

vom Berge, Philipp; Umkehrer, Matthias; Wanger, Susanne

## Article

# A correction procedure for the working hours variable in the IAB employee history

Journal for Labour Market Research

## Provided in Cooperation with:

Institute for Employment Research (IAB)

*Suggested Citation:* vom Berge, Philipp; Umkehrer, Matthias; Wanger, Susanne (2023) : A correction procedure for the working hours variable in the IAB employee history, Journal for Labour Market Research, ISSN 2510-5027, Springer, Heidelberg, Vol. 57, Iss. 1, pp. 1-16, <https://doi.org/10.1186/s12651-023-00331-0>

This Version is available at:

<https://hdl.handle.net/10419/278087>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

ORIGINAL ARTICLE

Open Access



# A correction procedure for the working hours variable in the IAB employee history

Philipp vom Berge<sup>\*</sup> , Matthias Umkehrer and Susanne Wanger

## Abstract

Administrative labour market data for Germany do not contain detailed information on working hours. This poses a serious challenge for many empirical research questions. Between 2010 and 2014, however, it is possible to merge a supplementary data source containing information on working hours reported by employers for each job to the German Social Accident Insurances. One complicating factor is that employers were allowed to report actual working hours, contractual working hours or a full-time worker reference value, and it is not obvious from the data which reporting scheme was actually used. In this report, we describe this supplementary data source and propose a correction procedure that maps all data entries so that they uniformly reflect contractual working time. After this correction, the distribution of contractual weekly working hours in the combined administrative labour market data closely resembles the distribution from the German Structural Earnings Survey.

**Keywords** Administrative labour market data, BeH, Correction procedure, SIAB, Working hours

**JEL Classification** C80, J30

## 1 Introduction

In Germany, employers have to notify social insurance institutions of each individual employment relationship (“job” from here on) at least once a year. These notifications are an important data source for scientific labour market research. The Institute for Employment Research (IAB) of the Federal Employment Agency (BA) consolidates and prepares these data in the so-called Employee History (Beschäftigtenhistorik, BeH) data set. The BeH provides detailed information on all jobs in Germany that are subject to social security contributions from the year 1975 onwards. Marginal part-time jobs, which are exempt from social security contributions, were added to the BeH in 1999, too. The BeH is an important building block of many data products of the IAB, including the Integrated Employment Biographies (IEB) and

standardised data products offered to external researchers via the Research Data Centre (Forschungsdatenzentrum, IAB-FDZ).

While the BeH includes the working-time status of a given job, i.e. full-time, part-time or marginal part-time, it does not provide exact information on the number of hours worked. The lack of working hours in these data, however, poses a challenge for many research questions, such as analyses of the gender wage gap, wage inequality and minimum wages, to name only a few. Fortunately, at least for the years 2010 to 2014, the number of total hours worked during a period of notification can be merged into the BeH from supplementary data for each job reported by an employer.

This supplementary data source originates from employers reporting to the German Social Accident Insurances (Deutsche Gesetzliche Unfallversicherung, DGUV), which was added to the regular social security notification process for the 2010 to 2014 period. One drawback of these reports was that employers were able to choose between several reporting schemes. These included actual working hours (excluding annual and sick

\*Correspondence:

Philipp vom Berge

philipp.vom-berge@iab.de

Institute for Employment Research (IAB), Regensburger Str. 104,  
90478 Nuremberg, Germany

leave, but including overtime), contractual or collectively agreed working hours (including annual and sick leave, but excluding overtime), a special full-time worker reference value that varies by year (in case none of the above were easily reportable), or an educated guess. Unfortunately, it is not possible to observe which scheme was chosen for a given job, leading to major inconsistencies in the raw data. It cannot be ruled out that some employers even chose different schemes for different jobs.

