2008), where $\amalg$ is used to symbolize stochastical independence; Reinkowski (2006) uses the symbol $\perp$ for statistical independence.

counterfactuals, on the one hand, and the artificially built simulacrum generated by applying one of diverse applicable matching techniques, on the other hand. Those matching procedures use and generally weight the non-treated people — concretely their values of covariates and outcomes — in order to build a control group that is a reflection of the treatment group which should preferably be as alike as possible, regarding a set of well-defined control variables $X$. It also has to be stressed that not the matched individuals themselves necessarily need to be as similar as possible, like the ideal-typical case of "statistical twins" would claim. In contrary, because of the weighting and potential re-using[46] of proper untreated matches, the similarity of treated and matched controls is not on the individual level but just on the group level. More precisely, both groups — treated and matched controls — are just *averagely* similar in the set of their control variables. And balancing scores, properly used for the matching procedure, do have the so-called "balancing property" which means that they lead to such balanced groups of treated and matched controls.[47]

According to Rosenbaum and Rubin (1983), balancing scores in large samples enable the assignment to treatment and control group without using the outcome variable.[48] In other words, balancing scores allow one to predict whether an individual is treated or not using only the information given by the vector $X$ of control variables.[49]

This also holds for the propensity score (PS), the most frequently used balancing score in empirical studies, which is defined as

$$(\textit{propensity score}) \qquad PS = P(X) = Pr(D = 1|X). \tag{8}$$

The *propensity score* (PS) is a one-dimensional matching parameter with balancing properties that can be interpreted as the predicted probability ($Pr$) of being treated ($D = 1$), given the vector of observed covariates ($X$).[50]

Rosenbaum and Rubin (1983) show that if the conditional independence assumption (CIA) —

[46]Matching procedures as a rule use sampling with replacement. Thus, controls can be re-used as proper matches for several treated individuals.

[47]See Rosenbaum and Rubin (1983).

[48]See Rosenbaum and Rubin (1983), cited by Reinkowski (2006).

[49]This clearly relates to the *initial* group of the untreated, whose members are just *potentially* used as controls, in contrast to the *matched* untreated which *actually* are used as controls after the matching.

[50]See Reinkowski (2006) and Caliendo and Kopeinig (2008).

also referred to as unconfoundedness — depends upon the covariates $X$, it is also conditional on the balancing score. The CIA based on the PS can thus be formalized as

$$(CIA/unconfoundedness\ given\ the\ PS) \qquad Y(0), Y(1) \amalg D|P(X), \quad \forall \quad X \tag{9}$$

which implies that given the propensity score — determined by the observable covariates $X$ which are not affected by the treatment — the outcomes are independent of the assignment into the treatment.[51]

A further assumption that has to be satisfied in order to properly implement PSM is the so-called 'common support condition', also referred to as the 'overlap condition'. The common support (CS) is met if

$$(common\ support) \qquad 0 < P(D = 1|X) < 1, \quad \forall \quad X \tag{10}$$

holds. The CS condition ensures that for all observed sets of $X$, and thus for all estimated $P(X)$, there is a positive probability of being treated as well as of staying untreated.[52] This excludes that there are values or value ranges of $X$ or $P(X)$ that allow for cases of so-called "perfect predictability" of the treatment $D$ given $X$ or $P(X)$.

According to Rosenbaum and Rubin (1983)[53], if both preconditions are satisfied — i.e. the conditional independence assumption (CIA) and the common support (CS) condition — then the condition of 'strongly ignorable treatment assignment' (SITA), or 'strong ignorability' for short, is fulfilled. And given strong ignorability, the *PSM estimator of the ATT* can be written as

$$\tau_{ATT}^{PSM} = E_{P(X)|D=1}\{E[Y(1)|D = 1, P(X)] - E[Y(0)|D = 0, P(X)]\} \tag{11}$$

which implies that the estimated average treatment effect on the treated (ATT) using PSM is calculated as the difference of mean outcomes between treated and matched untreated over the range of common support, where the outcome values of the control group are "appropriately weighted by the propensity score distribution" of the treated.[54]

[51]See Caliendo and Kopeinig (2008).
[52]See Heckman et al. (1999), cited by Caliendo and Kopeinig (2008).
[53]See Rosenbaum and Rubin (1983), cited by Reinkowski (2006) and Caliendo and Kopeinig (2008).
[54]See Caliendo and Kopeinig (2008).

### 4.3.1 Implementation steps

As mentioned above, PSM is quite a complex kind of analysis to carry it out in practice, consisting of the following *implementation steps*[55]:

1. *propensity score estimation*

2. choice of matching algorithm

3. check of common support

4. *matching* quality and *effect estimation*

5. sensitivity analysis

where (1) *estimating the propensity score* and (4) performing the chosen *matching* procedure with subsequent *effect estimation* are the two main steps in estimating the ATT applying propensity score analysis (PSA). Nevertheless, it is also essential to check the common support, the matching quality, and finally the robustness in order to evaluate the reliability of the results.

In cases of binary treatment variables, discrete decision models, like logit or probit models, are usually applied in order to *estimate the propensity score*.[56] As both models lead to quite similar results, the choice between them is not crucial.[57] For our analysis of the effects of benefit sanctions, we use a probit model to estimate the propensity score according to the standard setting of the Stata module *"psmatch2"*, developed and documented by Leuven and Sianesi (2014).[58]

A more important decision when performing PSA is the *choice of the matching algorithm*. An overview of diverse matching techniques and deeper insights into their advantages and drawbacks are given by various specialist literature[59] and shall not be discussed in-depth here.

For our analyses we predominantly conduct *kernel matching* (KM), as it applies a non-parametric matching estimator that is generated by using "weighted averages of all individuals in the control

[55] See Caliendo and Kopeinig (2008), slightly modified by Müller (2012).
[56] See Müller (2012).
[57] See Caliendo and Kopeinig (2008).
[58] See also Sianesi (2001) for a short overview of how to implement PSM estimators with Stata.
[59] See for example Guo and Fraser (2015), cited by Müller (2012); and see also Caliendo and Kopeinig (2008) and Müller (2012) for a brief overview and discussion.

group to construct the counterfactual outcome".[60] And thus KM leads to lower variance, because more information of the group of untreated is used, compared to another common matching technique called *nearest neighbor matching* (NNM).[61] We apply the two different variants of the latter (NNM) as a kind of robustness check.[62]

After the matching algorithm is chosen, the matching procedure is conducted under the restriction of common support. Subsequently, the ATT is estimated according to the above derived Formula 11 *(PSM estimator of the ATT)*, provided that the matching quality is adequately fulfilled and thus the balancing property of the PSM is given.[63] And finally sensitivity analyses have to be conducted in order to check the robustness of the estimation results.[64]

### 4.3.2 Details of the matching approach

For our analyses of the ex-post effects of benefit sanctions against employable welfare recipients in Germany, we apply two general kinds of matching procedures: *"nearest neighbor matching"* (NNM) and *"kernel matching"* (KM). We carry out the whole analysis separately for women and men, as well as for two distinct samples of inflow into welfare receipt — specifically, for the inflow samples of the years 2007 and 2008 — using results from 2008 for our main evaluation, as presented in Section 5, and the results from 2007 in one of our sensitivity analyses.

After checking out several variations of nearest neighbor matching (NNM) with $k$=3 (3-NNM) and $k$=5 (5-NNM) nearest neighbors, varying the caliper, as well as the kernel matching (KM) kernel and bandwidth, we decided to apply both 5-NNM with a caliper of 0.01 and KM using an *Epanechnikov kernel* (EKM) with a bandwidth of 0.06, with both the bandwidth of KM and the caliper of NNM chosen to avoid bad matches. The common support is a priori fulfilled for all of our estimations of ATT, as we use the appropriate option of the Stata module *psmatch2* that we apply for all of our PSM estimations. As is usual in PSM, we apply sampling with replacement also for the NNM, because it is beneficial for the matching quality to re-use good matches.

[60]See Caliendo and Kopeinig (2008).
[61]See Caliendo and Kopeinig (2008).
[62]See Section 4.3.2 for further details of the matching approaches we use and Section 5.4.2 for further details of the robustness checks we conduct by comparing the results of different matching algorithms.
[63]See Section 5.4.1 for more details about our checks of matching quality.
[64]For more details about the different kinds of robustness checks we conducted see Section 5.4.2.

As we explain in Section 4.4, we implement a dynamic matching approach with stratification in order to solve the missing start point problem for the untreated which arises from the flexible timing of the treatment. Additionally, we carry out the whole analysis for two kinds of outcome variables: firstly, for duration outcomes and secondly, for probability outcomes.[65] In the case of probability outcomes, we calculate monthly updated ATT for overlapping periods for each stratum of every group and subgroup of welfare recipients under study, as well as for direct and indirect sanctions. Regarding direct sanctions, we estimate monthly updated ATT over 24 months for each quarterly strata; concerning indirect sanctions, we estimate the ATT over 18 months for each half-yearly strata.[66] Hence we have a tremendous number of estimations to assess and interpret which we present in a cumulative form in tables, as well as in a more detailed form via a choice of graphs which show the development of accumulated ATT over time.

For our main evaluation, we use kernel matching (KM), while nearest neighbor matching (NNM) serves as one of our robustness checks. KM was chosen because of its advantages when compared to the also popular NNM. First of all, KM uses not only a few nearest neighbors like NNM, but the majority of the observations in the control group given the common support condition, whereby they are properly weighted conditionally on their similarity of propensity scores to the individuals in the treatment group who they are matched with. Secondly, bootstrapped standard errors can be computed by applying KM with Stata. However, as bootstrapping is extremely time-consuming and requires a large amount of computing capacity, especially when faced with a huge number of estimations, we mainly use the common calculation of standard errors as implemented by Leuven and Sianesi (2014) in their Stata module *psmatch2* which "does not take into account that the propensity score is estimated". In order to check the reliability of these results, we additionally carry out spot checks with bootstrapped standard errors.[67]

### 4.4 Stratification

Fredriksson and Johansson (2008) point out that in cases of flexible timing of treatments, it "makes

[65] For more information about the two kinds of outcome variables see Section 3.3.

[66] Detailed information about the stratification we provide in Section 4.4.

[67] As mentioned in Section 5.4.2, the results obtained with bootstrapped standard errors widely confirm the results of the commonly calculated estimations of the ATT.

more sense to think of the assignment to treatment as a dynamic process, where the start of treatment is the outcome of a stochastic process", instead of dealing with treatment assignment as a "static problem" whereby "the information contained in the timing of treatment is typically ignored". We implement a dynamic approach of propensity score matching (PSM), applying stratification in order to take the timing of the treatment at least roughly into account, and to solve the problem of missing start dates for the untreated.

Estimating the ex-post effects of sanctions on subsequent duration outcomes like unemployment duration or subsequent binary outcomes like the labor market status, we face an initially undetermined timing of treatment which can start at any point in time; thus the starting point to measure the subsequent outcomes for the untreated is also undetermined. For this so-called 'missing start date problem', Lechner (1999) provides several solutions in the context of matching approaches. Another prominent alternative to deal with flexible starting points of treatments would be the timing of events (ToE) approach, discussed in Section 4.1. Weighing up the advantages and disadvantages, we follow Sianesi (2004), Fitzenberger and Speckesser (2007), and Hofmann (2012) applying the practicable and feasible method of stratification. One of the advantages of using a stratified matching approach is that it "allows for heterogeneous treatment effects for the different duration intervals considered",[68] namely for the different strata of welfare duration.

We extend the above introduced static PSM with binary treatments to a dynamic setting. This dynamic PSM approach divides the duration of welfare receipt into a set of intervals, the so-called 'strata'. Based on these strata, a treatment group for every stratum is defined, and finally each treatment group is matched with a selected group of completely untreated, the matched control group.

### 4.4.1 Details of the stratification

The main decision when applying stratification is to find a proper length of the duration intervals. On the one hand, short intervals are preferable: "first, a relatively short observation window reduces the potential bias due to conditioning on future outcomes described in Fredriksson and Johansson (2008). Second, the shorter the duration intervals are defined, the better effect heterogeneity between

[68] See Hofmann (2012).

the duration intervals can be controlled for."[69] On the other hand, the shorter the chosen observation window, the fewer treatments are observed. Too few cases in the treatment group can lead to problems in the estimation process and prevent one from getting reliable and statistically significant results, even if the effect, in reality, is strong.

In our analysis of the ex-post effects of welfare sanctions, we apply quarterly strata of welfare duration in the case of direct sanctions and half-yearly strata in the case of indirect sanctions.[70] The latter choice is due to a lack of merely indirect sanctioned people.

Another choice when applying stratification is the number of observation intervals considered. As with increasing duration of welfare receipt, the number of sanctions per month decreases (i.a. because of an increasing number of exit events over time and hence a decreasing number of people at risk of being sanctioned), we restrict our analyses to sanctions within the first year of welfare receipt. In the case of *direct sanctions*, this means we apply *four quarterly strata* $S_i$ with $i = 1, ..., 4$, and in the case of *indirect sanctions* we apply *two half-yearly strata* $S_i$ of welfare duration with $i = 1, 2$.

A further decision when implementing a stratified matching approach concerns the point in time when the measurement of the outcome variable begins. We could potentially start to measure from the beginning of the stratum, or from the end of the stratum on. The latter option would have the disadvantage that the short-term effects within the stratum would be excluded and thus underestimated; using the beginning of the stratum as starting point has the advantage that the information on events during the stratum is included, but it generates a potential bias caused by late treatments[71] within the stratum, which also leads to an underestimation of the treatment effect. Hence, the potentially conceivable option of choosing the starting point for measuring the outcome somewhere within the stratum, e.g. in the middle of the stratum, would not solve the problem of potentially biased effect estimations, as both distortions work in the same direction and, thus, would

[69] See Hofmann (2012).

[70] Hofmann (2012) uses duration intervals of two months in her analysis of ex-post effects of unemployment insurance (UI) sanctions, based on administrative data of the Federal Employment Agency (FEA). We do not have access to complete administrative data sets of the German FEA, but use a data set based on the "Sample of Integrated Labour Market Biographies" (SIAB) which is a 2% sample of administrative data, provided by the Research Data Centre (FDZ), which prepares administrative data of the FEA in order to provide them for internal and external scientific research. Because we are not using a full data set but a 2% sample, two-monthly strata would be too short to guarantee a sufficient number of observations in the treatment group for several subgroups of welfare recipients under study.

[71] Further information on what is meant by "late treatments" and how they can cause a bias to the detriment of the treated is given below in Section 4.4.3.

not cancel out the bias. In order to avoid losing information on events during the stratum, we follow the usual approach in the literature,[72] and apply stratification with measuring the outcomes from the beginning of the strata.

### 4.4.2 Formalized PSM with stratification

Detailed formalized descriptions of the causal inference problem with endogenous treatments and the evaluation approach applying propensity score matching (PSM) in a stratified manner are provided by Sianesi (2004), Fitzenberger and Speckesser (2007), and Hofmann (2012), where Hofmann (2012) applies a more complex stratification using substrata.[73]

For reasons of understandability by a more general public, we use our own notation when presenting the results in Section 5 that should be intuitively understandable, even for readers who might not be that familiar with a high degree of formalization. As our notation in Section 5 diverges from Sianesi (2004), Fitzenberger and Speckesser (2007), and Hofmann (2012), we give, here, a brief overview of the formalized PSM with stratification adapted to our notation, and for more detailed information we refer the reader to the aforementioned literature.

Similar to Fitzenberger and Speckesser (2007) who follow Sianesi (2004), we apply "the standard static binary treatment approach recursively depending on the elapsed [...] duration" of welfare receipt. For this purpose, "the standard binary treatment approach" has to be extended "to a dynamic setting". Hence, the estimated ATT calculated with Formula 11 for the common approach with static treatments, presented in Section 4.3, "has to be interpreted in a dynamic context. We analyze treatment conditional upon the [...] [welfare] spell lasting at least until the start of the treatment and this being the first [...] treatment during the [...] [welfare] spell considered."[74] Hence, the *PSM estimator of the ATT for stratified matching*, taking into account dynamic treatments, can be

[72] For example see Fitzenberger and Speckesser (2007) and Hofmann (2012).

[73] As mentioned in Section 4.4.3 and Section 5.4.2 and following Hofmann (2012), we developed a procedure to deal with the similar problem of distortions caused by stratification that we call the 'adjustment procedure', and which we use for spot checks to reveal biased outcomes.

[74] See Fitzenberger and Speckesser (2007); we refer to "welfare" spells where they refer to "unemployment" spells; moreover they investigate the effects of a specific training program for unemployed as the treatment, while we investigate the ex-post effects of welfare sanctions.

formalized as

$$\begin{aligned} \tau_{ATT}^{PSM}(S_i, m_j) = E_{P(X)|D_i=1}\{&E[Y_j(1)|D_i = 1, W \geq m_0, D_0 = ... = D_{i-1} = 0, P(X)] \\ &- E[Y_j(0)|D_i = 0, W \geq m_0, D_0 = ... = D_{i-1} = 0, P(X)]\} \end{aligned} \tag{12}$$

where $D_i$ is the treatment dummy for sanctions starting in the $i$th interval of welfare duration $W$, namely in stratum $S_i$; $m_0$ is the month directly before the beginning of the stratum $S_i$; $D_0$ is the potential observation time of people at risk (i.e. in welfare receipt) previous to the considered welfare spell; equation $D_0 = ... = D_{i-1} = 0$ requires that no treatment has occurred before the beginning of stratum $S_i$, which for the treated means that $D_i = 1$ (imposed in stratum $S_i$) is the first treatment they experience; $P(X)$ is the propensity score (PS) (see Formula 8) given the vector of control variables $X$ with (at least during the welfare spell) time-invariant characteristics; the term $Y_j(1)|D_i = 1$ is the treatment outcome of the treated for period $P_j \equiv [m_1, m_2, \ldots, m_j]$, starting with the first month $m_1$ after the beginning of stratum $S_i$ and ending with final month $m_j$, where $j$ counts the months after the start of stratum $S_i$; and the term $Y_j(0)|D_i = 0$ is the non-treatment outcome of the matched untreated for period $P_j$ with final month $m_j$. Formula 12 is based on the PSM estimator for ATT in case of static treatments, according to Caliendo and Kopeinig (2008) (see Formula 11), supplemented and advanced according to the estimated treatment parameter for a stratified matching approach presented by Fitzenberger and Speckesser (2007), and finally adjusted to the notation we use for an illustrative way of presenting our results in Section 5.

It may catch an attentive reader's eye that the second part of the equation denotes the non-treatment outcomes according to Caliendo and Kopeinig (2008) as $Y_j(0)|D_i = 0$, instead of $Y_j(0)|D_i = 1$ according to Fitzenberger and Speckesser (2007); additionally, the latter ones do not include $P(X)$ in their formula.[75] The reason for these differences is that Fitzenberger and Speckesser (2007) present an estimated treatment parameter in an earlier state of the matching approach which represents the *true ATT* (according to Formula 2 in Section 4.2) with $Y_j(0)|D_i = 1$ as the non-observable counterfactual outcomes of the treated. But Formula 12 does not depict the true ATT but the *estimator of the ATT*, using the non-treatment outcomes of the matched untreated $Y_j(0)|D_i = 0$, appropriately weighted by the propensity score distribution $P(X)$ of the treated, as the counterfactual case.

[75] See Formula 2 in Fitzenberger and Speckesser (2007).

We estimate the average treatment effect on the treated (ATT) for two kinds of outcome variables: metric duration outcomes and monthly updated binary outcomes indicating the labor market status over time. Formula 12 is designed for the more complex case of monthly updated binary outcomes, but can easily be adapted to the case of metric duration outcomes. The latter ones do not require monthly updates, and hence the index $j$, counting the months after the beginning of the stratum $S_i$, as well as the final month $m_j$ are omitted for the estimation of duration outcomes.

As mentioned above and outlined in Section 4.1, a huge advantage of using propensity score matching (PSM) with stratification, instead of ToE models with a defined and hence inflexible distribution function, is that the PSM estimator in Formula 12 allows for “heterogeneity in the individual treatment effects and for an interaction of the individual treatment effects with dynamic sorting taking place”, provided that the “dynamic version of the conditional mean independence assumption (DCIA)” holds.[76] Concretely in our case, implementing PSM in a stratified manner allows for heterogeneous treatment effects for different intervals (strata $S_i$) of welfare duration $W$. Thus, our estimations for divers strata can reveal varying ex-post effects of sanctions depending on the previously elapsed duration of welfare receipt.

#### 4.4.3 Bias caused by stratification

There are two kinds of distortions caused by not considering the exact time of the treatment when applying PSM in a stratified manner. The first one is the core of the bias problem as a result of stratification. Thus we call it ‘the core bias’. Not only can it weaken the estimated treatment effect, but may even turn it into the opposite direction; if this kind of bias occurs, it affects the estimated ATT on probabilities as well as on durations.[77] Even if the core bias can decrease over time, it does not fully vanish. The second kind of bias we call the ‘time lag bias’, as it is caused by a time lag in the measurement of the treatment effect. It leads to a reduced effect estimation for probability outcomes as long as the treatment effect has not fully expired, namely if there is still a noticeable monthly effect. The time lag bias is less problematic than the core bias as it can merely weaken the estimated effect

[76]See Fitzenberger and Speckesser (2007) also for further details on the DCIA, the dynamic version of the CIA.

[77]As explained in Section Section 3.3, we carry out the whole analysis for two kinds of outcome variables: metric duration outcomes and monthly updated binary outcomes, indicating an exit event changing the labor market status; hence we get estimations of the ATT on durations as well as on probabilities.

but cannot turn it into the opposite direction and the distortion diminishes over time and vanishes when the treatment effect is expired. Hence, the time lag bias does not affect the ATT on probabilities in the long run and, moreover, it does not affect the ATT on durations at all. Therefore, we place the main focus in this section and in the discussion of the results in Sections 5 and 6 on the core bias. In the following sections, however, we do not distinguish between different kinds of biases, but usually refer to the entire distortion that we call "the bias", for short.

As we analyze the ATT based on two different kinds of outcome variables, it is important to take into account that the bias works in opposite directions depending on the type of outcome. In general, and if we do not explicitly mention the kind of outcome variables, we refer to probability outcomes. Concretely, a bias "to the detriment of the treated", or more precisely "to the detriment of the treatment effect", connotes a negative bias of the ATT on probabilities. And as lower probabilities correlate with higher durations, the bias of the ATT on durations would be accordingly positive; conversely, in case of a positively biased ATT on probabilities, the reverse is valid and the ATT on durations would be biased negatively. All this holds for both kinds of distortions, the core bias and the time lag bias, as well to the bias in total.

In the remainder of this section we highlight different aspects of the distortions and analyze the effects of the core bias in the context of different kinds of exit events.

### Origins and effects of the distortions

The origin of both kinds of distortions is that the outcome is measured from the start of the stratum on, while the treatment can occur at any time within the stratum. The observations with treatments after the beginning of the stratum, at first, cause a temporal shift in measuring the treatment effects on outcomes due to the *time lag between beginning of the stratum and treatment* which can distort ATT estimations on probabilities for a limited period of time. Secondly, and much more important, the time lag enables observations with *exit events before the treatment.* These events previous to the treatment are the origin of the more problematic core bias. The distortion due to *previous events in the control group* is the *typical case* of the bias arising from matching in a stratified manner that is regularly discussed in the literature. Furthermore, analyzing ex-post effects of benefit sanctions, the standard cases, predominantly regarded in the literature, are exit events that lead out of risk to be treated:

getting employed without receiving top-up benefits and exiting from benefit receipt comprising the non-employment option. Hence, exit events before the treatment in these standard cases can only occur in the group of non-treated. However, we additionally analyze exit events that do not transition from being at risk to getting sanctioned — i.e. taking up employment with supplementary welfare benefits and the reverse case of leaving employment with top-up benefits for mere welfare receipt. Both examples enable observations with events previous to the treatment, including those outside of the control group; they additionally allow for *previous events in the treatment group*. The latter one can generate *reverse causality*, and thus yield a further problem adding to the core bias and affecting it.

**Time lag bias:** The time lag bias results from the delay between the beginning of the stratum which is the starting point for measuring the outcome variable, and the average beginning of the treatment. Thus, a potential treatment effect occurs with a temporal shift according to the average time lag, while we implicitly take the start of the stratum to be the beginning of the treatments. This time lag weakens the effect estimations only if the ATT is measured on a specific reference date which is before the treatment effect ends. Therefore it arises in cases of monthly updated ATT on probabilities as long as the impact of the treatment is still effective; after the treatment effect has expired in the case of ATT on durations, the time lag bias does not become effective. The reason why the time delay does not affect the ATT on durations is that it *extends the durations* until the treatment, measured in absolute values, *equally for both groups*: treated and untreated. Thus, the difference between the durations of treatment and control group remains unaffected by the time lag, and hence, so does the estimated ATT based on durations. In the case of ATT on probabilities, however, and before the impact of the treatment is fully expired, the ATT is tendentially underestimated. The time lag bias is larger the later in the stratum the treatments averagely begin. It has merely a weakening effect, which is limited in time, but it can never turn the ATT into its opposite direction.

**Core bias:** The core bias arises from exit events prior to the treatment. Because we analyze an assumed causality between the treatment as the cause and the probability or the timing of the event as the effect, the cases with reverse order, namely with *exit events before the treatment*, are generally problematic. At best, reverse cases might cause a measured individual treatment effect of zero. This holds true if the probability of the exit events before the treatment would be identically distributed for the treated and the matched control group. Then, the estimated ATT is merely weakened by the observations with a measured 'zero effect'. The core bias, which in this case causes only a 'weakening effect', is higher the larger the share of observations with reverse order, and thus, the more often observations with a measured individual effect of zero occur. Only in the special case where there is actually no treatment effect, the distortion caused by the zero effect of the pre-treatment events would be zero. Otherwise, namely if there is a positive or negative treatment effect, the estimated ATT is weakened by the bias; but in this case the core bias cannot turn the effect into the opposite direction. In this particular constellation, facing identically distributed pre-treatment events in the treatment and in the matched control group, the core bias is least problematic. In contrast, much more problematic is the typical case most often discussed in the literature, where exit events lead 'out of risk' to be treated. In our study, this concretely refers to exit events that lead out of welfare receipt and, thus, out of risk to be sanctioned. This kind of exit event results in a probability of zero for the treatment group that the event occurs before the treatment. Hence, the higher the probability of exit events before the treatment of the matching partner is for the people in the control group, the stronger the core bias to the detriment of the treated, which causes a negative bias of the ATT on probabilities, and an accordingly positive bias of the ATT on durations. Thus, a negative treatment effect on probabilities would be strengthened by the negative core bias, and a positive treatment effect would be reduced or the estimations of the ATT would even turn into the opposite direction.

As we see already in this first glance at the core bias under different conditions, its effects are strongly determined by the kind of exit events under study, and by other specific circumstances. Therefore, in the following, we provide more detailed information about the core bias under various conditions and in the context of certain kinds of exit events which we analyze in our study:

**Exit events out of risk: exit from welfare and to mere employment**

Examples of exit events leading out of risk to be sanctioned are exit to mere employment *("only job")* ($O$), and exit from welfare ($ExWel$) which also comprises the non-employment option of leaving the labor market. As mentioned above, under these circumstances, the probability of exit events before the treatment is zero in the treatment group, while the probability of exit events prior to the treatment of their matching partner in the control group is usually non-zero. This causes a negative core bias, which weakens a potential positive treatment effect or may even turn the estimations of the ATT into the opposite direction, namely into negative values of the ATT on probabilities. In contrast, the negative core bias would strengthen a potential negative treatment effect, and thus result in even stronger negative values of the ATT. The more right-skewed the distribution of treatments within the stratum is, and the more likely early exit events in the group of non-treated are, the more severe the expected negative core bias will be.

**Exit event staying at risk: exit to employment with top-up benefits**

The exit event into employment with supplementary welfare receipt ($S$) induces a change of the labor market status accompanied with staying at risk to be sanctioned even after the event takes place. For this kind of transition from unemployed to employed welfare receipt, there are different constellations depending on the probability of pre-treatment events of the treatment group compared to the group of non-treated. The average difference between the pre-treatment outcomes of the two groups, in turn, depends on the probabilities of treatments in the two labor market states: before and after the event happens. Concretely, the average outcome difference before the treatment depends on whether unemployed compared to employed welfare recipients are more, less, or equally likely to be sanctioned and how large the potential individual differences averagely are.

**Reverse causality (indirect):** If the average probabilities of treatments in the pre- and post-event status are different, we are confronted with the problem of reverse causality: then the likelihood to be treated depends on the exit event. That is the reverse causality of what we initially intended to analyze. Even if the event — taking up employment with supplementary benefit receipt — generally does not directly cause a sanction, there may be an indirect causality, mediated by a potentially unequal probability to be sanctioned in the pre- and post-event status, concretely in the status of unemployed compared to employed welfare receipt.

**Equal probability of pre-treatment events for treated and non-treated:** As mentioned above, giving an example of the effects of the core bias under various circumstances, this constellation arises under the precondition of identically distributed exit events before the treatment for the treated and the matched control group. Under this precondition, the probabilities of pre-treatment events for treated and non-treated are equal and, hence, the maximal effect of the core bias is reduced to a distortion which can only weaken a potential positive or negative treatment effect but cannot turn it into the opposite direction. The reason for this is that the exit events prior to the treatment induce a measured individual treatment effect of zero, caused by the identical probability distribution of pre-treatment events for both groups: treated and matched controls. The larger the share of observations with reverse order, and thus with a measured individual treatment effect of zero, the stronger the weakening effect of the bias. This constellation with equal probabilities of pre-treatment events for both groups can only occur if the exit event does not affect the likelihood of the treatment. Concretely, in order to make this constellation feasible for exit into supplementary welfare receipt, the probabilities of being sanctioned must be equal for unemployed and employed welfare recipients.

**Lower probability of pre-treatment events for the treated:** Under these conditions, a bias to the detriment of the treated, or more precisely to the detriment of the treatment effect, occurs. Hence, the core bias negatively distorts the ATT on probabilities. If there is a huge number of pre-treatment events in the control group and only a few of these events in the treatment group, the negative bias can be strong enough to not only weaken a potential positive treatment effect, but to even turn it into negative effect estimations. An already negative treatment effect, however, would lead to even stronger negative values of the estimated ATT. This goes along with the statements to the core bias in case of an exit event out of risk causing a probability of pre-treatment events for the treated of zero, which is the extreme case — or technically speaking, the marginal case — of this category of circumstances causing a core bias.

**Higher probability of pre-treatment events for the treated:** In this case, a bias in favor of the treated, or more precisely in favor of the treatment effect, occurs. Hence, the core bias positively distorts the ATT on probabilities. In extreme cases with many pre-treatment events in the treatment group and considerably less of such events in the group of matched non-treated, the positive bias may be strong enough to not only reduce a potentially negative effect but to even turn it into a positive estimation of the ATT. In contrast, an already positive sanction effect would lead to even stronger positive estimations, and hence a positive ATT would be overestimated.

### Exit event staying at risk: exit from job to mere welfare

Another exit event not leading out of risk, apart from exiting to supplementary welfare receipt, is the converse case of exiting employment with top-up benefits to mere welfare receipt ($ExJob$). Theoretically, the same constellations with similar consequences are possible, like the previously described cases for exit to supplementary welfare receipt. In practice, however, exiting employment for mere welfare receipt is an event with a high probability of reverse causality directly caused by the exit event which tendentially leads to a strong positive core bias.

**Reverse causality (direct):** As 'culpably' losing a job is a compelling reason to impose sanctions against welfare recipients, the likelihood that, in the case of reverse order of treatment and exit event, the event of exiting employment is the reason for the sanction, and thus the direct cause of the treatment. Such cases of reverse causality lead to a higher probability of pre-treatment events for the treatment group compared to the matched controls. This causes a positive bias in the ATT on probabilities which might not just weaken a potential negative effect, but even turn it into positive effect estimations, while an already positive ATT would be strengthened and thus be overestimated. This constellation is similar to the above described case of higher probability of pre-treatment events for the treated compared to the non-treated for exits into supplementary welfare receipt. The main difference with exit into mere welfare is that the latter one is most probably distorted considerably more severely into the positive direction than in the case of exits to supplementary welfare receipt.

### Further thoughts to exit events staying at risk

After the previous overview on diverse constellations of the two analyzed exit events staying at risk — exit to supplementary, and exit to mere welfare receipt — we reflect about the factors that determines which kind of constellation occurs.

If the event, or more precisely the change of the labor market status caused by the event, does not affect the likelihood to be treated, the probability of the pre-treatment event should be equal for treatment and non-treatment group. The reason for this is that the matching procedure per definition targets at equal non-treatment outcomes for both groups: treatment and matched controls. In this case, the core bias is minimal and has only a weakening effect.

In contrast, if the probability of being sanctioned is affected by the exit event, the probabilities of pre-treatment events for treatment and control group are not equal. If the treatment is less likely after the event, pre-treatment events for the treatment group would be less likely than for the control group. The extreme case would be a probability of pre-treatment events of zero for the treatment group which leads to a maximal negative potential core bias if the probability of pre-treatment events is more than zero.

In the opposite case, if the treatment is more likely after the event, pre-treatment events for the treatment group would be more likely than for the matched non-treated. Thus the potential core bias would work in the positive direction. This is even more the case when the exit event directly causes the treatment which is quite probably the case for at least some of the observations of exiting employment before getting sanctioned. Then the probability of pre-treatment events for the treatment group tends to be much higher than for the controls. This causes a strongly positive core bias. The more cases with reverse causality, the stronger the positive core bias.

We have to take into account the effects of the aforementioned bias when we interpret the results in Chapter 5.

# 5 Results

In this section, we present the estimation results of propensity score matching (PSM), applying kernel-based matching for the inflow sample into welfare 2008.[78] Specifically, we report the estimated average treatment effects on the treated (ATT) of welfare sanctions on the probability of different outcome events, and on the duration until these events occur. For *unemployed* benefit recipients (Section 5.1), we analyze the sanction effect on taking up employment — differentiating between mere employment, and employment while receiving supplementary UB II, i.e. top-up welfare benefits —

[78]Further details of the matching approach, i.a. the matching algorithms and the inflow samples we apply, are given in Section 4.3.2; our findings from robustness checks based on comparing the results of different matching algorithms and inflow samples are addressed in Section 5.4.

and on completely leaving welfare receipt.[79] For *employed* benefit recipients (Section 5.2), we carry out basically the same analyses, but restrict the employment options to exit into mere employment, and consider the additional option of exiting employment with top-up benefits for mere welfare receipt. In Section 5.3, we present the results of PSM analyses for the case of being affected merely by *indirect* sanctions, i.e. by sanctions against related household members. And finally, in Section 5.4, we give information on our checks of matching quality and sensitivity analyses and the findings we get from these.

In addition to the two main groups of unemployed and employed welfare recipients, we separately analyze the effects of welfare sanctions on several subgroups: first, we define three age groupings: people *under 25 years* (*u25*), who face stricter monitoring and sanction conditions than the older ones, people aged *25 years and over* (*o25*), and the whole range of ages (*all*) comprising *u25* and *o25*; second, we differentiate between people who live in a federal state belonging to (the former) *West* or *East* Germany; and third, we differentiate between people of three different levels of labor market access: *low*, *middle*, *high*.[80]

As outlined in Section 3.3, our analyses are based on two distinct kinds of outcome variables: metric duration outcomes and monthly updated binary outcomes. The *metric* outcomes, measured in days from the beginning of each stratum ($S_i$) of welfare duration[81], represent the *durations* until a specific event takes place, e.g. employment entrance or exit from welfare. Thus the estimated ATT, calculated as the difference between mean outcomes for the treatment group and the matched control group, reveals the average number of days a sanctioned person is expected to stay longer (*positive ATT on durations*) or shorter (*negative ATT on durations*) in welfare receipt until the event happens, compared to the case of not having been sanctioned. This illustrative interpretation of the ATT, revealing the sanction effect as the difference of average welfare duration until a specific event occurs, is an advantage of the approach of metric outcomes. A disadvantage of using durations as outcome variables is that right-censored spells, i.e. cases without the specific (exit) event taking place in the observation period, drop out of the analysis of metric outcomes.

[79] The difference between 'taking up mere employment' and 'exit from welfare receipt' is that the latter one is more comprehensive as it additionally comprises the non-employment option (see Sections 5.1.2 and 5.2.2).

[80] Further information on the variable 'labor market access' are given in Section 3.1.

[81] In the case of direct sanctions, $S_i$ are quarterly strata of welfare duration with i=1–4; in the case of indirect sanctions, $S_i$ are half-yearly strata with i=1–2.

