ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2019)
182, *Part* 1, *pp.* 263–281

Participating in a panel survey changes respondents' labour market behaviour

Ruben L. Bach

University of Mannheim, Germany

and Stephanie Eckman

RTI International, Washington DC, USA

[Received July 2016. Final revision March 2018]

Summary. Panel survey participation can bring about unintended changes in respondents' behaviour and/or their reporting of behaviour. Using administrative data linked to a large panel survey, we analyse whether the survey brings about changes in respondents' labour market behaviour. We estimate the causal effect of panel participation on the take-up of federal labour market programmes by using instrumental variables. Results show that panel survey participation leads to an increase in respondents' take-up of these measures. These results suggest that panel survey participation not only affects the reporting of behaviour, as previous studies have demonstrated, but can also alter respondents' actual behaviour.

Keywords: Administrative data; Instrumental variable; Longitudinal data; Panel conditioning; Treatment effects

1. Background

Panel surveys are a key resource for researchers and policy makers who seek to understand dynamic processes, such as movements in and out of the labour force. Yet such surveys are also vulnerable to the critique that participation can distort respondents' behaviour and/or responses, making the collected data unrepresentative of the larger population. This phenomenon is referred to as *panel conditioning* (Halpern-Manners *et al*., 2017). Although concerns about panel conditioning first arose in the 1940s (Lazarsfeld, 1940), researchers from many disciplines still rely on panel data for causal analysis. In this study, we test whether participation in the large-scale German panel study 'Labour market and social security' alters respondents' labour market behaviour. We think of participation in any of the first three waves of the panel survey as a treatment and the panel conditioning effect as a treatment effect. The outcome variables of interest are take-up of federal labour market programmes and job search behaviour, i.e. we use techniques of causal analysis to study whether panel survey participation makes respondents more or less likely to take part in the labour market programmes and whether it helps respondents to find a job faster.

Our study faces two major methodological challenges. First, we need to disentangle the effect of survey participation on changes in behaviour from changes in reporting. Administrative labour market data, which are independent of respondents' reporting, make this disentangle-

Address for correspondence: Ruben L. Bach, Professorship for Statistics and Methodology, University of Mannheim, Mannheim 68131, Germany.
E-mail: r.bach@uni-mannheim.de

© 2018 The Authors Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/19/182263
Published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

ment possible. Second, we need to control for confounding effects of survey non-response and attrition from the estimates of panel conditioning. We use an instrumental variable (IV) approach and select a second sample of people who were eligible for selection into the panel survey but were not selected. Thus, our data consist of two random subsamples—one selected for the survey; the other not. Instrumenting actual participation in several waves of the survey, i.e. the treatment, with the (random) invitation to participate in the survey, we adjust for bias due to non-response and attrition and estimate the causal effect of panel survey participation on respondents' labour market behaviour.

Previous studies of behavioural changes due to panel survey participation are rare and most suffer from design flaws. Many do not disentangle changes in reporting from changes in behaviour due to limitations of the data; others do not untangle panel conditioning effects from other effects such as attrition. We contribute to this sparse literature and tackle the methodological challenges that previous studies have encountered.

Before we turn to the data, methods and results of our study, we discuss panel conditioning and its two forms in more detail. We also explain why we expect panel participation to alter survey respondents' labour market behaviour.

The programs that were used to analyse the data can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

1.1. Panel conditioning

If participation in a panel survey induces changes in behaviour, then the survey's sample becomes less representative of the population over time, and estimates based on the data will be biased (Yan and Eckman, 2012). The classic finding of this type of panel conditioning, from the field of political science, is that participation in a pre-election survey increases voter turnout in upcoming elections (Clausen, 1969; Kraut and McConahay, 1973; Yalch, 1976; Traugott and Katosh, 1979; Granberg and Holmberg, 1992). However, not all studies have detected this effect (Smith *et al.*, 2003). Panel participation can also affect other types of behaviour: water treatment product use and purchases of health insurance (Zwane *et al.*, 2011); purchases of automotive services (Borle *et al.*, 2007), automobiles (Morwitz *et al.*, 1993; Chandon *et al.*, 2005) and computers (Morwitz *et al.*, 1993); saving for retirement (Crossley *et al.*, 2017); cheating in examinations (Spangenberg and Obermiller, 1996). For an extensive review of relevant studies in consumer behaviour and marketing research, see Dholakia (2010).

The literature proposes several theoretical explanations for changes in behaviour panel conditioning (Warren and Halpern-Manners, 2012). Two mechanisms best explain why we expect changes in behaviour due to panel conditioning to arise in our labour market survey data. Cognitive stimulus theory holds that repeatedly being asked the same questions makes respondents more aware of the topic of the survey, raises their consciousness of the issues and motivates them to engage in the behaviour under study (Sturgis *et al.*, 2009; Zwane *et al.*, 2011; Warren and Halpern-Manners, 2012). This mechanism probably explains panel conditioning in studies of voting behaviour: a pre-election interview may stimulate interest and participation in the election and thereby increase voter turnout among respondents (Clausen, 1969). The second explanation is concerned with stigmatized or socially non-normative behaviour. When survey questions force respondents to confront a conflict between their behaviour and society's norms, they might bring their future behaviour in line with social norms to avoid dissonance (Williams *et al.*, 2006).

On the basis of these two theoretical explanations, we hypothesize that repeatedly answering questions in a panel survey about whether or not one has participated in active labour market

policies (ALMPs) increases the likelihood that respondents will participate in those measures. ALMPs are programmes that are administered by the German Government that aim to reduce unemployment and to increase participation in the labour market (Crépon and van den Berg, 2016): they may include measures such as job application training or continuing education courses. Participation in such measures is often mandatory for recipients of unemployment benefit, and failure to participate may result in sanctions, i.e. temporary cuts in benefit receipt (Jacobi and Kluve, 2007). Taking part in a survey that includes several questions about such programmes may stimulate respondents to think more about the ALMPs, enhance their awareness of them and thereby increase the likelihood that respondents will participate in these ALMPs. Additionally, respondents may feel embarrassed to report that they have not participated in such measures and change their future behaviour accordingly. In this way, panel participation could change some respondents' behaviour. Furthermore, we expect that the size of the effect of panel survey participation on respondents' labour market behaviour will increase with each wave.

Participating in the panel may also increase respondents' likelihood of finding employment, i.e., because ALMPs are designed to help unemployment benefit recipients to find a new job, we expect that participating in the panel also reduces the search time that respondents need to find a job. It is also possible that survey participation has an independent effect on job search success, perhaps due to the stigma explanation that was discussed above.

Panel conditioning can also induce changes in how respondents report their behaviour, which we refer to as changes in reporting (Waterton and Lievesley, 1989; Sturgis *et al.*, 2009; Cantor, 2010; Warren and Halpern-Manners, 2012). For example, respondents may report more accurately in later waves of a panel survey, because of increased trust in the interviewing process. In contrast, respondents may also learn from prior participation how the questionnaire is structured and then falsify their answers to reduce the length of the interview. For a detailed review of theoretical explanations for changes-in-reporting panel conditioning and an extensive literature review see, for example, Warren and Halpern-Manners (2012).

Both forms of panel conditioning can occur at the same time (Halpern-Manners and Warren, 2012; Yan and Eckman, 2012). For example, unemployed respondents of a panel survey on labour market behaviour may under-report unemployment because of social desirability bias in early waves of the survey but report more truthfully in later waves as they become more trustful of the interviewer. Yet, at the same time, (repeatedly) asking respondents about labour market topics could stimulate respondents' job search activities and thereby lead to changes in their actual behaviour (the cognitive stimulation hypothesis). Researchers need to be aware of the various ways that panel conditioning can occur and how it can bias inference: working with data affected by panel conditioning, researchers risk mischaracterizing the existence, magnitude and correlates of changes across survey waves in respondents' attitudes and behaviour, which are the main estimates made from panel data (Clinton, 2001; Halpern-Manners *et al.*, 2017). In addition, as mentioned above, the conclusions that are drawn from the panel participants may not generalize to the larger population.