In this report, we develop a correction procedure to deal with these different reporting schemes. First, we estimate the probabilities that a given value of working time reflects actual hours, contractual hours or the full-time reference values. Second, based on these estimated probabilities, we propose ways of adjusting all working hours reports so that they uniformly reflect contractual working time. We argue that these corrected working hours are much more useful for empirical research than the uncorrected ones in most applications. After correction, the distribution of contractual weekly working hours closely resembles the distribution from the German Structural Earnings Survey (SES).<sup>1</sup>

The rest of the report is structured as follows. In Sect. 2, we provide a brief institutional background for the data module “German Social Accident Insurance” of German social security notifications and discuss the limitations of its working hours information. In Sect. 3, we describe the algorithm behind our correction procedure. In Sect. 4, we compare the distributions of uncorrected and corrected working hours in the BeH to the distribution of working hours in the SES and show that these distributions resemble each other quite closely after correction—but not before. In Sect. 5, we discuss strengths and limitations of our approach. In Sect. 6, we comment on data availability, before we conclude in Sect. 7.

## 2 Working hours data from the module “German Social Accident Insurance”

### 2.1 Institutional background

DGUV is a branch of the German social security system that insures employees against the consequences of work and commuting accidents and occupational diseases. Within the German social security system, the notification procedure of the statutory accident insurance occupies a special position. Unlike the other branches of social insurance—pension, health, nursing care and

unemployment insurance—the DGUV is financed solely by employer contributions. In return, the DGUV releases the employer from liability for accidents at work and occupational diseases (DGUV 2016). Since only the employer pays the contributions, up to the year 2008 establishments reported accumulated information on annual wage totals, annual working hours and risk classes directly to the DGUV via an annual wage statement and not via the regular social security notification process.

In 2009, this procedure was replaced by an extended notification procedure. Instead of the (establishment-related) cumulative notification via the wage statement sent directly to the DGUV, the notification was now integrated into the regular social security notification process (Höller 2010). In 2016, this procedure was modified again. Since then, there has been a two-stage procedure: in addition to an (employee-related) annual notification of the employers to the statutory accident insurance in the regular social security notification process, the (establishment-related) wage statement directly to the DGUV—as common before 2009—was reintroduced again. As a result, working hours are again no longer part of the regular social security notification process after 2014 (DGUV 2018).

Our supplementary data source therefore dates from the years 2010 to 2014, when the wage statement and the reporting of working hours were integrated into the regular social security notification process.

### 2.2 Different reporting schemes

When assessing the quality of the variable “working hours”, it should be noted that the primary purpose of reporting was the implementation of the statutory accident insurance, i.e. mapping the accident risk of employees present at the workplace. For this purpose, the determination of working hours by the employer should, as far as possible, be based on the information already available in the establishment in order to limit the additional bureaucratic effort. The DGUV therefore proposed the following procedure for determining annual working hours (DRV 2008):

- Insofar as the actual hours worked per employee are recorded, e.g. by a time recording system or hourly logs, these should be reported by the establishment.
- If the actual working time is not recorded in this way, the target hours without overtime, i.e. the contractually or collectively agreed working time of the employees, should be reported.
- If this is not available either, the so-called “full-time worker reference value” or, in the case of part-time work, the corresponding proportion thereof should be reported. Instead of using the full-time worker ref-

<sup>1</sup> We choose contractual working time as our target measure, both because this makes comparison to the SES easier and because the project started out with minimum wage research in mind, where contractual working time is more useful. With the supplementary data described in this report, however, researchers can in principle reweight the corrected working hours to reflect actual working hours.

erence value, the employer may also provide a conscientious estimate of the hours worked.

The scheme according to which an establishment reports working hours also depends on the contribution standard to which the DGUV has assigned the establishment. It differs in part for the three branches of DGUV—the accident insurance institutions for trade and industry, those for the public sector, and those for the agricultural sector (DGUV 2016). The contribution standard is determined according to the company’s sector and also indicates the basis on which the contribution to statutory accident insurance is calculated. Contribution standards during the period 2010 to 2014 were for example remuneration, number of insured persons or number of inhabitants. In particular, many employers in the public sector report according to the latter two criteria.

If the contribution standard is not based on remuneration, no wage statement and therefore no reporting of hours worked is required from the establishment (DRV 2011). As a result, the variables for working hours in these establishments contain missing values (see also Sect. 3.1 and Appendix A.4).