The *monthly updated binary* outcome variables, by contrast, indicate whether a specific event occurs within a defined time period $P_j$, beginning at the start of stratum $S_i$ and ending on the last day of the final month $m_j$. Hence, the estimated ATT represents the difference in average *probabilities* that the event takes place within the time period $P_j$, for the sanctioned, and matched non-sanctioned groups. One advantage of this approach is that it shows the development of the cumulative ATT over time, revealing the effects of welfare sanctions in the short, medium and long run,[82] which we exemplarily illustrate by a wide selection of *plotted ATT on probabilities* and their 90% confidence intervals[83], shown in the appendix, Figures1–37. In the case of *direct* sanctions, we estimate the ATT over two years time — that is for final months $m_j$ with j=1–24, and in the case of *indirect* sanctions over a period of one and a half year — that is for final months $m_j$ with j=1–18.

But even the *tables* with condensed information on the estimation results of *binary* outcomes provide additional information compared to the ATT on durations. These summaries expose the *number of* (overlapping) periods with *significant (positive/negative) ATT on probabilities*, summarized over a certain number of overlapping periods, $P_j$, which we have restricted to periods with final months $m_j$ subsequent to the end of stratum $S_i$.[84] Beyond the different definition and interpretation of binary outcome variables, and thus of their derived ATT, a critical reason for the increase of informative value is that we need not restrict the analysis to uncensored spells, which is another major advantage of using binary outcomes.

Let us illustrate this by an example: in the case of the exit event 'employment entrance', welfare spells of people who don't find a job within the observation period cannot be a part of the analysis of metric outcomes, as their welfare durations until employment entry are unknown. However, these so-called 'right-censored' spells are included in the analysis based on binary outcomes. This might lead to significant sanction effects on the probability to take up employment, even if the ATT on the

[82]In this paper *short-term* effects refer to a period of three months maximum, *medium-term* effects to more than three until twelve months time, and *long-term* effects to periods of more than twelve months.

[83]We illustrate the ATT plots with confidence intervals (ci) of 90% corresponding to the significance level of $\alpha$=0.1 used for the tables with binary outcomes.

[84]We initially calculate the monthly updated ATT of binary outcomes for periods, $P_j$, starting with final month, $m_j$, after the beginning of the stratum, i.e. starting with $m_j$=$m_1$. Due to the bias during the strata caused by not taking into account the exact time of treatment within the stratum (see Section 4.4.3), however, we restrict the evaluation of our results to the time periods $P_j$ with final months $m_j$ subsequent to the end of strata. Specifically, we consider final months $m_j$ with j=4–24 following the end of quarterly strata in case of direct sanctions; in the case of indirect sanctions, we consider final months $m_j$ with j=7–18 following the end of half-yearly strata.

duration until employment entrance is insignificant.

A further reason why using binary outcomes could reveal significant effects, even if the analysis of durations does not, is that the extent of the treatment effect as well as its direction can vary over time in a way that leads to such seemingly contradictory results. This could be the case if the sanction effect on the probability of a specific event turns out to be, for example, positive in the first few months and changes direction in the following months, and thus the later monthly effect (at least partly) neutralizes the earlier one, which might lead to an insignificant effect on the duration as a whole.

It is important to recognize that, in contrast to a *positive ATT of metric outcomes* which, for sanctioned people implies a politically *unintended increase in average duration* until a specific event takes place (e.g. employment entry or exit from welfare), a *positive ATT of binary outcomes* indicates an *increase in average probability* that the specific event takes place within a defined time period and, thus, works in the politically *intended* direction; for *negative* ATT, of course, the reverse is valid.

As explained in Section 4, we have to consider that the matching procedure combined with stratification may cause a bias to the disadvantage of the treated, which could lead to negatively biased estimations of the ATT on probabilities and positively biased ATT on durations. Even if the bias arises from exit events during the stratum, the estimations for final months $m_j$ after the end of stratum (i.e. with $j > 3$ in case of direct sanctions, respectively with $j > 6$ in case of indirect sanctions) can also be negatively affected, as — calculating the ATT on probabilities for overlapping periods $P_j$ — the exit events within the stratum are still contained in $P_j$. However, the later the final month occurs, the higher the number of cases with exit events after the end of stratum is; and therefore, the share of possibly biased cases diminishes, concurrently the entire potential bias shrinks.

As we cannot verify or quantify the potential bias, we initially present the pure numeric results as a first step, then relativize the results below by discussing the insights that we gain from the plotted ATT regarding the possible bias as a second step; finally, we discuss our findings more comprehensively, reflecting on the plausibility in general, and comparing them with previous studies, in Section 6.

## 5.1 Unemployed welfare recipients

Initially, we present the estimation results for unemployed welfare recipients who are the main target group of activation policy in the German welfare system under 'Hartz IV'. Unemployed people, in particular, are at the focus of the monitoring and sanction regime, with the goal of speeding up their employment entrance, and exiting their receipt of welfare benefits.

### 5.1.1 Transition from unemployment to employment

Investigating the sanction effects on transition from unemployment to employment, we distinguish between three exit events: first, exit to *mere employment* which means *without welfare receipt (O: 'job only')*; second, exit to *employment with top-up benefits (S: 'supplementary')*; and third, exit to *employment in general (G: 'job in general')* which consists of 'job only' ($O$) and job with additional welfare receipt ($S$) without any distinction.

Table 3 shows the estimation results of *metric* outcomes, namely the average treatment effects on the treated (ATT) of benefit sanctions on welfare *duration* until employment entrance, for the first two quarterly strata for unemployed men who entered welfare (UB II) in 2008. For women, the estimated ATT of the metric outcomes turned out to be either not significant or not sufficiently reliable — the latter one dominating because of too few observations in the treatment group, which is defined in cases of less than 50 treated persons.

Within the reliable results (black figures) in Table 3 for men, the ATT of just the exit into mere employment case for the second stratum ($S_2$) turned out to show significant sanction effects. Concerning the whole range of ages from 15 to 56 years (*all*), those men who are sanctioned during the second stratum of their welfare receipt need, on average, 70.14 days longer to find a job with sufficient earnings to bring them out of welfare receipt than the unsanctioned control group. For men aged 25 years and over (*o25*), the sanction effect on the duration until leaving welfare for mere employment is a bit smaller: more precisely, the sanctioned need, on average, 57.02 days longer than if they would not have been sanctioned. Also, for the sub-groups of unemployed men in Western Germany, and male unemployed with mid-level labor market access, we get significant positive effects on welfare durations in the case where sanctions were imposed during the second stratum ($S_2$). These results, being counter to the political intention of speeding up employment entrance, could — at least partly — be caused by the above mentioned potential bias to the detriment of the treated, which we briefly elucidate at the end of this section and discuss more detailed at the end of Section 5.1.2 and in Section 6.

As explained above, there are substantial merits to extend the analysis of durations by additionally calculating the *ATT on probabilities* using (monthly updated) *binary* outcomes which indicate different labor market states: firstly, right-censored observations with unknown duration until employment

**Table 3:** Unemployment duration until employment entrance — men 2008

| Age | Region | Market access | Exit to job (G/O/S)[2] | ATT[1] $S_1$ | ATT[1] $S_2$ |
|---|---|---|---|---|---|
| all | | | G | 3.49 | 21.30 |
| all | | | O | -18.47 | 70.14*** |
| all | | | S | 20.41 | 4.71 |
| all | West | | G | 0.05 | 7.55 |
| all | West | | O | -25.35 | 65.06** |
| all | West | | S | 20.96 | 6.01 |
| all | | middle | G | -7.92 | 13.70 |
| all | | middle | O | -19.61 | 84.37*** |
| all | | middle | S | 1.88 | -4.32 |
| all | | high | G | 13.03 | 177.93* |
| all | | high | O | 4.95 | 240.59*** |
| all | | high | S | 38.44 | 202.91* |
| u25 | | | G | 0.81 | 29.25 |
| u25 | | | O | 12.51 | 92.33** |
| u25 | | | S | 6.07 | 1.03 |
| u25 | West | | G | -1.69 | 28.40 |
| u25 | West | | O | -16.81 | 139.99** |
| u25 | West | | S | -5.61 | -69.42 |
| o25 | | | G | 8.72 | 10.64 |
| o25 | | | O | -22.15 | 57.02** |
| o25 | | | S | 25.97 | 7.44 |
| o25 | West | | G | 7.42 | 1.82 |
| o25 | West | | O | -33.78 | 40.14 |
| o25 | West | | S | 34.21 | 10.69 |

[1] ATT of *metric outcomes:* difference of mean durations until employment entrance, measured in days from the start of quarterly stratum of welfare duration ($S_i$).
[2] *Events*: exit to: job only (O), i.e. without welfare receipt, job with supplementary welfare receipt (S), job in general (G), i.e. comprising (O) and (S).
*Subgroups:* age-group in years: all=15–56, u25=15–24, o25=25–56; region: West/East German states; labor market access: low, middle, high.
*Significance levels:* $\alpha$=0.1*, $\alpha$=0.05**, $\alpha$=0.01***.
*Gray figures:* not reliable results (<50 treated cases).

**Table 4:** Unemployed's employment entrance — men 2008

| Age | Region | Market access | Exit to job (G/O/S)[2] | Number of periods with sign. ATT[1] $S_1$ | $S_2$ | $S_3$ | $S_4$ |
|---|---|---|---|---|---|---|---|
| all | | | G | n.s. | n.s. | 5(+) | n.s. |
| all | | | O | n.s. | 21(−) | 20(−) | 2(−) |
| all | | | S | n.s. | 3(+) | 15(+) | n.s. |
| all | West | | G | n.s. | n.s. | 9(+) | 3(−) |
| all | West | | O | n.s. | 16(−) | 17(−) | n.s. |
| all | West | | S | n.s. | 15(+) | 16(+) | n.s. |
| all | East | | G | 3(−) | 2(−) | | |
| all | East | | O | 1(+) | 14(−) | | |
| all | East | | S | 13(−) | n.s. | | |
| all | | low | G | 2(−) | | | |
| all | | low | O | 17(−) | | | |
| all | | low | S | n.s. | | | |
| all | | middle | G | n.s. | n.s. | 3(+) | 2(−) |
| all | | middle | O | n.s. | 18(−) | 20(−) | 2(−) |
| all | | middle | S | 1(−) | 3(+) | 4(+) | n.s. |
| all | | high | G | n.s. | | | |
| all | | high | O | 1(+) | | | |
| all | | high | S | 1(−) | | | |
| u25 | | | G | n.s. | 1(−) | n.s. | |
| u25 | | | O | n.s. | 9(−) | 7(−) | |
| u25 | | | S | n.s. | n.s. | 8(+) | |
| u25 | West | | G | n.s. | n.s. | 2(+) | |
| u25 | West | | O | 6(−) | 1(−) | n.s. | |
| u25 | West | | S | 3(−) | 1(−) | 1(−) | |
| o25 | | | G | n.s. | n.s. | 2(+) | 3(−) |
| o25 | | | O | n.s. | 21(−) | 20(−) | 13(−) |
| o25 | | | S | 7(−) | 4(+) | n.s. | 1(−) |
| o25 | West | | G | n.s. | n.s. | 2(+) | 7(−) |
| o25 | West | | O | n.s. | 13(−) | 15(−) | 14(−) |
| o25 | West | | S | n.s. | 17(+) | 3(+) | 5(−) |
| o25 | East | | G | 3(−) | 2(−) | | |
| o25 | East | | O | 1(+) | 9(−) | | |
| o25 | East | | S | 4(−) | 1(−) | | |

[1] ATT of *binary outcomes:* difference of mean probabilities of employment entrance until the end of month $m_j$ after start of quarterly stratum of welfare dura tion ($S_i$); number of positive (+) and/or negative (–) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=4–24, or no significant (n.s.) ATT in any of these period s, on a significance level of $\alpha$=0.1.
[2] *Events*: exit to: job only (O), i.e. without welfare receipt, job with supplementary welfare receipt (S), job in general (G), i.e. comprising (O) and (S).
*Subgroups:* age-group in years: all=15–56, u25=15–24, o25=25–56; region: West/East German states; labor market access: low, middle, high.

entrance are also included in the analysis, which is expected to reduce problems caused by too few observations in the treatment group. Secondly, the monthly updated binary outcomes, at least for several periods $P_j$, can show significant effects on the probability of getting a job, even if the effect on the duration until employment entrance is insignificant. Indeed the monthly updated ATT on probabilities based on binary outcomes, presented in Table 4 (men) and Table 5 (women), yield significant results in many more cases than the ATT based on durations.

On the one hand, the significant *positive* ATT on *durations* for men, presented in Table 3, are confirmed by the corresponding significant *negative* ATT on *probabilities* in Table 4, which indicates lower mean probabilities of getting a job for sanctioned men compared to the fictive case of not having been sanctioned. According to this, the ATT of binary outcomes for exit to *'job only'* ($O$) show significantly *negative* effects of sanctions imposed during the second stratum ($S_2$) for all 21 considered periods $P_j$ with final months $m_j$ after end of stratum $S_i$ (i.e. for $j = 4, \ldots, 24$) for the following two groups: unemployed men of the whole range of ages (*all*), that is 15–56 years, and unemployed men aged 25 years and above (*o25*). These significantly negative ATT of binary outcomes imply that the average probability of sanctioned people leaving welfare receipt for *mere employment* ($O$) before the end of the final month $m_j$ is significantly lower than for the unsanctioned control-group. Hence, the negative ATT of binary outcomes confirm the economically and politically unintended direction of sanction effects on taking up employment without benefit receipt for unemployed men, which we likewise get from the metric outcomes.

On the other hand, the binary outcomes reveal a much more differentiated picture than the metric outcomes, showing plenty of periods $P_j$ with significant negative as well as positive ATT, even if the durations do not yield reliable and significant estimates: in contrast to duration outcomes, binary outcomes for unemployed men lead to significant ATT not only for sanctions in the second stratum of welfare receipt ($S_2$), but — depending on the subgroup — for at least a few periods, $P_j$, of each of the first four quarterly strata ($S_1, \ldots, S_4$). Moreover, they yield significant treatment effects for further subgroups of unemployed men (e.g. living in East Germany, having a low level of labor market access), as well as for unemployed women. Furthermore, significant effects of welfare sanctions are not only shown for the exit to 'job only' ($O$), but also to employment with supplementary benefit receipt ($S$) and employment in general ($G$).

Considering only cases — where 'cases' here refers to (sub)groups combined with strata $S_i$ — with at least 14 of the regarded 21 periods $P_j$ with significant ATT in Table 4 for unemployed men, reveals *negative* sanction effects only for the probability to enter *mere employment* ($O$), whereas *positive* effects appear only for entering *employment with top-up benefits* ($S$). In other words, regarding cases with a high persistence of significant ATT, sanctions against unemployed men on average lower their chance of finding a job which is well-paid enough to bring them out of benefit receipt, but raise the

**Table 5:** Unemployed's employment entrance — women 2008

| Age | Region | Market access | Exit to job (G/O/S)[2] | Number of periods with sign. ATT[1] $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|---|---|---|
| all | | | G | n.s. | n.s. | 19(+) |
| all | | | O | 1(−) | 7(−) | n.s. |
| all | | | S | n.s. | n.s. | 21(+) |
| all | West | | G | 1(+) | n.s. | 21(+) |
| all | West | | O | n.s. | 3(−) | 1(−) |
| all | West | | S | n.s. | n.s. | 21(+) |
| all | | middle | G | n.s. | n.s. | 21(+) |
| all | | middle | O | 2(−) | n.s. | n.s. |
| all | | middle | S | 1(−) | n.s. | 21(+) |
| u25 | | | G | n.s. | | |
| u25 | | | O | 4(−) | | |
| u25 | | | S | n.s. | | |
| u25 | West | | G | 21(+) | | |
| u25 | West | | O | n.s. | | |
| u25 | West | | S | n.s. | | |
| o25 | | | G | 2(−) | n.s. | 21(+) |
| o25 | | | O | n.s. | n.s. | 1(+) |
| o25 | | | S | 10(−) | n.s. | 21(+) |
| o25 | West | | G | 1(−) | | |
| o25 | West | | O | n.s. | | |
| o25 | West | | S | 4(−) | | |

[1] ATT of *binary outcomes:* difference of mean probabilities of employment entrance until the end of month $m_j$ after start of quarterly stratum of welfare duration ($S_i$); number of positive (+) and/or negative (–) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=4–24, or no significant (n.s.) ATT in any of these periods, on a significance level of $\alpha$=0.1.
[2] *Events*: exit to: job only (O), i.e. without welfare receipt, job with supplementary welfare receipt (S), job in general (G), i.e. comprising (O) and (S).
*Subgroups:* age-group in years: all=15–56, u25=15–24, o25=25–56; region: West/East German states; labor market access: low, middle, high.

probability of getting a job needing supplementary welfare benefits.[85]

A high persistence of *positive* sanction effects for *male* unemployed occurs only in a few cases. To be precise, cases with at least 14 of 21 periods $P_j$ of positively significant ATT show up only for the transition into *job with top-up benefits* ($S$), only for sanctions imposed during the second and/or third stratum ($S_2$, $S_3$) of their welfare receipt, and only for the following three groups: the total of unemployed men (*all*), male unemployed living in Western Germany (*West*), and those who live in West Germany and are 25 years and older (*o25/West*).

For *female* unemployed (Table 5), in contrast, *positive* sanction effects dominate the results. Focusing on the cases with more than 10 periods $P_j$ with significant ATT, positive sanction effects on the probability of getting a job were observed in all cases, for at least 19 of the 21 periods considered. All these cases with a high persistence of positive ATT occur for exit into *employment with top-up benefits* ($S$) and for exit into *job in general* ($G$). Similar to male unemployed, persistent *positive* sanction effects do *not* appear for women taking up *mere employment* ($O$). It is striking that, with one exception, unemployed women show more than 10, namely 19 or all 21 of the regarded periods with significant ATT only for the *third* stratum ($S_3$) — i.e. for sanctions imposed during the 7th to the 9th month of their welfare receipt. The sole exception are female unemployed under 25 years living in Western Germany (*u25/West*) who show all 21 considered periods $P_j$ with significant positive effects on taking up employment with supplementary welfare benefits ($S$) for sanctions imposed during the *first* stratum ($S_1$).

Rounding out the numerical results for the probability of transition from unemployment to employment presented in Table 4 and 5, we find distinct and persistent *positive* effects of welfare sanctions only for the probability of taking up employment with *supplementary welfare receipt* ($S$). However, the probability of entering *mere employment* ($O$), i.e. without needing top-up benefits, is either *not affected* significantly and persistently, or is mainly affected *negatively*. Whereas persistent and significant *positive* sanction effects on the probability to enter employment show up primarily for unemployed *women*, concerning exit into employment with top-up benefits ($S$), distinct *negative* sanction effects appear almost exclusively for unemployed *men*, concerning exit into mere employment

[85] The extent to which these results could be caused by a relatively high negative bias for the exit into job only ($O$) and a lower (or even no) bias for the exit into employment with supplementary welfare receipt ($S$) is discussed in Section 5.1.2 and Section 6, using the findings from the graphs with plotted ATT, which give us deeper insights.

($O$). As mentioned above, these results may be caused by a relatively high negative bias for the exit into job only ($O$) and a lower (or even no) bias for the exit into employment with supplementary welfare receipt ($S$). Methodological reasons for a tendentially lower negative bias in the case of exits into employment with supplementary welfare receipt ($S$) compared to entering mere employment ($O$) are outlined in Section 4.4.3; detailed discussions about the bias and the plausibility of these results are provided in Section 5.1.2 and Section 6, using the findings from the graphs with plotted ATT presented below.

Deeper insights into the results based on binary outcomes, namely the ATT on probabilities, are obtained from the graphs depicting the monthly updated ATT of welfare sanctions and its 90% confidence interval, which are included as figures in the appendix for selected groups and subgroups as well as for different exit events. Figures 1 to 15 show the monthly updated ATT (and its 90% confidence intervals) of benefit sanctions against unemployed UB-II-recipients on their probability to enter employment, separately for women (red) and men (blue), for the various subgroups, and for different categories of employment as exit events.[86] These graphs illustrate the development of the cumulative sanction effect on the probability to take up a job for overlapping periods, $P_j$, each starting with the beginning of the quarterly stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of consecutive final months $m_j$, representing the months after the beginning of stratum (with j=1–24).

According to the summarized information in Tables 4 and 5, Figure 1 for the third stratum ($S_3$) of *men* shows 20 significantly *negative* ATT for overlapping periods $P_j$ with final months $m_j$ after the stratum, that is for $P_4$–$P_{24}$ with $m_4$–$m_{24}$, regarding exit to *mere employment* ($O$), which means without receiving additional welfare benefits. At the same time, Figure 2 shows a significantly *positive* effect for *women* sanctioned in their third quart of welfare receipt ($S_3$) for all 21 periods $P_j$ with final months after the stratum ($P_4$–$P_{24}$).[87] An ATT of around 0.15, for example, as is depicted in Figure 2 for stratum $S_3$ of women, reveals a 15 percentage points higher average probability of taking up employment with supplementary UB-II-receipt ($S$) for unemployed who are sanctioned in the third quarter of welfare receipt compared to the case of not having been sanctioned. As shown, the plotted

[86] An overview of the figures with plotted ATT in the appendix is given in Table 15.

[87] Cases with confidence intervals (CI) that overlap the zero line do *not* reveal a *significant* treatment effect (ATT) on the significance level of $\alpha = 0.1$.

ATT, of course, contains the same information, just in another form compared to the tables with condensed results. But the graphs additionally depict the exact values of the ATT and — at least as importantly — show the development of the ATT over time for overlapping periods $P_j$, which could give us further insights, possibly also about the potential bias.

Indeed, it is salient that most of the cases with *persistent* negative significant ATT show strongly negative significant ATT already during the stratum, that is, for periods $P_1$–$P_3$ with final months $m_1$–$m_3$. This can be seen as a strong hint that, at least in several cases, we are confronted with a considerable bias to the detriment of the treated that might be high enough to possibly turn positive effects into negative results. Because we can see such cases with presumably strong negative distortions even more distinctly for the transition from unemployment to exiting welfare, we discuss this phenomenon in some detail in Section 5.1.2, and more comprehensively in Section 6.

As we calculate the ATT for *overlapping periods* $P_j$, the plots of the ATT for final months $m_1$–$m_{24}$ reveal the development of the *cumulative sanction effect over time*, specifically over two years after the beginning of the stratum. In the case of *increasing (decreasing)* ATT over time, the *monthly* sanction effect turns out to be *positive (negative)*. Hence, a nearly *monotonic negative slope* of the plotted ATT exposes that individuals with later exit events (after the stratum) are on average negatively affected by sanctions, independently of whether the total effect might be positive because of a possibly high enough positive effect shortly after the sanction. Put differently, even if the sanction effect is initially positive for people with a relatively high probability of getting a job, people with worse chances on the labor market are on average even less likely to become employed if they are sanctioned than if they would not have been sanctioned. In these cases with nearly monotonic negative slopes of cumulated ATT over a longer period, the *monthly* sanction effect in the medium and long run is negative, regardless of a possibly positive effect in the short run, or even in total.

Indeed, we find such shapes of *decreasing ATT* over a longer time frame (of at least more than one or even two years) for a number of cases, for example, for strata three and four ($S_3$, $S_4$) of unemployed men's exit into mere employment (Figure 1) as well as for strata $S_3$ and $S_4$ of the following subgroups: male unemployed living in Western Germany (*West*) (Figure 4), men with medium-level market access (*middle*) (Figure 7), and men aged over 25 years (*o25*) (Figure 11), each in terms of exit into job without needing top-up benefits ('*job only*' ($O$)). These graphs with almost monotonous negative slopes imply that the individual's chances of finding a job that brings them out of benefit receipt are, on average, negatively affected by sanctions, at least if they do not belong to the possibly positively affected people with exit events immediately, or short after the sanction.[88] Hence, no matter what the extent and direction of the short term effects, in these cases the *monthly* sanction effects are negative in the medium and long run.

Ordinary economic thinking, however, would lead us to expect a cumulated ATT that is positive and *monotonously increasing* with declining slope. Simplifying and neglecting other aspects, it could be assumed that sanctions must have a distinct positive effect on taking up employment as more pressure on unemployed people would increase their search effort, lower their reservation wages, and let them accept worse job conditions; all of this should lead to a higher probability of taking up a job. This assumed clearly positive effect of sanctions would be expected to be high at the beginning and should decrease with the passing of time, but never become less than zero; hence, the cumulated ATT should increase, or possibly stagnate after a while, but never decrease or even become negative.

Our results, indeed, show a few examples of graphs with plotted ATT which reveal nearly this kind of shape: starting with a positive ATT that is almost monotonously increasing with declining slope. This applies to the third stratum ($S_3$) of men regarding exit to employment with top-up benefits ($S$) (Figure 2), as well as for men in Western Germany (*West*) under the same conditions (Figure 5).

But more often, we find shapes with initially increasing, and later slightly decreasing, alternating, or stagnating ATT, which are still not too far away from the above mentioned ideal-typically expected form. Such patterns of plotted ATT for overlapping periods we find, for example, for stratum $S_3$ of unemployed women (*all*) (Figures 2 and 3), of female unemployed living in Western Germany (*West*) (Figures 5 and 6), of women with medium-level access to the labor market (*middle*) (Figures 8 and 9), and of women aged over 25 years (*o25*) (Figures 12 and 13), all in terms of exit to work with supplementary benefits ($S$) and to job in general ($G$). These cases are examples of unambiguously positive sanction effects already within the strata, despite the most likely considerable negative distortions during the stratum. In most of these cases, there is a noticeable shift upwards of around

[88] Even if our numeric results and the plots of ATT display negative ATT of sanctions for the main part of the overlapping periods $P_j$ — as is the case for many of the previously mentioned subgroups of unemployed men regarding exit into mere employment — we cannot exclude the possibility that the total sanction effect is nevertheless positive, because of a (possibly high enough) bias to the detriment of the treated that we cannot quantify.

10 percentage points immediately after the stratum, which is a strong hint of a negative bias of about 0.1 (see stratum $S_3$ of Figures 2, 3, 5, 6, 8, and 9). Hence, for example, unemployed women living in Western Germany who are sanctioned in their third quarter of welfare receipt are 35 percentage points more likely to take up employment with top-up benefits than if they would not have been sanctioned (see stratum $S_3$ of Figure 5).

We have shown that on the one hand, a number of (sub)groups and strata reveal shapes of cumulated ATT over time that are at least close to the shape that is expected by classical economic theory; on the other hand, there is, nevertheless, also a huge bundle of different patterns of plotted ATT that are partly far away from the expected one: there are n-shaped plots, as well as u-shaped, zigzag- or s-shaped plots with nearly monotonic negative slope as mentioned above. Obviously, there is a vast variety of shapes of *cumulated* ATT over time for diverse groups and different strata of welfare receipt, which implies that the *monthly* effects of welfare sanctions on the probability of entering employment are neither clearly positive, nor distinctly negative, but depend on the individuals and their situations concerning their labor market access, how long their welfare receipt already lasts, the timing of the sanction, and many other factors.

#### 5.1.2 Transition from unemployment to welfare exit

Beyond the aim to expedite the unemployed's taking up employment, another purpose of German welfare policy under 'Hartz IV' — mainly targeting fiscal objectives — is to speed up the unemployed's exit from welfare receipt *(ExWel: 'exit from welfare')*, here namely from UB-II-receipt. The main difference between transition to mere employment and transition to welfare exit is that the latter one also includes the option to leave the labor market, the so-called non-employment option.[89]

The effects of sanctions on the transition from unemployment to welfare exit ($ExWel$), measured in days of welfare duration, and presented in Table 6 for unemployed men who entered UB-II-receipt in 2008, go widely along with the results for transition to mere employment ($O$) shown in Table 3. The same subgroups of unemployed men who are sanctioned in their second stratum ($S_2$) of welfare receipt and show significant positive ATT on durations for exit into 'job only' ($O$) (Table 3) reveal significant

[89]For more details about the non-employment option, which means neither being (self-)employed nor receiving unemployment benefits, see Section 1 and Hillmann and Hohenleitner (2015).

positive ATT also in terms of duration until exiting welfare (Table 6). That means, sanctioned unemployed men need averagely more time than the matched unsanctioned until they get a job which is well-paid enough to bring them out of benefit receipt, or until they leave UB-II-receipt including the non-employment option. For example, male unemployed with medium level of labor market access (*middle*) need on average 108.64 days longer to leave welfare receipt if they are sanctioned in the second stratum than without being sanctioned. For women there is no significant ATT on durations within the reliable results for any of the four strata, as it is also the case for exit to employment (see Section 5.1.1).

It is remarkable that for each of these cases with significant positive ATT on durations in Table 3 *and* Table 6, the ATT for exiting welfare ($ExWel$) is larger than for entering mere employment ($O$). So the whole range of ages (15–56 years) (*all*) of unemployed men who are sanctioned in their second quarter of UB-II-receipt ($S_2$) need, on average, 94.71 days longer to leave welfare receipt (Table 6), but just 70.14 days longer to enter mere employment (Table 3), each compared to the matched unsanctioned

**Table 6:** Unemployed's welfare duration — men 2008

| Age | Region | Market access | ATT[1] S1 | S2 | S3 | S4 |
|---|---|---|---|---|---|---|
| all | | | 29.61 | 94.71*** | 126.59*** | 69.45** |
| all | West | | 43.34* | 81.79*** | 151.65*** | 98.17** |
| all | East | | -54.76 | 147.21** | 59.26 | 103.04 |
| all | | low | 114.98** | 13.91 | 22.38 | 177.87* |
| all | | middle | 24.15 | 108.64*** | 138.65*** | 45.51 |
| all | | high | -42.25 | 128.16 | 165.23 | n.r. |
| u25 | | | 87.02** | 118.65** | 132.95** | 66.62 |
| u25 | West | | 112.89** | 80.38 | 78.41 | -1.58 |
| o25 | | | 16.91 | 89.19*** | 139.58*** | 83.43** |
| o25 | West | | 27.97 | 83.33** | 157.07*** | 110.13*** |
| o25 | | low | 96.04* | 15.51 | 25.98 | 345.31*** |
| o25 | | middle | 12.88 | 110.25*** | 154.79*** | 57.46 |
| o25 | | high | -74.35* | 99.07 | 146.07 | n.r. |

[1]ATT of *metric outcomes:* difference of mean durations until exit from welfare receipt, measured in days from the beginning of quarterly stratum of welfare duration ($S_i$).
*Subgroups:* age-group in years: all=15–56, u25=15–24, o25=25–56; region: West/East German states; labor market access: low, middle, high.
*Significance levels:* $\alpha$=0.1*, $\alpha$=0.05**, $\alpha$=0.01***.
*Gray figures:* not reliable results (<50 treated cases); n.r.: no results.

control group. Hence, according to these numeric results, welfare sanctions seem to not only lower male unemployed's chances of entering mere employment, but also their probability of leaving welfare receipt via the non-employment option. However, it has to be emphasized that these results are most likely distorted by a bias to the detriment of the sanctioned, which tends to be relatively high for exit events which exclude continued welfare receipt, i.e. exit to mere employment ($O$) and exit from welfare ($ExWel$), as explained in Section 4.4.3, discussed in more detail, based on the plotted ATT, later in this section, and finally examined from a more comprehensive point of view in Section 6.

Analyzing unemployed's transition to welfare exit reveals more significant results than transitions into mere employment. A simple reason is that examining 'exit from welfare' ($ExWel$) produces (at least slightly) less right-censored spells than examining 'exit to job only' ($O$), because exiting welfare includes the non-employment option, and hence includes additional exit events. Thus, we get significant positive ATT for exit from UB-II-receipt ($ExWel$) not only for the second stratum ($S_2$) as for entering 'job only' ($O$) in Table 3, but for quite a few cases for each of the four examined strata ($S_1$–$S_4$).

One prominent result of Table 6 is the significant negative ATT for unemployed men aged over 25 years with a high degree of labor market access (*o25/high*), which is the only ATT on durations that we have got in the politically intended direction. People of this group who are sanctioned in their first quarter of welfare duration ($S_1$) leave UB-II-receipt on average 74.35 days earlier than members of the matched control group. For this subgroup of men with good chances on the labor market, and who sanctioned in an early state of their unemployment, an outstanding effect in the intended direction to shorten welfare receipt seems quite plausible.

The results for the binary outcomes with predominantly significant *negative* ATT *on probabilities* presented in Table 7 for men largely confirm the significant *positive* ATT *on durations* in Table 6. But there are also a few cases with significant *positive* ATT *on probabilities* in Table 7, mainly for men and with at least medium or even high level of labor market access (*middle*, *high*), and for men living in Eastern Germany (*East*), in all cases exclusively for sanctionees in the first stratum ($S_1$). For women, the results were predominantly not significant, except for a few cases in the first stratum Whereas women living in Eastern Germany (*East*) sanctioned in their first quarter of receiving welfare benefits

**Table 7:** Unemployed's exit from welfare — 2008

| | | | Men | | | | Women |
|---|---|---|---|---|---|---|---|
| Age | Region | Market access | Number of periods with sign. ATT[1] | | | | |
| | | | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_1$ |
| all | | | 2(+)/3(–) | 20(–) | 21(–) | 6(–) | n.s. |
| all | West | | 5(–) | 12(–) | 21(–) | 15(–) | n.s. |
| all | East | | 5(+) | 9(–) | | | 10(–) |
| all | | low | 12(–) | n.s. | | | 3(–) |
| all | | middle | 2(+) | 20(–) | 21(–) | 1(–) | n.s. |
| all | | high | 4(+) | 2(–) | | | 3(+) |
| u25 | | | 16(–) | 14(–) | 10(–) | | n.s. |
| u25 | West | | 18(–) | 2(–) | | | n.s. |
| o25 | | | 3(+)/2(–) | 13(–) | 21(–) | 17(–) | n.s. |
| o25 | West | | 3(+)/2(–) | 5(–) | 21(–) | 16(–) | n.s. |
| o25 | East | | 9(+) | 1(–) | | | |
| o25 | | low | 7(–) | n.s. | | | |
| o25 | | middle | 3(+) | 11(–) | 21(–) | 6(–) | 2(–) |
| o25 | | high | 8(+) | | | | |

[1] ATT of *binary outcomes:* difference of mean probabilities of exit from welfare receipt until the end of month $m_j$ after start of quarterly stratum of welfare duration ($S_i$); number of positive (+) and/or negative (–) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=4–24, or no significant (n.s.) ATT in any of these periods, on a significance level of $\alpha$=0.1.
*Subgroups:* age-group in years: all=15–56, u25=15–24, o25=25–56; region: West/East German states; labor market access: low, middle, high.

($S_1$) show relatively persistent *negative* ATT on probabilities,[90] women with early sanctions ($S_1$) and a high degree of labor market access show just a few overlapping periods $P_j$ with *positive* ATT on probabilities.

In order to get deeper insights into the results for the binary outcomes, we focus on the plotted ATT shown in Figures 16 to 27 of the appendix for the transition from unemployment to exit from welfare ($ExWel$).[91] As seen above for the transition into employment presented in Section 5.1.1, the graphs for exiting welfare also reveal a huge variety of diverse shapes of plotted ATT. And also similar to the results for employment entrance, there are several cases with indications of high negative distortions which emerged during the stratum, at least if the results are persistently negatively significant. Such

[90] As further discussed in Section 6, the relatively persistent negative ATT for the first stratum ($S_1$) of women living in Eastern Germany could be caused by a relatively high bias to the detriment of the treated because of a higher propensity to work of women in the Eastern part of Germany, and thus a higher probability for early exit events of the untreated.

[91] An overview of the figures with monthly updated ATT on probabilities in the appendix is given in Table 15.

patterns in the graphs we find mainly for transition into mere employment ($O$) and exiting welfare ($ExWel$). However, they occur more often and intensely for exit from welfare, as for example, in the second and third stratum ($S_2$, $S_3$) of men, presented in Figures 16, 17, 20, 22, 24, and 26 which, for male unemployed and several subgroups, show predominantly persistent negative significant (cumulated) ATT over the full considered period of two years from the beginning of the strata.

Nevertheless, there are also cases without stable negative significant results, which show similar patterns with clearly negative deviations during the stratum compared to the following periods $P_j$ with final months $m_j$ (immediately) after the stratum (i.e. with $j > 4$). Such kinds of patterns occur for men as well as for women, often with no or just a few significant results for periods with final months after the stratum. So the second stratum ($S_2$) of unemployed women (Figure 16), the second until forth stratum ($S_2$–$S_4$) of female unemployed in Western Germany (*West*) (Figure 17), and the second stratum ($S_2$) of women with medium-level labor market access (*middle*) (Figure 20), of women aged 25 years or older and living in Western Germany (*o25/West*) (Figure 25), and of women over 25 years with medium-level market access (*o25/middle*) (Figure 26) have notably lower ATT during the stratum compared to the following periods $P_j$ with $j > 4$ which mainly show insignificant results.