1.2. Methodological challenges

Studies of panel conditioning are confronted with two major methodological challenges. The first is that disentangling changes in behaviour from changes in reporting is difficult or impossible to do without validation records for the answers that are given in the survey (Waterton and Lievesley, 1989; van der Zouwen and van Tilburg, 2001). Indeed, most studies of changes in reporting do not distinguish between the two types of panel conditioning (for exceptions see Pennell and Lepkowski (1992), van der Zouwen and van Tilburg (2001), Duan *et al.* (2007) and

Halpern-Manners and Warren (2012)). To isolate changes-in-behaviour panel conditioning, we need data that are unaffected by respondents' reporting, such as administrative process data. Using such data, researchers can study changes in behaviour by, for example, comparing respondents' behaviour in the administrative data with the behaviour of those who were not interviewed. This is the approach that we follow in our study. If one was interested in changes in reporting, comparisons of respondents' survey answers with administrative records may be used (e.g. Yan and Eckman (2012)). If only survey data are available, we may compare experienced respondents with novel respondents in panel studies with rotating designs to study changes in reporting, but this approach does not fully separate the two forms of panel conditioning (Halpern-Manners and Warren, 2012).

The second major challenge to estimating panel conditioning error is adjustment for confounding sources of error in panel studies (Williams and Mallows, 1970; van der Zouwen and van Tilburg, 2001; Sturgis *et al.*, 2009; Das *et al.*, 2011; Halpern-Manners and Warren, 2012). The most common confounding source of error is non-response and panel attrition. Initial non-response, for example, will result in systematic differences between respondents of a survey and the originally selected sample. If these differences are not properly adjusted for, analysis of panel conditioning may be confounded by non-response bias. Similarly, compositional changes in the survey sample due to non-response in later waves of a panel survey, i.e. panel attrition, may be mistaken for panel conditioning. For example, almost all studies of panel conditioning in the Current Population Survey do not distinguish between attrition and conditioning (Halpern-Manners and Warren, 2012). Some researchers, using other data sets, do attempt to control for the effects of non-response and attrition by conditioning on covariates related to non-response and attrition; others exclude attrition by relying on panel surveys with rotating designs. Comparing (reporting) behaviour between respondents who participate in all waves but have different levels of panel tenure enables us to exclude panel attrition (see, for example, Halpern-Manners and Warren (2012), for details on this approach). Two other potential sources of error that can easily be mistaken for panel conditioning are interviewer effects and mode effects. van der Zouwen and van Tilburg (2001), for example, showed that changes in reported network size over time are due to changes in interviewer between two waves of a panel survey and not panel conditioning. Similarly, Halpern-Manners and Warren (2012) warned that changes over time may also result from a change in the data collection mode: the Current Population Survey, for example, uses personal interviews in the first and telephone interviews in subsequent rounds.

We note that, among large socio-economic panel surveys, only one study of changes in behaviour does not suffer from the problems that were discussed above. Crossley *et al.* (2017) employed a quasi-experimental design to identify a survey participation effect in a large-scale panel survey in the social sciences. Analysing administrative wealth data from respondents of a Dutch on-line panel, they found that participating in an interview about saving for retirement has, on average, a negative effect on respondents' future saving behaviour. Although Crossley *et al.* (2017) estimated the effect of participation in a survey module that was fielded only once, their results lend support to the hypothesis that (repeated) survey participation can alter respondents' behaviour. Moreover, their identification strategy, which uses an IV approach, can be applied to the estimation of panel conditioning effects, and that is the approach which we take in this study.

2. Data

Data for our analysis come from the German panel study the *Panel Arbeitsmarkt und soziale Sicherung* (PASS) ('Labour Market and Social Security') survey, which is a large-scale yearly panel survey that is conducted by the Institute for Employment Research on labour market

topics. It uses a mixed mode design of computer-aided telephone interviews and computer-aided personal interviews. The survey consists of a household interview and additional individual interviews with all household members who are at least 15 years old. The main topics of the interviews are pathways into and out of unemployment benefits receipt type II (long-term benefits due to unemployment, disability or employment that do not reach a minimum standard of living), dynamics of the material and social situation of benefit recipients, changes in recipients' behaviour and attitudes over time, and interactions between recipients and the benefit-providing agencies (Trappmann *et al.*, 2013).

The PASS sample consists of two subsamples. The *recipient subsample* ($n = 29\,309$) is a representative sample of all unemployment benefit units in Germany drawn from a register that is maintained by the Federal Employment Agency. The sample was drawn from the most recent available administrative records in July 2006. Unemployment benefit units are in most cases identical to households (see Trappmann *et al.* (2013)). Variables on the register, however, concern the individuals residing in these households. The second subsample is a *general population sample* of households, selected from a commercial data set of addresses in Germany. We do not use this subsample in our analysis and do not describe it further (see Trappmann *et al.* (2013) for details).

Both samples were drawn by using a multistage sampling design (Schnell, 2007; Rudolph and Trappmann, 2007). In the first stage, 300 postcode areas were selected as primary sampling units. In the second stage, unemployment benefit units in the postcode areas selected were drawn from the administrative register of unemployment benefit recipients. Since the recipient sample was drawn from the administrative data, respondents and non-respondents can easily be identified in the administrative records. We discuss linkage between the survey and administrative data in more detail below.

We consider data from the first three waves of the PASS survey excluding refreshment samples that were introduced in several waves (Trappmann *et al.*, 2013). Data were collected annually between winter 2006 and spring 2009. In each of these waves, respondents were asked the same set of questions about their participation in ALMPs (see the on-line supplementary materials for the text of the questions). These questions may provide a cognitive stimulus that will increase respondents' awareness and take-up of the programmes, and for this reason we hypothesize changes-in-behaviour panel conditioning.

Household response rates were 35% in wave 1, 51% in wave 2 and 65% in wave 3 (wave 2 and wave 3 response rates were conditional on participation in previous waves; Christoph *et al.* (2008), Gebhardt *et al.* (2009) and Berg *et al.* (2010)). However, in our data set we cannot distinguish between households that were selected and fielded ($n = 23\,736$) and households that were selected and not fielded ($n = 5573$). Therefore, the household response rate that was calculated with our larger data set is 22.8%. Since the subset of households used during fieldwork is a random subsample of the original sample, we do not expect the additional households to bias our estimates.

2.1. Administrative data and linkage

The administrative data that we use to investigate changes-in-behaviour conditioning are the 'Integrated employment biographies' (IEBs), which include records of all spells of employment subject to social security, unemployment benefit receipt, participation in ALMPs or spells of job search (Jacobebbinghaus and Seth, 2007; Institute for Employment Research, 2013). These records can be aggregated to the person level and contain histories of employment, unemployment, job search and benefit receipt, as well as records of ALMP participation. Although the administrative data are not free of error, their overall quality is very high, and they are used

by the German Government to calculate pension claims, to administer benefit claims and to make payments (Jacobebbinghaus and Seth, 2007; Köhler and Thomsen, 2009; Kreuter *et al.*, 2010). In general, data on benefits, ALMP and job search are of the highest quality, because they are generated by activities of the Federal Employment Agency itself (Jacobebbinghaus and Seth, 2007). A variety of sociodemographic variables such as gender, date of birth, citizenship, education and place of residence are also included in the data set.

Because the PASS recipient sample is selected from the same source of data, we can easily identify members of the households sampled for the PASS survey in the IEBs. Overall, we can identify and obtain full information on 28 377 selected households, i.e. 96.8% of the original household sample. Cases that do not have sociodemographic variables that are needed for our analysis (see Section 3.5) are not included (3.2%). We do not expect this restriction of the sample to bias our estimates of panel conditioning because households with similar scarce administrative records will also not be included in our control group, for the same reason (see Section 3). For the 28 377 households with full information, we obtain individual records on 38350 household members.