### 2.3 Limitations of the uncorrected variable “working hours”

#### 2.3.1 Lack of comparability of the different reporting schemes

When reporting actual hours worked, e.g. via time recording systems, it is possible to measure very precisely how many hours an employee has worked or was present at the establishment. However, many types of wage analyses require information on hours paid, i.e. including in particular additional information on annual and sick leave.

When reporting the target working time, i.e. the collectively agreed or contractual working hours, the working time agreed in the contract is indicated. However, overtime is not taken into account. This might be problematic for groups for which no contractual working hours are fixed. These can be, for example, employees with performance pay or trust-based working hours and managers.

Due to a number of weaknesses, the full-time worker reference value has the least informative value. It functions as a simplified auxiliary value for the specification of working time and specifies a flat number of yearly working hours.

The full-time worker reference value applies uniformly to all sectors and is recalculated annually by the DGUV using statistics from the Federal Statistical Office and the health insurance companies (cf. Lehner and Rupert 2009 and the sample calculation in Appendix A.1). It takes calendar working days, public holidays, average

days of annual and sick leave and paid weekly hours into account. It does this, however, not on the basis of data from the current year, but from the calendar year two years ago. Due to an overestimation of sick days, the full-time worker reference value is regularly biased downward (for detailed information see also Appendix A.1). According to the DGUV definition, the reference value for a “full-time worker” corresponds to the average number of hours actually worked per year by a fully employed person in the commercial economy. The full-time worker reference value therefore only reflects a macroeconomic average that does not differentiate by sector, which might be problematic in many types of analysis. Thus, the statutory accident insurance itself is actually critical about using the value and advises to use it only if time recording systems or a target working time are not available.

Because of these identified weaknesses, we filter out the information that is presumed to be based on the full-time worker reference value from the BeH working hours variable and adjust it in the proposed correction procedure to uniformly reflect a contractual workweek of 39 h in the case of full-time work (cf. comments in Sect. 3.5).

#### 2.3.2 Different reporting periods of employment notifications

Employers were obliged to report working hours for the following employment notifications:

- Annual notifications: notifications for all employees subject to social security or in marginal part-time employment on December 31st.
- Deregistrations: notifications due to end of employment.
- Employment interruption notifications: deregistrations due to interruption of employment for more than one month because the entitlement to remuneration ceases.

Accordingly, notifications of working hours are irregularly scattered throughout the year, and the notifications on working hours always refer to a specific employment period. These can cover the entire calendar year or, in extreme cases, just one day. In order to obtain comparable data on working hours, the different lengths of the spells must be normalized to a uniform period. This is done by converting all reports to average weekly working hours (cf. comments in Sect. 3.4).

#### 2.4 Missings or cases in which reporting is not required

The share of employees with missing values for the working time variable is comparatively high at nearly 16 percent. Several factors lead to this high proportion of missing values. First, as described above, not all sorts of

spells contain working hours information, for example spells reflecting separate reports of one-time-payment. Second, there are a number of groups of employees for whom certain parts of the declaration, including hours worked, do not need to be filled. This is always the case if the DGUV does not require any data from the employment notification procedure for the calculation of contributions. Thus, for example, the person groups maritime pilots, unsteady workers, marginal part-time workers (household cheque) or people in short-term employment (household cheque) are exempt from reporting hours worked and therefore notifications are missing by design.<sup>2</sup> Third, there are establishments that do not have to provide information because they are exempt from the obligation to report to DGUV.<sup>3</sup> This is relevant for the employees of the DGUV and for employees who are insured with agricultural accident insurance institutions. Furthermore, this is the case when the contribution is calculated according to a criterion other than remuneration for work (see Sect. 2.2) or there is no accident insurance obligation due to employment abroad. And last, there is a proportion of establishments that do not report for unknown reasons.

As a result, there is some selectivity of missing values in the working hours variable (see Appendix A.4 for more information on this subject).

### 3 Correction procedure

As discussed in the previous section, the variable “working hours” can be reported using different schemes. These measure different aspects and vary in their degree of validity. Unfortunately, it is not possible to observe which reporting scheme was chosen for a given job, leading to major inconsistencies in the raw data. It cannot be ruled out that some employers even choose different schemes for different jobs. Therefore, the uncorrected variable “working hours” is hardly usable for labour market research in its original form. A correction procedure must therefore consider the different data collection methods. The aim of our correction process is to propose a procedure that creates a harmonised measure of contractual working hours from the supplementary hours data. This new variable will be useful for many types of econometric analyses, such as wage analyses.