These graphs revealing strongly negative average treatment effects during the stratum — in most cases *significantly* negative with an ATT (on probabilities) of at least around minus 10 percentage points — provide severe indications of a negative bias, especially if the following ATT with final months immediately after the stratum are distinctively higher. Hence, such patterns of plotted ATT apparently confirm the suspicion, mentioned in Section 5.1.1 for exit to mere employment, that there are cases with negative distortions high enough to turn possibly positive effects into insignificant or even negative results. To put it differently, even if Table 7 shows mainly negative or insignificant results, this could be caused by a negative bias that emerged during the strata, which possibly turns strong positive effects to insignificant results, and weak positive or insignificant effects to negative results.

But even if the total effect of welfare sanctions should mainly be positive or at least insignificant, there are also people being confronted with a negative impact of sanctions. Again similarly to Section 5.1.1, there are several graphs of plotted ATT over time that show remarkably strong negative slopes for periods $P_j$ with final months $m_j$ after the stratum (i.e. for $j > 4$), which indicates notably negative *monthly* sanction effects for people with later exit events. These people can be assumed to have tendentially worse chances on the labor market, or have other reasons for being prone to a later exit from welfare. And obviously those people are (additionally) negatively affected by welfare sanctions, even if the majority would be positively affected to an extent that the total effect might be positive.

Examples of initially strongly increasing and partly even significant positive treatment effects (mainly within the stratum) and subsequently strongly decreasing cumulated ATT are the first stratum ($S_1$) of the following subgroups of male unemployed: men living in Eastern Germany (*East*) (Figure 18), men with mid- or even high-level of labor market access (*middle*, *high*) (Figures 20 and 21), male unemployed aged 25 or over (*o25*) (Figure 24), men over 25 years and living in West Germany (*o25/West*) (Figure 25), and men over 25 with medium- or high-level labor market access (*o25/middle*, *o25/high*) (Figures 26 and 27). On this, the shapes of plotted ATT for the first stratum ($S_1$) of men with mid- or even high level of labor market access (*middle*, *high*) are particularly striking with a very abrupt turn from a steep positive slope within the stratum to a steep negative slope afterwards. Hence, there are subgroups of unemployed people which are quite heterogeneous concerning their reaction to sanctions. In these cases, people with already high chances for an early exit event get pushed by sanctions and are more likely to leave welfare receipt. But people with tendentially later exit events get even worse chances for an early exit from welfare if they are sanctioned. Put differently, these are examples of strongly positive short-term effects of benefit sanctions on the one hand, and strongly negative sanction effects in the medium and long run, on the other hand.

Analogically to the graphs of cumulated ATT over time for employment entrance presented in Section 5.1.1, the shapes of plotted ATT for exiting welfare are likewise very heterogeneous. We again find, for example, u-, n-, s- and zigzag-shaped graphs. So we cannot make one clear statement about the impact of welfare sanctions on leaving UB-II-receipt for different groups and subgroups of unemployed. Instead, the results show us how diverse people react on sanctions, which can lead them to earlier or later exits from welfare, depending on the specific conditions and circumstances.

### 5.2 Employed people with supplementary welfare receipt

Despite the term '*unemployment* benefit II', it is not just unemployed, but also *employed* people who are eligible for the German UB II, if their household's income doesn't meet the legally defined minimum subsistence level. The share of employed people receiving supplementary UB II — the so-called *'Aufstocker'* — grew from about 23% of the employable UB-II-recipients in 2007, to around 30% in 2012 this number stabilized, and even decreased slightly from around 1.3 to 1.2 million people between 2007 and 2012 and still remains significantly above one million people. Thus, this is a notable and important group that cannot be neglected in a comprehensive analysis of welfare sanctions.

For employed welfare recipients, we consider the following exit events: entering mere employment *(O: 'job only')*, i.e. without receiving top-up benefits, presented in Section 5.2.1; and exiting the receipt of welfare benefits *(ExWel: 'exit from welfare')*, comprising the alternative to leave the labor force, shown in Section 5.2.2, and additionally, the option of exiting employment with top-up benefits for mere welfare receipt *(ExJob: 'exit from job')* discussed in Section 5.2.3.

#### 5.2.1 Transition from supplementary welfare receipt to mere employment

The aim of German welfare policy regarding employed people needing top-up benefits to cover their household's subsistence minimum is to increase the so-called *Aufstocker's* family income in order to totally leave welfare receipt. Beyond possibly increasing the income of a related household member, the employed welfare recipients themselves could either try to increase their working hours, negotiate with the current employer to raise their salary, or find a better paid job — each in order to enhance the household's income sufficiently to leave welfare receipt. Even if sanctions against employed welfare recipients may not be primarily intended to foster this aim, it could nevertheless be seen as a welcome side-effect of the sanction policy of German Jobcenters. And based on the previous studies of benefit and welfare sanctions which found such kinds of 'pushing effects' for sanctioned unemployed, it seems natural to suppose similar impacts for employed welfare recipients.

Our estimations of the average treatment effects on the treated (ATT) of welfare sanctions on the probabilities of transitions from employment with supplementary welfare receipt to mere employment presented in Table 8, however, show either no significant ATT or — mostly even persistent — negative sanction effects, especially for the first two strata ($S_1$, $S_2$) of employed men.[92] Additionally, taking into account the plotted ATT over two years time, as shown in Figure 28 until Figure 30, we get deeper information about the possible bias to the detriment of the treated; furthermore, we see the development of the effects of benefit sanctions against employed welfare recipients, revealing long-term effects.

For example, the first and second stratum ($S_1$, $S_2$) of *all* employed men receiving top-up benefits (Figure 28), of employed men living in *Western* Germany (Figure 29), and of male employed with *medium*-level labor market access (Figure 30) reveal negative ATT on probabilities of around -0.05 to -0.10 within the strata, and steeply downwards pointing shapes of monthly updated ATT for at least one year after the strata. On the one hand, the already considerable negative ATT within the strata can be seen as a strong indication of a negative bias; on the other hand, the ongoing strong negative progression of the *cumulated* ATT reveals a remarkable negative *monthly* ATT during about one year after the sanction. This means that even if the possible distortion during the strata is high enough to turn insignificant or positive short-term effects into negative ATT, there are considerable negative effects of welfare sanctions in the medium and long run — at least in some cases, and for individuals with tendentially later exit events, and thus with already worse chances of obtaining a job which gets

**Table 8:** Top-up benefits to mere employment — 2008

| | | | Men | | Women |
|---|---|---|---|---|---|
| Age | Region | Market access | Number of periods with sign. ATT[1] | | |
| | | | $S_1$ | $S_2$ | $S_1$ |
| all | | | 21(−) | 21(−) | 14(−) |
| all | West | | 21(−) | 21(−) | |
| all | | middle | 21(−) | 19(−) | |
| all | | high | 8(−) | | |
| o25 | | | 21(−) | 21(−) | |
| o25 | West | | 20(−) | | |

[1]ATT of *binary outcomes:* difference of mean probabilities of exit to mere job until the end of month $m_j$ after start of quarterly stratum of welfare duration ($S_i$); number of positive (+) and/or negative (−) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=4–24, or no significant (n.s.) ATT in any of these periods, on a significance level of $\alpha$=0.1.
*Subgroups:* age-group in years: all=15–56, o25=25–56; region: West/East German states; labor market access: low, middle, high.

[92]Our estimations of the ATT on *durations* until the exit event, for the transition from employment with receiving top-up benefits *('supplementary')* ($S$) to mere employment *('job only')* ($O$), yield no reliable results because of too few exit events in the treatment group.

them out of needing supplementary welfare benefits.

Altogether, the plots for *employed* welfare recipients' exit to employment without benefit receipt are much more homogeneous than the very heterogeneous shapes of ATT plots for *unemployed* people's exit to mere employment. On the whole, the plots of employed's cumulated ATT are mostly downwards heading and show either insignificant or significant negative effects of benefit sanctions against female and male employed welfare recipients. Thus, even if the short term effects might be non-negative, the monthly effects in the medium and long run tend to be clearly negative, in most cases for at least one year after the strata.

### 5.2.2 Transition from supplementary welfare receipt to welfare exit

As already mentioned in Section 5.1.2, the exit event 'leaving welfare receipt' ($ExWel$) comprises 'taking up mere employment' (*'job only'*) ($O$) as well as the possibility of leaving the labor force, the so-called 'non-employment option'. Consequently, there are more exit events in the analysis than if we only consider entering mere employment, which is an advantage of estimating the ATT of duration outcomes. In contrast to the analysis of the durations until entering mere employment presented in Section 5.2.1, where we have too few cases with exit events in the treatment group in order to get reliable results, the estimations for exiting welfare ($ExWel$) yield several reliable, and even a few significant results also for duration outcomes which we present in Table 9 for male employed receiving supplementary welfare benefits.

**Table 9:** Duration of top-up benefits until exit from welfare — men 2008

| Age | Region | Market access | ATT[1] $S_1$ | ATT[1] $S_2$ |
|---|---|---|---|---|
| all | | | 67.62* | 102.75** |
| all | West | | 43.17 | 151.95** |
| all | | middle | 92.44* | 114.95* |
| o25 | | | 32.00 | 158.64** |
| o25 | West | | 7.93 | 182.02* |

[1]ATT of *metric outcomes:* difference of mean durations of (supplementary) welfare receipt until welfare exit, measured in days from the start of quarterly stratum of welfare duration ($S_i$).
*Subgroups:* age-group in years: all=15–56, u25=15–24, o25=25–56; region: West/East German states; labor market access: low, middle, high.
*Significance levels:* $\alpha$=0.1*, $\alpha$=0.05**, $\alpha$=0.01***.
*Gray figures:* not reliable results (<50 treated cases).

According to the results in Table 9, employed men receiving top-up benefits who are sanctioned in their first quarter of welfare receipt need more than two months (namely 67.62 days) longer until they leave welfare receipt than if they would not have been sanctioned. These results, being in contrast to the expected direction of sanction effects, may at least be partly due to a potential negative bias to the detriment of the treated which we cannot quantify, as mentioned on several occasions before. Hence, we focus the main part of our interpretation on the probability outcomes, from which we expect to derive further insights.

Table 10 shows a few estimations with positive ATT on probabilities, almost exclusively for the first stratum ($S_1$) of employed men living in Western Germany and aged over 25 years (*o25/ West*). The second strata ($S_2$) of several subgroups, such as men over 25 years (*o25*) or men with medium-level labor market access (*middle*), are dominated by negative ATT. Clues as to the extent to which these persistent negative results could be caused by a bias to the detriment of the sanctioned were obtain from the plotted ATT. Indeed, Figures 31 and 32 reveal strongly negative ATT — in many cases as early as the second and third strata ($S_2$, $S_3$) — of about 10 until 20 percentage points, while in a few cases an upwards shift shortly after the strata is an additional indication of a probable bias of around -0.1 to -0.2.

The fact that we do not get reliable, or even significant results for other subgroups of employed men with top-up benefits, like men under 25 years (*u25*) or men living in Eastern Germany (*East*), is most probably caused by a lack of exit events in the treatment group, especially as these subgroups

**Table 10:** Employed's exit from welfare — men 2008

| Age | Region | Market access | Number of periods with sign. ATT[1] S$_1$ | S$_2$ | S$_3$ |
|---|---|---|---|---|---|
| all | | | n.s. | 17(–) | 3(–) |
| all | West | | 1(+) | | 3(–) |
| all | | middle | 1(–) | 20(–) | |
| o25 | | | n.s. | 21(–) | |
| o25 | West | | 8(+) | 21(–) | |

[1]ATT of *binary outcomes:* difference of mean probabilities of exit from welfare receipt until the end of month $m_j$ after start of quarterly stratum of welfare duration ($S_i$); number of positive (+) and/or negative (–) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=4–24, or no significant (n.s.) ATT in any of these periods, on a significance level of $\alpha$=0.1.
*Subgroups:* age-group in years: all=15–56, o25=25–56; region: West/ East German states; labor market access: low, middle, high.

in general consist of notably fewer individuals. For women, a lower sanction rate can be an additional important reason for the scarcity of reliable results, at least if there is a low number of observations anyway, as is the case for transition from supplementary welfare receipt.

Comparing Tables 8 and 10, it is salient that exiting welfare (Table 10) leads to less significant negative results and even a few significant positive ATT, while entering mere employment (Table 8) shows just negatively significant ATT. Such a discrepancy between the two exit events must be due to the non-employment option which obviously tend to be affected positively by welfare sanctions. In other words, benefit sanctions against employed welfare recipients tendentially raises their probability to leave the labor market. That is valid at least for the subgroups and strata which show the mentioned discrepancy between the events 'welfare exit' with less negative or even positive ATT, and 'exit to mere employment' with clearly negative estimations of the sanction effect.

### 5.2.3 Transition from supplementary to mere welfare receipt

Examining the effects of benefit sanctions on transitions from employment with supplementary welfare receipt, there is one further possible event that has to be taken into account: the option to exit employment (*ExJob*) in order to live on welfare benefits exclusively.

Looking at the purely numerical results of our analysis of sanction effects on duration outcomes presented in Table 11, we see a strongly significant negative effect on durations until exiting employment for the first stratum ($S_1$) of men receiving top-up benefits. More precisely, male employed

**Table 11:** Duration of top-up benefits until mere welfare receipt — men 2008

| Age | Region | Market access | ATT[1] $S_1$ |
|---|---|---|---|
| all | | | -88.13*** |
| all | West | | -90.42*** |
| all | | middle | -66.13* |
| o25 | | | -100.23*** |
| o25 | West | | -107.08*** |

[1]ATT of *metric outcomes:* difference of mean durations of supplementary until mere welfare receipt, measured in days from the beginning of quarterly stratum of welfare duration ($S_i$).
*Subgroups:* age-group in years: all=15–56, o25=25–56; region: West/East German states; labor market access: low, middle, high.
*Significance levels:* $\alpha$=0.1*, $\alpha$=0.05**, $\alpha$=0.01***.
*Gray figures:* not reliable results (<50 treated cases).

receiving supplementary unemployment benefits II (UB II), who are sanctioned in their first quarter of welfare receipt leave their job on average nearly three months (exactly 88.13 days) earlier than if they would not have been sanctioned. However, in contrast to previously discussed exit events, for exiting employment (*ExJob*) we face a potential bias not to the detriment, but in *favor* of the treated — in terms of a higher probability of the exit event, and thus of a shorter duration until the event occurs. The monthly updated ATT on probabilities, presented in Table 12 and plotted in Figures 33 to 35 shall give us further hints on whether, and to what extent this could be the case.

We see in Table 12 that the first stratum ($S_1$) of employed women and men, as well as of various subgroups of employed men, reveals persistent significant positive sanction effects on the probability of leaving employment for mere welfare receipt (*ExJob*). Even the relatively small subgroup of employed men with a high level of labor market access (*high*) shows a quite large number of overlapping periods $P_j$ with positive ATT on the probability of quitting their job and living only on welfare. Nevertheless, as outlined in Section 4.4.3, the possibility of a positive bias must be taken into account.

Indeed, there are a few shapes of plotted ATT over time for certain strata and (sub-)groups, as shown in Figures 33 through 35, which provide indications of positive distortions, at least in some cases. For example, the first stratum ($S_1$) of employed women depicted in Figure 33 shows a steeply upwards pointing shape of monthly updated *cumulated* ATT during the stratum that is staying quite constant around approximately 15 percentage points afterwards. This is a strong indication of a positive bias of

**Table 12:** Employed's exit to mere welfare — 2008

| | | | Men | Women |
|---|---|---|---|---|
| Age | Region | Market access | Number of periods with sign. ATT[1] | |
| | | | $S_1$ | $S_1$ |
| all | | | 21(+) | 21(+) |
| all | West | | 21(+) | |
| all | | middle | 21(+) | |
| all | | high | 10(+) | |
| o25 | | | 21(+) | |
| o25 | West | | 21(+) | |

[1]ATT of *binary outcomes:* difference of mean probabilities of exit to mere welfare receipt until the end of month $m_j$ after start of quarterly stratum of welfare duration ($S_i$); number of positive (+) and/or negative (–) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=4–24, or no significant (n.s.) ATT in any of these periods, on a significance level of $\alpha$=0.1.
*Subgroups:* age-group in years: all=15–56, o25=25–56; region: West/East German states; labor market access: low, middle, high.

around +0.15, because there is no plausible reason for a factual, strong positive *monthly* effect during the stratum which ends abruptly directly afterwards — not least because sanctions are usually spread randomly within the strata and are not imposed only at the beginning and middle of the strata.[93]

Similar to the women, employed men (Figure 33), men with a high level of labor market access (*high*) (Figure 34), and men over 25 years living in Western Germany (*o25/West*) (Figure 35) exhibit relatively high positive ATT during their first stratum ($S_1$) which are mainly upwards heading quite steeply. Even though the initial trend continues pointing strongly upwards for three further months subsequent to the stratum, lasting until period $P_6$ with final month $m_6$, the plotted ATT subsequently turns direction and tends to head downwards. The interpretation of such shape of cumulated ATT — in contrast to the straightforward example of female employed — is rather ambiguous.

The monthly updated cumulated ATT for the mentioned groups of men is already quite high in the first month after the beginning of stratum $S_1$, that is, in period $P_1$ with final month $m_1$. They start with an ATT of more than 20 (Figure 33), more than 10 (Figure 34), and more than 15 percentage points (Figure 35), respectively. Additionally, the ATT mostly increases distinctly in the following two periods ($P_2$, $P_3$) with final months $m_2$ and $m_3$ within the stratum. Such kinds of patterns during the stratum could give a good reason for suspecting a positive bias, at least if the ATT would stay constant or point downwards directly after the end of the stratum.

However, the upwards pointing trend of the cumulated ATT exceeding the duration of the stratum for about three further months implies that the *monthly* ATT is still positive during months $m_4$ to $m_6$. Such positive monthly effects after the stratum, though, cannot be caused by the potential positive bias which could emerge exclusively during the strata.[94] Accordingly, we have an indication of factual positive effects of welfare sanctions against the mentioned groups of employed men — male employed of any age (*all*: 15–56 years), men with a high level of labor market access (*high*), and men over 25 years living in Western Germany (*o25/West*) — sanctioned in their first quarter of welfare

[93] Although the distribution of welfare sanctions in the first stratum ($S_1$) often tends to be skewed to the left and thus is less uniform than within later strata, the hypothetical possibility that sanctions could be imposed exclusively at the beginning and middle of the first stratum is neither likely nor plausible. Otherwise it could have been an explanation for such a pattern of plotted ATT on condition that the welfare sanctions have only very short-term effects not lasting longer than one month maximum which is also very unlikely, given that sanctions regularly last three months.

[94] As a consequence of how the treatment and the outcome variables are generated (see Sections 3.2 and 3.3), and as explained in Section 4.4.3, in contrast to the *negative* potential bias on probability outcomes of other exit events, the *positive* potential bias for transitions from supplementary to mere welfare receipt can emerge only *during* the strata, and not also shortly afterwards.

receipt ($S_1$) which increases their probability to give up employment for mere welfare receipt ($ExJob$). Later on — after about six months of increasing ATT — the curves change direction, mainly heading downwards. This implies negative *monthly* effects of welfare sanctions in the medium and long run for the aforementioned groups of employed men.

Altogether, such kinds of cumulated ATT — increasing for the first six months and decreasing afterwards — that we found for the first stratum of several groups, imply that employed welfare recipients who tend to quit or lose their job and live just on welfare already during their first half-year of benefit receipt would be more likely to experience this if they are sanctioned. In contrast, people who tend to exit their jobs later than half a year after starting to get top-up benefits are even less likely to exit from employment if they are sanctioned.

However, these effects predominantly tend to be not significant, at least with our data set based on a 2% sample of administrative data which, in cases of small treatment groups, tend to suffer from a scarcity of rare exit events. Taking a markedly larger sample, or using the administrative data as a whole, could lead to more significant results also for rare exit events and small treatment groups, as is the case with transitions from employment with supplementary welfare receipt into mere welfare receipt.

### 5.3 Indirect sanctions

In a comprehensive analysis of benefit sanctions' impact on employment and welfare receipt, the effect on indirect sanctioned — that is, people who are affected by sanctions against related household members — is a further important research topic. This is particularly the case if families or households are entitled to the benefits rather than individuals, which in many European countries applies for tax-based welfare payments in contrast to unemployment insurance payments. Likewise, the tax-funded welfare benefits for employable people in Germany, named unemployment benefits II (UB II), are granted to needy workers and job-seekers along with their related household members; in contrast, the insurance based unemployment benefits I (UB I) are granted on an individual basis and independently of need.

Although a vast majority of unemployed people in Germany do not receive unemployment insurance benefits (UB I) but welfare payments (UB II), and although a big share of the employable welfare recipients do not live alone but together with related household members, the question of how sanctions influence the employment-related behavior and labor force decisions of indirectly affected household members is an issue almost entirely neglected by previous economic research.[95] In order to account for the importance of this topic, we carried out the whole analysis for *direct sanctions*, referring to sanctions against the employable individuals themselves, as well as for *indirect sanctions*, referring to sanctions against household members of the employable welfare recipients.

Nonetheless, the outcome of the analysis based on indirect sanctions is relatively scarce, primarily due to a lack of cases in the treatment group. The main reason for this shortage is that we have to impose strong restrictions in order to identify individuals who are suitable for the treatment group of the indirectly sanctioned. On the assumption that the effect of direct sanctions exceeds the effect of indirect sanctions, we can only use people for the treatment group of indirect sanctioned if they have never been directly sanctioned either before, or after the indirect sanction until the exit event occurs.[96] Moreover, we had to broaden the strata to six months in order to get a notable amount of treatments per stratum. This, however, entails other drawbacks like a longer period that cannot properly be used to reveal short-term effects. Notwithstanding the difficulties, we got some reliable and significant results that we present and discuss in Section 5.3.1 for unemployed, and in Section 5.3.2 for employed welfare recipients who are affected by sanctions exclusively against related household members.

### 5.3.1 Impact of indirect sanctions on transition from unemployment

Estimating the effects of indirect sanctions on the transition from unemployment, we got reliable and significant results only for women and their probabilities to enter employment; for men, for exits from welfare (*ExWel*) comprising the non-employment option, and for duration outcomes, we did not get reliable and significant results.

The estimation results presented in Table 13 reveal 10 of 12 considered overlapping periods $P_j$ with significant negative ATT for entering mere employment ($O$: *'only job'*), and 7 of 12 considered periods

[95] As mentioned in the introduction, Section 1, we already considered indirect sanctions in our previous study on the impact of welfare sanctions on labor market decisions based on survey data (Hillmann and Hohenleitner (2015)), but we could not disentangle the effects of direct and indirect sanctions because of far too few cases of indirect sanctioned in the survey data.

[96] See Section 3.2. For individuals who are directly sanctioned, in contrast, later indirect sanctions or repeated direct sanctions are no obstacles to stay in the treatment group — we just have to make sure that we only use the first sanction of an individual for reasons of comparability.

with significant negative ATT for taking up any kind of employment ($G$: *'job in general'*).[97] This applies for unemployed women living in Western Germany (*West*), sanctioned in their first half-yearly stratum of welfare receipt ($S_1$). By contrast, unemployed men do not reveal significant effects of welfare sanctions against related household members. The plotted ATT, exemplarily presented in Figure 36 for transition into mere employment ($O$) and Figure 37 for entering employment in general ($G$), each for unemployed living in Western Germany (*West*), confirm that only women react to family member's sanctions. That especially women in Western Germany react significantly to other household member's sanctions might be due to a historically conditioned and still existing lower propensity to participate in the labor market when compared to men, and to women in Eastern Germany.

The negative direction of the effect, however, seems less plausible and may be caused by a negative distortion to the detriment of the treated. On the one hand, the graphs for entering mere employment (Figure 36) show downwards heading cumulated ATT for women, wherein the first few months tend to be increasingly significant also because of very narrow confidence intervals, later changing abruptly to wider confidence intervals which let the ATT initially become insignificant. This pattern of plotted ATT and its confidence intervals speaks for a possibly negative bias. However, the ATT in all cases and strata, for women and men, starts close to zero in the first month. Nevertheless, for women the cumulated ATT is slowly but noticeably downwards heading during the first few overlapping periods, predominantly for exit into mere employment ($O$). These facts together can be seen as indications of

**Table 13:** Unemployed to employment — indirect sanctions

| | | | Women 2008 |
|---|---|---|---|
| Age | Region | Exit to job (G/O/S)[2] | Number of periods with sign. ATT[1] $S_1$ |
| all | West | G | 7(–) |
| all | West | O | 10(–) |
| all | West | S | n.s. |

[1]ATT of *binary outcomes:* difference of mean probabilities of employment entrance until the end of month $m_j$ after start of half-yearly stratum of welfare duration ($S_i$); number of positive (+) and/or negative (–) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=7–18, or no significant (n.s.) ATT in any of these periods, on a significance level of $\alpha$=0.1.
[2]*Events*: exit to: job only (O), i.e. without welfare receipt, job with supplementary welfare receipt (S), job in general (G), i.e. comprising (O) and (S).

[97]For indirect sanctions, we carried out the analysis for two half-yearly strata ($S_1$, $S_2$), each for 18 overlapping periods $P_j$ with final months $m_j$ with $j = 1, 2, .., 18$. Because of a potential bias emerging predominantly during the strata, we consider only the periods $P_j$ with final months $m_j$ after the end of the half-yearly strata, i.e. with $j = 7, 8, .., 18$.

a notable but small negative bias in the first few months.

On the other hand, the downwards trend of the cumulated ATT continues for the whole observation period of 18 months for the first stratum of women in Western Germany (*West*) and their exit into employment without receiving top-up benefits ($O$) (Figure 36). Even if the emergence of a negative bias exceeds the duration of the stratum, this cannot explain the clearly long-term downwards trend of the cumulated ATT, as the emergence of the bias after the stratum is diminishing with progressing final months. Hence, there indeed seems to be a negative *monthly* effect of indirect sanctions against women in Western Germany, at least in the medium and long run.

Yet, at first sight, a negative effect of indirect sanctions seems less plausible if we assume that the partner of the woman is sanctioned. A sanction against the partner might be more likely a reason for increasing the effort to find a job in order to compensate for the financial loss. But a high proportion of women receiving UB II are single mothers living with their children. And as people aged under 25 years generally are not allowed to leave their parents' household if they cannot fully live on their own income[98] — which in Germany is still more likely for children whose parents are already depending on welfare benefits — it is plausible that a considerable share of sanctioned household members of women are not their partners but their children. This holds even more as young people under 25 years are sanctioned considerably more frequently than people aged 25 years or older.

If the negative effect of indirect sanctions on women's probability to take up mere employment in the medium and long run is mainly due to sanctioned children, the negative direction of the impact seems more plausible than in the case of sanctioned partners. A reason for this is that people under 25 are not only sanctioned more frequently, but also more severely. If their fault is not just having been unpunctual or having missed an appointment which is punished with a 10% cut of the base benefit, their so-called 'major breach of duty' is initially sanctioned with a 100% cut of the base benefit. And in the first case of recurrence they already get a 100% cut of the whole UB II including accommodation costs, where each sanction generally lasts three months, regardless of the amount of benefit cut.[99] As a consequence, mothers — and single mothers even more — may have less time and energy to find a new

[98] 'Own income', in this case, includes all income or assets that brings them out of welfare receipt in terms of UB II. Hence, receiving students assistance wouldn't be an obstacle to move out of the parents' household.

[99] People aged 25 years or over are sanctioned for the same 'major faults' with a 30% cut of the base benefit initially, and in case of recurrence with a 60% cut of the base benefit in the second step, and with a 100% cut including costs for accommodation and mostly even health insurance only in the third step.

job if they have to struggle with their sanctioned children and with the huge loss of household's income over three months. Furthermore, it also seems plausible that the negative effects last considerably longer than just the three months time during the sanction, because compensating a 100% benefit cut in the household may cause debts of a size that must be repaid over many months, given that the family is living at the bare subsistence level anyway, even without being sanctioned or repaying debts.

All in all, the result that women's probability of finding a job which brings them out of benefit receipt is affected negatively even in the long run, seems more plausible if they are affected by sanctioned children than by sanctioned partners. Firstly the loss of household's income is more severe, secondly the threat of repeated sanctions is larger, and finally they might feel more responsible for their children's future than for a partner's behaviour and thus — instead of increasing the effort to find a job for themselves — possibly increase the effort to support their children to get a job or an apprenticeship.

As we can not differentiate between sanctions against partners, children, or even parents, because of too few cases of merely indirect sanctioned in general, we just give a first glimpse showing that there is much more to explore in this research field. But our data set, based on a 2% sample of administrative data, is too limited for more detailed research on this topic.

### 5.3.2 Impact of indirect sanctions on transition from supplementary welfare receipt

Analyzing the effects of indirect sanctions on the transition from employment with supplementary benefit receipt, we got reliable and significant results only for women and their probability of leaving welfare receipt (*ExWel*). For men, for exits to employment ($G/O/S$), for exits from employment to mere welfare receipt (*ExJob*), and for duration outcomes, we did not get reliable and significant results.

The estimation results presented in Table 14 show significant negative ATT on the probabilities of exiting welfare (*ExWel*), only for two overlapping periods $P_j$ with final months $m_j$ after the end of the first half-yearly stratum ($S_1$), i.e. for $P_j$ with $j > 6$. This applies for employed women receiving top-up benefits who are affected by sanctions exclusively against related household members. As discussed in detail in Section 5.3.1 above, it is more likely that negative effects of indirect sanctions are caused by sanctioned children than by sanctioned partners. It seems plausible to transfer those findings about unemployed women to employed women with supplementary welfare receipt as well— at least for their entry to mere employment which is also a part of exiting welfare.

It is striking that we get insignificant ATT for taking up mere employment ($O$) despite the negative significant ATT on the probabilities of exiting welfare ($ExWel$). Even if the insignificant results for entering employment are caused by too few exit events in the treatment group, the number of exit events for leaving welfare is obviously high enough to obtain at least a few significant results. As the only difference between entering mere employment and leaving welfare is the non-employment option, the significant negative ATT must be caused by the additional exit events of leaving the labor market. Hence, employed women with top-up welfare benefits are less likely to leave the labor market if they are affected by a sanction against a related household member in their first half-yearly stratum of benefit receipt ($S_1$). This seems plausible independent of whether the sanctioned family member is a child or a partner or even a parent, as the family's income is reduced by the sanction, making it is less affordable to leave the labor force.

As mentioned above, for the analysis of the effects of indirect sanctions, a much larger sample than our data set based on a 2% sample of the administrative data is necessary in order to get more detailed results in this research field.

### 5.4 Matching quality and robustness

In order to evaluate the reliability of our results, we checked the matching quality and the robustness of our estimations in several kinds of ways.

**Table 14:** Supplementary welfare receipt — indirect sanctions

| | | Women 2008 |
|---|---|---|
| Age | Exit from/to | Number of periods with sign. ATT[1] |
| | | $S_1$ |
| all | from welfare (ExWel) | 2(−) |
| all | to job only (O) | n.s. |
| all | to welfare only (ExJob) | n.s. |

[1]ATT of *binary outcomes:* difference of mean probabilities of the exit event until the end of month $m_j$ after start of half-yearly stratum of welfare duration ($S_i$); number of positive (+) and/or negative (–) significant ATT, summarized over periods $P_j$ with final months $m_j$, considering j=7–18, or no significant (n.s.) ATT in any of these periods, on a significance level of $\alpha$=0.1.

### 5.4.1 Matching quality

After the matching process, the quality of the matching has to be examined. Specifically, we have to figure out whether the estimated propensity score is appropriate to balance the covariates between treatment and control group. If that is the case, the so-called 'balancing property' is nearly satisfied. This means that individuals with the same propensity score have almost identical distributions of observed and unobserved characteristics, independently of being a member of the treatment or control group; in other words, the selection to treatment or control group can be assumed to be random, and hence treated and untreated on an average can be assumed as nearly identical.[100]

As suggested by Müller (2012), we explored the matching quality in two ways. Firstly, we checked whether significant differences in the covariates between the treated and the untreated are still significant after the matching procedure. Following Rosenbaum and Rubin (1985), we applied two-sided $t$-tests to ensure that no significant mean differences between the treatment and control groups occurred after the matching procedure for all covariates.

Following Sianesi (2004) and Müller (2012), we additionally checked the pseudo $R$-squared (pseudo $R^2$), and applied the likelihood ratio chi-square test (LR-$\chi^2$-test), both before and after the matching procedure to ensure a high matching quality. As the pseudo $R^2$ is a measure of the heterogeneity between treatment and control group, it should be very low, preferably close to zero, after the matching and clearly lower than before the matching. Conversely, the LR-$\chi^2$-test which checks whether at least one single covariate has a significant impact on the probability to be treated, must not be significant *after* the matching procedure in order to guarantee a good balance between matching and control group. To put it differently, the null hypothesis ($H_0$) that the common effect of the covariates on the treatment is zero must not be rejected after the matching, i.e. at least $p > 0.1$ must be satisfied; for a high matching quality, however, the $p$-value should be close to 1. If these requirements are met, the 'balancing property' is approximately satisfied.[101]

In the vast majority of our estimations, the balancing property was clearly satisfied. For samples with a lack of observations in the treatment group, however, there were estimations with worse matching quality which we excluded from the further analysis. Concretely, we checked for pseudo

[100]See Becker and Ichino (2002) and Müller (2012).
[101]See Müller (2012) and Heinrich et al. (2010).

$R^2 < 0.1$, $p$-value of LR-$\chi^2$ test close to 1, i.e. LR-$\chi^2$-test must be (highly) insignificant, and finally, for all covariates, the mean differences between treatment and control group must not be significant *after* matching. Additionally, as a rule we rejected estimations with less than 50 treated cases, and in just a few exceptional cases we used estimations with less than 50 but at least 25 treated individuals if the balancing property was clearly satisfied.[102]

### 5.4.2 Robustness

In order to check the estimated ATT for its sensitivity to the matching algorithm, we carried out the entire analyses with two different matching procedures: firstly, *'nearest neighbor matching'* (NNM) with $k$=5 nearest neighbors (5-NNM) and a caliper of 0.01 in order to exclude bad matches, and secondly, *'kernel matching'* (KM), specifically kernel matching using an *'Epanechnikov kernel'* (EKM) with a bandwidth of 0.06, with and without bootstrapped standard errors.[103] Additionally, we checked out several variations of both of these algorithms. For example we also tried out nearest neighbor matching with $k$=3 nearest neighbors (3-NNM) and varied the caliper of NNM and the bandwidth of EKM. The results are very robust against these variations. Amongst 3-NNM and 5-NNM, there were nearly no differences, and between 5-NNM and EKM there were only small differences which, in the overwhelming majority of cases didn't change the significance levels of the ATT. Furthermore, we made spot checks with a special self-written adjustment procedure which is only feasible for NNM that excludes the problem of the bias caused by the stratification, but yields other drawbacks which prevented us from using it more widely. Even those checks generally brought out astonishingly low differences in the results of using 5-NNM with and without adjustment. Hence, on the whole, our findings seem to be very robust against variations of the matching algorithm.

Additionally, we did the whole analysis — including the variants of the matching algorithms —

[102]It has to be stressed that even in cases with just a few observations in the treatment group ($25 \leq n_t < 50$), there are still more than 1,000 individuals in the control group, because benefit sanctions are rare events from a statistical point of view. Hence, there is still a large number of untreated people available for the matching procedure which can yield reliable results with proper matching quality even if the number of treated is scarce.