Table 1 shows the number of households with at least one realized interview per wave and the number of households that were successfully identified in the administrative records. Overall, we can identify and obtain full information on 6415 households that responded in at least one of the first three waves of the PASS survey (95.3%; the last column of Table 1). Likewise, we obtain full information on 21 962 households that did not respond to any of the first three waves of the PASS survey (97.3%). Thus, the linkage rate among all households is very high.

Because of German privacy regulations, we cannot link individual response indicators from the survey to the administrative data without respondents' consent. Restricting our analysis sample to consenting respondents may introduce bias if consenters are different from non-consenters. To avoid such bias, we do not identify individuals within households and treat *all* household members as respondents if at least one person in the household responded to the survey. This issue might lead us to underestimate the true panel conditioning effect because some household members have not received the stimulus of responding to the survey (within-household response rates at the individual level are 85.6% in wave 1, 85.5% in wave 2 conditionally on the response in wave 1, and 83.5% in wave 3, conditionally on the response in wave 2; Christoph *et al.* (2008), Gebhardt *et al.* (2009) and Berg *et al.* (2010)). However, it avoids any bias due to non-consent in our estimates. On the basis of these definitions, about 22.8% of

Table 1. Number of households with at least one realized interview and number of successfully linked households per wave

	Numbers in the following waves:			*Response in any wave*
	Wave 1	*Wave 2*	*Wave 3*	
Households with at least 1 realized interview	6691	3391	3615	6734
Responding households successfully linked to administrative records	6371 (95.2%)	3236 (95.4%)	3445 (95.3%)	6415 (95.3%)
Individuals within linked responding households	8647	4337	4626	8728

all individuals who were identified in the administrative data responded to at least one wave of the PASS survey.

Identifying the administrative records for respondents and non-respondents is only the first step, however: to obtain an unbiased estimate of the effect of repeated survey participation on ALMP take-up, we need to control for the confounding effects of non-response and panel attrition.

3. Methods

We regard participation in at least one of the first three waves of the PASS survey as a treatment, and the panel conditioning effect as a treatment effect: we are interested in estimating how receiving the treatment (participating in the PASS survey) changes respondents' behaviour. The PASS respondents form the treatment group. To estimate the treatment effect, we selected a second random sample from the IEB administrative data set ($n = 38\,350$) to form the control group. This sample consists of unemployment benefit recipients who were eligible for the first wave of the PASS survey but were not selected in the first wave nor in any later waves. Thus, our data consist of two random samples: one selected for the survey, i.e. the respondents and non-respondents to the PASS survey; the other, the control group, not.

In formal terms, $Z \in [0, 1]$ defines an indicator of whether one was *assigned* to treatment ($Z = 1$) or control ($Z = 0$), i.e. selected for the survey or not. Moreover, $D \in [0, 1]$ denotes the treatment indicator, i.e. whether one actually *participated* in the survey ($D = 1$) or not ($D = 0$). If everyone complied with the treatment status that was assigned, then $D = Z$. Assignment to treatment (Z) is fixed over time. Actual treatment (D), however, may change over time; for example, a household that did not participate in the first wave ($D = 0$) may participate in the second wave, thus changing to $D = 1$. Let Y be the outcome of interest (take-up of ALMPs and job search behaviour). If everyone complied with the treatment, the treatment effect would simply be the difference between the average outcomes of the treated individuals $E(Y|D = 1)$ and the control group $E(Y|D = 0)$.

However, as shown in Section 2, many people who were selected for the survey did not participate: $Z = 1$, but $D = 0$. These people are *non-compliers*, i.e. they did not respond to the survey, their assigned treatment status. Non-response and attrition in surveys have many causes and can lead to bias or endogenous selection into treatment (Groves *et al.*, 1992, 2000; Abadie, 2003; Kreuter *et al.*, 2010). For example, people who agree to participate in three waves of a survey may be more compliant and thus more likely to participate in ALMPs, even without the treatment of the survey (e.g. Zabel (1998), Rizzo *et al.* (1996) and Lepkowski and Couper (2002)). Thus, comparing the outcomes of the treated cases with the outcomes of the control group may bias our analysis of panel conditioning, because survey respondents are different from non-respondents in many ways.

One method to address the problem of non-compliance with treatment assignment is to estimate an intention-to-treat (ITT) effect. In this approach, rather than comparing individuals with different treatment statuses, we compare those assigned to different treatments, conditionally on a vector of predetermined covariates X:

$$\text{ITT} = E[Y|Z = 1, X] - E[Y|Z = 0, X]. \tag{1}$$

Since Z was randomly assigned, the ITT estimates the causal effect of the offer of treatment. However, because of non-compliance with the treatment status that was assigned (the take-up rate of the treatment, $E(D|Z = 1)$, across all waves is about 22.8%), the effect that is estimated via equation (1) will be too small relative to the average causal effect on the treated (Angrist

and Pischke (2009), chapter 4). A more powerful class of methods uses the randomization of Z in an indirect way to adjust for the bias due to non-compliance and to estimate the treatment effect.

3.1. Instrumental variables

The IV approach is based on the following idea: if an instrument Z is available that induces exogenous variation in the treatment variable D, then instrumenting D with Z enables us to estimate the treatment effect of D (Imbens and Angrist, 1994; Angrist *et al.*, 1996; Abadie, 2003).

Following the potential outcomes framework (Rubin, 1974, 1977), we define two potential outcomes: Y_1 is the outcome that occurs when a case receives treatment (participates in the survey) and Y_0 is the outcome without treatment. Obviously, we observe only $Y = DY_1 + (1 - D)Y_0$ for a given individual, i.e. either Y_1 or Y_0. Furthermore, let D_z represent the potential treatment status given $Z = z$. If, for a given case $D_1 = 0$, then that case would not participate if selected; $D_1 = 1$ means that a case would participate if selected. Analogously to the potential outcomes set-up, we observe only $D = ZD_1 + (1 - Z)D_0$, but never both potential treatments for any individual.

Following Angrist *et al.* (1996), we divide the population into four groups:

(a) *compliers*, $D_1 > D_0$ or, equivalently, $D_0 = 0$ and $D_1 = 1$;
(b) *always-takers*, $D_1 = D_0 = 1$;
(c) *never-takers*, $D_1 = D_0 = 0$;
(d) *defiers*, $D_1 < D_0$ or, equivalently, $D_0 = 1$ and $D_1 = 0$.

In our framework, the group of survey respondents are the compliers: they were assigned to take the treatment, i.e. selected for the survey, and complied with the treatment assignment, i.e. responded to the survey. There are no always-takers, i.e. people who take the treatment irrespectively of their treatment assignment status, since participation in the survey is possible only for cases who are selected for participation. However, we do have never-takers: people who do not participate in the survey when they are selected (and also when they are not selected). Non-compliance in our set-up is one sided because people who are not assigned to the survey cannot decide between response and non-response. In other words, the probability that a case who is assigned to control does not take the treatment equals 1 ($\Pr(D_0 = 0) = 1$). There are no defiers for the same reason.

Angrist *et al.* (1996) showed that instrumenting D with Z estimates the local average treatment effect (LATE) for compliers under certain assumptions that we discuss in the next section. Moreover, since there are no always-takers and no defiers, the group of compliers and the group of treated are identical, and the LATE for compliers equals the average treatment effect on the treated.

3.2. Identification assumptions

To state the assumptions that are needed for the IV approach, we need to include Z in the definition of potential outcomes. Let Y_{zd} represent the potential outcome if $Z = z$ and $D = d$, and let X be a vector of known characteristics. Then, with the following non-parametric assumptions, we can use IV techniques to estimate the LATE for compliers.

(a) Independence of the instrument: conditionally on X, the random vector $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, D_0, D_1)$ is independent of Z.