Our correction algorithm proceeds in the following four steps:

Step 1: Estimate the probability of each reporting scheme (actual/contractual/reference value) for a given establishment.

Step 2: Estimate the probability of each reporting scheme for a given value of working hours (irrespective of the establishment).

Step 3: Construct a combined probability of each reporting scheme for a given observation combining the probabilities estimated for a given establishment and the probabilities estimated for a given value of working hours.

Step 4: Correct for annual and sick leave and adjust the full-time worker reference values.

Before we explain these steps in detail, we describe our basic preparation of the BeH data in the next subsection.<sup>4</sup>

#### 3.1 Basic data preparation

We start from all job notifications (“spells” from here on) in the BeH for the years 2010 to 2014 and merge the raw working hours information to each spell. As discussed in Sect. 2.3, some spells have no working hours information, like spells reflecting separate reports of one-time payments, sectors for which no working hours have to be reported or cases where employers failed to report. These spells make up roughly 16 percent of all selected spells and we drop all of them.<sup>5</sup>

##### 3.1.1 Imputation of the working-time status variable

For our correction procedure, we rely on the working-time status included in the BeH. This variable identifies a job as full-time, part-time, or marginal part-time. However, this information is not without its own problems. Most importantly, a 2011 revision of occupational codes was accompanied by a significant increase in the number of missing values in the working-time status variable in 2011. To fill these gaps, we rely on the adjustment implemented by Ludsteck and Thomsen (2016). Additionally, we fill remaining gaps by writing backwards and forwards, prioritizing information after 2011, to make the time series more consistent. By doing so, the following special case has to be highlighted.

##### 3.1.2 Dealing with misreported shifts in working-time status within jobs

Especially when comparing 2011 with 2012, not all shifts from full-time to part-time, and vice versa, within the same job appear to reflect true changes but rather a delay in employers updating the working-time status

<sup>2</sup> These person groups of the employment notification procedure are coded in the variable “Employment status (erwstat)”.

<sup>3</sup> According to the DGUV, it is not possible to delimit these establishments according to economic activity or person group keys in the employment statistics.

<sup>4</sup> An alternative correction procedure is briefly described in Appendix A.2. This alternative procedure was used in Dustmann et al. (2022).

<sup>5</sup> See Appendix A.4 for an analysis of the selectivity of those missing values.

information.<sup>6</sup> We are suspicious because in some of these cases the earnings do not change plausibly with the working-time status shifting. We therefore assume a shift in working-time status within the same person, establishment and occupation to be real if average daily earnings decline by at least 15 percent (full-time to part-time shift) or increase by at least 10 percent (part-time to full-time shift), respectively. We arrive at these thresholds because full-time job stayers usually do not have wage increases exceeding 20 percent and part-time job stayers usually do not have wage decreases below 25 percent and we also allow for a buffer of 10 percentage points. Finally, if a working-time status shift within a job is not accompanied by a plausible—according to our thresholds—earnings shift, we replace the working-time status before 2012 with that from 2012.<sup>7</sup>

Finally, we calculate working hours per week, accounting for the different lengths of the spells, as follows:

$$h_{week} = \frac{h_{total} * d_y * 5}{d_{total} * wd_y} \quad (1)$$

where  $h_{total}$  depicts total working hours in the spell,  $d_{total}$  the length of the spell in days,  $d_y$  calendar days in year  $y$  (365 or 366) and  $wd_y$  official working days in year  $y$ . We take the number of working days per year from the IAB's working time measurement concept (Wanger et al. 2016; IAB 2022).

### 3.2 Step 1 of 4: estimate the probability of each reporting scheme for a given establishment

The idea behind the first step is that relatively high values of weekly working hours will more likely reflect contractual working hours than actual working hours or the relatively low full-time worker reference value (keeping in mind that actual working hours *exclude* annual and sick leave in the DGUV definition, see Sect. 1).<sup>8</sup> We use this information to estimate how likely it is that a

particular establishment reports either contractual or actual working hours (or the full-time worker reference value). We perform this step using only spells of full-time work lasting the whole year, because other spells make the selection much more complicated.<sup>9</sup> Furthermore, we ignore weekly working hours below 25 and above 50 in this step because these might be misreported. This step involves about 30 percent of all spells and 35 percent of all establishments.