[103]In Section 4.3, we explain, im more details, the decision process of choosing the matching algorithms and their concretizations — as bootstrapping is extremely time-consuming and needs tremendous computing capacity, we carried out just spot-checks with bootstrapped standard errors in order to compare the significance levels with the results gained from the calculation of standard errors as it is implemented by Leuven and Sianesi (2014) in their Stata module *'psmatch2'* which "does not take into account that the propensity score is estimated". The spot-checks with bootstrapped standard errors widely confirm our results gained from the standard calculation implemented by *psmatch2*.

for two different inflow samples: people who started to receive welfare benefits in 2007, or in 2008. We analyzed the inflow samples of 2007 and 2008 separately to figure out the dependency of our findings on the year of entering welfare receipt. In spite of self-evidently occurring variations in the concrete values of the ATT and standard deviations at large, the two inflow samples don't reveal fundamentally different results. Hence, our findings are not strongly dependent on the year that the observation starts. This holds at least for the period of 2007 and 2008.[104]

Another kind of sensitivity analysis investigates unobserved confounding factors. These so-called confounders are variables that affect the treatment variable — more precisely, the likelihood of being treated — as well as the outcome variables, and thus they can cause a hidden selection bias which distorts the estimations of the treatment effects.[105]

As explained in Section 4.3, the 'conditional independence assumption' (CIA), also called 'unconfoundedness', has to be satisfied in order to obtain robust results. The CIA claims that differences in the outcome between treatment and control group must be independent of the selection process, and thus be caused only by the treatment.[106] To satisfy this, it is favorable to use as many potential confounders as possible as control variables. Nevertheless, investigating humans and their behavior, we generally have to reckon with unobserved confounders, and hence we have to deal with possible unobserved heterogeneity. One proper way to assess the quality of the estimated treatment effects, even for the case that the CIA is not fully satisfied, are the sensitivity analyses based on the so-called 'Rosenbaum bounds' and the 'Hodges-Lehmann point estimators', suggested by Rosenbaum (1993).[107] Following Liu et al. (2013) and Müller (2012), we carried out spot-checks based on Rosenbaum bounds and Hodges-Lehmann point estimators in order to explore the impact of potential unobserved confounding factors on our estimations of the treatment effects (ATT). All in all, we find the estimation results of the probability outcomes mainly robust against potential unobserved heterogeneity caused by unobserved confounders. The duration outcomes, however, reveal even more estimations whose

[104] Variations in the quarter of the year in which the individual starts to receive welfare do most probably influence the results because of seasonal effects. Hence, in our estimations we control for the inflow quarters within the inflow years; see Section 3.4.

[105] A good introduction to sensitivity analysis for unobserved confounding is given by Liu et al. (2013).

[106] See Müller (2012), Rosenbaum and Rubin (1983), and D'Orazio et al. (2006).

[107] See Rosenbaum (2002), DiPrete and Gangl (2004), and Müller (2012). Rosenbaum's approaches are the most frequently used method to deal with unobserved heterogeneity associated with matching methods like propensity score matching (PSM); see Liu et al. (2013).

unconfoundedness could not be affirmed.[108] These results of the robustness checks go often, but not always, along with high levels of significance or insignificance of the estimated ATT.

# 6 Discussion and assessment

The results of our comprehensive analyses presented in Section 5 agree, in part, with previous studies on benefit and welfare sanctions in Europe, differ, in places, from other studies' findings and, for certain aspects, no comparison is currently possible, due to the uniqueness of our analyses. At a minimum, those parts of our findings which contradict corresponding studies need to be discussed and evaluated. Furthermore we have to take into account the potential bias arising from stratification, introduced with its possible variations on a theoretical base in Section 4.4.3; in Section 5, we mention it on various occasions, presenting our numerical results and partly discussing and assessing the bias in greater detailed while interpreting our graphical results. In this section, we discuss and evaluate our results in the context of related studies and the potential bias.

The unique aspects of our analyses are, first, that we consider not only unemployed but also employed welfare recipients receiving supplementary welfare benefits and, second, that we do not only consider sanctions against employable welfare recipients themselves (direct sanctions) but also analyze the effect of imposed sanctions upon other members of the employable individuals' households (indirect sanctions). All other well-known European studies on welfare sanctions are restricted to *unemployed* welfare recipients and to sanctions *directly* imposed upon them. Hence, our results for *employed* welfare recipients, in Germany colloquially called *'Aufstocker'* presented in Section 5.2, and our results for welfare recipients *indirectly* affected by sanctions upon their household members, presented in Section 5.3, cannot be benchmarked against other studies, however we still considering them due to the potential bias.

But let us first focus on the bulk of our results, presented in Section 5.1, which reveal how unemployed welfare recipients respond to sanctions imposed upon them directly. In this field, there already exists some partly corresponding studies which can be used to validate, as well critical question

[108] The null hypothesis supposing that the estimated effects are exclusively due to confounding factors, can be rejected up to the maximum value of $\Gamma_{max}$ with a $p$ value of $p$<0.1; evaluating the results of the sensitivity analysis with Rosenbaum bounds, we follow Aakvik (2001) who suggests that a $\Gamma$ of 2.0 with still significant treatment effects can be seen as a 'very large' robustness.

our results. The most striking fact when comparing our findings with the results of corresponding studies is that not only we do estimate significant positive effects of welfare sanctions on entering employment and leaving welfare receipt, but we also obtain a lot of cases with significant negative estimations of the average treatment effect on the treated (ATT).

From the perspective of the sanctioned individuals who may have experienced not only the pushing effects of sanctions, but also their paralyzing and debilitating impacts, this might not be surprising. Also, several studies on the impact of welfare sanctions, mostly qualitative surveys or paraphrasing them, are consistent with that or give at least indications of adverse individual effects that can be detrimental to taking up employment (see e.g. Ames (2009), Götz et al. (2010), Ehrentraut et al. (2014), Wolff (2014), and van den Berg et al. (2015, 2017)). Such detrimental conditions could adversely affect the physical as well as the mental condition of the sanctioned as a result of financial and existential pressure, examples of which include: an increasing burden of debt; being threatened with, or being disconnected from electricity and heating services; and losing, or being in danger of losing one's home. These existential threats can easily be seen as factors that are not only detrimental to the individual's well-being, but also to the probability of achieving employment.

Nevertheless, the overwhelming majority of other studies' results — not only about UIB sanctions but also about welfare sanctions — reveal positive effects of sanctions on the transition rate from unemployment to employment. At first glance, the following reasons may be responsible for this seemingly contradiction between qualitative and quantitative surveys: either the mentioned conditions, detrimental for the individual, are less effective than the pushing effects, at least for the majority of sanctioned people, such that the effect of individuals who react positively to sanctions, in terms of taking up employment, dominates the effect of individuals who are impeded or even prevented from getting employed by (possibly severe) adverse individual sanction effects. Or the vast majority of estimated positive sanction effects on entering employment can be caused by time-related factors: it seems plausible that negative effects like rising debt and health problems can accumulate and increase over time and, hence, be more effective in the medium- and long-runs, while the pushing effects of sanctions work primarily in the shorter run. And if individuals who predominantly respond positively to pressure are in the majority — at least in the time horizon of the survey — the impact on people with predominantly adverse effects is outweighed.

Now let us turn the focus from those general reflections back to the information we gain from our data analyses considering the potential bias, and the findings of corresponding studies in order to assess and finally conclude our results.

## 6.1 Unemployed welfare recipients

Although we cannot quantify the negative bias, presumably distorting our results for unemployed welfare recipients in terms of entering employment and leaving welfare, we can assess our findings by means of the curve shape of the plotted ATT, and against the backdrop of related results from previous studies. As already partly discussed and evaluated in Section 5.1, there are hints as to the magnitudes of such negative distortions. Here we refer to the outcomes which measure the probabilities of exit events.[109] Hence, the ATT could range from $-1.0$ to $+1.0$ which implies that the mean difference of the probability of experiencing the exit event between sanctioned and matched non-sanctioned can theoretically range from a minimum of $-100$ to a maximum $+100$ percentage points. The hints that we get from the graphs with the plotted ATT show that, for the cases of the most extreme and persistent negative ATT of men, we get negative values in the range of $-0.15$ to $-0.25$ during the strata and shortly afterwards, which is the period during which the bias arises. This means that the estimated sanction effect shows the sanctioned to have a 15 to 25 percentage points lower probability of experiencing the exit event than the non-sanctioned control group. And if, additionally, the curve of plotted ATT shortly after the stratum skips or steeply slopes upwards, it is very likely that the initially negative values are strongly biased. Nevertheless, we cannot limit the bias to those maximal negative values during the strata, as the factual treatment effect could be positive, and according to previous studies, should be estimated positively. Hence, the bias may, in some cases, even exceed the 15 to 25 percentage points mentioned above.

All in all, we must evaluate the absolute values especially of our negative outcomes very critically, as they are most likely caused by a considerable negative bias as a result of the stratification explained in Section 4.4. The extent of the bias, however, depends strongly on different circumstances, as explained

[109] The other class of outcome variables we use for our analyses measures the durations until the exit events occur. As outlined in Section 3.3 and explained in detail at the beginning of Section 5, the findings for these kinds of outcomes can be interpreted mostly in accordance with the outcomes of probabilities, but because of the reversed signs they must be interpreted in the opposite direction.

in detail in Section 4.4.3. Amongst other things, it strongly depends upon the kind of exit events, the distribution of the treatments within the strata, and the likelihood and distribution of the exit events of the matched control group during the strata. Hence, the bias can vary strongly between different exit events, and it can vary still noticeably also between different groups and sub-groups and between the strata.

In spite of these difficulties in interpreting and assessing the absolute values of our results, the shape of the ATT's curve, which reveals the development of the sanction effects over 24 months, can be interpreted independently of a potential or actual bias. As already explained while presenting the results in Section 5, we show the *cumulated* ATT over time and, thus, the differences of estimated ATT between the overlapping periods with consecutive final months reveal the *monthly* effects. Hence downward slopes after the stratum reveal negative monthly effects, while upward slopes reveal positive monthly effects. This is valid independently of the bias which can arise only within the strata, or possibly short afterwards. As we thus cannot reliably interpret the slope within the strata, and possibly shortly afterwards, we refrain from interpreting these very early periods with the help of slopes. But we can decently interpret the slope after the strata — at least within a short distance of the strata — in order to get information about the monthly effects in the medium and long term.

To summarize: even if the absolute values of our results are contestable, the development of the ATT over time still provides reliable insights. Taking all of this into account, and reviewing the plausibility while also being aware of the findings of previous studies, we come to the following conclusions by assessing and evaluating our results.

### 6.1.1 Divergences between different strata

It is evident that the ATT for the first strata ($S_1$) tend to differ, and are often less negative or more positive than for later strata ($S_2$ to $S_4$). This could be due to two reasons: firstly, early sanctions could be more effective in the positive sense, namely that they increase the transition rate into employment and out of welfare receipt. This would go along with previous studies, mainly on UI sanctions, but also partly on welfare sanctions. secondly, the discrepancies between the first and the later strata also give a strong hint that the bias arising in the first strata could be lower than in the later ones. This also seems plausible, as there are several factors that account for a tendentially more left skewed distribution of sanctions within the first stratum compared to later strata. And as explained in detail in Section 4.4.3, a left skewed distribution of sanctions within the stratum reduces the potential bias as it lowers the probability of exit events of the matched non-treated before the sanctions of their matching partners occur.

One of the factors which may cause a left skewed distribution of sanctions is that a substantial proportion of people entering UB-II-receipt are initially also receiving unemployment insurance benefits (UIB), which is called unemployment benefits I (UB I) in Germany. If the amount of monthly UB-I payments does not cover the minimum subsistence level of their families, they are eligible for supplementary UB II. This can either be the case because UB I, depending on the previous income, is too low, or because the individuals are sanctioned within the UIB system, and thus become eligible to apply for UB II. If people are sanctioned in the UB I system, their breach of duty automatically causes a sanction within the UB II receipt, too. Furthermore, sanctions at the beginning of UB-I-receipt are disproportionately more likely, as part of the people do not fulfill, or even know their obligation to start job search three months before the end of their previous employment.

Moreover, the case workers at the German Jobcenters responsible for UB II are generally required to make initial offers to the clients in order to check their willingness to work, even at the beginning of UB-II-receipt. Because of a lack of proper jobs, such offers are mostly not job offers but ‘offers’ to participate in a measure of the ALMP which, at least for UB-II-recipients, is often some standardized training on how to prepare job applications that does not necessarily fit the clients’ requirements. Refusing or culpably dropping out of such measures results in strong sanctions lasting three months.[110] Thus, the sanction probability at the start of UB-II-receipt, independently of also getting UB-I-payments, is disproportionately high.

All these factors count for early sanctions within the first stratum causing a left skewed sanction distribution, and thus lead to less negatively biased ATT. Nevertheless, we cannot identify, for certain, whether the more positive ATT in the first strata ($S_1$) come from the lower negative bias, or from the possibility that early sanctions might be more effective than later ones. But as the progression of the later strata (from $S_2$ to $S_4$) do not give clear hints supporting the hypothesis that earlier sanctions

[110] More detailed information about the sanction regime under UB II are provided by Hillmann and Hohenleitner (2015) and van den Berg et al. (2014, 2015).

are more effective, it seems more likely that the outlying results for the first strata ($S_1$) are mainly caused by a lower bias.

### 6.1.2 Divergences between men and women regarding employment entry

Comparing the estimated treatment effects of men and women upon the rate of entry into different types of employment, it is striking that negative ATT occurs more often for men with respect to mere employment, and positive ATT occur more often for women with respect to employment with supplementary welfare receipt. Besides, all other outcomes tend to be mostly insignificant. Consistent with that, for the concise exit event of 'entering employment in general', which comprises mere employment as well as employment with top-up benefits, men tend to show negative significant ATT, and women tend to show positive significant treatment effects.

This all together points to the fact that the estimated negative ATT for men concerning employment in general are mainly driven by the ATT on entering mere employment, while the positive effects for women regarding employment in general are mainly driven by the ATT on entering subsidized work, concretely employment with top-up welfare benefits. Additionally, these results seem to imply that women's responses to welfare sanctions are stronger towards employment with supplementary benefit receipt, while men tend to respond more strongly with regard to mere employment. That seems plausible as still more women than men work in lower paid or part-time jobs, especially if they raise children.

However, the fact that the estimated ATT for men in terms of entering mere employment show mainly negative results, and the estimations for women in terms of entering subsidized work are mainly positive, does not seem to be plausible at the first sight. Indeed this is a strong hint for biased results. Recalling the explanations about the bias in various constellations in Section 4.4.3, the potential negative bias works strongest for exit events out of risk — specifically, out of welfare receipt — and tends to be weaker for exit events staying at risk, like taking up employment while receiving top-up welfare benefits. As there seems to be no other plausible explanation for the divergence between men and women in terms of negative respectively positive treatment effects, it seems most likely, that the negative estimations for men are indeed caused by a substantial negative bias.

Hence, the absolute values of the estimations for exits into mere employment must be severely questioned. As other studies which distinguish between exits into subsidized versus unsubsidized work tend to find stronger positive sanction effects for unsubsidized work than for subsidized work, it can also be asked why for women, we estimate positive effects in terms of entering employment with supplementary welfare receipt, but mainly insignificant effects in terms of entering mere employment.

Taking all of this into account, it seems most likely that the effects for women on taking up mere employment are, in fact positive, as well as for men, and that the effects for men are even stronger. The reason for our diverging results in terms of mere employment in this scenario would be that the estimations are most probably strongly negatively biased, for men even more than for women. This assessment seems to be plausible, as men in terms of taking up mere employment presumably tend to even more early exit events than women do. And hence, the share of non-treated controls with exit events occurring before the sanction of their matching partners, and thus the negative bias, for men is tendentially higher than for women.

Nevertheless, it is not possible to assess, with any certainty, whether the sanction effects are stronger for regular or subsidized employment. Practical plausibility considerations can point in both directions. On the one hand, sanctioned people who want to escape from the pressure of future repeated sanctions and intensified monitoring by the Jobcenters, can be assumed to have stronger incentives to find a way out of welfare receipt, be it taking up employment or leaving the labor force. It seems plausible that entering mere employment would be preferred to starting a job needing top-up welfare benefits, because the latter option would not get them out of risk to be sanctioned again. At least people with good chances on the labor market, who do have such kind of choices, may prefer taking up mere employment to subsidized work.

On the other hand, the share of people with worse chances on the labor market within the quite heterogeneous group of welfare recipients is considerably higher than in the group of UIB recipients. Thus, it might be that even if the majority of sanctioned people would prefer to get out of welfare receipt, they cannot find a job which pays enough to cover their and their family's (if any) minimum subsistence level. These people may respond to sanctions with an increased probability of taking up employment even if this does not lead them out of welfare receipt. Although they are still at risk of being sanctioned because of the continued welfare receipt, they could expect to be less severely monitored if they start a new job, as they have demonstrated their willingness to work.

Hence, in practice, it is ambiguous whether a majority of sanctioned would respond with a higher probability either to take up mere employment, or to start a job with top-up welfare receipt. Referring to corresponding studies like, e.g. Schneider (2008, 2010), which tendentially find unsubsidized work positively, and more strongly affected than subsidized work, previous findings point more in the direction of stronger positive sanction effects resulting in the take up of mere employment than for supplementary welfare receipt. Following these findings, it can be suspected that the negative bias for men in our results regarding taking up mere employment would be even higher than assessed above; it would not just move an insignificant effect in the negative direction, but turn even a significant positive effect into negative effect estimations. This scenario cannot be ruled out — rather, it must be seen as a not unlikely possibility. If this is the case, the stronger negative ATT estimations for men just reflect a higher probability of early exit events in the group of non-treated which causes a bias high enough to not only outweigh, but even dominate the actual positive sanction effect. Following these thoughts, women obviously have lower probabilities of early exits into mere employment, and thus show a lower negative bias. In terms of employment with top-up benefits, however, sanctioned women are more likely to respond positively than sanctioned men. But as mentioned above, the bias for this exit event which does not lead out of risk is not as strong or, at the best, is even negligible.

Following the scenario of the previous paragraph, and in accordance with corresponding studies, we can assess and summarize our results as follows: the absolute values of our results concerning exit to employment are, most likely, negatively biased in a substantial extent, at least regarding entrance to mere employment. It can be assumed that the total effects of sanctions in terms of taking up mere employment are actually positive for men, possibly a bit stronger than for women, as men are generally more likely to experience early exit events to mere employment and, thus, they may be more responsive with regards to this exit event. Accordingly, their negative bias is stronger which turns their positive effects into negative estimations of the ATT. The lower negative bias of women turns their real, probably also positive, effects into insignificant estimations concerning exits into mere employment.

As the bias for exits into employment with supplementary welfare receipt seems to be considerably lower, it can be assumed that the stronger estimated effects for women reflect the real effects, considering the fact that women respond more positively to sanctions in terms of taking up subsidized work than men. Additionally, it can be presumed that a possibly still working negative bias is much weaker than for exit to mere employment. Hence, the factual treatment effects may even be positive for men, while a weak negative bias turns them to insignificant results. Whereas, for women, the effect is clearly positive and, given a still existing but small bias, presumably to a bit larger extent than the estimated ATT depict.

### 6.1.3 Divergences between entering mere employment and exiting welfare

On the whole, the estimated effects of sanctions on exiting welfare are stronger than on entering mere employment. As mentioned above in Section 5, the difference between them is that exiting welfare also includes the option of leaving the labor force, namely the non-employment option. Although in this study, we do not explicitly analyze the exit event of leaving the labor force — we have done this already in our previous study on welfare sanctions (Hillmann and Hohenleitner (2015)) — the findings of these analyses also provide implicit insights into the non-employment option. As the estimated treatment effects on exiting welfare point in the same direction as on entering mere employment, but are stronger, we can conclude that exiting the labor market is affected by sanctions in a similar way as entering mere employment. This does not necessarily mean to a similar *extent*, but in a similar direction. That conclusion goes along with our previous study analyzing the exit from labor force (Hillmann and Hohenleitner (2015)) and with other related studies on welfare sanctions which also consider the non-employment option (Busk (2014) for Finland and van den Berg et al. (2015) for young welfare recipients in Germany).[111]

According to the line of arguments and the final conclusive assessment for exit into mere employment above, the absolute values of the estimated ATT for exit from welfare must likewise be interpreted under the assumption of being negatively biased in a substantial, presumably even stronger, extent. The reason for this is that exit from welfare leads out of risk, just as exit to mere employment does, but to a stronger extent as it additionally comprises exit from labor force which also leads out of risk to be sanctioned. And as explained above in this section, and in more detail in Section 4.4.3, exits out of risk imply the highest potential negative bias. Hence, according to the above concluding assessment

[111] Even the negative results dot not contradict that, as they are most likely caused by a partly strongly negative bias, which likewise distorts the outcome for leaving the labor market.

concerning exit to unsubsidized work, we can assume that the actual effect of welfare sanctions on exiting welfare — as well as the effect on both of its components, entering mere employment and leaving the labor force — is positive.

### 6.1.4 Highlights of the subgroups

As the sanction effects on various subgroups are described in detail in Section 5, here we shall highlight just a few apparent patterns. There are indications that people living in Western Germany tend to have stronger positive sanction effects in terms of taking up employment in general while, in terms of leaving welfare receipt, people in Eastern Germany tend to reveal stronger positive sanction effects. The latter may be driven by a higher proportion of people leaving the labor market. These findings hold more strongly for men than for women.

People with a high level of labor market access respond to welfare sanctions with a stronger enhanced transition rate out of welfare receipt than people with a medium level of labor market access, while the lowest increase of transition rates out of welfare reveals sanctioned with a low level of labor market access. Whether these findings are mainly driven by entering mere employment or leaving the labor market can not clearly be assessed by our estimations because the subgroups with low and high levels of labor market access are too small in order to get enough exit events in the treatment group to obtain more reliable results.[112]

### 6.1.5 Development over time

Even if the absolute values of our results are strongly biased in some cases and under certain circumstances, the curve progressions of the plotted ATT reveal insights that are valid independently of the bias which almost exclusively arises during the strata. Although the *absolute* values are also biased after the strata as depicted by the *cumulated* ATT, the *slope* after the strata is no longer affected by the bias, at least more than a short distance from the end of the strata. The progression of the slope reveals the development of the *monthly* treatment effects over time. As mentioned above in the introductory part of this subsection (Section 6.1), downward slopes after the stratum reveal

[112] As our data set is based on the SIAB data, a 2% sample of the IEB data set (see Section 3); in case of small subsamples, we occasionally suffer a lack of treated cases and, thus, a lack of exit events in the treatment group. This could be improved by using a larger sample, or even the full IEB data set combined with administrative data.

negative monthly effects, while upward slopes reveal positive monthly effects, that are both stronger, the steeper the curve runs.

In our presentation of the plotted ATT in Section 5, we have seen very diverse curve shapes. A substantial proportion of these plots of cumulated ATT, though, depict either a quite continuous downwards trend in the curve, or at least a downwards trend for a considerable time before the end of the observation period. In these cases, we can suspect that the downwards slope would continue for a while longer if the observation period is extended. But also, if this latter possibility for prolonged observation periods might not hold, the visible long periods of downwards slope do reveal considerable negative monthly sanction effects in the medium and long run.

To reiterate, this is not a general assertion, as there are also various other shapes of the cumulated ATT, but it holds for the described negative curve progression, of which there are numerous examples. Nevertheless, we emphasize especially these kinds of shapes with longer periods of negative slope, and thus negative monthly effects, as they are the problematic ones in practice.

Long-term or steeply downward heading slopes of cumulated ATT can lead to negative sanction effects as a whole, even if the initial and short-term effects are strongly positive. Whether, for some cases, negative long-term effects dominate positive short-term effects depends, amongst other factors, vastly on the time horizon of the analyses. But even if such negative effects do not fully outweigh the positive ones, including in the long-run, it is a fact that there are considerable numbers of sanctioned people whose probabilities of entering employment or leaving welfare are clearly negatively affected by welfare sanctions. And even if the negative impact on them would never outweigh the positive impact on others, they cannot be neglected.

### 6.2 Employed people with supplementary welfare receipt

Concerning the effects of sanctions on employed welfare recipients, we cannot compare our results with other studies as this is still a relatively new field of research, and we did not find any corresponding studies. Scientific literature on employed welfare recipients in Europe is scarce, and the effects of sanctions against this group are still almost unexplored.

As in Section 5.2, we have already discussed our findings for employed welfare recipients in quite some detail; also, regarding the bias and the plotted ATT, we mainly summarize our findings here in brief, and partly complement them.

### 6.2.1 Entering mere employment

In contrast to unemployed welfare recipients, there are no systematic differences evident between the results of the first strata ($S_1$) and the following strata ($S_2$ until $S_4$). This can be an indication that the distribution of the sanctions within the first stratum of employed welfare recipients is neither left skewed nor systematically differently distributed compared to later strata.

Similar to the results for unemployed welfare recipients discussed above, we must also reckon with a probably substantial, negative bias for the employed welfare recipients in terms of entering mere employment. Thus, we cannot assess with any certainty, whether the absolute values of treatment effects are indeed negative, especially in the short run.

Nevertheless, it stands out that the slope of the cumulated ATT is strongly and quite persistently negative. Hence, the above findings and assessments for unemployed welfare recipients concerning the development of the monthly treatment effect over time can be transferred to the group of employed people. Also, in the medium and longer terms, the probability that possible initially positive effects are outweighed, or finally dominated by negative effects in the mid- and long-term is seemingly even higher than for unemployed people.

Employed people receiving top-up welfare benefits with a higher level of labor market access seem to be less negatively affected by sanctions than those with medium level market access.

### 6.2.2 Leaving welfare receipt

Compared to entering mere employment, the effects of sanctions upon employed people on exiting welfare are considerably less negative, or even positive. This rather strong discrepancy between both the exit events can either be caused by a lower negative bias concerning leaving welfare, or by a strong positive effect of sanctions on leaving the labor market.

On the one hand, the first strata differ from later strata, especially from the second strata, and are tendentially insignificant. Thus, they seem to be less negatively biased. In contrast, the second strata seem to be most severely negatively biased, more so than the estimations for the third and forth strata. Nevertheless, the strong downward trends that we see for entering mere employment cannot be observed for leaving welfare.

All together, the less negative ATT for the first strata regarding exiting welfare compared to entering employment seems to be mainly caused a lower negative bias. However, the curve progressions of the later strata point to an increased probability of leaving the labor market. Should this assessment hold true, later sanctions on employed people would have stronger adverse impacts, namely increasing the probability of leaving the labor market, than sanctions in the first quarter of welfare receipt.

Nevertheless, the evaluation of the results concerning the outcomes for employed welfare recipients with respect to exit from welfare is quite complicated to interpret, and thus rather speculative at this point. Further research with a larger sample is necessary to get more detailed and reliable results in order to properly evaluate these outcomes.

### 6.2.3 Leaving employment for mere welfare receipt

Regarding the exit event of quitting employment in order to merely live on welfare receipt, we face the phenomenon of positively biased ATT, which we explained in detail in Section 4.4.3, and already referred to and discussed in the presentation of our results in Section 5.2.3. Concluding these explanations, we can detect that the overwhelming positive estimated treatment effects are mainly due to a strong positive bias; as a positive bias can only arise during the strata, the cases with upwards heading slopes during a few months after the strata reveal an actual positive monthly effect of welfare sanctions against employed people in terms of quitting their job in order to live merely on welfare payments. This kind of response to sanctions is clearly politically unintended, and can be seen as an adverse effect.

Nevertheless, the vast majority of the curve progressions are downwards heading (or at least not upwards heading), even if the total effect is estimated to be positive. This implies that even if the actual effect minus the positive bias is still positive, the monthly effects tend to reduce this initially adverse effect. But as the downwards slope is mostly not very steep, a potential actual positive effect in the short run may most probably not be outweighed in the medium or even long run.

However, whether the initial effect in the short run is indeed positive, can hardly be assessed with an adequate certainty. So also in this case, further analyses with a larger sample are necessary to be able to evaluate these outcomes properly.

### 6.3 Indirect sanctions

Although we have done all the analyses for indirect sanctions as comprehensively and as well as for direct sanctions, reliable results for indirect sanctions are very scarce. This seems to be not mainly a result of factual scarce effects, but seems predominantly caused by a lack of cases of merely indirectly sanctioned welfare recipients and, as a consequence, a lack of exit events in the treatment group. An analysis of the impact of indirect sanctions requires a data set with many more cases of merely indirect sanctioned people — ideally, a full dataset based on administrative data which would provide an optimal data base.

Notwithstanding the difficulties due to a lack of cases, we get at least interesting hints as to which direction the results of an analysis with a proper data set could point. We clearly see that men do not respond significantly to indirect sanctions in our data set, but women seem to respond tendentially negatively. This holds for the transition from unemployment to mere employment, as well as for the transition from employment with top-up benefits to exit from welfare.

As these assessments are not only based on the absolute values which can be negatively distorted, but are also and mainly based on the negatively sloped curve progressions for women, this can be seen as a first glance showing that it should be worth the effort to undertake further research on this topic.

## 7 Conclusion

For our comprehensive analyses on welfare sanctions in Germany, we use the inflow samples into welfare for the years 2007 and 2008 from an exclusively prepared rich data set, based on a 2% sample of administrative data of the German Federal Employment Agency (FEA), and provided by the Research Data Centre (FDZ) of the Institute for Employment Research (IAB), covering the years 2004 until 2010. We present and discuss our results based on the inflow sample 2008 and use the results for 2007 as one of several kinds of robustness checks.

We conduct our analyses for diverse outcomes according to the main initial samples of unemployed and employed welfare recipients. The latter are colloquially called *'Aufstocker'* in Germany, which roughly means 'top-up recipients'. The outcomes are mainly the probabilities of various exit events changing the individual labor market status (entering/exiting employment, leaving welfare), and secondarily the corresponding durations until these events occur. For reasons of simplicity and clarity, in the summarizing presentation of our results, we only refer to the outcomes with respect to probabilities.

In addition to analyzing the sanction effects for the initial samples as a whole, we also investigate various subsamples categorized by age, gender, education (which is indirectly contained in a variable comprising the individual 'labor market access', or LMA), and finally by regional differences (living in Eastern or Western Germany). Furthermore, we carry out our whole analyses of the main and the subsamples, not only for people who are directly treated by sanctions imposed against themselves (direct sanctions), but also for employable people who are merely indirectly affected by sanctions upon their related household members (indirect sanctions).

Employing a dynamic approach of propensity score matching (PSM) with stratification is, on the one hand, appropriate and especially favorable for analyzing an extremely heterogeneous population such as employable welfare recipients. On the other hand, we have to deal with biased effect estimations which must be considered when interpreting the absolute values of our numerical and graphical results. Nevertheless, we get a huge amount of results by analyzing the monthly updated average treatment effects on the treated (ATT) over 24 months after the sanction, which reveal the development of the cumulated, as well as of the monthly ATT over two years time. — More precisely, we use quarterly strata of the individuals' welfare durations in order to cover the dynamic setting of the treatment in the case of direct sanctions. In contrast,, regarding people merely indirectly affected by sanctioned household members, our time horizon is 18 months, using half-yearly strata. — The development of the ATT over time, in fact, gives insights that are valid, independent of a possible bias of the absolute values.

The findings we get from our analyses of direct sanctions against *unemployed* welfare recipients are the most extensive ones, as this group is by far the largest of the main groups of directly and indirectly sanctioned people, and thus we have a large number of exit events, not only in the control but also in the treatment group. Hence, we also get reliable and exploitable estimations for many of the smaller subgroups, which is often not the case for smaller main samples and, thus even smaller subsamples. For these results alone, there are partly corresponding studies that we can use in order to reflect our findings against the background of previous research. This is because previous studies on welfare sanctions, as a rule, only consider direct sanctions against unemployed welfare recipients.

Though receiving several negative treatment effects, especially in terms of exit to mere employment (unsubsidized work), and primarily for men, while predominantly getting insignificant ATT for women in terms of mere employment, we assess these absolute values of our numerical results as negatively biased to a substantial extent. Taking this into account, we assume tendentially positive sanction effects on employment, at least in the short run. The same is true, but even stronger for exiting welfare, and thus, this holds also for exiting the labor force which implicitly is revealed by the divergence between the ATT for mere employment and exiting welfare. Men are more responsive to welfare sanctions in terms of taking up unsubsidized work while, in terms of entering employment with top-up benefits (subsidized work), women were found to be more responsive than men.

Still with respect to unemployed welfare recipients, we get additional findings by analyzing various subsamples. For instance, we find indications of the following differences, at least as slight tendencies: in terms of entering employment in general — comprising subsidized as well as unsubsidized work — people living in Western Germany tend to show stronger positive sanction effects, at least in the short run. In contrast, regarding exit from welfare, people in Eastern Germany show tendentially stronger positive effects in the short run. The latter may be driven mainly by a higher share of people who leave the labor force. These findings are tendentially stronger for men than for women.

Concerning the timing of the sanctions, we see differences that account for time dependence, however, we cannot identify clear patterns that are independent of the individual factors. On the whole, we find hints that early sanctions — that is those that occur within the first quarter of welfare receipt — cause rather positive effects than later sanctions. But we cannot verify this with sufficient certainty.

An outstanding and unique achievement of our analyses is that we reveal the monthly progression of the various effects of welfare sanctions over two years time. We find the *development of sanction effects over time* for diverse exit events as extremely versatile, depending on the analyzed groups and subgroups, their individual characteristics, regional differences, and the timing of the sanction.

There are some upwards heading cumulated effects, mostly with decreasing slope, which imply positive but decreasing monthly effects. Such patterns of development of the sanction effects with mainly non-negative slopes are unproblematic in terms of the considered exit events like entering employment or leaving welfare receipt. But they are, by far, not the majority. The persistent and steeply downwards trends in the progression of the cumulated effects, however, are quite problematic in practice, and thus need special attention.

Indeed, there are a considerable number of cases with long periods of downwards heading cumulated ATT, often with steep slopes, which reveal considerable negative monthly sanction effects in the medium and long run. Whether these scenarios actually outweigh positive sanction effects in the short run, we cannot assess for certain. But also if the negative effects in the medium and long run do not dominate the initially positive effects, we clearly find negative effects for a considerable number of the sanctioned people that cannot be neglected. The fact that the adverse effects harm people who already have lower probabilities of early exit events (and thus, worse chances on the labor market), while the positive sanction effects predominantly work for people with higher probabilities of early exit events (and thus, with initially better chances on the labor market) is especially problematic.

Regarding direct sanctions against *employed* people receiving supplementary welfare benefits, the so-called *'Aufstocker'*, we find patterns of sanction effects on taking up employment and on leaving welfare receipt that are quite similar to the effects for unemployed people. However, we have no corresponding studies to compare our results and assess their absolute values, which most likely are negatively biased to a certain extent. Altogether, it is quite difficult and rather speculative to evaluate to what extent a likely negative bias distorts the absolute values of the estimated ATT. Therefore, we cannot asses whether the effects on employed people are higher or lower than on unemployed. Also we do not know whether the estimated negative treatment effects are mainly due to a strong negative bias, or rather reveal factually negative sanction effects.

However, the development of the ATT over time, which we can interpret independently of a possible bias, shows strongly and even steeper downwards heading slopes compared to the unemployed. This implies that — even if the initial and maybe also the total effects are positive — there is a considerable number of employed people who are severely negatively affected by sanctions in terms of taking up mere employment (unsubsidized work) and leaving welfare receipt in general.

A totally unexplored research question is whether, and if so, to what extent, sanctions against employed welfare recipients affect their probabilities of quitting the job and merely living on welfare. Regarding this only for *employed* welfare recipients possible exit event, we get positive sanction effects. This implies that sanctions raise the probability of quitting the job and living solely on welfare receipt, which can be judged as a politically undesirable adverse effect. It has to be taken into account that, for this kind of exit event, the bias works in the opposite direction, that is, towards positive distortions. Hence, these findings are most likely distorted by a positive bias. And as mentioned, we cannot assess the extent of the bias, in particular not for *employed* people. Thus, the absolute values of these results are contestable.

Nevertheless, we can adequately interpret the slopes of the plotted ATT which reveal the development of the monthly sanction effects over time. Concerning the exit event 'quitting employment for mere welfare receipt', the slopes are mainly slightly negative, and only in a few cases — mostly for a short time — we find positive slopes, which are the problematic ones. The time periods with negative slopes of the cumulated ATT for this exit event can be judged as desirable, because they lower a possibly enhanced transition rate from subsidized employment to unemployment, whereby the negative slopes are rather flat and thus the diminishing effect is rather small. The positive slopes for this exit event are the problematic ones, as they reveal a factual positive effect independently of a possible bias. This implies that, for a part of the employed welfare recipients, sanctions indeed enhance the probability of leaving employment in order to merely live on welfare payments. But such upwards heading curves of cumulated ATT, revealing undesirable effects, are rather rare and occur only during short periods of about three months after the strata — i.e. after the quarter during which the sanction was imposed.

Although we carry out the whole analyses for *indirect sanctions* as well as for direct sanctions, we only got very few reliable results for indirectly sanctioned people. This is mainly due to a lack of cases of merely indirectly sanctioned welfare recipients and, as a consequence, a lack of exit events in the treatment group. In order to get more reliable estimations, further research with a data set including many more cases of merely indirectly sanctioned people is necessary.