(b) Exclusion of the instrument: $\Pr(Y_{1d} = Y_{0d}|X) = 1$ for $d \in \{0, 1\}$.
(c) First stage: $0 < \Pr(Z = 1|X) < 1$ and $\Pr(D_1 = 1|X) > \Pr(D_0 = 1|X)$.
(d) Monotonicity: $\Pr(D_1 \geqslant D_0|X) = 1$.

Assumption (a) means that treatment assignment Z is ignorable or as good as randomly assigned, conditionally on X. This assumption is justified in our study. The cases that are selected for the PASS survey and the control data set are equal size random samples of people who were registered as unemployment benefit recipients in the IEB at the date of sample selection of the PASS survey. As we would expect from random assignment of Z, there are only few significant correlations between Z and a set of covariates that are derived from the administrative data (see Section 3.5 for details on these covariates and the on-line supplementary materials for supporting analysis).

Assumption (b) states that variation in Z, i.e. assignment to treatment or control, does not influence potential outcomes except through D, i.e. through response or non-response. Moreover, the assumption enables us to define potential outcomes in terms of D alone, i.e. $Y_0 = Y_{00} = Y_{10}$ and $Y_1 = Y_{01} = Y_{11}$. Taken together, assumptions (a) and (b) guarantee that the only effect that Z has on Y is through D, i.e. that being selected for the survey does not affect participation in ALMPs except through taking part in the PASS survey. It may still be possible that being selected for the survey, receiving the advance letter announcing the survey and perhaps receiving a few contact attempts could influence ALMP behaviour without survey participation. However, we believe that the chances of that happening are very small, and we are not aware of any literature showing that survey contact attempts and advance materials change behaviour.

Assumption (c) states that D and Z must be correlated, conditionally on X. In addition, the support of X conditionally on $Z = 1$ must coincide with the support of X conditionally on $Z = 0$. Since 8728 people (about 22.8%) responded in at least one of the first three waves of the PASS survey, and no non-selected cases participated, this assumption appears to be satisfied. In addition, tests indicate that we do not need to worry about weak identification (F-statistic from an IV regression with a linear first stage, $F_{1,76678}$, 11 187.81; partial R^2 of the treatment indicator, 0.13; Stock and Yogo (2005) and Angrist and Pischke (2009), chapter 4).

Assumption (d), i.e. monotonicity, holds trivially because only people who are selected for the survey could participate in the survey.

In addition to the four assumptions that were discussed above, we need to make the stable unit treatment value assumption (Rubin, 1978; Angrist *et al.*, 1996). It holds that potential outcomes for each treated case are unrelated to the treatment status of others. This assumption would be violated if, for example, a responding household interacts with a non-responding household. Given the number of households with benefit receipt in Germany in July 2006 (about 4.01 million; Bundesagentur für Arbeit (2017)) and the sample size of the PASS survey (29 309), the chances that households would interact are very small. Thus, we do not see any reason why this assumption would not hold.

In general, these assumptions are similar to the assumptions that are needed for the popular Wald estimator. In our case, however, the assumptions are more flexible as they allow conditioning on X. Since there are minor differences between the group that was invited to participate in the survey (see above and the on-line supplementary materials for supporting analysis) and the control group, we choose the estimator that is presented below and control for X.

Having discussed the assumptions in detail, we next turn to the identification of the LATE or, in the terminology of Abadie (2003), the local average response function (LARF).

3.3. Identification

Define the object of interest, the LARF for compliers, as $E[Y|D, X, D_1 > D_0]$. The challenge for the estimation of this LARF is the identification of the compliers (those who respond when invited to participate and do not respond when not invited). Among the cases that were invited to participate in the survey, the compliers are easy to identify, of course: these are the respondents. In the control group, however, we cannot identify compliers individually because we cannot distinguish whether a case complies with being assigned to the control group condition or is a never-taker. Using Abadie's κ and theorem 3.1 of Abadie (2003), pages 236–237, however, we can express expectations for compliers with the LATE assumptions in the following way (see Section 3.2). Let $g(\cdot)$ be any measurable real function of (Y, D, X) with finite expectation. Furthermore, define

$$\kappa = 1 - \frac{D(1-Z)}{\Pr(Z=0|X)} - \frac{(1-D)Z}{\Pr(Z=1|X)}. \tag{2}$$

The second term of κ is greater than 0 only for the unselected always-takers (in our cases, the second term equals 0, because people who were not invited to participate in the survey could not participate, i.e. there are no unselected always-takers). The third term is greater than 0 only for the selected never-takers. Thus, the only cases with $\kappa > 0$ are the compliers and the selected always-takers because there are no defiers by definition (see Section 3.1). Therefore, we can think of κ as a case level weight that identifies the complier population, enabling us to estimate the LARF for compliers. Abadie showed that $E[\kappa] = \Pr(D_1 > D_0)$. Interestingly, this procedure simplifies to the popular Wald estimator when X contains a constant only (Angrist and Pischke (2009), chapter 4). Since $\Pr(Z=1|X)$ in equation (2) is unknown because of X, we estimate it by using a probit model.

3.4. Estimation

When Y is continuous, estimation of the LARF, $E[Y|D, X, D_1 > D_0]$, is straightforward with linear regression. Suppose that $E[Y|D, X, D_1 > D_0] = h(D, X, \theta)$, with $h(D, X, \theta) = \alpha d + x'\beta$ and parameter vector $(\theta = \alpha; \beta)$. Using the results from above, Abadie (2003), page 239, proposed the following least squares estimator for continuous outcomes:

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(\alpha,\beta)\in\theta} \frac{1}{n} \sum_{i=1}^{n} \kappa_i (y_i - \alpha d_i - x_i'\beta)^2. \tag{3}$$

When $\Pr(Z=1|X)$ is estimated with least squares, equation (3) produces the traditional two-stage least squares estimator (Angrist and Pischke (2009), chapter 4).

When the outcome Y is binary, a probit transformation of the linear part of equation (2) can be used. The estimator is then

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(\alpha,\beta)\in\theta} \frac{1}{n} \sum_{i=1}^{n} \kappa_i \{y_i - \Phi(\alpha d_i - x_i'\beta)\}^2. \tag{4}$$

In both cases, variances are estimated according to theorem 4.2 of Abadie (2003), page 242.

Both estimators are implemented in the LARF package (version 1.4) by An and Wang (2016) in R (R Core Team, 2017). We next define the outcome variables, participation in ALMP and job search after participation in the survey, and the set of covariates X in more detail.

3.5. Outcomes and control variables

We consider two variables that are related to ALMP participation as outcomes. The first is the number of ALMPs taken. The second is a simple indicator of whether a person participated in any ALMP (1) or did not (0). If panel conditioning has led to changes in behaviour, we should see that respondents participate in more programmes and are more likely to participate in ALMPs. In building these outcome variables, we consider only those spells that started after the first day of fieldwork (because the survey cannot affect ALMP spells before it started) and those that occurred before January 31st, 2010, the day before fieldwork of wave 4 started (because later spells may have been influenced by later waves of the survey and in this study we consider only the first three waves). Furthermore, we consider both outcomes at three different periods in time. The first period begins after the first day of fieldwork of wave 1 and ends just before the beginning of wave 2. The second period begins after wave 2 and ends before wave 3 and the third, finally, after wave 3 and before wave 4. However, we must then also define the treatment accordingly. Table 2, showing possible response patterns, defines the treatment indicator for each outcome period.

Estimating the treatment effect for each of these treatments separately enables us to study whether panel conditioning effects become stronger over time, i.e. whether the effect size increases with each additional wave. However, our approach may underestimate the true panel conditioning effect because we treat all members of a responding household as respondents (Section 2.1) and because the definitions of the treatments in waves 2 and 3 also include people who responded in only one wave (Section 3.2).

Regarding the question of whether panel survey participation increases the probability of employment, we again consider two outcome measures. First, we simply calculate the duration from the beginning of the fieldwork of the survey until the occurrence of the first job spell in the records (in days). Second, we create an indicator of whether a person found a job after the beginning of the fieldwork of the survey or not. Both variables consider only data until the beginning of wave 4. For brevity, we do not consider these two outcomes at different points in time.