We choose a threshold of 35 working hours per week and (provisionally) define all full-time year-round spells as contractual if they exceed this value. The threshold is justified by the fact that contractual hours of full-time work are almost never below 35 according to both collective agreements (WSI 2014) and official statistics (Statistisches Bundesamt 2016, Sect. 1.6.5). Note that actual weekly working hours of a worker might also be above 35 if contractual working hours are high while days of absence are low or accumulated net overtime work is high.<sup>10</sup> Therefore, the threshold is chosen to strike a balance between two concepts that will somewhat overlap in reality.

Figure 1 depicts the distribution of working hours per week for year-round spells of full-time workers in our raw data in 2014, showing the 35 working hours per week threshold as well as the clustering of notifications at the full-time worker reference value of 31.7 h per week (calculated as  $1580/249.3*5$ ).

We then calculate from all full-time year-round spells the following two elements to estimate the probabilities of each reporting scheme for each establishment:

- a) The proportion of spells with weekly working hours exceeding 35 (likely contractual), not exceeding 35 (likely actual) or equalling exactly the year-specific full-time worker reference value.
- b) A weighting factor that reflects our confidence in the establishments' reporting and its relevance for subsequent calculations.

In Fig. 2, we show how the average proportions according to a) vary with establishment size. Establishments with only one single employee (in full-time year-round employment) are somewhat more likely to report contractual hours than actual hours or a reference value,

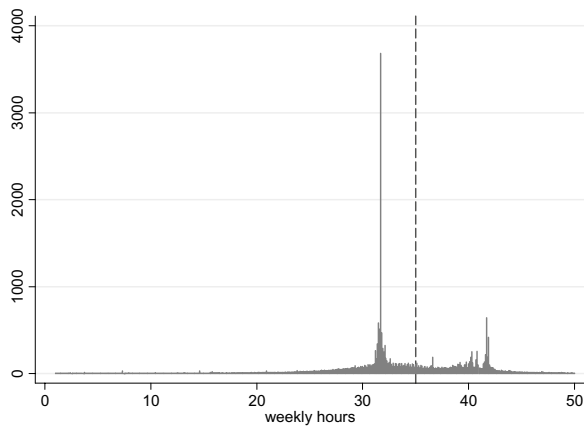
<sup>6</sup> Problems stemming from the working-time status not being properly updated have been recognized in the literature for quite a while. These include breaks in time series for wages and wage inequality (Fitzenberger and Seidlitz 2020) and women being mislabelled as still working full-time after returning from maternal leave (Frodermann et al. 2013). The correction we use here is not intended to make the data on working-time status consistent for long-run analysis, but only for the 2010–2014 window we use for our correction procedure. We acknowledge that some misclassification might still be present after these steps.

<sup>7</sup> There are very few occasions (<0.1 percent) with missing information on the working-time status variable even after those imputation steps. We drop those in what follows.

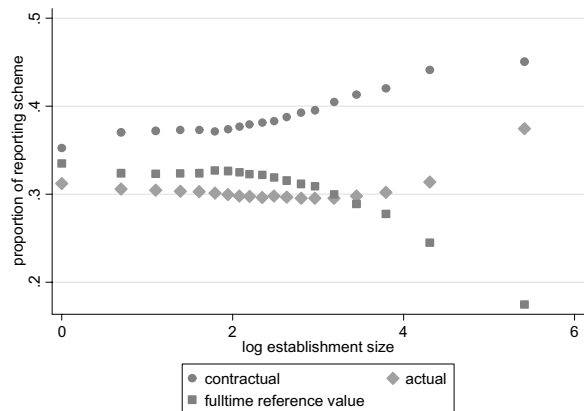
<sup>8</sup> In the IAB-working time measurement concept (IAB 2022) for example, actual working days excluding annual and sick leave usually vary somewhere between 205 and 210 days per year, significantly below potential working days that range between 248 and 252 days per year. Overtime is generally not high enough to completely compensate for this discrepancy.