Nevertheless, we get at least a few hints as to which direction the results of an analysis with a larger and proper data set would probably point. We find men not significantly responding to sanctions against their household members; but women seem to respond tendentially negatively. This holds for *unemployed* women in terms of entering mere employment (unsubsidized work) as well as for *employed* women in terms of exiting welfare receipt.

In conclusion, we find *highly heterogeneous effects* of welfare sanctions in terms of total effects, as well as in their progression over time. The initial effects and their development over time depend on several conditions, specifically on individual factors like age, gender, and education, on regional differences (between Eastern and Western Germany), and on the timing of the sanction.

Generally, the negative effects tend to work stronger in the medium and long term, and the positive effects tend to work stronger in the short term. Hence, the shorter the *time horizons* of studies on welfare sanctions are, the more the positive effects are overrated systematically. Especially the frequently occurring cases with strongly negative slopes of cumulated ATT indicate that the early positive effects, mainly driven by people with good labor market chances, are at the expense of people with strongly negative sanction effects, even in the long run. These detrimental sanction effects are supposed to be driven mainly by people with worse labor market perspectives.

Therefore, the observation periods for studies on the effects of welfare sanctions should be *as long term as possible*, or at least as long as any notable effects can be measured. But nevertheless, it has to be considered that — estimating the average treatment effect on the treated (ATT) — positive sanction effects even in the long run are still only *average* effects. As long as the distribution of the sanciton effects is not known, we must reckon with a (possibly wide) range of sanction effects for different people, depending on their labor market chances. Thus, even in case of positive sanction effects in total, these effects might be accompanied by negative sanction effects for (possibly a minority of) sanctioned people which are detrimentally affected by santions. In order to clearly identify (or exclude) negative sanction effects on various (sub-)groups, *further research* is necessary, especially with focus on people with worse labor market chances.

To investigate the effects of sanctions against employed welfare recipients more deeply, and to better explore the effects also on employable people who are indirectly affected by sanctions against their household members, *further research* is also necessary for those groups. We have done a first step to also investigate these two, as yet, almost unexplored groups of welfare recipients affected by sanctions.

However, for these rather small groups of welfare recipients, our 2% sample of administrative data does not provide a sufficient number of observations with exit events in the treatment group. Using a much larger sample of administrative data should lead to a larger number of reliable and exploitable results, for *employed* welfare recipients as well as for *indirectly sanctioned*.

For further research, it is desirable to avoid the problem with biased estimations as a result of the stratification. In order to solve this problem, we developed a procedure to employ nearest neighbor matching in a way that excludes observations with exit events occurring before the treatment of the matching partner. This program, which we call the 'adjustment procedure', avoids the bias due to stratification, however, it requires a higher computer capacity than we had access to. And because of additionally strong time limitations concerning the data access, we could only use this most time-consuming procedure for spot-checks to get an approximate idea of the size of the bias problem. Nevertheless, computer routines to avoid the bias due to stratification are generally feasible, and hence should be employed whenever possible.

With our comprehensive study on the impact of welfare sanctions, we provide a huge range of results giving important new insights across the spectrum of ex-post effects on the individual labor market status. The limitations, however, should encourage further research on a still quite selectively explored research field.

## References


Aakvik, A., Feb. 2001. Bounding a matching estimator: The case of a Norwegian training program. Oxford Bulletin of Economics and Statistics 63 (1), 115–143.

Abbring, J. H., van den Berg, G. J., van Ours, J. C., July 2005. The effect of unemployment insurance sanctions on the transition rate from unemployment to employment. Economic Journal 115 (505), 602–630.

Ames, A., 2009. Ursachen und Auswirkungen von Sanktionen nach §31 SGB II, Hans-Böckler-Stiftung Edition. Vol. 242 of Arbeit und Soziales. Hans-Böckler-Stiftung, Düsseldorf.

Arni, P., Lalive, R., van Ours, J. C., Dec. 2013. How effective are unemployment benefit sanctions? Looking beyond unemployment exit. Journal of Applied Econometrics 28 (7), 1153–1178.

Bacher, J., 2002. Statistisches Matching: Anwendungsmöglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS. Zentralarchiv für Empirische Sozialforschung 51, 38–66.

Becker, S. O., Ichino, A., 2002. Estimation of average treatment effects based on propensity scores. The Stata Journal 2 (4), 358–377.

Boockmann, B., Thomsen, S. L., Walter, T., Dec. 2014. Intensifying the use of benefit sanctions: an effective tool to increase employment? IZA Journal of Labor Policy 3 (1), 1–19.

Busk, H., Nov. 2014. Search in the labour markets: Empirical evidence of the role of technology and sanctions. Ph.D. thesis, Jyväskylä Studies in Business and Economics, No. 51, University of Jyväskylä.

Caliendo, M., Kopeinig, S., Feb. 2008. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys 22 (1), 31–72.

DiPrete, T. A., Gangl, M., Dec. 2004. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. Sociological Methodology 34 (1), 271–310.

D'Orazio, M., Zio, M. D., Scanu, M., 2006. Statistical Matching: Theory and Practice. Wiley Series in Survey Methodology. John Wiley & Sons.

Drepper, B., Effraimidis, G., Oct. 2016. Identification of the timing-of-events model with multiple competing exit risks from single-spell data. Economics Letters 147, 124–126.

Ehrentraut, O., Plume, A.-M., Schmutz, S., Schüssler, R., Mar. 2014. Sanktionen im SGB II: Verfassungsrechtliche Legitimität, ökonomische Wirkungsforschung und Handlungsoptionen. WISO Diskurs, Friedrich Ebert Stiftung, Bonn.

Fitzenberger, B., Speckesser, S., May 2007. Employment effects of the provision of specific professional skills and techniques in Germany. Empirical Economics 32 (2), 529–573.

Fredriksson, P., Johansson, P., Oct. 2008. Dynamic treatment assignment: The consequences for evaluations using observational data. Journal of Business & Economic Statistics 26 (4), 435–445.

Gangl, M., DiPrete, T. A., Feb. 2004. Kausalanalyse durch Matchingverfahren. DIW Discussion Papers 401, German Institute for Economic Research (DIW), Berlin.

Gaure, S., Røed, K., Zhang, T., Dec. 2007. Time and causility: A Monte Carlo assessment of the timing-of-events approach. Journal of Econometrics 141 (2), 1159–1195.

Götz, S., Ludwig-Mayerhofer, W., Schreyer, F., 2010. Sanktionen im SGB II: Unter dem Existenzminimum. IAB Brief Report 10/2010, Institute for Employment Research (IAB), Nuremberg.

Guo, S. Y., Fraser, M. W., 2015. Propensity Score Analysis: Statistical Methods and Applications, 2nd Edition. Vol. 11 of Advanced Quantitative Techniques in the Social Sciences. Thousand Oaks.

Heckman, J., Summer 1997. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. The Journal of Human Resources 32 (3), 441–462.

Heckman, J., Ichimura, H., Smith, J., Todd, P., Sep. 1998. Characterizing selection bias using experimental data. Econometrica 66 (5), 1017–1098.

Heckman, J. J., LaLonde, R. J., Smith, J. A., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, 1st Edition. Vol. 3, Part A. Elsevier, Ch. 31, pp. 1865–2097.

Heckman, J. J., Smith, J. A., Spring 1995. Assessing the case for social experiments. The Journal of Economic Perspectives 9 (2), 85–110.

Heinrich, C., Maffioli, A., Vázquez, G., Aug. 2010. A primer for applying propensity-score matching. Impact-Evaluation Guidelines, Technical Notes IDB-TN-161, Inter-American Development Bank.

Hillmann, K., Hohenleitner, I., Sep. 2015. Impact of welfare sanctions on employment entry and exit from labor force — Evidence from German survey data. HWWI Research Paper 168, Hamburg Institute of International Economics (HWWI), Hamburg.

Ho, D. E., Imai, K., King, G., Stuart, E. A., Jan. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis 15, 199–236.

Hofmann, B., Feb. 2012. Short- and long-term ex-post effects of unemployment insurance sanctions. Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik) 232 (1), 31–60.

Kumpmann, I., Jun. 2009. Im Fokus: Sanktionen gegen Hartz-IV-Empfänger: Zielgenaue Disziplinierung oder allgemeine Drohkulisse? Wirtschaft im Wandel 15 (6), 236–239.

Lalive, R., van Ours, J. C., Zweimüller, J., Dec. 2005. The effect of benefit sanctions on the duration of unemployment. Journal of the European Economic Association 3 (6), 1386–1417.

Lechner, M., Jan. 1999. Earnings and employment effects of continuous off-the-job training in East Germany after unification. Journal of Business & Economic Statistics 17 (1), 74–90.

Leuven, E., Sianesi, B., Oct. 2014. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Last download: 2017-08-26.

Liu, W., Kuramoto, S. J., Stuart, E. A., Dec. 2013. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. Prevention Science 14 (6), 570–580.

Müller, C. E., 2012. Quasiexperimentelle Wirkungsevaluation mit Propensity Score Matching: Ein Leitfaden für die Umsetzung mit Stata. Working Paper 19, University of the Saarland — Center for Evaluation, Saarbrücken.

Reinkowski, E., 2006. Mikroökonometrische Evaluation und das Selektionsproblem — Ein anwendungsorientierter Überblick über nichtparametrische Lösungsverfahren. Zeitschrift für Evaluation 5 (2), 187–226.

Rosenbaum, P. R., Dec. 1993. Hodges-lehmann point estimates of treatment effect in observational studies. Journal of the American Statistical Association 88 (424), 1250–1253.

Rosenbaum, P. R., 2002. Observational Studies, 2nd Edition. Springer Series in Statistics. Springer.

Rosenbaum, P. R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.

Rosenbaum, P. R., Rubin, D. B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. American Statistician 39 (1), 33–38.

Rosholm, M., Svarer, M., June 2008. The threat effect of active labour market programmes. Scandinavian Journal of Economics 110 (2), 385–401.

Roy, A. D., Jun. 1951. Some thoughts on the distribution of earnings. Oxford Economic Papers, New Series 3 (2), 135–146.

Rubin, D. B., Oct. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66 (5), 688–701.

Rubin, D. B., Autumn 2004. Teaching statistical inference for causal effects in experiments and observational studies. Journal of Educational and Behavioral Statistics 29 (3), 343–367.

Schneider, J., Apr. 2008. The effect of unemployment benefit II sanctions on reservation wages. IAB-Discussion Paper 19/2008, Institute for Employment Research (IAB), Nuremberg.

Schneider, J., 2010. Activation of welfare recipients: Impacts of selected policies on reservation wages, search effort, re-employment, and health. Ph.D. thesis, Department of Economics, Free University of Berlin (FUB).

Sianesi, B., 2001. Implementing propensity score matching estimators with stata. Tech. rep., University College London and Institute for Fiscal Studies.

Sianesi, B., Feb. 2004. An evaluation of the swedish system of active labor market programs in the 1990s. The Review of Economics and Statistics 86 (1), 133–155.

Stürmer, T., Joshi, M., Glynn, R. J., Avorn, J., Rothman, K. J., Schneeweiss, S., May 2006. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. Journal of Clinical Epidemiology 59 (5), 437–447.

Svarer, M., 2010. The effect of sanctions on exit from unemployment: Evidence from Denmark. Economica 78 (312), 751–778.

Urkaregi, A., Martinez-Indart, L., Pijoán, J. I., Jul.–Dec. 2014. Balancing properties: A need for the application of propensity score methods in estimation of treatment effects. Statistics and Operations Research Transactions 38 (2), 271–284.

van den Berg, G. J., Uhlendorff, A., Wolff, J., 2014. Sanctions for young welfare recipients. Nordic Economic Policy Review (1), 177–210.

van den Berg, G. J., Uhlendorff, A., Wolff, J., Dec. 2015. Under heavy pressure: Intense monitoring and accumulation of sanctions for young welfare recipients in Germany. IAB Discussion Paper 34/2015, Institute for Employment Research (IAB), Nuremberg.

van den Berg, G. J., Uhlendorff, A., Wolff, J., 2017. Schnellere Arbeitsaufnahme, aber auch Nebenwirkungen: Wirkungen von Sanktionen für junge Alg-II-Bezieher. Tech. Rep. 5/2017, Institute for Employment Research (IAB), Nuremberg.

van den Berg, G. J., van der Klaauw, B., van Ours, J. C., Jan. 2004. Punitive sanctions and the transition rate from welfare to work. Journal of Labor Economics 22 (1), 211–210.

van den Berg, G. J., Vikström, J., Apr. 2014. Monitoring job offer decisions, punishments, exit to work, and job quality. The Scandinavian Journal of Economics 116 (2), 284–334.

van der Klaauw, B., van Ours, J. C., Mar. 2013. Carrot and stick: How reemployment bonuses and benefit sanctions affect exit rates from welfare. Journal of Applied Econometrics 28 (2), 275–296.

vom Berge, P., König, M., Seth, S., Jan. 2013. Sample of Integrated Labour Market Biographies (SIAB) 1975–2010. FDZ Data Report, Research Data Centre (FDZ) of the Institute for Employment Research (IAB) at the German Federal Employment Agency (FEA), Nuremberg.

Wolff, J., May 2014. Sanktionen im SGB II und ihre Wirkungen. IAB Comments 2/2014, Institute for Employment Research (IAB), Nuremberg.

Zhang, Z., Jan. 2017. Propensity score method: a non-parametric technique to reduce model dependence. Annals of Translational Medicine 5 (1).

## Figures

Table 15 gives an overview of the following figures with plotted ATT, referred to in Section 5 for different groups and subgroups of welfare recipients. The first two parts (Figure 1 until Figure 15 and Figure 16 until Figure 27), described in Section 5.1, give an overview of the graphs for *unemployed* welfare recipients' *(UE)* transition to exiting welfare *(ExWel)* and to employment, distinguishing transition into job only ($O$), job with supplementary welfare receipt ($S$), and job in general ($G$). The following three parts (Figure 28 until Figure 35), described in Section 5.2, give an overview of the graphs for transition from employment with supplementary welfare receipt to entering mere employment ($O$), to exiting welfare *(ExWel)*, and to exiting Job *(ExJob)*, all for *employed* welfare recipients *(Emp)*. And finally, the last part (Figure 4 until Figure 6), gives an overview of the graphs for people, effected by *indirect (ind)* sanctions, i.e. caused by a sanctioned household member, described in Section 5.3. The graphs show the plots of monthly updated cumulated ATT for four quarterly strata ($S_1$–$S_4$) in case of *direct* sanctions and for two half-yearly strata ($S_1$–$S_2$) in case of *indirect* sanctions.[113]

[113]Some graphs depict less strata due to convergence problems, caused by too few cases in the treatment group.

**Table 15:** Overview of the figures with plotted ATT[1]

| Figure | Table | Section | Sanction (dir/ind) | Status (UE/Emp) | Exit to job (O/S/G) / from welfare (ExWel) / from job (ExJob) | Group / Subgroup |
|---|---|---|---|---|---|---|
| 1–3 | 4–5 | 5.1.1 | dir | UE | O/S/G | all |
| 4–6 | 4–5 | 5.1.1 | dir | UE | O/S/G | West |
| 7–9 | 4–5 | 5.1.1 | dir | UE | O/S/G | medium |
| 10 | 4–5 | 5.1.1 | dir | UE | G | u25/West |
| 11–13 | 4–5 | 5.1.1 | dir | UE | O/S/G | o25 |
| 14–15 | 4–5 | 5.1.1 | dir | UE | O/S | o25/West |
| 16 | 7 | 5.1.2 | dir | UE | ExWel | all |
| 17 | 7 | 5.1.2 | dir | UE | ExWel | West |
| 18 | 7 | 5.1.2 | dir | UE | ExWel | Ost |
| 19 | 7 | 5.1.2 | dir | UE | ExWel | low |
| 20 | 7 | 5.1.2 | dir | UE | ExWel | medium |
| 21 | 7 | 5.1.2 | dir | UE | ExWel | high |
| 22 | 7 | 5.1.2 | dir | UE | ExWel | u25 |
| 23 | 7 | 5.1.2 | dir | UE | ExWel | u25/West |
| 24 | 7 | 5.1.2 | dir | UE | ExWel | o25 |
| 25 | 7 | 5.1.2 | dir | UE | ExWel | o25/West |
| 26 | 7 | 5.1.2 | dir | UE | ExWel | o25/medium |
| 27 | 7 | 5.1.2 | dir | UE | ExWel | o25/high |
| 28 | 8 | 5.2.1 | dir | Emp | O | all |
| 29 | 8 | 5.2.1 | dir | Emp | O | West |
| 30 | 8 | 5.2.1 | dir | Emp | O | medium |
| 31 | 10 | 5.2.2 | dir | Emp | ExWel | all |
| 32 | 10 | 5.2.2 | dir | Emp | ExWel | medium |
| 33 | 12 | 5.2.3 | dir | Emp | ExJob | all |
| 34 | 12 | 5.2.3 | dir | Emp | ExJob | high |
| 35 | 12 | 5.2.3 | dir | Emp | ExJob | o25/West |
| 36–37 | 13 | 5.3 | ind | UE | O/G | West |

[1] *Figures:* Plotted ATT on probabilities for transition from un-/employment to different exit events, corresponding to the tables with condensed results based on binary outcomes in Section 5.
[2] *Exit events: exit to:* job only (O), job with supplementary welfare receipt (S), job in general (G); *exit from:* welfare (ExWel), Job (ExJob).
*Subgroups: age-group in years:* all=15–56, u25=15–24, o25=25–56; *region:* West/East German states; level of *labor market access* (LMA): low, medium, high.

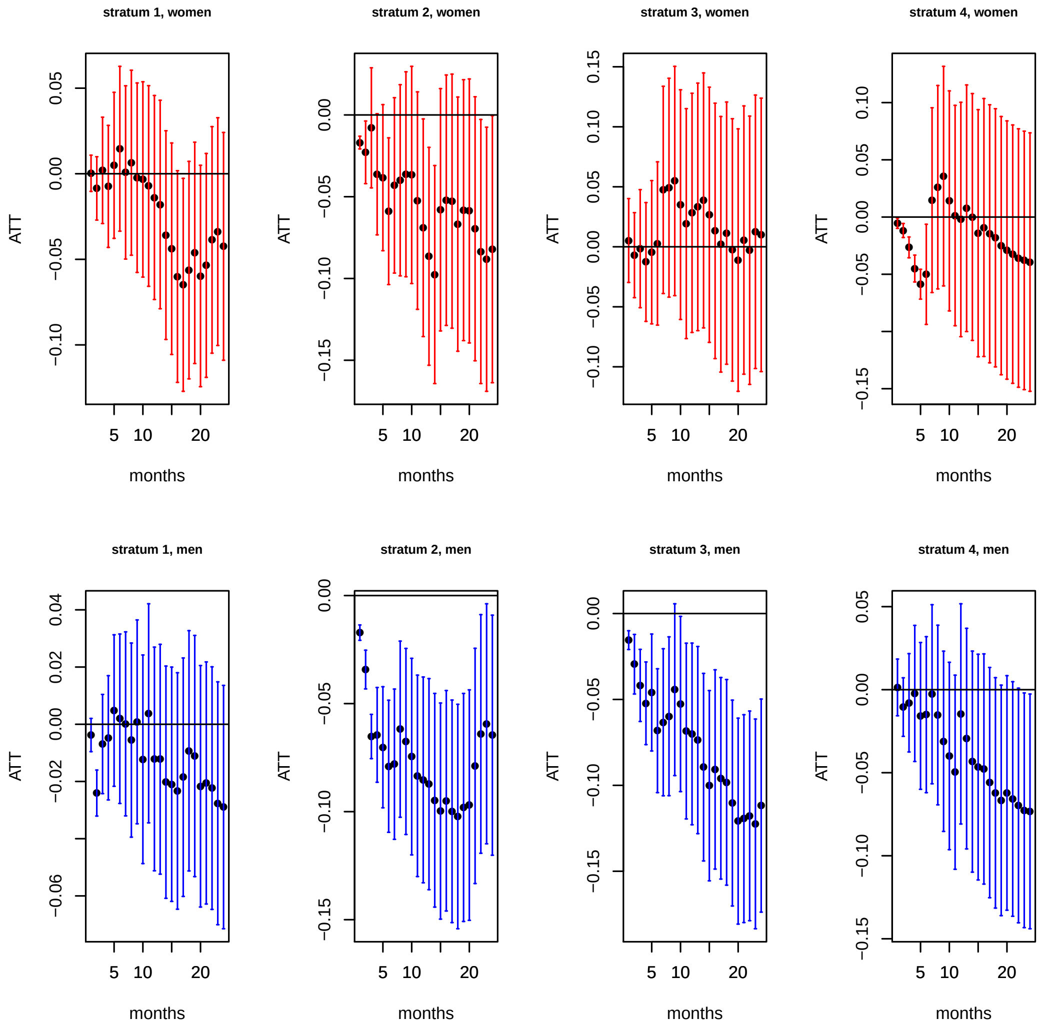


**Figure 1:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

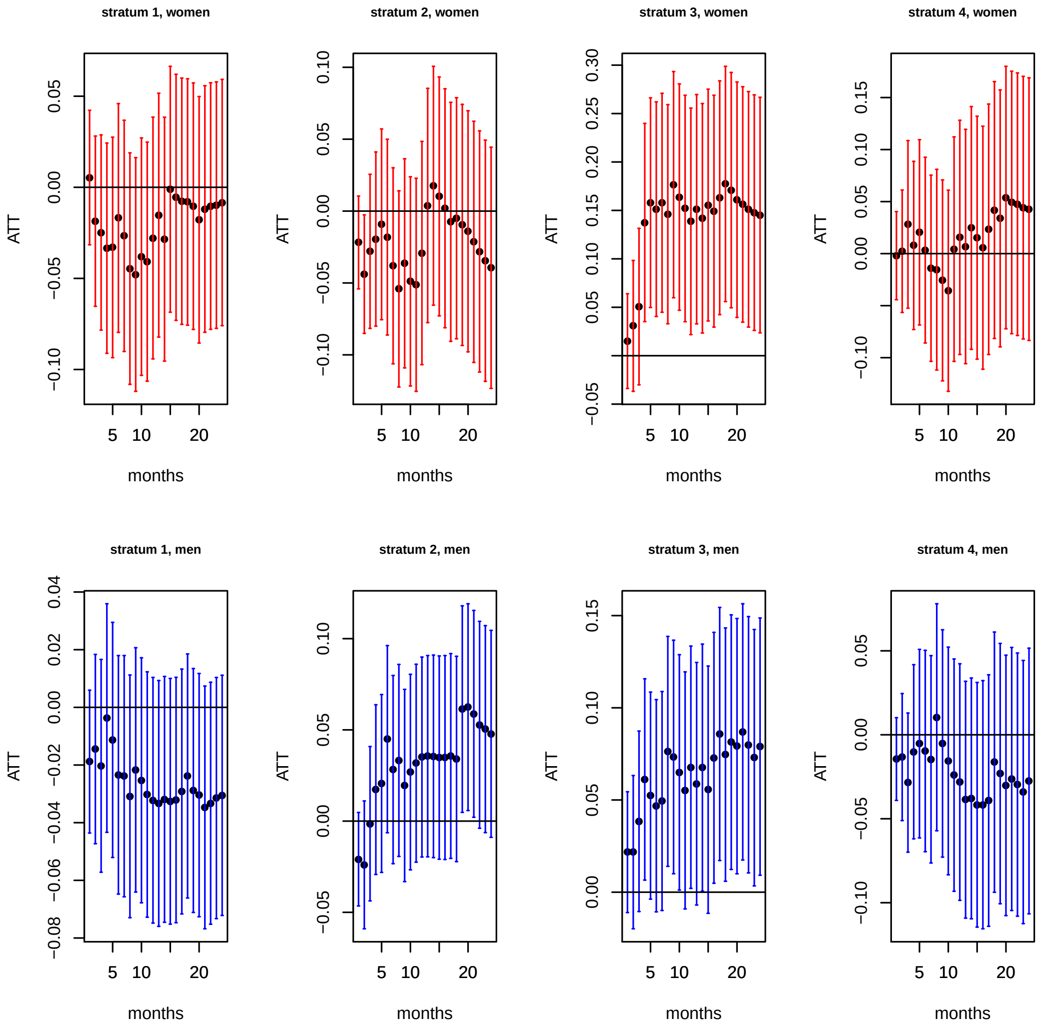


**Figure 2:** The plots show the monthly updated ATT on the probability of exit into *employment with top-up benefits ("supplementary" (S))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

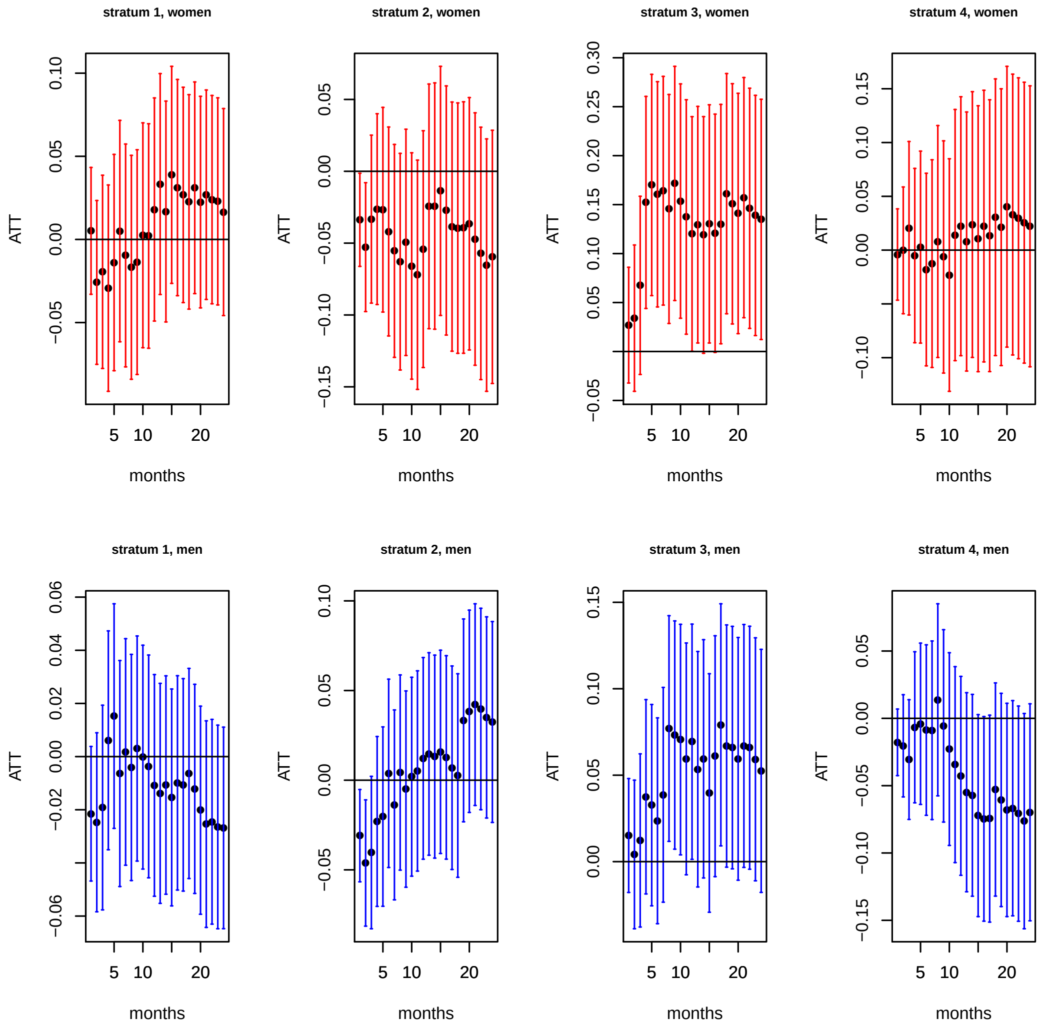


**Figure 3:** The plots show the monthly updated ATT on the probability of exit into *employment ("job in general" (G))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

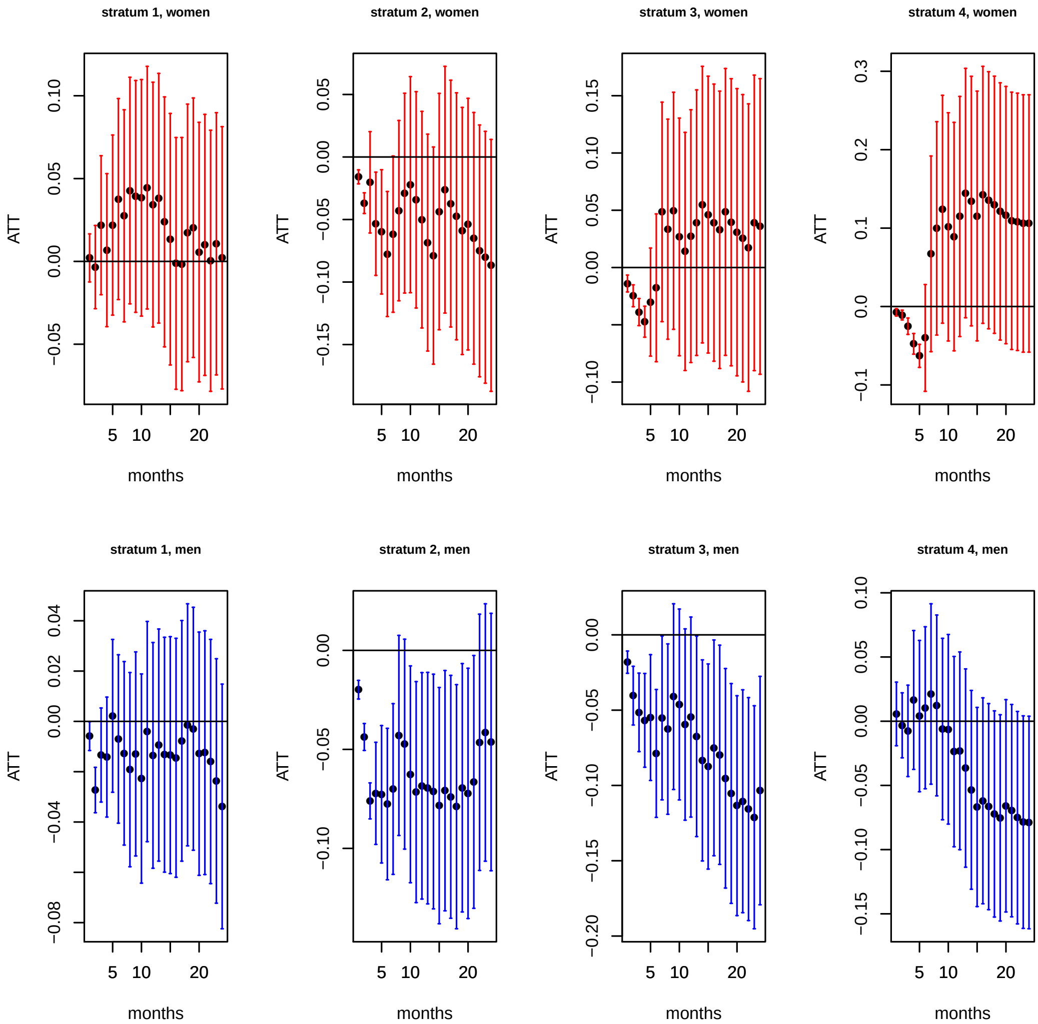


**Figure 4:** The plots show the monthly updated ATT on the probability of exit into *mere employment (“job only” (O))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

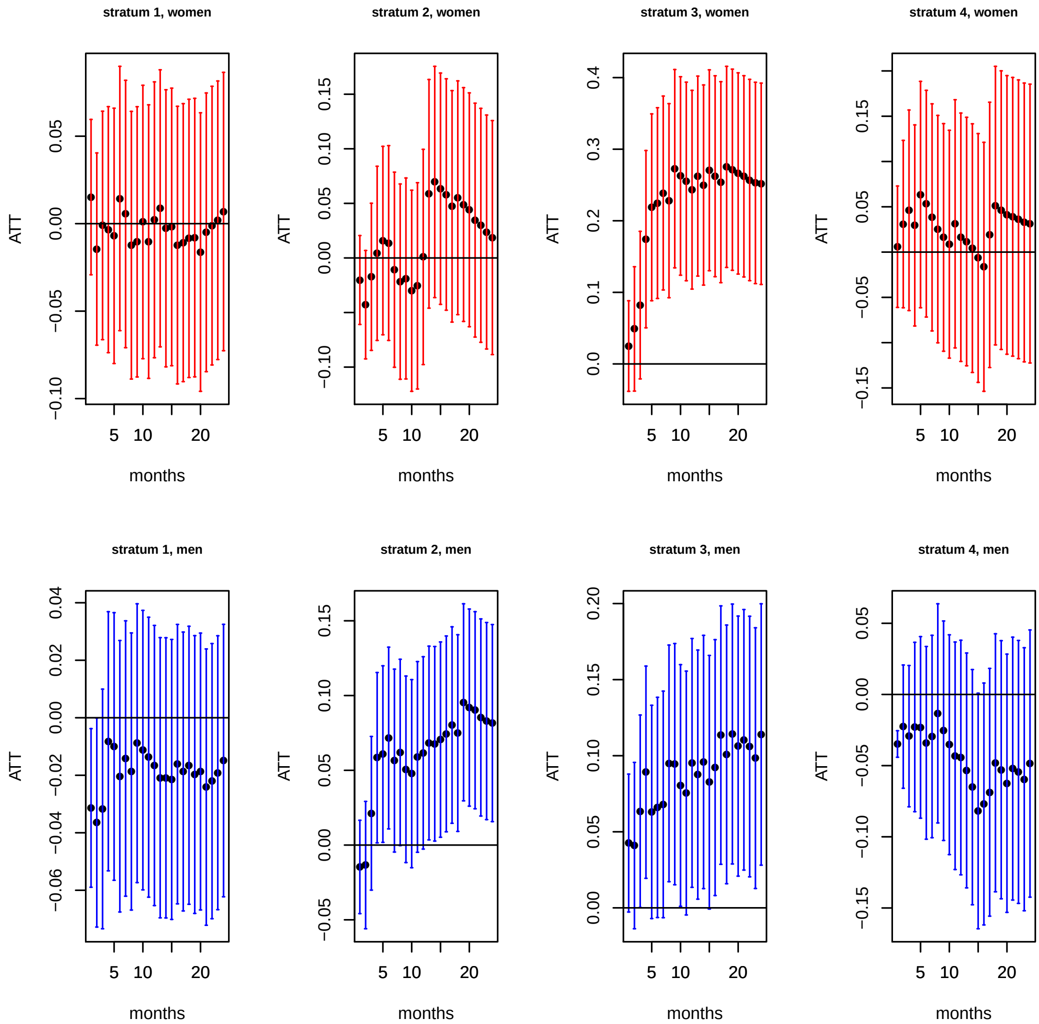


**Figure 5:** The plots show the monthly updated ATT on the probability of exit into *employment with top-up benefits ("supplementary" (S))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