For each outcome, we first estimate the ITT that is described in equation (1). Second, we estimate the LARF with the estimator that is defined in equation (3) for the number of ALMP

Table 2. Treatment variable definitions and response patterns

Outcome period	*Treatment*	*Responses to survey in the following waves:*			*N*	*Total*
		Wave 1	*Wave 2*	*Wave 3*		
After wave 1 and before wave 2	$D=1$	×			8647	8647
After wave 2 and before wave 3	$D=1$	×	×		4310	8674
		×			4337	
			×		27	
After wave 3 and before wave 4	$D=1$	×	×	×	3295	8728
		×	×		1015	
		×		×	1259	
			×	×	18	
		×			1683	
			×		9	
				×	54	

participations and the time to find a job and the estimator that is defined in equation (4) for the two binary participation indicators.

As a falsification or 'placebo' test, we create the same two ALMP outcome variables for earlier spells of programme participation: those that ended before the first day of fieldwork, as suggested by Athey and Imbens (2017). If the assumptions that were discussed above are met and the model is correctly specified, we should not find a significant treatment effect on pretreatment outcomes, because the survey cannot affect ALMP spells that occurred before the survey started. We do not perform this test for the job search outcomes because it is not clear how we should define the search time for the last job held before the survey started. In addition, most of our cases were unemployed at the time of selection and had been unemployed for a while.

Furthermore, we derive a set of covariates X from the administrative data. These covariates cover sociodemographics (age, gender, education, nationality and place of residence) as well as labour market histories (past employment, unemployment and unemployment benefit receipt). We include these covariates in all of our analyses, for two reasons: first, to control for any differences that might arise by chance between the group that is assigned to treatment and the group that is assigned to control; second, because including covariates, even if they do not differ between $Z=1$ and $Z=0$, can reduce some of the variability in the outcome variable and thereby increase the precision of the estimates (Angrist and Pischke (2009), chapter 4). For a complete list of the covariates that we use, see the on-line supplementary materials.

In sum, with the methods that are discussed here, we can estimate the causal effects of repeated participation in the PASS survey on the take-up of ALMPs and employment. The results enable us to answer our research question: whether panel participation leads to changes in respondents' labour market behaviour.

4. Results

In the first step, we present descriptive statistics of the outcome variables. Then, we present the results of the ITT analysis and our main results, the estimated treatment effects of participation in several waves of the survey. At each step, we also check the results from the falsification test: if our models are working well, they should not detect any treatment effect in the ALMP variables before wave 1, i.e. before treatment began.

4.1. Descriptive statistics

Table 3 shows descriptive statistics of the outcome variables at four different periods in time. The first set of rows show that the two ALMP outcomes do not differ much between the group that is selected for the survey ($Z=1$) and the control group of unselected recipients ($Z=0$) before the start of the survey. This result is expected, given the random selection process. The last two columns, however, show that respondents to the survey ($D=1|Z=1$), i.e. selected cases who will respond to at least one wave of the survey, participate more often in ALMPs and in more ALMPs than non-respondents ($D=0|Z=1$). These results support the claim that actual participation in the survey, i.e. take-up of the treatment, is selective, which inspired our use of the IV approach in the first place. We observe similar patterns after waves 1, 2 and 3 (the third–eighth rows).

In the last two rows, we see that, after participating in the survey, the mean number of days to find a job as well as the percentage of people who found a job do not differ much between $Z=1$ and $Z=0$. Differences between respondents and non-respondents, however, are more pronounced. Respondents need less time to find a job and are more likely to find a job. In the next section we evaluate whether these differences are due to survey participation.

Table 3. Descriptive statistics of outcome variables

	Outcomes	*Mean (standard deviation)*			
		Selected *(Z=1)*	*Not selected* *(Z=0)*	*Respondents*† *(D=1|Z=1)*	*Non-respondents* *(D=0|Z=1)*
Before wave 1	Number of ALMP participations	1.71 (2.05)	1.70 (2.03)	1.79 (2.07)	1.68 (2.04)
	ALMP participation	63%	62%	64%	62%
	N	38350	38350	8728	29622
After wave 1 and before wave 2	Number of ALMP participations	0.35 (0.71)	0.34 (0.70)	0.37 (0.71)	0.35 (0.71)
	ALMP participation	25%	24%	27%	25%
	N	38350	38350	8647	29703
After wave 2 and before wave 3	Number of ALMP participations	0.69 (1.11)	0.66 (1.11)	0.71 (1.12)	0.68 (1.11)
	ALMP participation	37%	36%	39%	37%
	N	38350	38350	8674	29676
After wave 3 and before wave 4	Number of ALMP participations	1.01 (1.48)	0.98 (1.49)	1.04 (1.51)	1.00 (1.47)
	ALMP participation	45%	44%	46%	45%
	N	38350	38350	8728	29622
After wave 1 and before wave 4	Days until new job	1333.74 (928.40)	1324.26 (928.52)	1280.28 (926.07)	1335.26 (928.60)
	Found job	54%	55%	58%	54%
	N	38350	38350	8728	29622

†See Table 2 for definitions.

4.2. Treatment effects

Fig. 1 shows the results of the ITT analysis, i.e. the differences in the ALMP outcomes between the group that is assigned to survey participation and the group that is assigned to control as shown in equation (1). Results of this analyses are similar to the third and fourth columns of Table 3, but the results that are shown in Fig. 1 control for the covariates that were introduced in Section 3.5. Before we discuss the ITT estimates, we first check that there are no significant differences between the selected and unselected cases with respect to the two ALMP outcomes that were measured *before* the start of the survey, i.e. we run the falsification test that was described in Section 3.5. The first row in Fig. 1 shows that the method estimates no significant differences in the number of ALMPs (Fig. 1(a)) nor for the ALMP participation indicator (Fig. 1(b)). This finding supports the claim that assignment to treatment or control was in fact random, conditionally on X (assumption (a)).

After the first wave (the second row of Fig. 1), the ITT value is positive and significant for both ALMP outcomes, though very small. Unemployment benefit recipients who were selected for the survey participate in about 0.01 more ALMPs than do those in the control group (Fig. 1(a)), controlling for X. In addition, the take-up of ALMPs is about 1 percentage point larger among the selected cases (Fig. 1(b)). Both effects become stronger after waves 2 and 3 (the third and fourth rows of Fig. 1), i.e. when respondents may have been exposed to the survey more than once. These results are evidence that participation in the PASS survey leads to a small increase in ALMP participation, and the effect becomes stronger over the three waves.

Regarding the time to find a job, the ITT value is 3.98 days with a standard error of 6.12, controlling for X. The ITT value for the indicator of whether a person found a job is −0.17

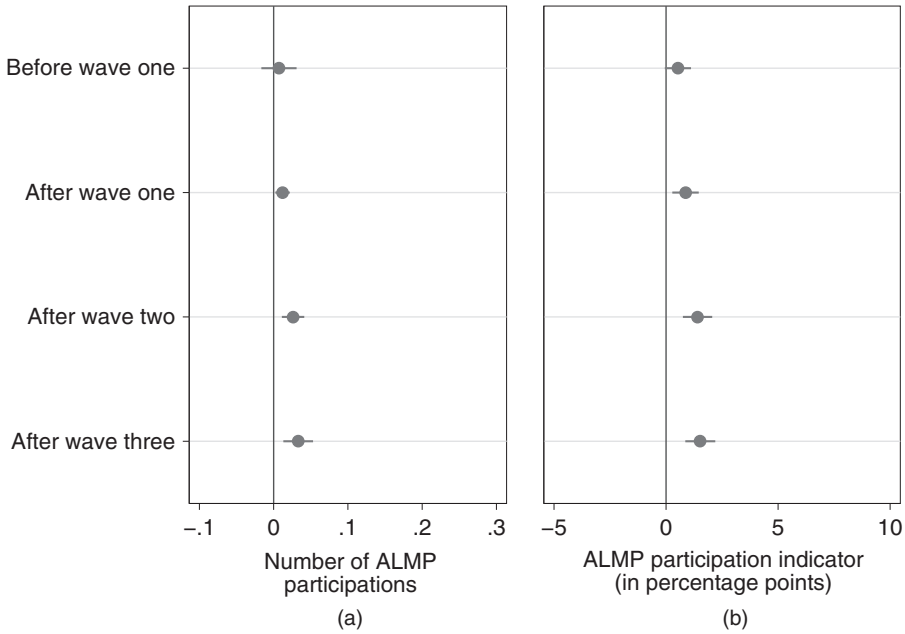

Fig. 1. Point estimates of panel conditioning effects with 95% confidence intervals: estimates from the ITT analysis, controlling for covariates X (see Section 3.5); $N = 76700$

percentage points with a standard error of 0.33. Both ITT values are insignificant. Thus, ITT results do not seem to support the hypothesis that respondents need less time to find a job or are more likely to find a job than people who were not interviewed.