<sup>9</sup> For part-time workers, we do not observe the proportion of a full-time job. For spells that do not span the whole year, fractions of the full-time worker reference value become blurry due to rounding.

<sup>10</sup> As an example, with contractual working hours set at 40, 250 working days a year, annual leave at the legal minimum of 20 and no sickness leave during the year, a worker would already have  $40*230/250=36.8$  actual working hours per week.



**Fig. 1** Distribution of raw weekly working hours in 2014—full-time workers. The figure shows the distribution of raw weekly working hours in 2014 for year-round spells of full-time workers; limited to the 0–50 h window. 35 working hours per week threshold marked as dashed line. Full-time worker reference value is the big spike at 31.7 h. Source: IAB Beschäftigtenhistorik (BeH) V10.06.00-202012; working hours from DGUV notifications



**Fig. 2** Probability of reporting schemes at the establishment level. The figure shows the proportion of spells likely to reflect contractual hours, actual hours or a full-time worker reference value among each establishments’ full-time year-round notifications by establishment size (in logs), without confidence weights. Taking logs for the x-axis is chosen because the establishment size distribution is highly skewed. Source: IAB Beschäftigtenhistorik (BeH) V10.06.00-202012; working hours from DGUV notifications

by about four percentage points, but these differences are not particularly pronounced. On average, about 42 percent of notifications involved in this step are likely to reflect contractual working hours. The likelihood to report a reference value, in turn, declines substantially with the size of establishments increasing (21 percent on average), while the likelihood to report actual working hours increases substantially in establishment size (38 percent on average). Overall, very large establishments

are more likely to report actual hours and contractual hours. They relatively rarely report reference values.

The second element (b) accounts for the fact that many establishments in our data do not follow one clear reporting scheme according to the proportions from (a). Others are small and we therefore observe only a small number of reports for calculating the proportions. We want to account for this uncertainty and consider it in our further calculations. To do so, we model a weight for each establishment  $e$  in a given year via the entropy of a Dirichlet distribution where the three concentration parameters  $\alpha$  are chosen to be the respective proportions  $s$  of the three groups  $i$  according to a), each supplemented by the inverse of establishment size<sup>11</sup>:

$$w_e = -\frac{1}{2} \left( \log \mathbf{B}(\boldsymbol{\alpha}) + (\alpha_0 - 3)\psi(\alpha_0) - \sum_{i=1}^3 (\alpha_i - 1)\psi(\alpha_i) \right) \quad (2)$$

where  $\mathbf{B}(\boldsymbol{\alpha}) = \prod_{i=1}^3 \Gamma(\alpha_i) / \Gamma(\alpha_0)$ ;  $\alpha_0 = (\alpha_1 + \alpha_2 + \alpha_3)$ ;  $\Gamma$ : the gamma function;  $\psi$ : the digamma function.

Ultimately, the weights have the following features:

- Small establishments get little weight, even if reporting clearly, as this might be a coincidence. For example, a one-worker establishment gets a weight of roughly 0.5. A clearly reporting three-worker establishment gets a weight of roughly 1.35.
- Establishments not reporting clearly also get little weight. A three-worker establishment reporting all three schemes gets a weight of roughly 0.42. The weight tends towards 1.077 as the establishment size grows as long as all reporting scheme are observed with equal proportions.
- Large and clearly reporting establishments receive a large weight. The weight tends towards the size of the establishment the clearer the reporting scheme.

We use these weights to weight the proportions estimated at the establishment level up or down, accordingly. These weighted proportions also enter the following steps. We will calculate similar weights for estimates at the level of each specific working hours value, too, which we outline next.