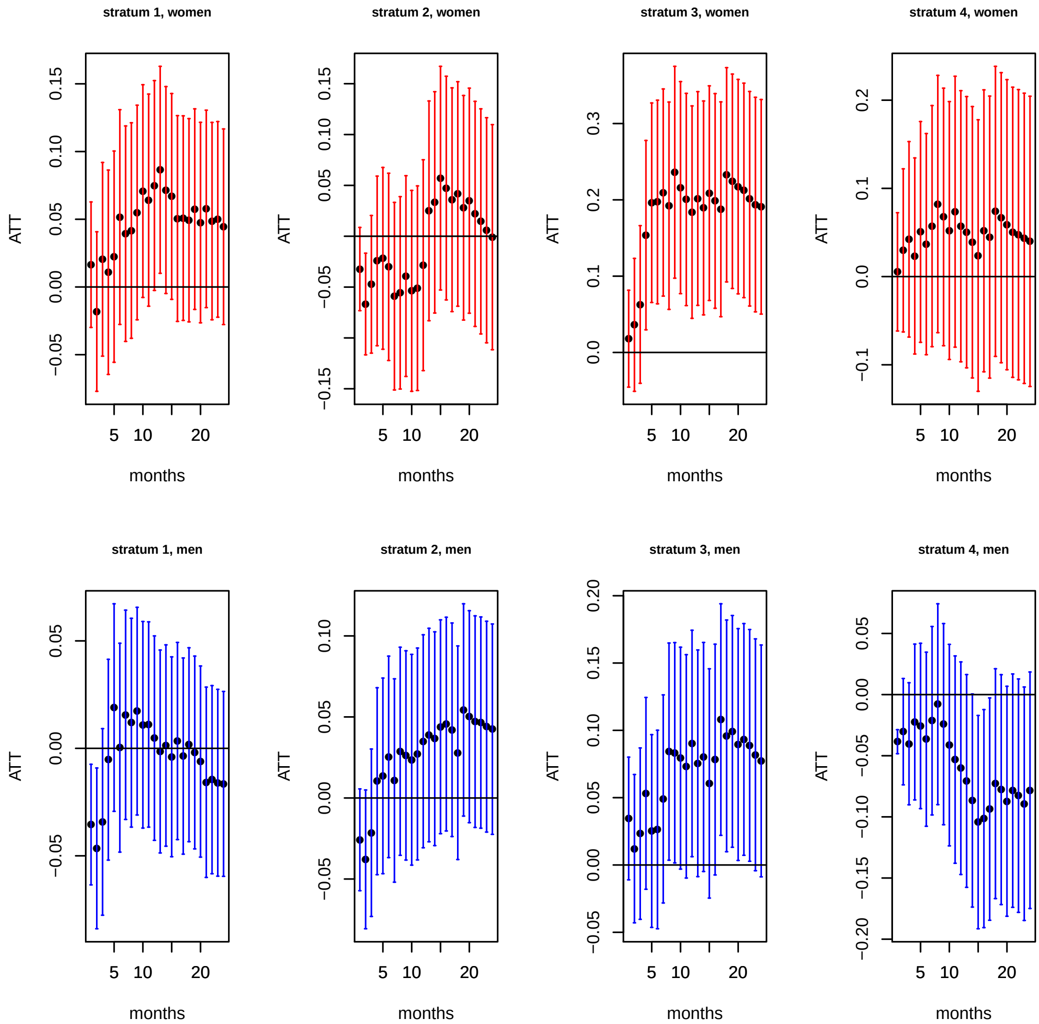


**Figure 6:** The plots show the monthly updated ATT on the probability of exit into *employment ("job in general" (G))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

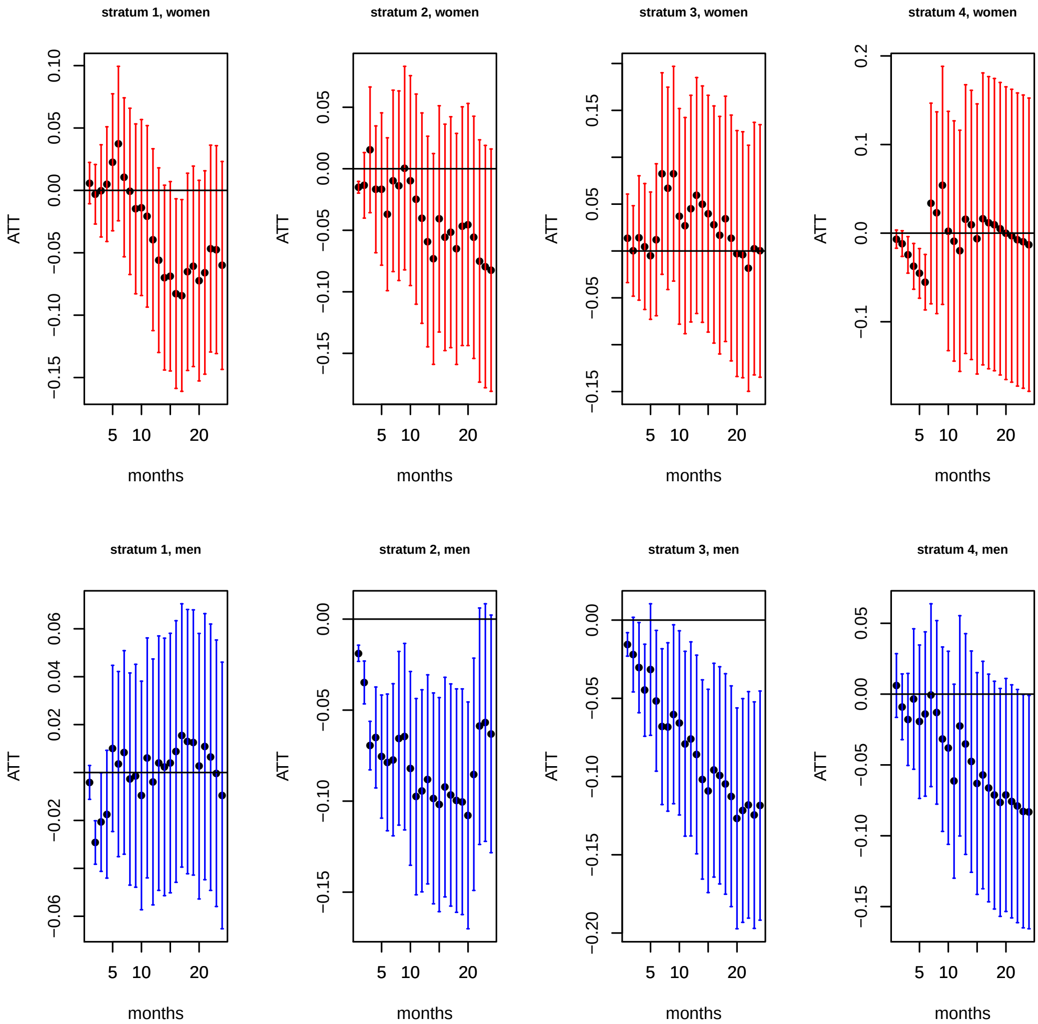


**Figure 7:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients with *medium-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

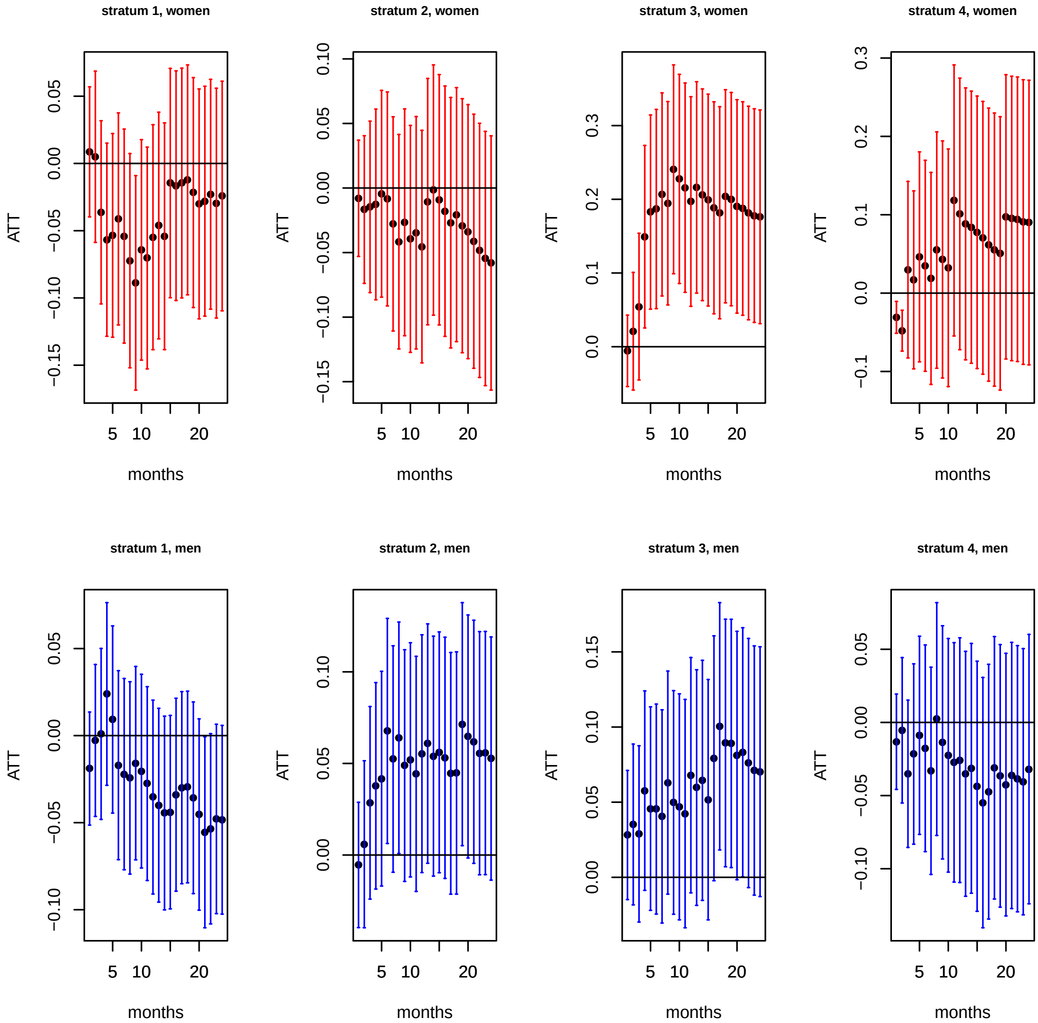


**Figure 8:** The plots show the monthly updated ATT on the probability of exit into *employment with top-up benefits ("supplementary" (S))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients with *medium-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

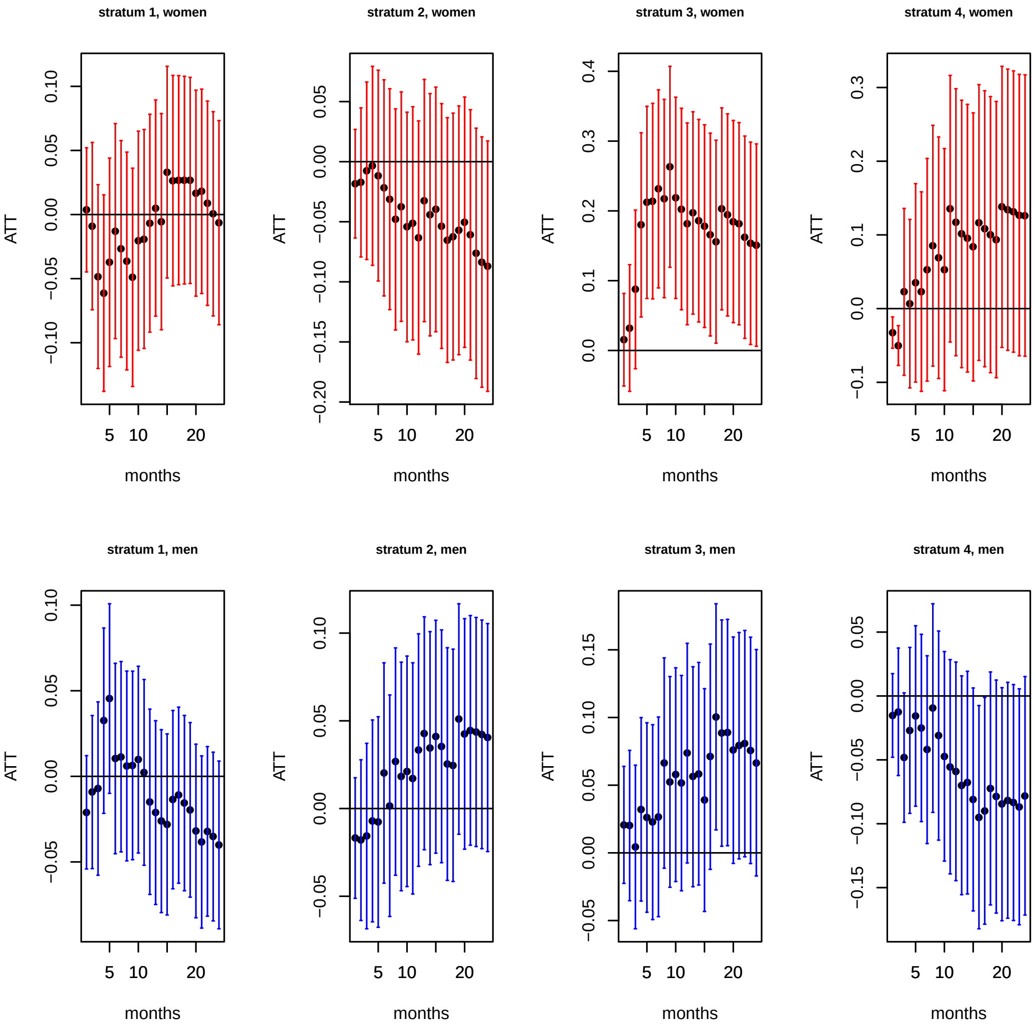


**Figure 9:** The plots show the monthly updated ATT on the probability of exit into *employment ("job in general" (G))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients with *medium-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

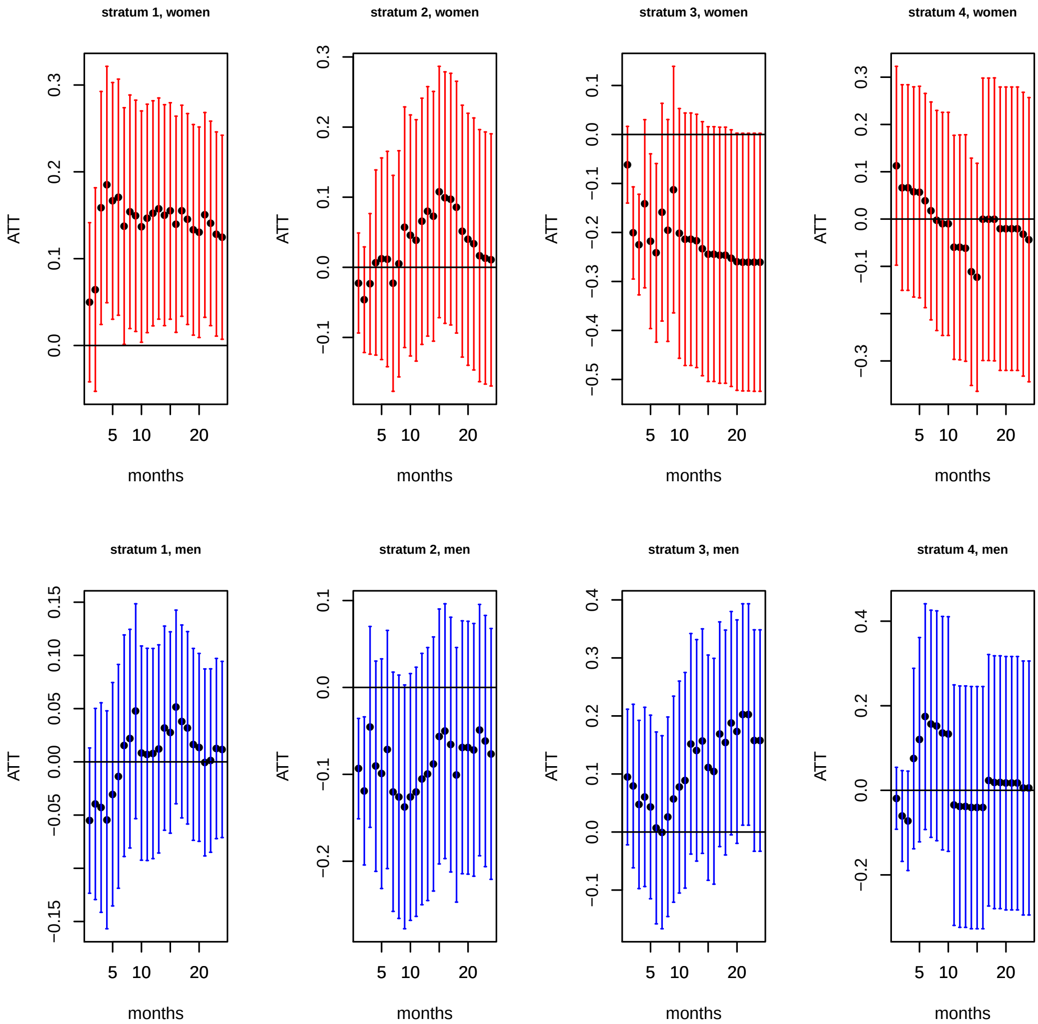


**Figure 10:** The plots show the monthly updated ATT on the probability of exit into *employment ("job in general" (G))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *under 25 years* (u25) in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

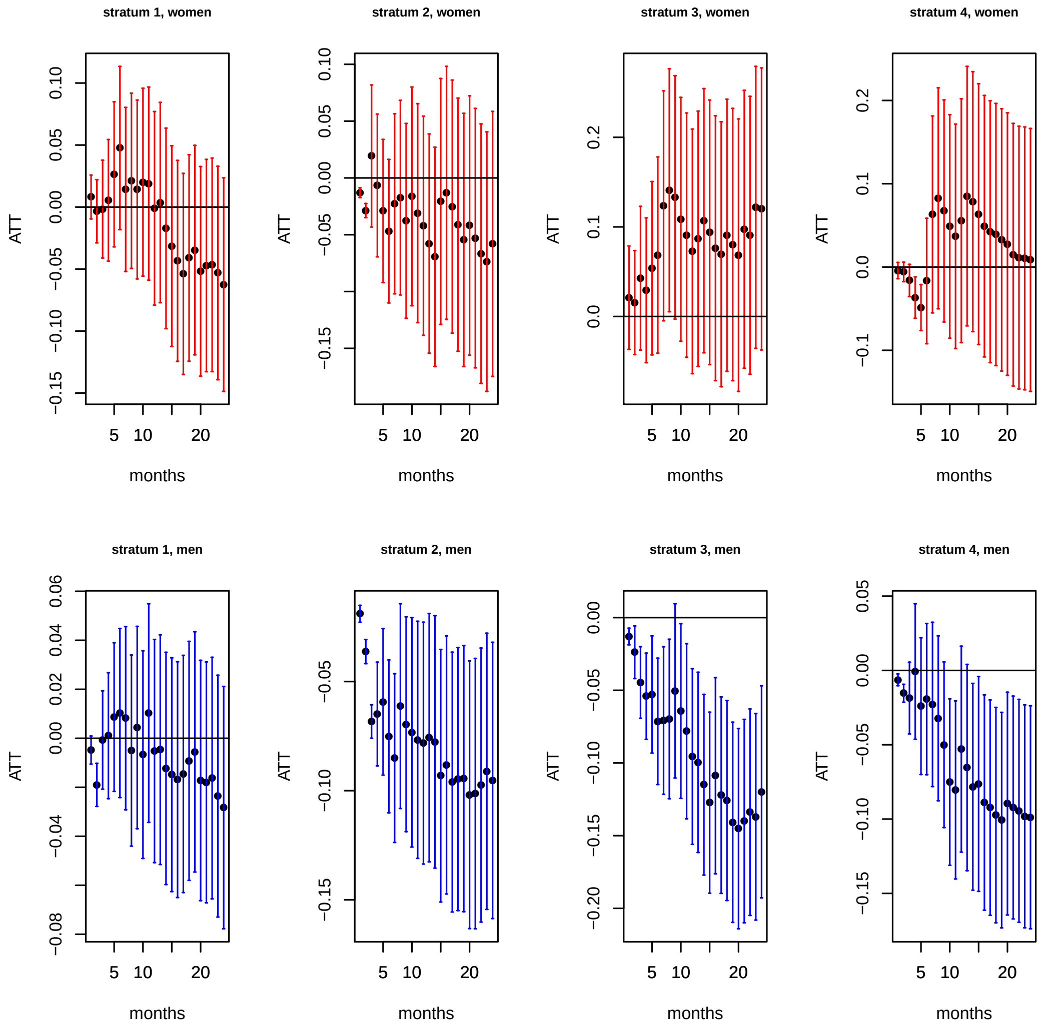


**Figure 11:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

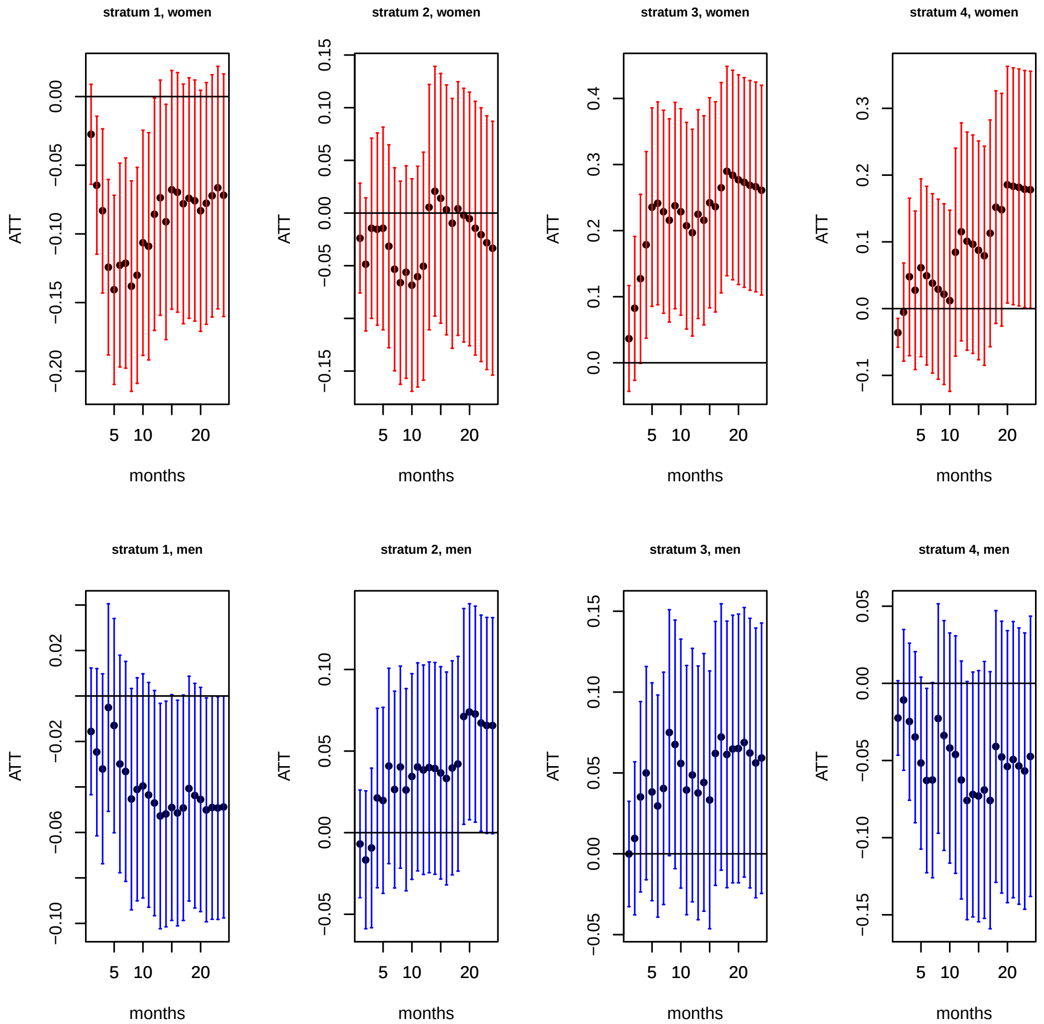


**Figure 12:** The plots show the monthly updated ATT on the probability of exit into *employment with top-up benefits ("supplementary" (S))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

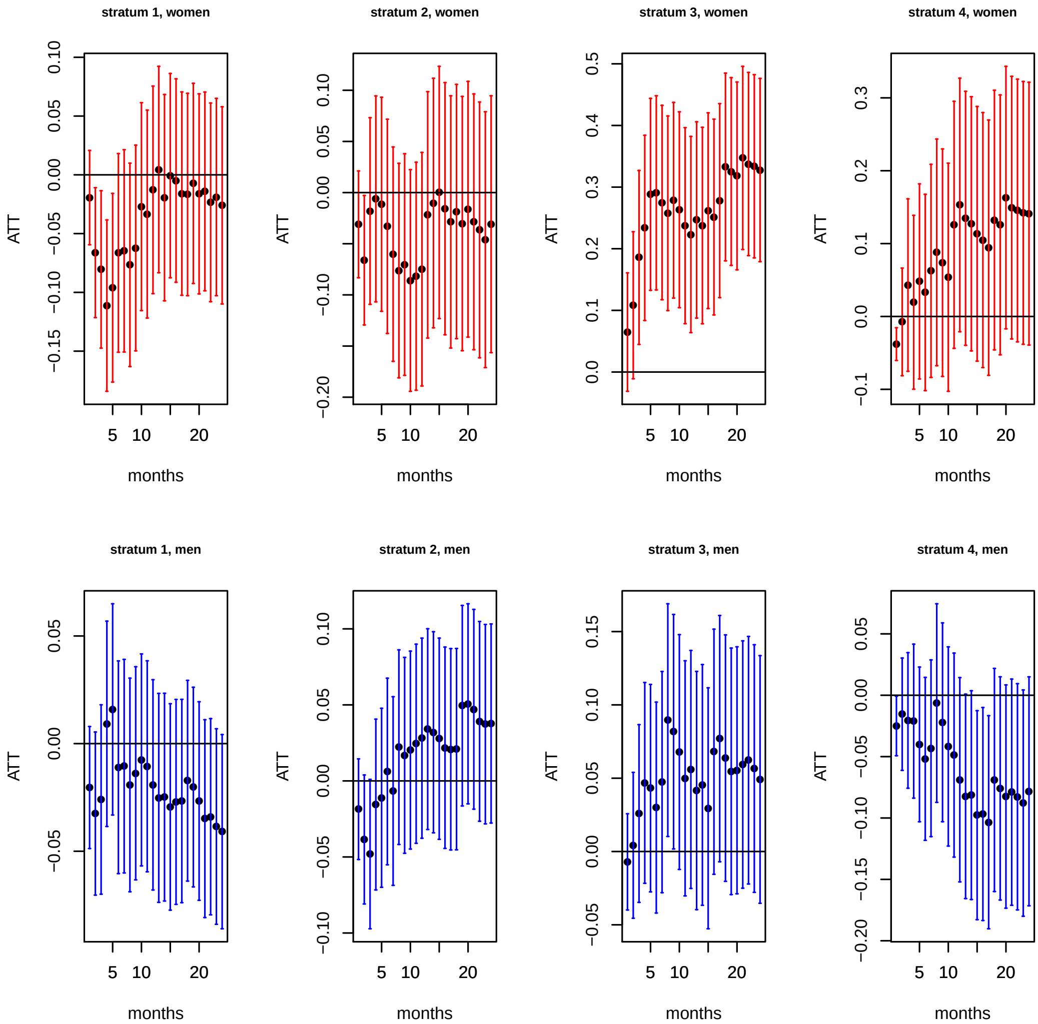


**Figure 13:** The plots show the monthly updated ATT on the probability of exit into *employment ("job in general" (G))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

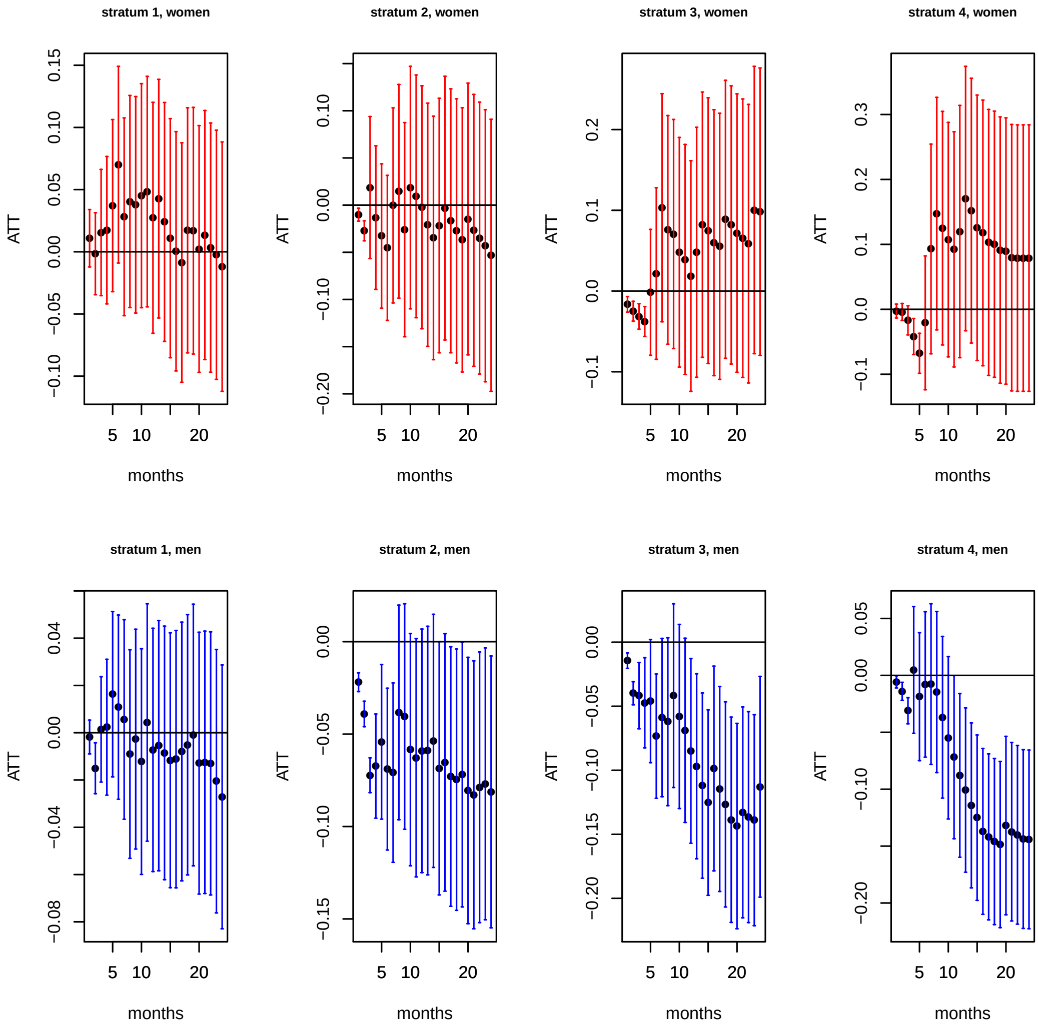


**Figure 14:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

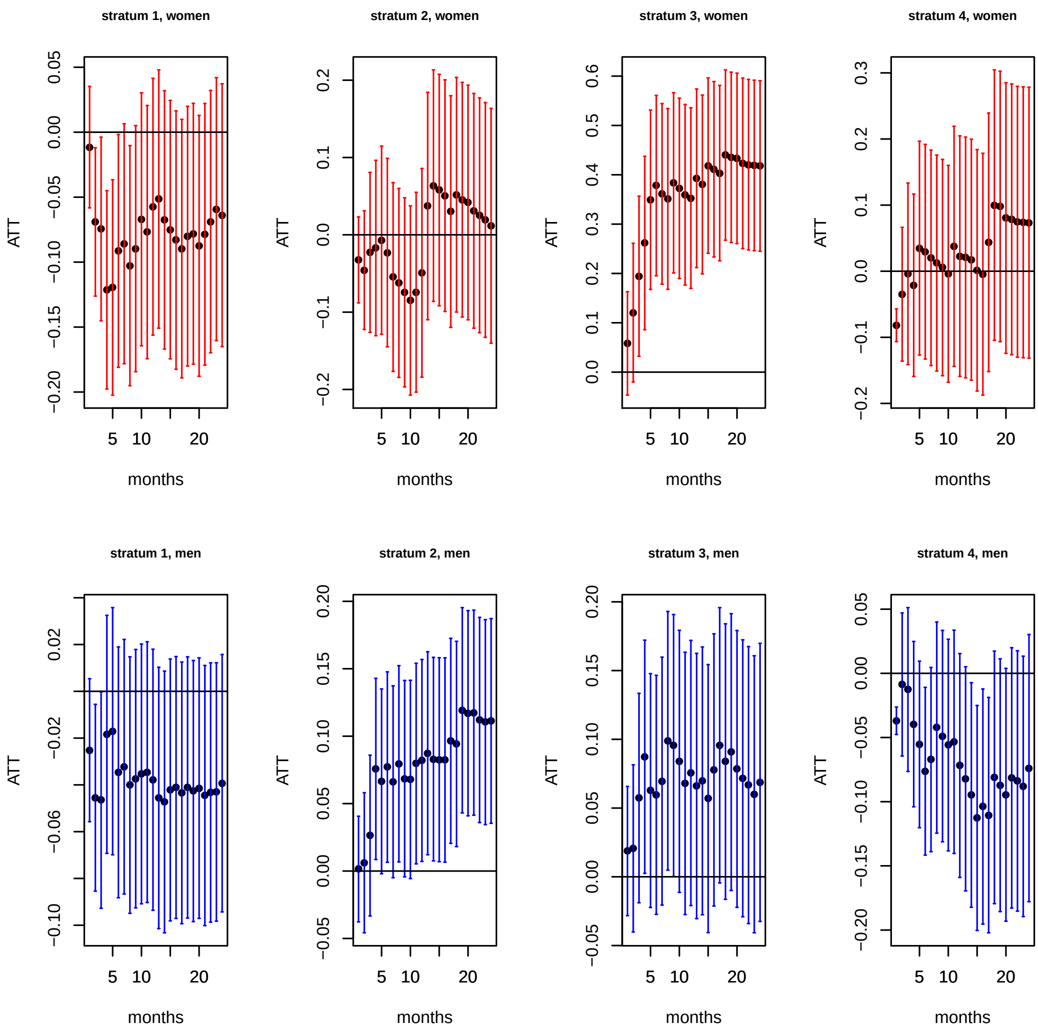


**Figure 15:** The plots show the monthly updated ATT on the probability of exit into *employment with top-up benefits ("supplementary" (S))* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