However, all ITT effects that were described above underestimate the average causal effect on the treated, because of substantial non-compliance with the treatment assignment status. Using the LARF estimator that was described in Section 3, we can obtain a clearer picture of the actual effect size of (repeated) survey participation.

Fig. 2 shows the main results of our study; the LARF estimates from the IV models. Again, we control for the set of covariates X (see Section 5) to ensure that treatment assignment Z is in fact random (conditionally on X) and to reduce some of the variability in the outcome variable and thereby to increase the precision of the estimates. Once we include covariates in the estimation, the LARF estimator that was described in Section 3 still equals the average treatment effect on the treated under one-sided non-compliance. The average treatment effect on the treated, however, is then specific to these covariates.

Again, before turning to the outcomes after survey participation, we first look at the take-up of ALMPs *before* the start of the survey to assess the model specification with a falsification test. If the model is correctly specified and if the assumptions that were discussed in Section 3 hold, then there should be no significant effect on ALMP participation before treatment. The first row of Fig. 2 reports the results of this analysis. We do not find a significant treatment effect for the number of ALMPs (Fig. 2(a)) or for the ALMP participation indicator (Fig. 2(b)). These results, together with the results from the falsification test in Fig. 1, suggest that our IV model is correctly specified, and the assumptions that were discussed in Section 3 are met and thus that we can estimate the causal effect of survey participation on respondents' take-up of ALMPs.

Turning to the other rows in Fig. 2, the effect of participation in the survey on ALMP take-up after the start of the survey, we find a significant positive causal effect of participation in the first

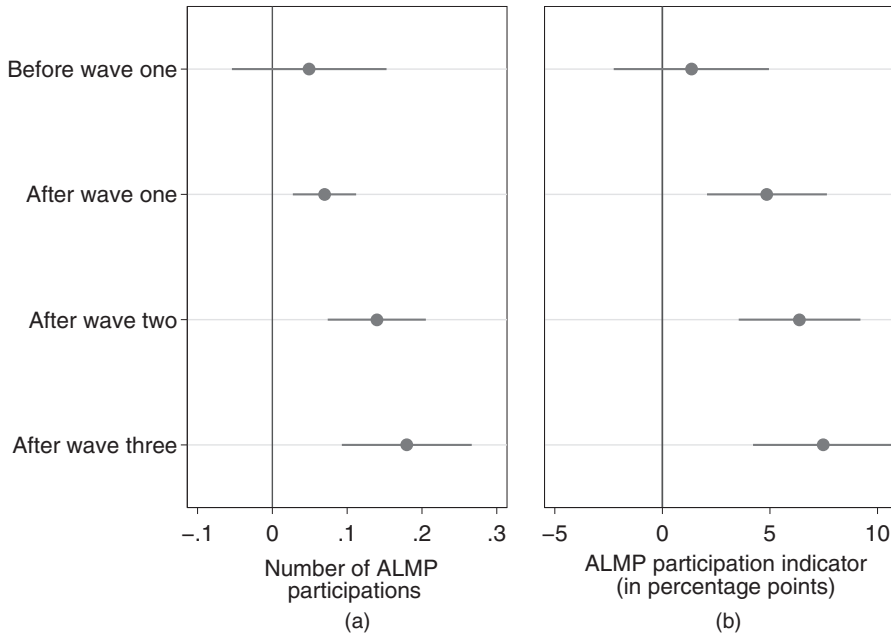

Fig. 2. Point estimates of panel conditioning effects with 95% confidence intervals: estimates from the IV analysis using the LARF estimator and controlling for covariates X (see Section 3.5); $N = 76700$

wave of the survey on respondents' take-up of ALMPs (second row). Respondents participate in about 0.07 more programmes (Fig. 2(a)) and participation increases the probability of taking up an ALMP by about 4.8 percentage points (Fig. 2(b)). After wave 2, the magnitude of both effects increases to about 0.14 programmes and 6.4 percentage points. Wave 3 leads to a further increase in effect size to about 0.18 and 7.5. Comparing the LARF estimates (Fig. 2) with the ITT estimates (Fig. 1), we see the expected increase in effect size (recall that the ITT underestimates the true effect because it does not adjust for non-compliance). However, we also see that, mainly because of non-compliance among participants who were invited to participate in the survey, the confidence intervals of the LARF estimates are much wider.

Next, we analyse whether the survey also affects respondents' employment probabilities. Because ALMPs are designed to increase employment probabilities for the unemployed, we expect that the increase in ALMP participation leads to a decrease in the time that respondents need to find a job and that respondents find a job more often than people who were not interviewed. Using the same LARF estimator, we do not find that respondents, after participating in up to three waves of the panel, find a job faster (17.22 days with a standard error of 26.80). Similarly, respondents do not find a job more often than people who were not interviewed (−0.73 percentage points with a standard error of 1.43). Thus, we do not find evidence that the increase in ALMP participation results in a significant change in employment probability.

Taken together, these results are strong evidence for the presence of panel conditioning effects in the PASS survey. Survey participation seems to increase the number of programmes that people participate in and to increase the likelihood of participating in ALMPs. Thus, repeated participation in the PASS survey changes respondents' actual labour market behaviour, i.e. respondents are more likely to participate in ALMPs, and to participate in more programmes, than similar people who are not exposed to the survey. Moreover, these effects seem to increase with the number of survey waves. Contrary to our expectations, survey participation does not

affect respondents' probabilities of employment. However, taking all the results and the falsification test together, we find strong evidence of changes-in-behaviour panel conditioning in the PASS survey.

5. Discussion

In this study, we have regarded selection and participation in one or more of the first three waves of a large panel survey as a treatment and have used techniques of causal analysis to estimate changes in respondent behaviour due to panel conditioning as a treatment effect. The results revealed that respondents of the survey are more likely to participate in ALMPs, and to participate in more programmes, than a group of eligible but unselected unemployment benefit recipients. Each additional exposure to the treatment, i.e. each additional wave, intensifies this effect. We do not see an increase in respondents' probabilities of employment.

Because many people who were selected for the survey did not respond, our analysis accounts for non-compliance with the treatment that was assigned. We relied on an IV approach and instrumented actual participation in the survey with the random assignment of people to the survey. In addition, we discussed the assumptions that are necessary to identify the LARF with this instrument in detail and addressed potential concerns about model misspecification or violated assumptions with a falsification test.

Two theoretical mechanisms offer plausible explanations why panel participation can change respondents' behaviour. We hypothesized that asking people (repeatedly) about a specific behaviour works as a stimulus, which increases respondents' awareness and motivates them to take up this behaviour. Another hypothesis holds that, feeling embarrassed about reporting non-normative behaviour, people bring their behaviour in line with society's norms. In line with these hypotheses, we found that respondents' labour market behaviour is changed by (repeated) participation in the survey. However, we cannot distinguish between these two mechanisms.

The true panel conditioning effect may be larger than that shown in Fig. 2. We probably underestimated the effects of repeated participation in the survey for two reasons. First, we treated all members of a responding household as respondents because of privacy regulations that prevent us from identifying individuals within households. Not all individuals in responding households responded to the survey, however, and therefore did not receive the stimulus. Thus, our treatment group included some untreated cases. Second, the definitions of the treatments after waves 2 and 3 also include people who responded in one wave only and therefore received only a 'reduced' stimulus (see Table 2).

Our findings may have interesting implications for research on ALMPs, labour market policy and labour economics in general. Several studies in the field have found that assigning unemployment benefit recipients to an ALMP has a 'threat effect' (e.g. Van den Berg and Bergemann (2009), Fitzenberger *et al*. (2010) and Graversen and Larsen (2013)), i.e. unemployed individuals who are assigned to participate in an ALMP increase their job search activity and/or lower their reservation wages to find a job before the programme starts, to avoid having to participate in the programme. Such an effect may exist because individuals dislike the participation experience or because it reduces their leisure time or time available for job search. Participating in a panel survey that makes ALMPs more salient by asking questions about them, however, seems to drive people into programme participation. Whether it is preferable to bring people into jobs that pay less than their original reservation wages, or to stimulate them to participate in ALMPs through survey participation, however, is a question that we cannot answer in this paper, especially because we do not find evidence that the increase in programme participation that is caused by participating in the survey affects respondents' employment probabilities.

Future work should expand our results to other outcomes for which administrative data are available. We would also like to see our results replicated with other data sets and variables. However, we note that, in scenarios where external validation records are not available, a clear distinction between the two forms of panel conditioning (and elimination of other confounding sources of error) will be difficult. Unfortunately, few studies will have external validation data to hand.

Our results also suggest that the PASS recipient sample is no longer representative of all recipients in Germany (at least in terms of ALMP participation), because participation in the survey over the waves has changed respondents' behaviour. As a consequence, inference made from PASS data may be biased if it includes ALMP participation either as a dependent or independent variable. For example, assessments by using the PASS data of whether ALMP programmes help the unemployed to find a new job, which is an important public policy question, may apply only to PASS respondents and not to the larger recipient population, because the respondents have been changed by the survey.

We note that participation in federal labour market programmes is a rather specific form of (labour market) behaviour, and we cannot generalize these findings to other behaviours of interest in panel studies. Yet, with our example, we hope to raise researchers' attention to the fact that repeated participation in panel surveys can change respondents' behaviour, which is a fact that is often not acknowledged by researchers working with panel data.

The possibility that panel data such as from the PASS survey are biased because of changes-in-behaviour and/or changes-in-reporting panel conditioning has been acknowledged for a long time. However, to date, panel conditioning has been primarily studied by researchers from survey methodology or survey research, and the majority of work has been published in corresponding journals. Yet, as panel data and panel methods have become more popular in recent years with social scientists and economists as tools to uncover causal effects, applied researchers need to be aware that such data can come with new sources of error. Other panel-specific sources of error, such as attrition, have been widely acknowledged by researchers and are addressed, for example, by weighting methods or by introducing refreshment samples. Panel conditioning, by contrast, is often ignored. Our results suggest that this strategy is unwise, because panel conditioning can have strong effects on substantively important variables.

Acknowledgements

We are especially grateful to Mark Trappmann, Steffen Kaimer and Alexander Mosthaf who helped us to obtain the data for this paper. We are also grateful to Katharine G. Abraham, Johannes Bauer, John Czajka, Andreas Diekmann, David Drukker, Georg-Christoph Haas, Monika Jungbauer-Gans, Florian Keusch, Frauke Kreuter, Gerhard Krug, Johannes Ludsteck, Austin Nichols, Joseph Sakshaug, Malte Schierholz, members of the Department of Sociology's Research Colloquium at Ludwig Maximilians University, Munich, the Associate Editor and two reviewers who provided helpful comments on earlier drafts of this paper. Ruben Bach gratefully acknowledges financial support from the graduate programme of the Institute for Employment Research and the University of Erlangen-Nuremberg.

References

Abadie, A. (2003) Semiparametric instrumental variable estimation of treatment response models. *J. Econmetr.*, **113**, 231–263.

An, W. and Wang, X. (2016) LARF: instrumental variable estimation of causal effects through local average response functions. *J. Statist. Softwr.*, **71**, 1–13.

Angrist, J., Imbens, G. and Rubin, D. (1996) Identification of causal effects using instrumental variables. *J. Am. Statist. Ass.*, **91**, 444–472.

Angrist, J. and Pischke, J.-S. (2009) *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton: Princeton University Press.

Athey, S. and Imbens, G. W. (2017) The state of applied econometrics: causality and policy evaluation. *J. Econ. Perspect.*, **31**, 3–32.

Berg, M., Cramer, R., Dickmann, C., Gebhardt, D., Gilberg, R., Jesske, B., Marwinski, K., Wenzig, C. and Wetzel, M. (2010) Codebook and documentation of the panel study 'Labour Market and Social Security' (PASS)—datenreport wave 3. *FDZ Datenreport 06/2010*. Nuremberg: Institute for Employment Research.

Borle, S., Dholakia, U. M., Singh, S. S. and Westbrook, R. A. (2007) The impact of survey participation on subsequent customer behavior: an empirical investigation. *Marktng Sci.*, **26**, 711–726.

Bundesagentur für Arbeit (2017) Bedarfsgemeinschaften und deren Mitglieder—Deutschland mit Ländern und Kreisen (unrevidiert)—Juli 2006. Bundesagentur für Arbeit, Nürnberg. (Available from `https://statistik.arbeitsagentur.de/Statistikdaten/Detail/200607/iiia7/gs-asu-sgbii-rev/gs-asu-sgbii-rev-d-0-xls.xls`.)

Cantor, D. (2010) A review and summary of studies on panel conditioning. In *Handbook of Longitudinal Research: Design, Measurement and Analysis* (ed. S. Menard), pp. 123–138. Amsterdam: Elsevier.

Chandon, P., Morwitz, V. G. and Reinartz, W. J. (2005) Do intentions really predict behavior?: Self-generated validity effects in survey research. *J. Marktng*, **69**, no. 2, 1–14.

Christoph, B., Müller, G., Gebhardt, D., Wenzig, C., Trappmann, M., Achatz, J., Tisch, A. and Gayer, C. (2008) Codebook and documentation of the panel study 'Labour Market and Social Security' (PASS). *FDZ Datenreport 05/2008*, vol. 1, *Introduction and Overview, Wave 1 (2006/2007)*. Nuremberg: Institute for Employment Research.

Clausen, A. R. (1969) Response validity: vote report. *Publ. Opin. Q.*, **32**, 588–606.

Clinton, J. D. (2001) Panel bias from attrition and conditioning: a case study of the knowledge networks panel. *Technical Report*. Department of Political Science, Stanford University, Stanford.

Crépon, B. and van den Berg, G. (2016) Active labor market policies. *A. Rev. Econ.*, **8**, 521–546.

Crossley, T., de Bresser, J., Delaney, L. and Winter, J. (2017) Can survey participation alter household saving behavior? *Econ. J.*, **127**, 2332–2357.

Das, M., Toepoel, V. and van Soest, A. (2011) Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociol. Meth. Res.*, **40**, 32–56.

Dholakia, U. M. (2010) A critical review of question-behavior effect research. *Rev. Marktng Res.*, **7**, 147–199.

Duan, N., Alegria, M., Canino, G., McGuire, T. and Takeuchi, D. (2007) Survey conditioning in self-reported mental health service use: randomized comparison of alternative instrument formats. *Hlth Res. Educ. Trust*, **42**, 890–907.

Fitzenberger, B., Osikominu, A. and Paul, M. (2010) The heterogeneous effects of training incidence and duration on labor market transitions. *Discussion Paper 5269*. Institute for the Study of Labor, Bonn. (Available from `ftp://repec.iza.org/RePEc/Discussionpaper/dp5269.pdf`.)