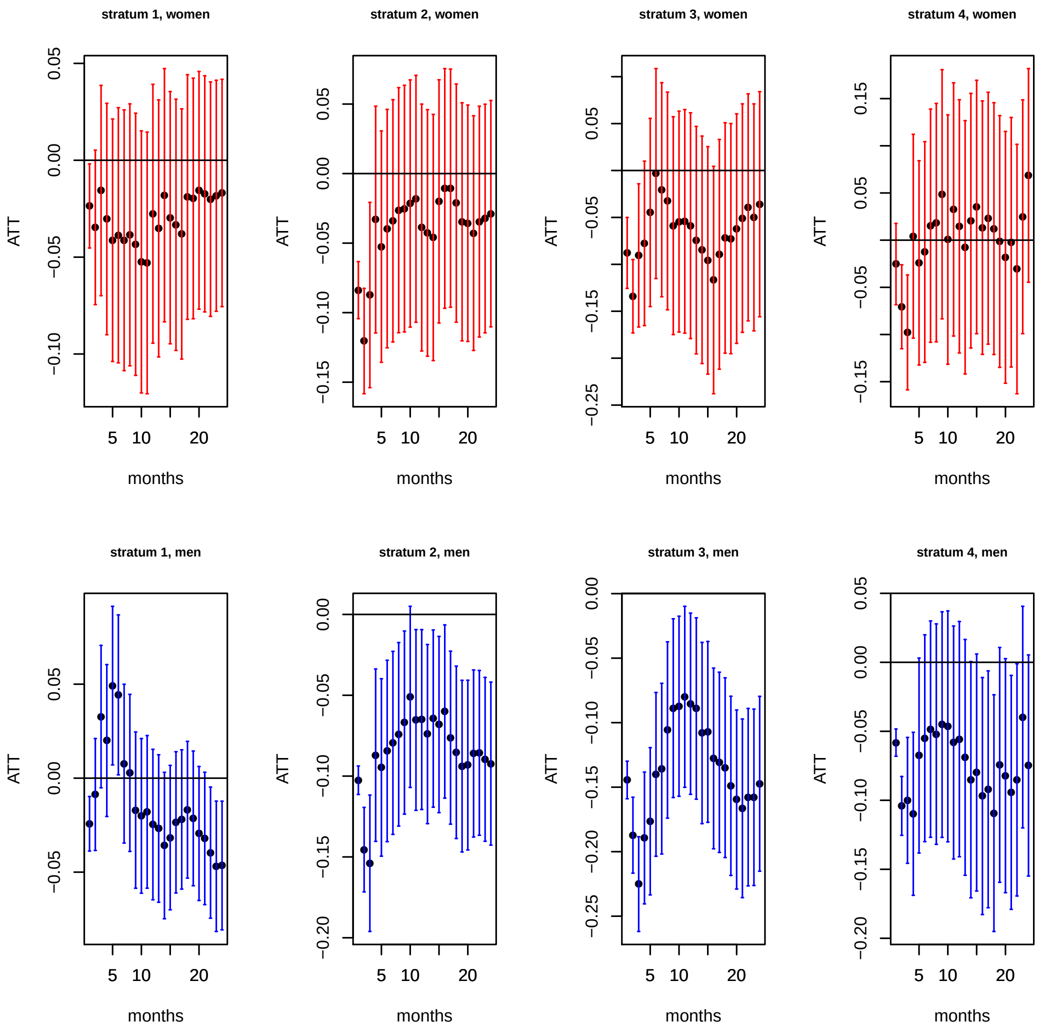


**Figure 16:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

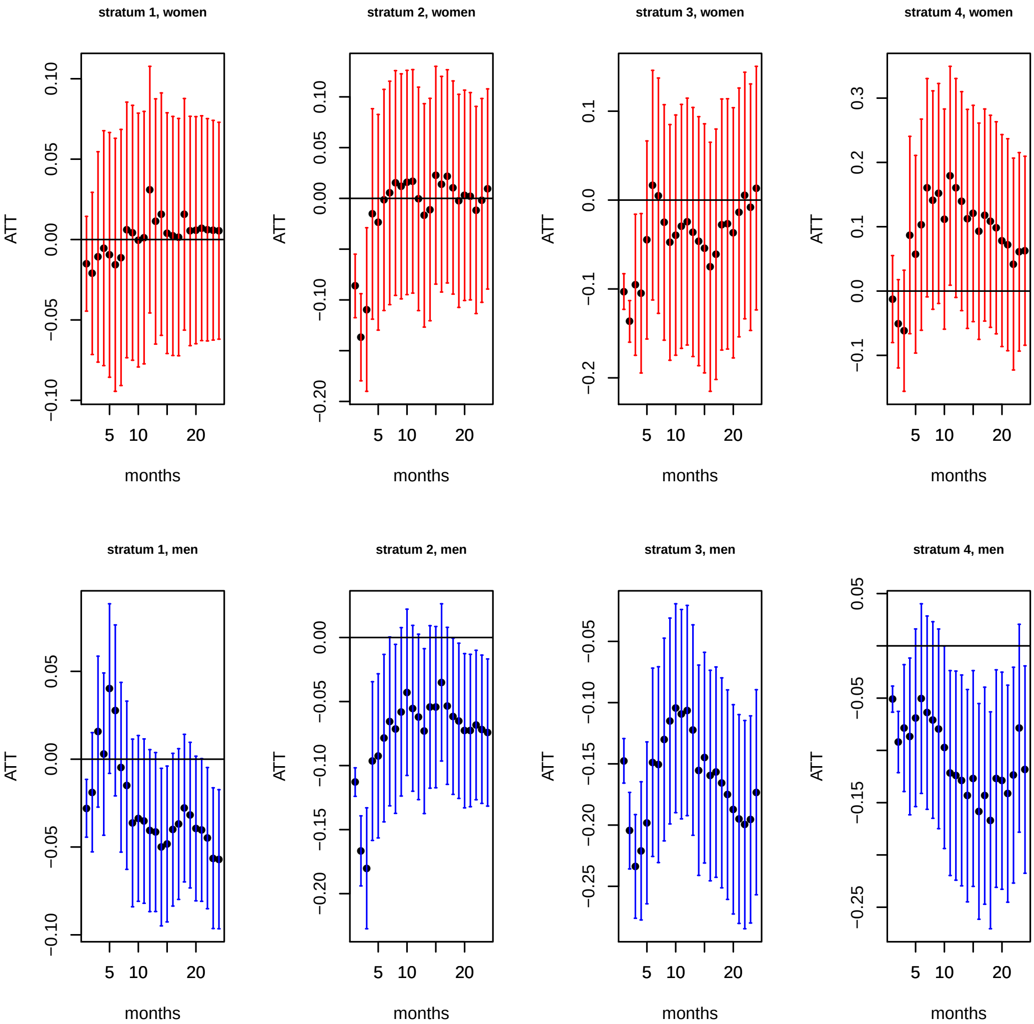


**Figure 17:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

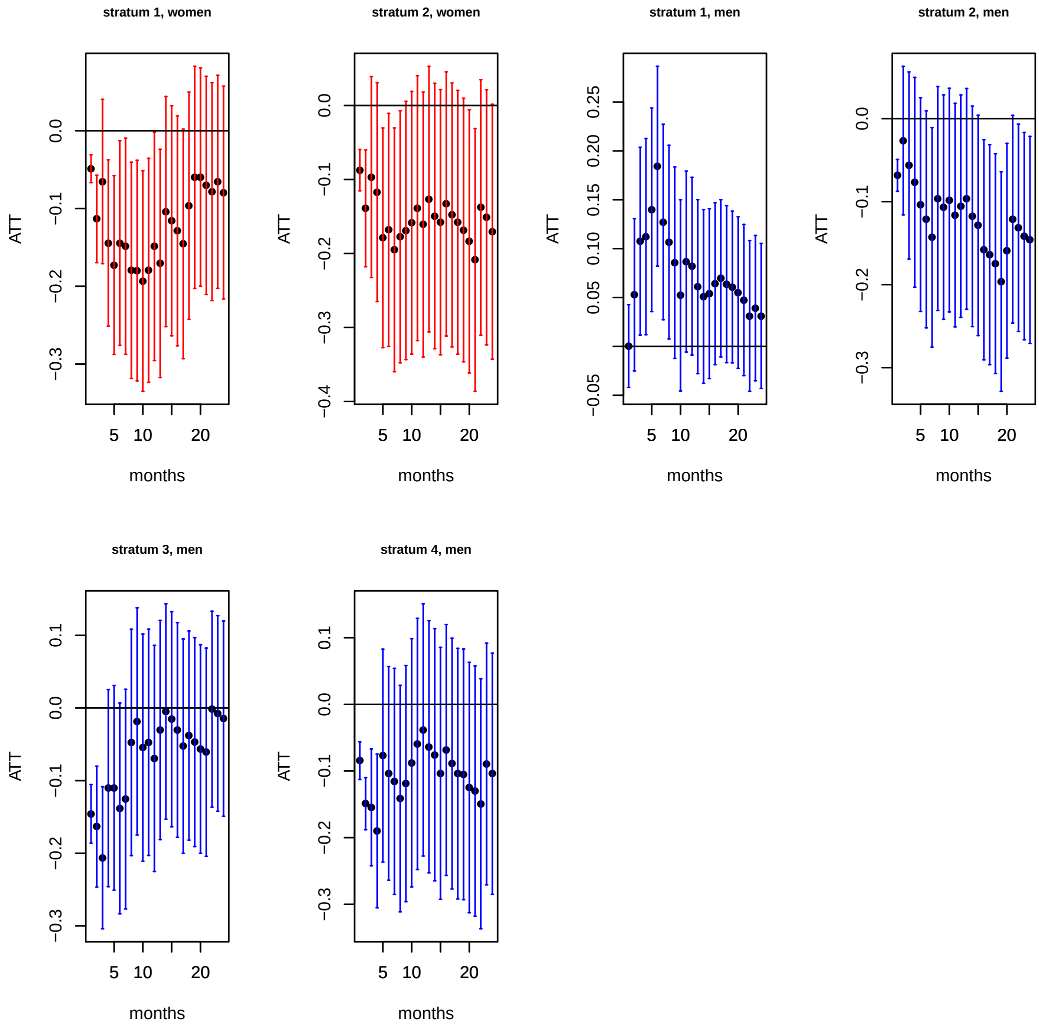


**Figure 18:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients in *Eastern Germany* (EG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

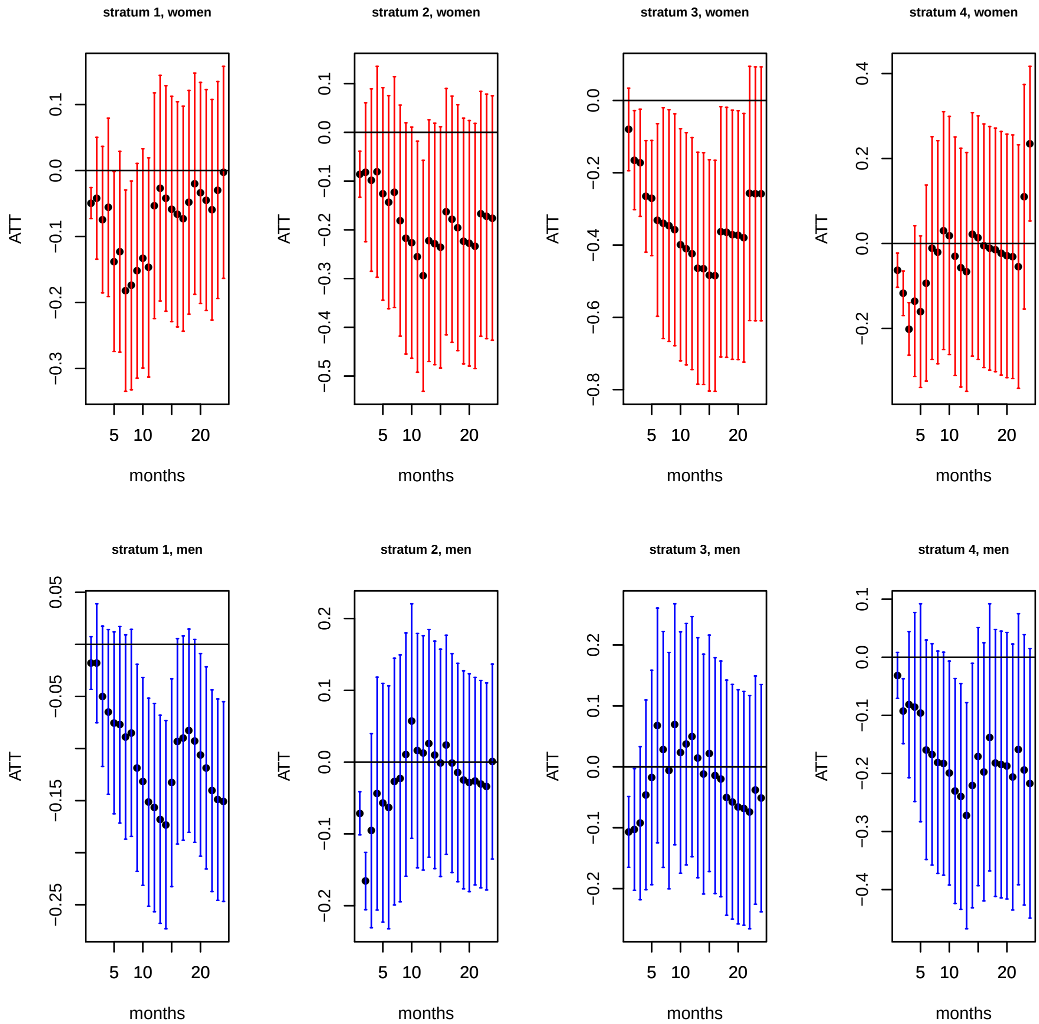


**Figure 19:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients with *low-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

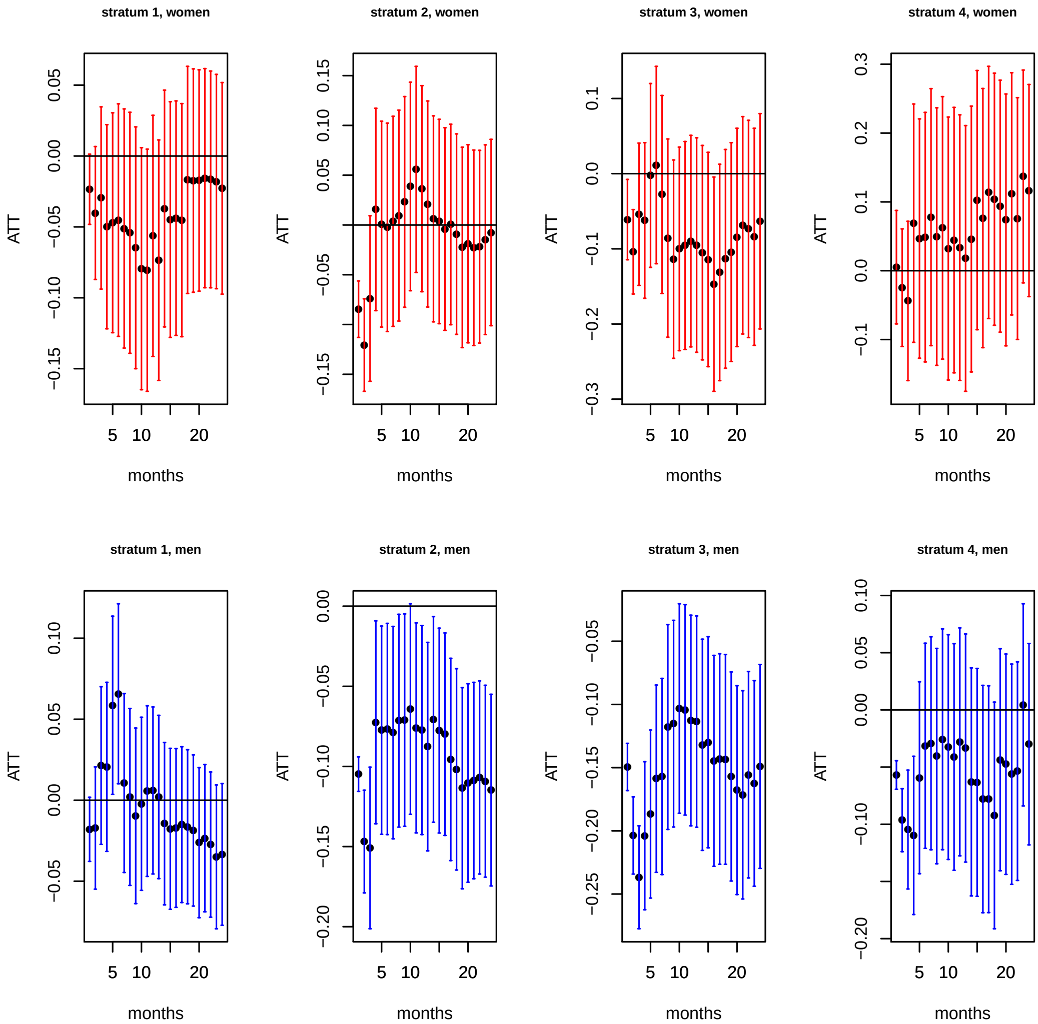


**Figure 20:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients with *medium-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

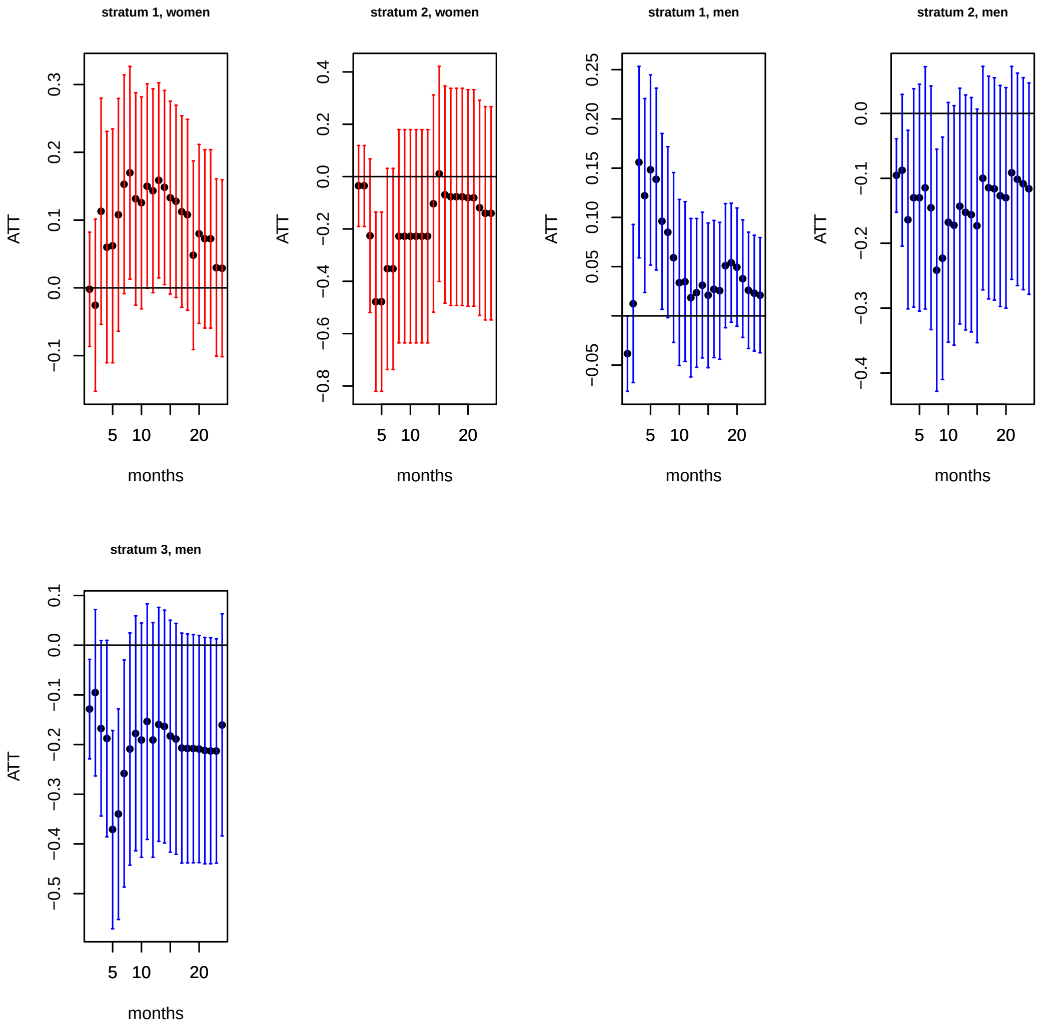


**Figure 21:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients with *high-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

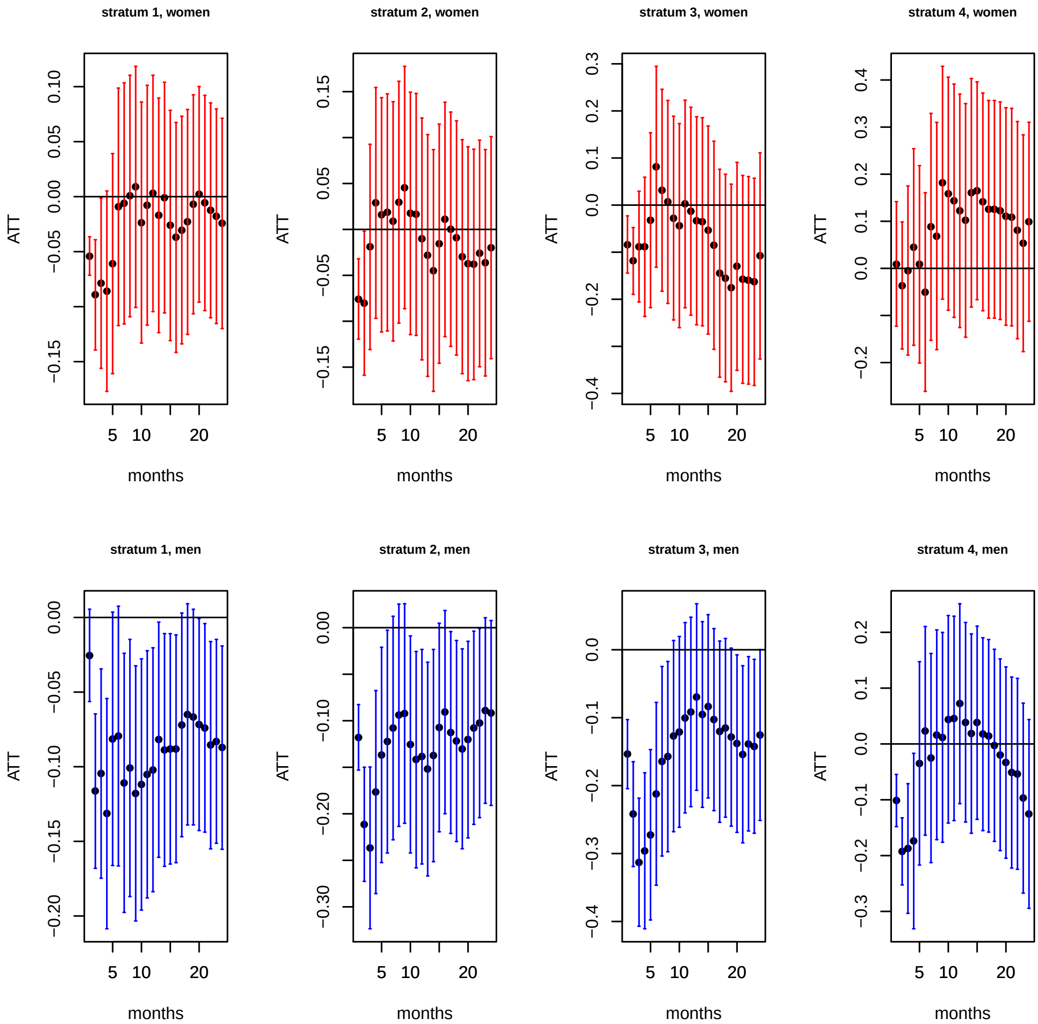


**Figure 22:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *under 25 years* (u25) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

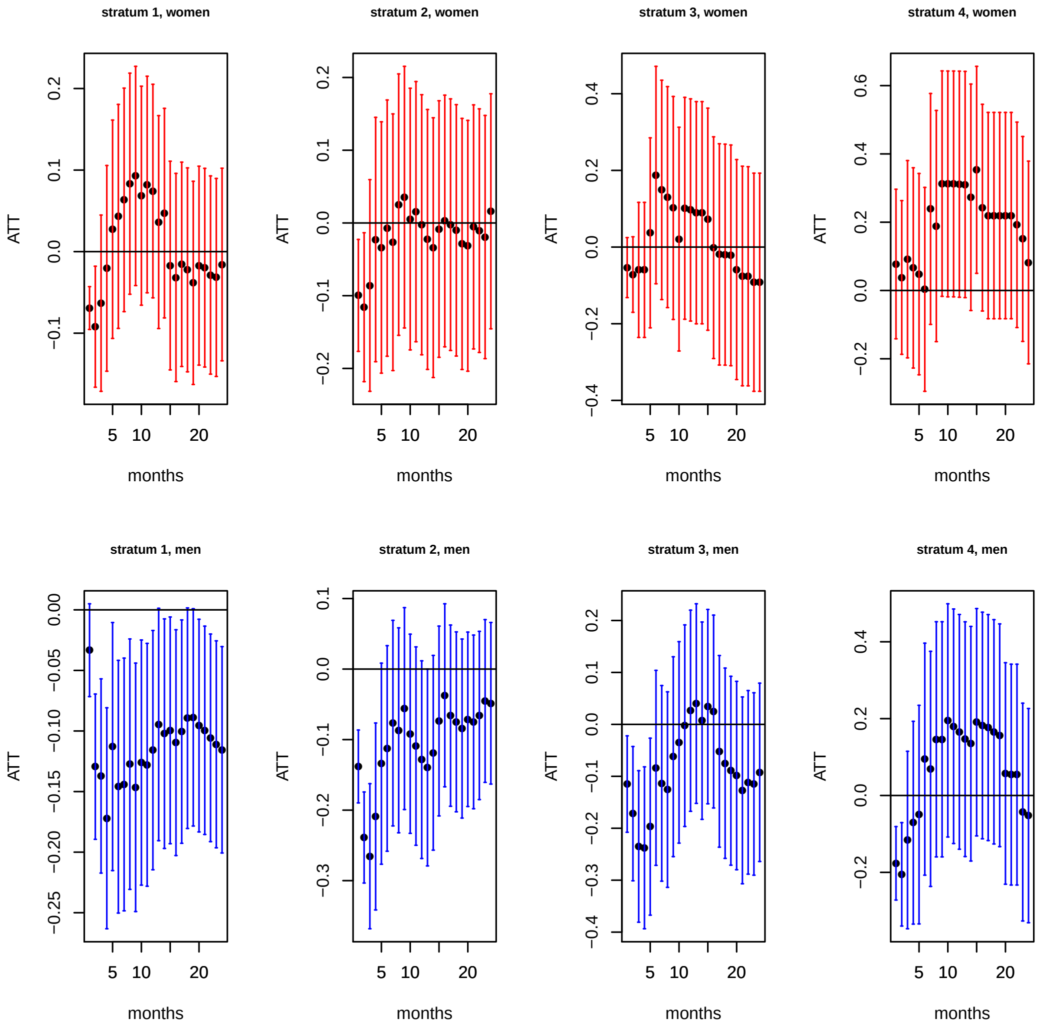


**Figure 23:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *under 25 years* (u25) in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

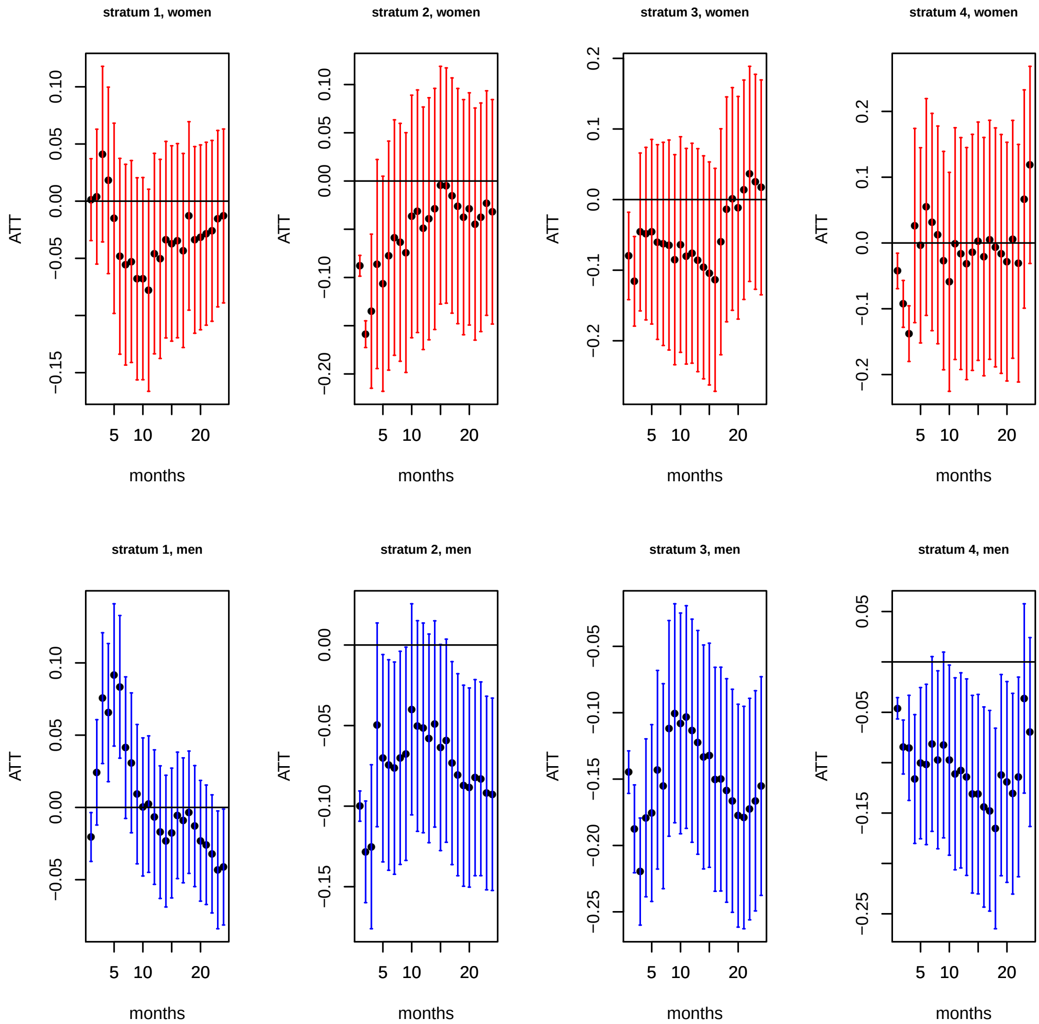


**Figure 24:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

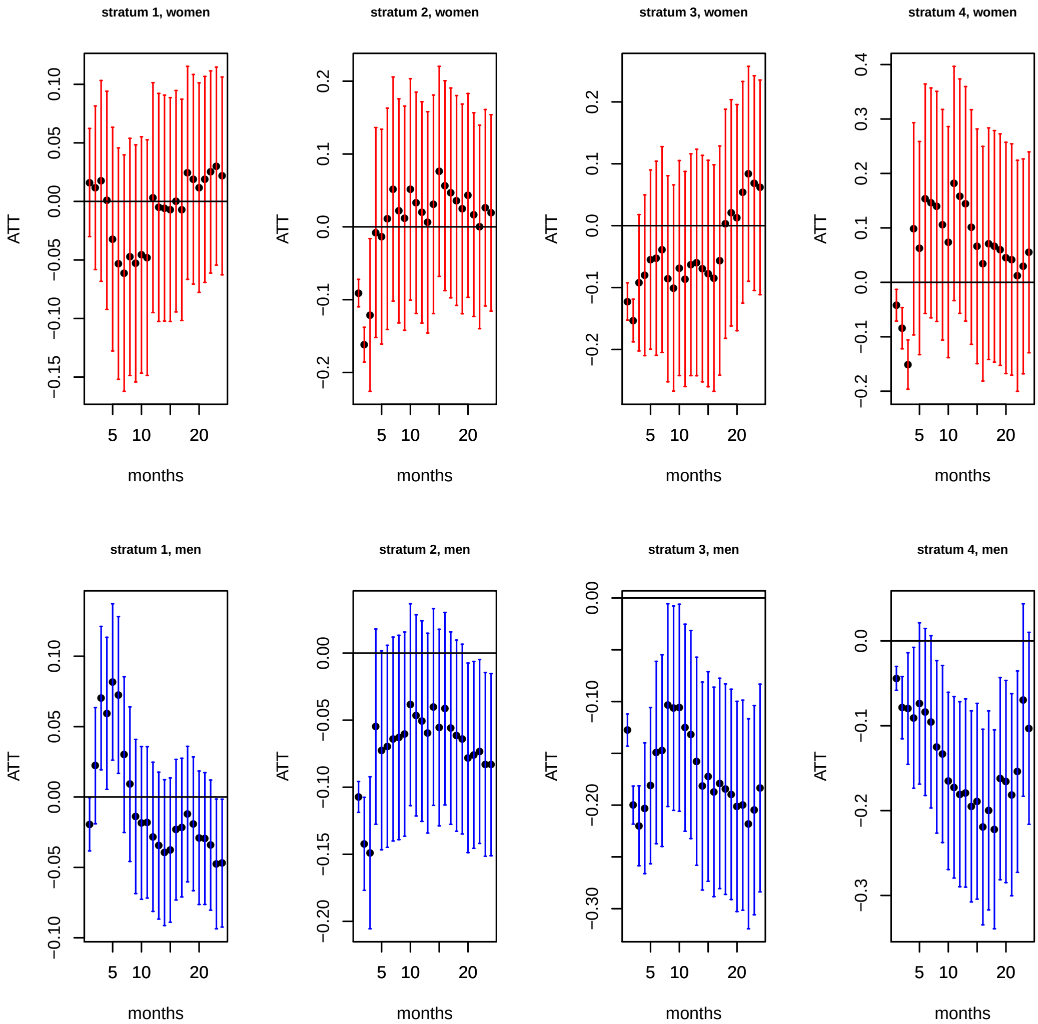


**Figure 25:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

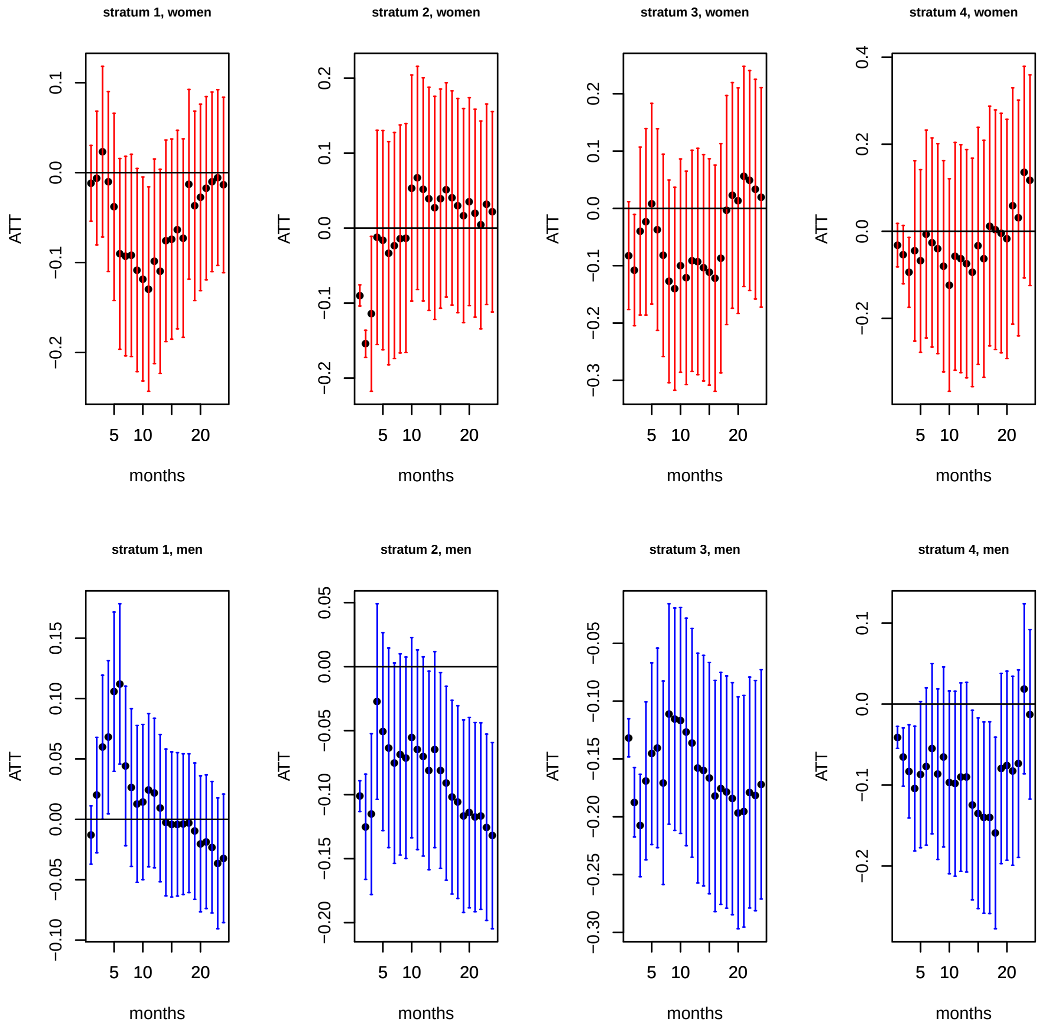


**Figure 26:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* with *medium-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

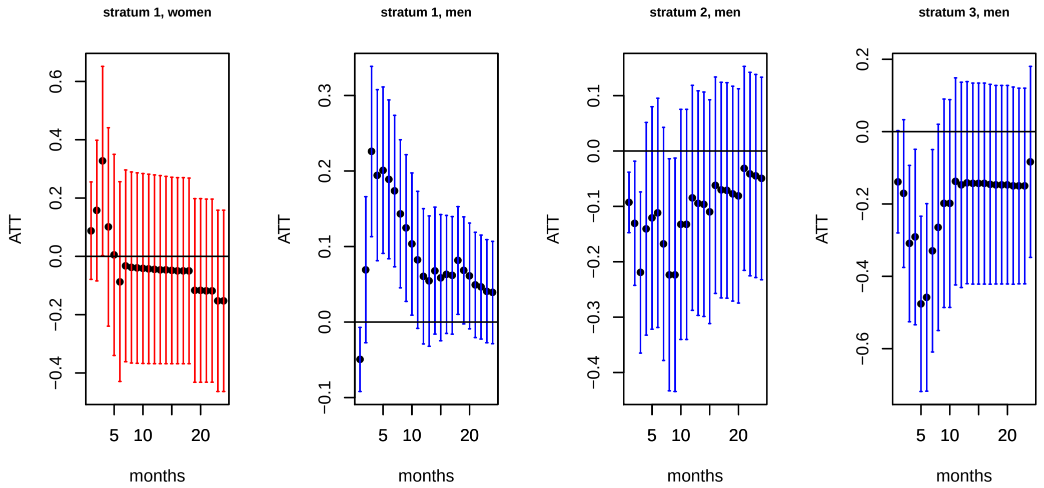


**Figure 27:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *unemployed (UE)* welfare recipients *over 25 years* (o25) with *high-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

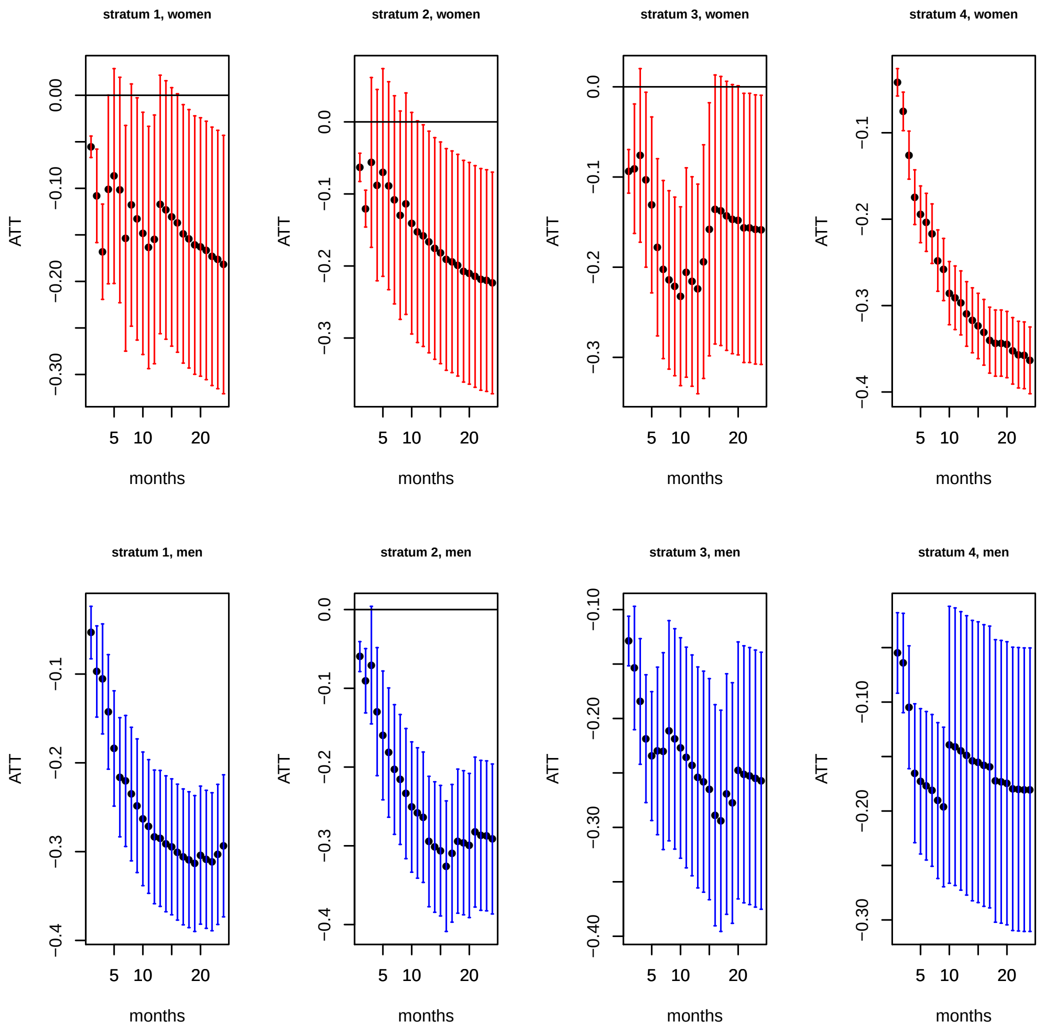


**Figure 28:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

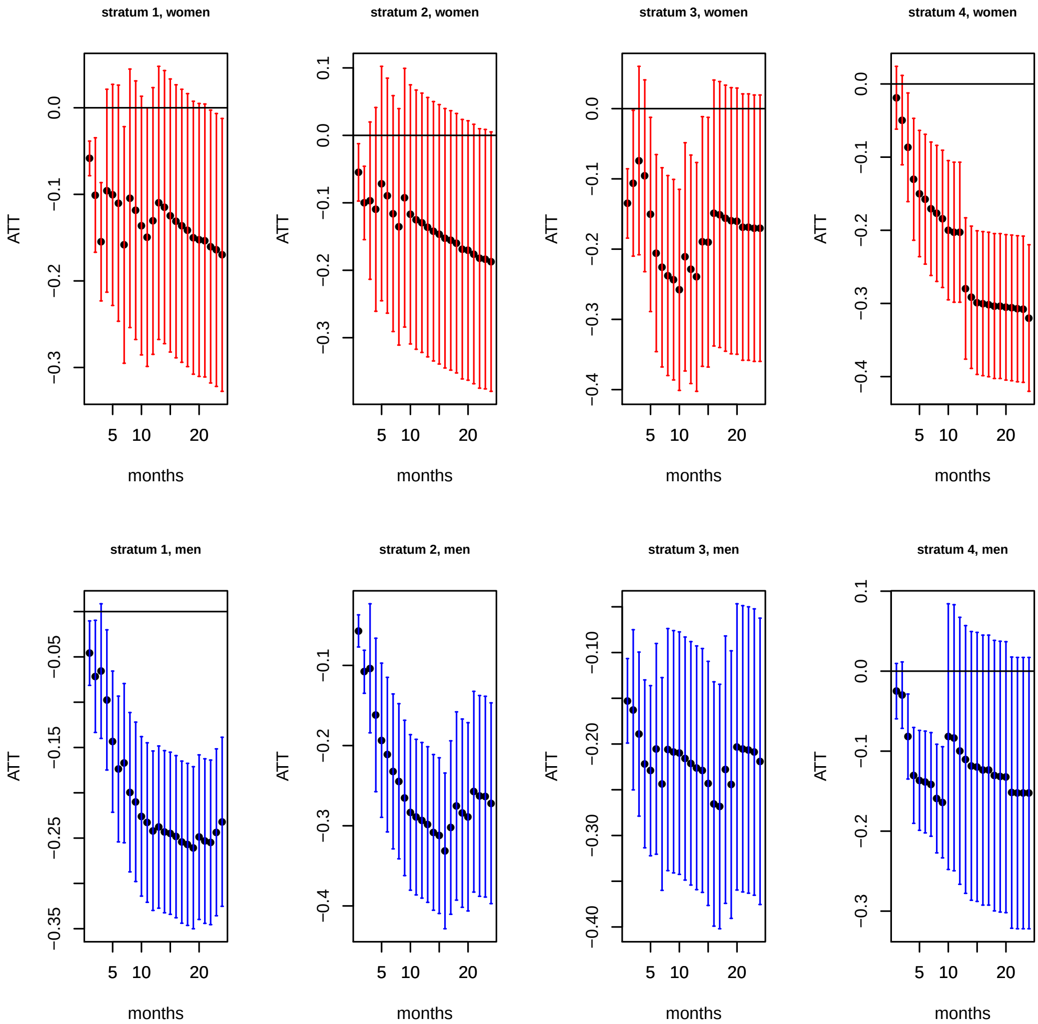


**Figure 29:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

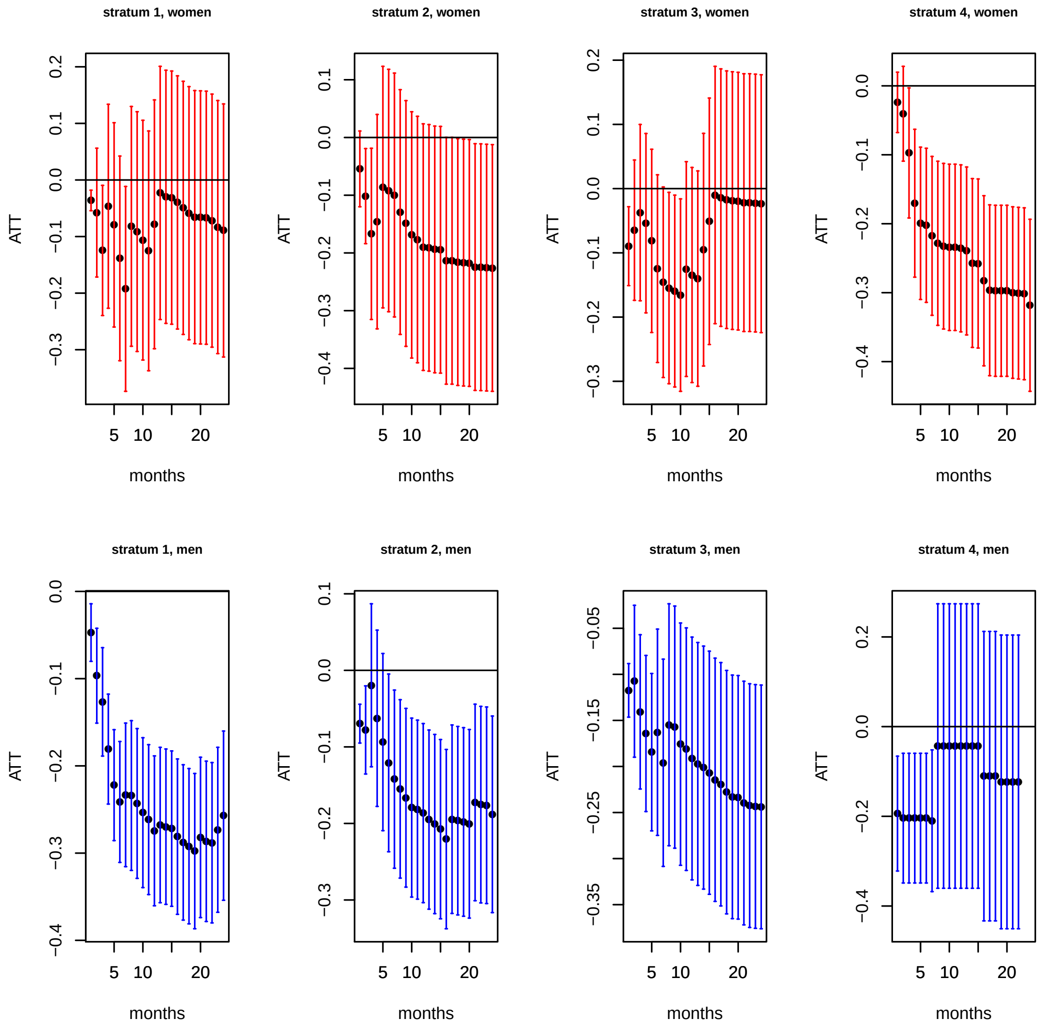


**Figure 30:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients with *medium-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

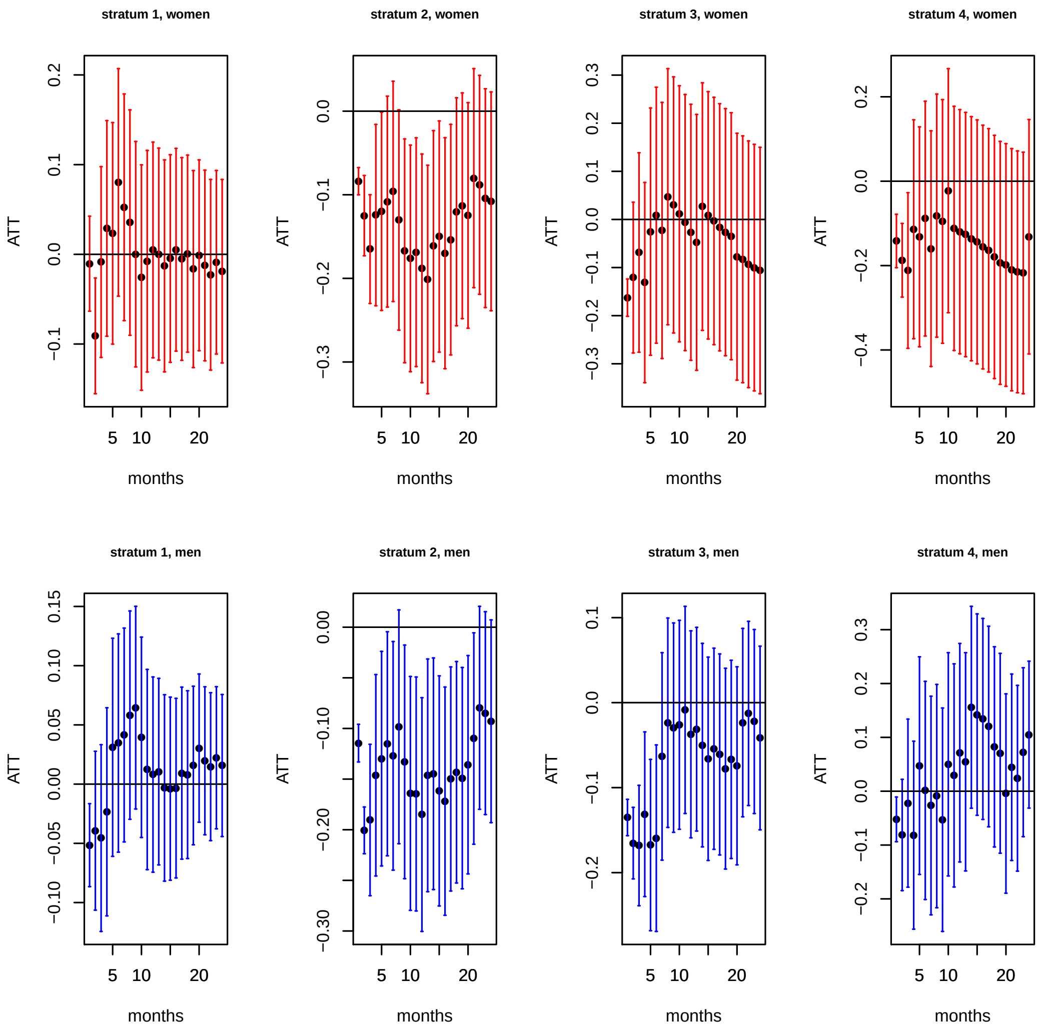


**Figure 31:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

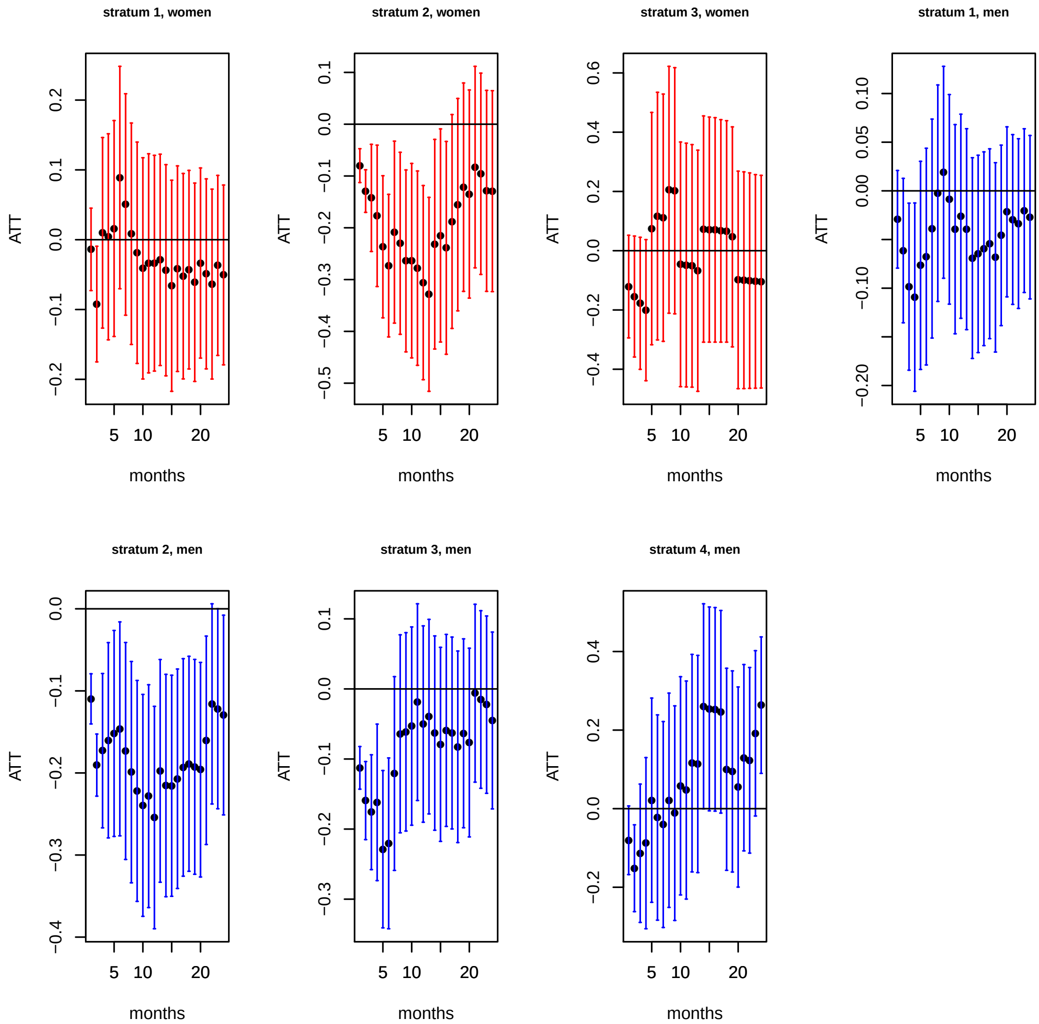


**Figure 32:** The plots show the monthly updated ATT on the probability of *exiting welfare (ExWel)* and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients with *medium-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

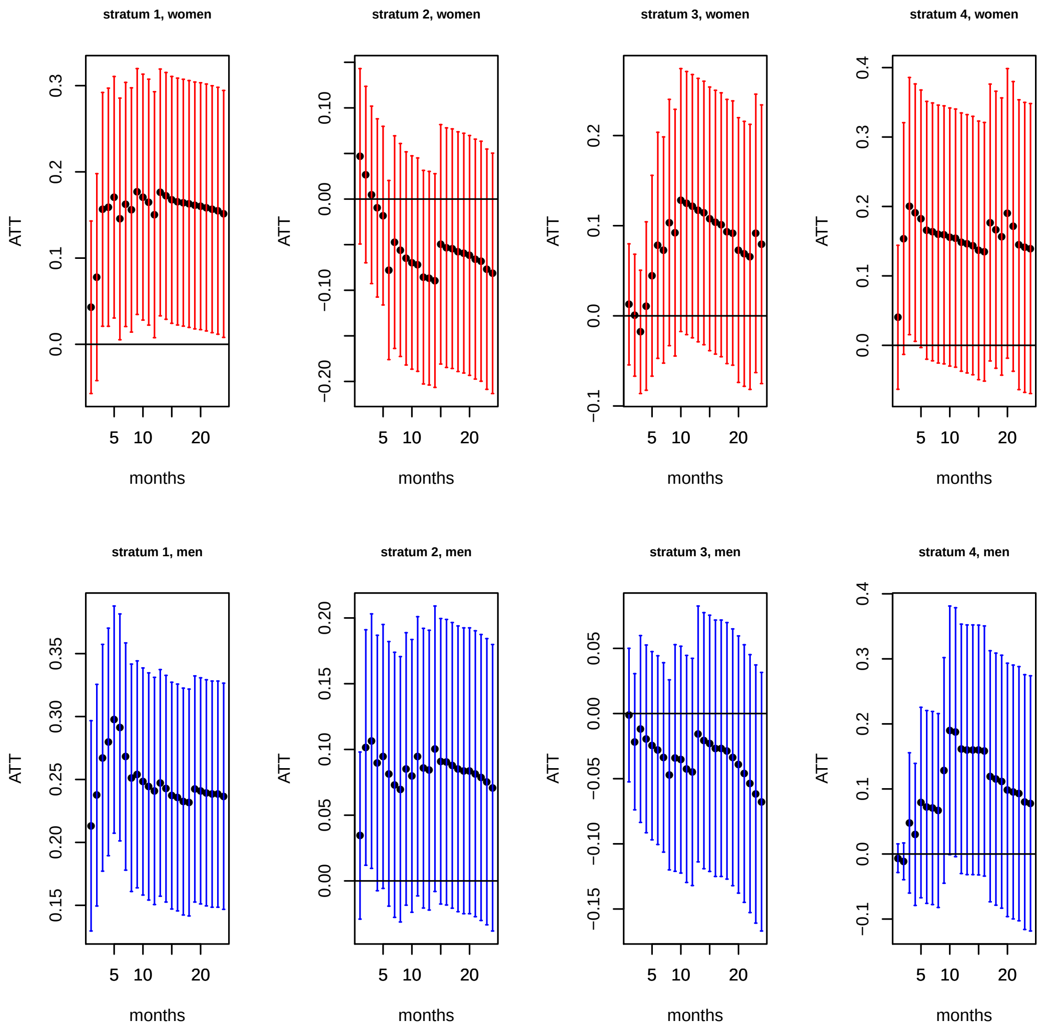


**Figure 33:** The plots show the monthly updated ATT on the probability of *exiting employment (ExJob)* for mere welfare receipt and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

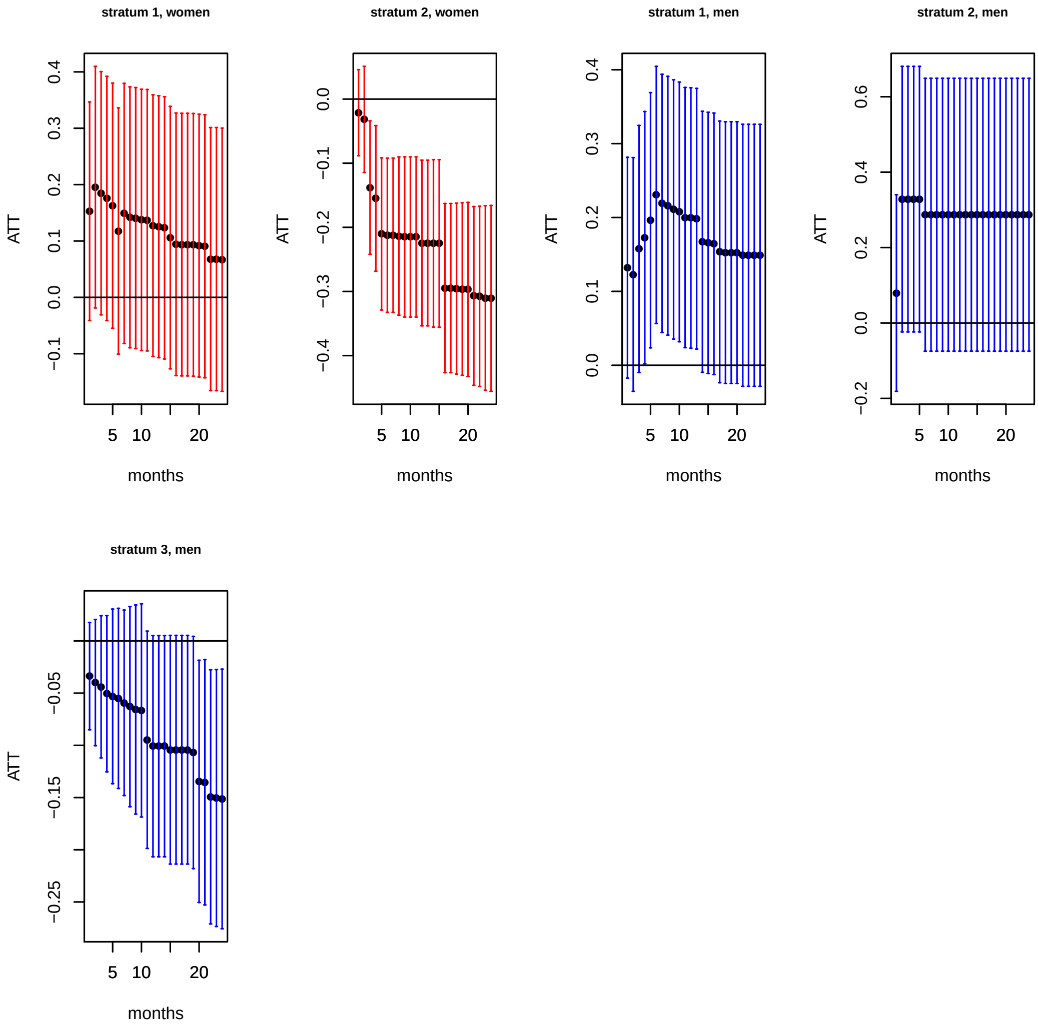


**Figure 34:** The plots show the monthly updated ATT on the probability of *exiting employment (ExJob)* for mere welfare receipt and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients with *high-level labor market access* (LMA) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

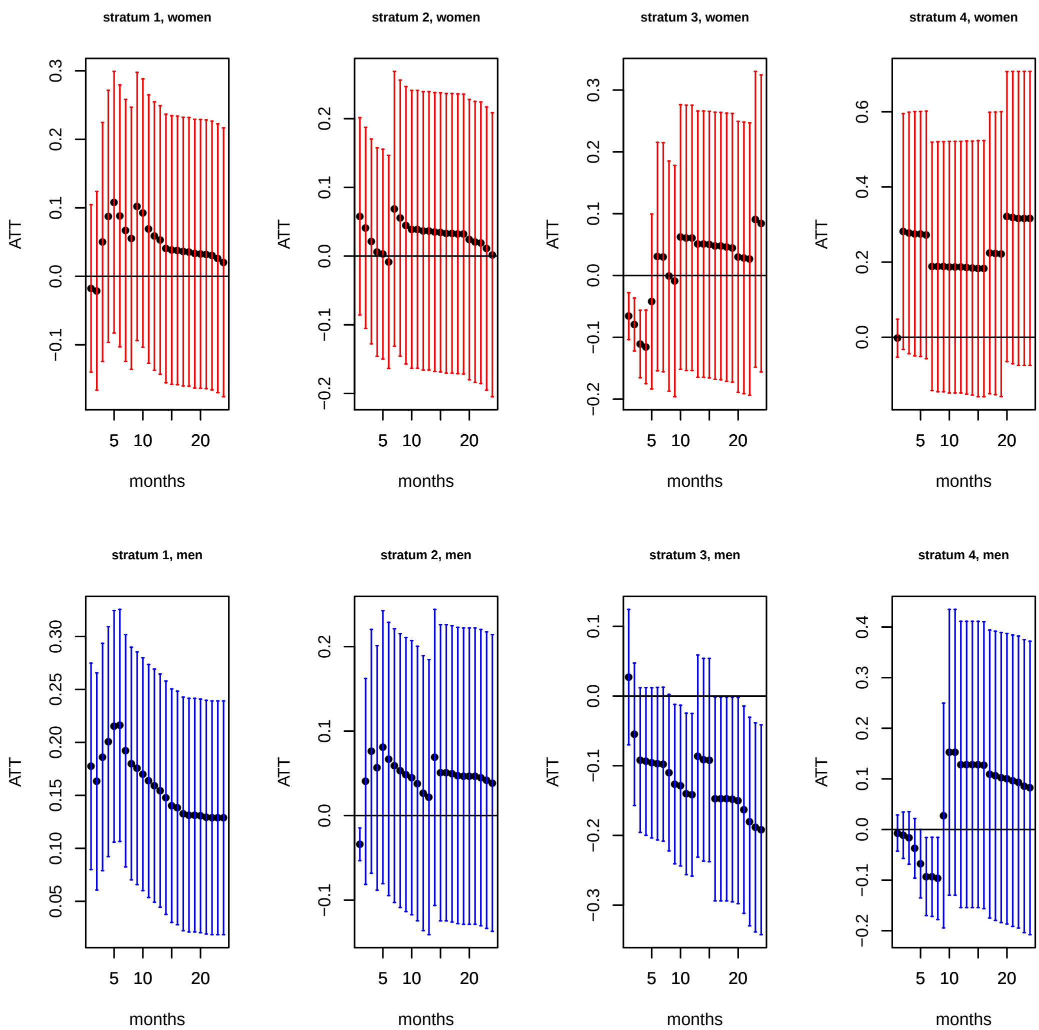


**Figure 35:** The plots show the monthly updated ATT on the probability of *exiting employment (ExJob)* for mere welfare receipt and its 90% confidence interval of *direct* sanctions for *employed (Emp)* welfare recipients *over 25 years* (o25) in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (quarterly) stratum $S_i$ of welfare duration (with i=1–4) and finishing with the end of ongoing final months $m_j$, with j=1–24 counting the months after the beginning of stratum.

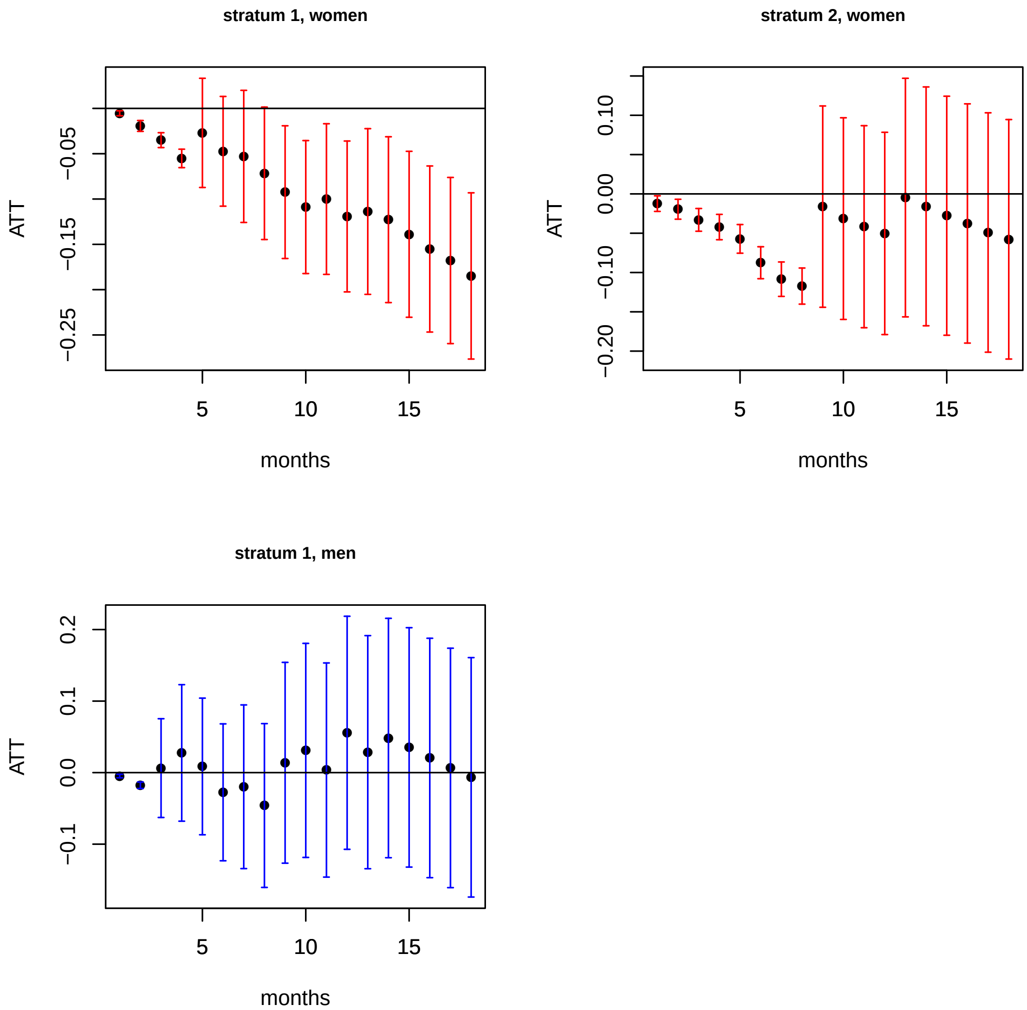


**Figure 36:** The plots show the monthly updated ATT on the probability of exit into *mere employment ("job only" (O))* and its 90% confidence interval of *indirect (ind)* sanctions for *unemployed (UE)* welfare recipients in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (half-yearly) stratum $S_i$ of welfare duration (with i=1–2) and finishing with the end of ongoing final months $m_j$, with j=1–18 counting the months after the beginning of stratum.

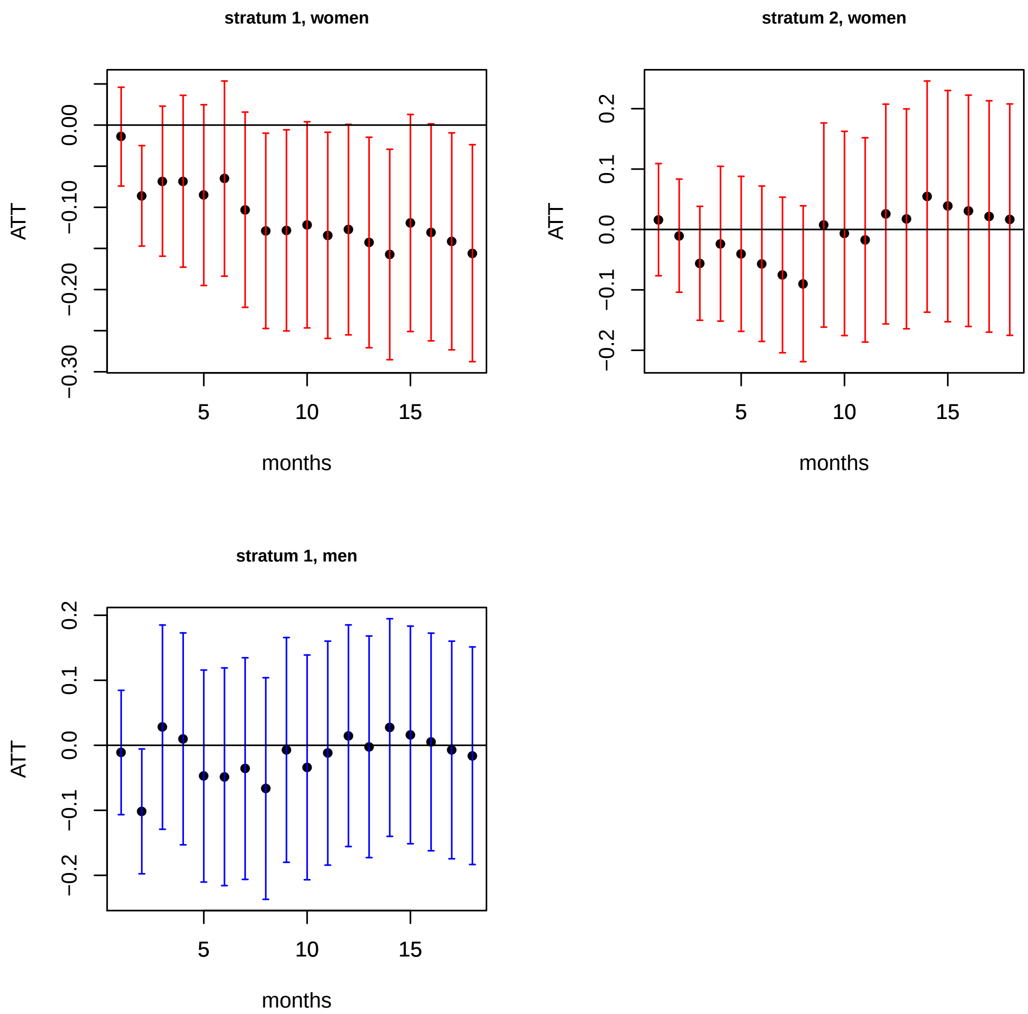


**Figure 37:** The plots show the monthly updated ATT on the probability of exit into *employment ("job in general" (G))* and its 90% confidence interval of *indirect (ind)* sanctions for *unemployed (UE)* welfare recipients in *Western Germany* (WG) of the inflow cohort 2008, separately for women (red) and men (blue). They illustrate the development of the sanction effect for overlapping periods $P_j$, each starting with the beginning of the (half-yearly) stratum $S_i$ of welfare duration (with i=1–2) and finishing with the end of ongoing final months $m_j$, with j=1–18 counting the months after the beginning of stratum.

The **Hamburg Institute of International Economics (HWWI)** is an independent economic research institute that carries out basic and applied research and provides impulses for business, politics and society. The Hamburg Chamber of Commerce is shareholder in the Institute whereas the Helmut Schmidt University / University of the Federal Armed Forces Hamburg is its scientific partner. The Institute also cooperates closely with the HSBA Hamburg School of Business Administration.

The HWWI's main goals are to:
• Promote economic sciences in research and teaching;
• Conduct high-quality economic research;
• Transfer and disseminate economic knowledge to policy makers, stakeholders and the general public.

The HWWI carries out interdisciplinary research activities in the context of the following research areas:
• Digital Economics
• Labour, Education & Demography
• International Economics and Trade
• Energy & Environmental Economics
• Urban and Regional Economics

Hamburg Institute of International Economics (HWWI)

Oberhafenstr. 1 | 20097 Hamburg | Germany
Telephone: +49 (0)40 340576-0 | Fax: +49 (0)40 340576-150
info@hwwi.org | www.hwwi.org