Gebhardt, D., Müller, G., Bethmann, A., Trappmann, M., Christoph, B., Gayer, C., Müller, B., Tisch, A., Siflinger, B., Kiesl, H., Huyer-May, B., Achatz, J., Wenzig, C., Rudolph, H., Graf, T. and Biedermann, A. (2009) Codebook and documentation of the panel study 'Labour Market and Social Security' (PASS). *FDZ Datenreport 06/2009*, vol. I, *Introduction and Overview, Wave 2 (2007/2008)*. Nuremberg: Institute for Employment Research.

Granberg, D. and Holmberg, S. (1992) The Hawthorne effect in election studies: the impact of survey participation on voting. *Br. J. Polit. Sci.*, **22**, 240–247.

Graversen, B. K. and Larsen, B. (2013) Is there a threat effect of mandatory activation programmes for the long-term unemployed? *Empir. Econ.*, **44**, 1031–1051.

Groves, R. M., Cialdini, R. M. and Couper, M. P. (1992) Understanding the decision to participate in a survey. *Publ. Opin. Q.*, **56**, 475–495.

Groves, R. M., Singer, E. and Corning, A. (2000) Leverage-saliency theory of survey participation: description and an illustration. *Publ. Opin. Q.*, **64**, 299–308.

Halpern-Manners, A. and Warren, J. R. (2012) Panel conditioning in longitudinal studies: evidence from labor force items in the Current Population Survey. *Demography*, **49**, 1499–1519.

Halpern-Manners, A., Warren, J., and Torche, F. (2017) Panel conditioning in the general social survey. *J. Am. Statist. Ass.*, **46**, 103–124.

Imbens, G. and Angrist, J. (1994) Identification and estimation of local average treatment effects. *Econometrica*, **62**, 467–475.

Institute for Employment Research (2013) *Integrated Employment Biographies: V11.00.00*. Nuremberg: Institute for Employment Research.

Jacobebbinghaus, P. and Seth, S. (2007) The German integrated employment biographies sample IEBS. *Zeits. Wirts. Sozlwissens.*, **127**, 335–342.

Jacobi, L. and Kluve, J. (2007) Before and after the Hartz reforms: the performance of active labour market policy in Germany. *J. Lab. Markt Res.*, **40**, 45–64.

Köhler, M. and Thomsen, U. (2009) Data integration and consolidation of administrative data from various sources: the case of Germans' employment histories. *Hist. Socl Res.*, **34**, 215–229.

Kraut, R. E. and McConahay, J. B. (1973) How being interviewed affects voting: an experiment. *Publ. Opin. Q.*, **37**, 398–406.
Kreuter, F., Müller, G. and Trappmann, M. (2010) Non-response and measurement error in employment research: making use of administrative data. *Publ. Opin. Q.*, **74**, 880–906.
Lazarsfeld, P. (1940) 'Panel studies'. *Publ. Opin. Q.*, **4**, 122–128.
Lepkowski, J. and Couper, M. (2002) Nonresponse in the second wave of longitudinal household surveys. In *Survey Nonresponse* (eds R. D. Groves, J. E. Dillman and R. Little), pp. 259–271. New York: Wiley.
Morwitz, V. G., Johnson, E. and Schmittlein, D. (1993) Does measuring intent change behavior? *J. Consmr Res.*, **20**, 46–61.
Pennell, S. G. and Lepkowski, J. N. (1992) Panel conditioning effects in the survey of income and program participation. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 566–571.
R Core Team (2017) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Rizzo, L., Kalton, G. and Brick, M. (1996) A comparison of some weighting adjustment methods for panel nonresponse. *Surv. Methodol.*, **22**, 43–53.
Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.*, **66**, 688–701.
Rubin, D. B. (1977) Assignment to treatment group on the basis of a covariate. *J. Educ. Statist.*, **2**, 1–26.
Rubin, D. B. (1978) Bayesian inference for causal effects: the role of randomization. *Ann. Statist.*, **6**, 34–58.
Rudolph, H. and Trappmann, M. (2007) Design und Stichprobe des Panels "Arbeitsmarkt und soziale Sicherung" (PASS). In *Neue Daten für die Sozialstaatsforschung: zur Konzeption der IAB-Panelerhebung "Arbeitsmarkt und Soziale Sicherung"* (ed. M. Promberger), pp. 60–101. Nuremberg: Institute for Employment Research.
Schnell, R. (2007) Alternative Verfahren zur Stichprobengewinnung für ein Haushaltspanelsurvey mit Schwerpunkt im Niedrigeinkommens- und Transferleistungsbezug. In *Neue Daten für die Sozialstaatsforschung: zur Konzeption der IAB-Panelerhebung "Arbeitsmarkt und Soziale Sicherung"* (ed. M. Promberger), pp. 33–59. Nuremberg: Institute for Employment Research.
Smith, J. K., Gerber, A. S. and Orlich, A. (2003) Self-prophecy effects and voter turnout: an experimental replication. *Polit. Psychol.*, **24**, 593–604.
Spangenberg, E. and Obermiller, C. (1996) To cheat or not to cheat: reducing cheating by requesting self-prophecy. *Marktng Educ. Rev.*, **6**, no. 3, 95–103.
Stock, J. and Yogo, M. (2005) Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (eds D. Andrews and J. Stock), pp. 80–108. New York: Cambridge University Press.
Sturgis, P., Allum, N. and Brunton-Smith, I. (2009) Attitudes over time: the psychology of panel conditioning. In *Methodology of Longitudinal Surveys* (ed. P. Lynn), pp. 113–126. New York: Wiley.
Trappmann, M., Beste, J. Bethmann, A. and Müller, G. (2013) The PASS panel survey after six waves. *J. Lab. Markt Res.*, **46**, 275–281.
Traugott, M. W. and Katosh, J. P. (1979) Response validity in surveys of voting behavior. *Publ. Opin. Q.*, **43**, 359–377.
Van den Berg, G. J. and Bergemann, A. H. (2009) The effect of labor market programs on not-yet treated unemployed individuals. *J. Eur. Econ. Ass.*, **7**, 606–616.
Warren, J. R. and Halpern-Manners, A. (2012) Panel conditioning in longitudinal social science surveys. *Sociol. Meth. Res.*, **41**, 491–534.
Waterton, J. and Lievesley, D. (1989) Evidence of conditioning effects in the British Social Attitudes panel. In *Panel Surveys* (eds D. Kasprzyk, G. J. Duncan, G. Kalton and M. P. Singh), pp. 319–339. New York: Wiley.
Williams, P., Block, L. G. and Fitzsimons, G. J. (2006) Simply asking questions about health behaviors increases both healthy and unhealthy behaviors. *Socl Inflnce*, **1**, 117–127.
Williams, W. H. and Mallows, C. L. (1970) Systematic biases in panel surveys due to differential nonresponse. *J. Am. Statist. Ass.*, **65**, 1338–1349.
Yalch, R. F. (1976) Pre-election interview effects on voter turnout. *Publ. Opin. Q.*, **40**, 331–336.
Yan, T. and Eckman, S. (2012) Panel conditioning: change in true value versus change in self-report. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 4726–4736.
Zabel, J. (1998) An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. *J. Hum. Resour.*, **33**, 479–506.
van der Zouwen, J. and van Tilburg, T. (2001) Reactivity in panel studies and its consequences for testing causal hypotheses. *Sociol. Meth. Res.*, **30**, 35–56.
Zwane, A. P., Zinman, J., van Dusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D. S., Hornbeck, R., Giné, X., Duflo, E., Devoto, F., Crepon, B. and Banerjee, A. (2011) Being surveyed can change later behavior and related parameter estimates. *Proc. Natn. Acad. Sci. USA*, **108**, 1821–1826.

Supporting information
Additional 'supporting information' may be found in the on-line version of this article.