### 3.3 Step 2 of 4: estimate the probability of each reporting scheme for a given value of total working hours

In this second step, we determine—for each specific value of total working hours separately—the probabilities for

<sup>11</sup> This means that for each group  $i = 1, 2, 3$  we choose  $\alpha_i = s_i + 1/size$ . The supplementary term is chosen ad hoc and plays a role comparable to the hyper-parameters in a Bayesian prior. It expresses “ignorance” before we observe the proportions, but it also ensures a favourable scaling for our weights.

an assignment to the three different reporting schemes. We allow these probabilities to vary by the duration of the spell, working-time status and year. To do so, we take the reporting probabilities and weighting factors calculated for establishments in the previous step and transfer them to all spells of these establishments, i.e. also to spells not used previously, like part-time, marginal part-time, and full-time spells outside the primary selection window. The final estimates of the hours-specific reporting probabilities  $p_h$  are then the weighted averages of the establishment-specific probabilities for all  $n = 1, \dots, N$  relevant spells.<sup>12</sup>

$$p_{ih} = \frac{\sum_{n=1}^N (s_{ihn} * \omega_{en})}{\sum_{i=1}^3 \sum_{n=1}^N (s_{ihn} * \omega_{en})} \quad (3)$$

Next, we apply these estimated probabilities to all observations with this specific total working hours value, assuming that the hours-specific measure is a good approximation even for establishments without any full-time employees. Additionally, and analogous to Eq. (2) in the first step, we determine a year-specific weighting factor  $\omega_h$  to reflect our confidence in the reporting logic behind each specific working hours value, where the number of reports behind each value takes the role of establishment size in determining the concentration parameters.<sup>13</sup>

### 3.4 Step 3 of 4: construct a combined measure of the probability of each reporting scheme for a given observation

As our final measure of the probabilities on which reporting scheme a particular observation was based on, we combine the probabilities estimated for both establishments (step 1) and specific working hour values (step 2) and weight them with their respective confidence weights:

$$p_i = (s_{ie}\omega_e + p_{ih}\omega_h)/(\omega_e + \omega_h). \quad (4)$$

The latter assures that the information that we have more confidence in receives a higher weight. For instance, for very small establishments more importance is given to the probabilities estimated for the specific working hours value. Conversely, hour-specific measures with an unclear reporting scheme are sharpened considerably in large establishments, if these establishments report relatively clearly overall. If no establishment weight is

**Table 1** Relevant values for working hours correction in Eq. 5

	$wd_y^{act}$	$wd_y$	$r_y$
2010	213.0	253.0	1,610
2011	211.3	251.7	1,570
2012	209.6	249.6	1,600
2013	208.2	249.0	1,590
2014	208.6	249.3	1,580

Source: IAB working time measurement concept (IAB 2022), see also Wanger et al. (2016); full-time worker reference value from DGUV (2014)

available, we use only the hours-specific measure and vice versa. Finally, if neither an hours-specific nor an establishment-specific weighting factor can be determined for an observation, we use an average value that depends on the size of the establishment, the working-time status and the year.

### 3.5 Step 4 of 4: correct for annual and sick leave and adjust the full-time worker reference values

In this final step, we use our combined probability measure from step 3 to adjust the working hours for lacking annual leave and paid days in sick leave, but only to the extent that they reflect actual working hours reports or reports of the full-time worker reference value. We implement this adjustment by taking the raw total working hours reported for each spell and convert them into weekly working hours while using different numbers of workdays for the different reporting schemes (note that this differs from Eq. (1) where we only used workdays reflecting potential working days). Because we can only assign a given observation to one of the three reporting schemes with a certain probability, we calculate the final hours variable  $h_{week}^{corr}$  as a weighted average:

$$h_{week}^{corr} = \frac{h_{total} * d_y}{d_{total}} * \left( \frac{5}{wd_y^{act}} * p_{act} + \frac{5}{wd_y} * p_{con} + \frac{39}{r_y} * p_{ref} \right) \quad (5)$$

where  $h_{total}$ ,  $d_{total}$ ,  $d_y$  and  $wd_y$  are defined as before in Eq. (1).  $wd_y^{act}$  denotes average actual annual working days taken from the IAB's working time measurement concept (IAB 2022), where we derive actual working days by subtracting average days of absence due to annual and sick leave from potential annual working days.  $p_{act}$ ,  $p_{con}$  and  $p_{ref}$  are our combined probability measures for actual, contractual or reference value reporting. Finally,  $r_y$  is the year-specific full-time worker reference value, for which we assume a working week of 39 h. We depict the values used for our proposed adjustment in Table 1.

With the probabilities and confidence weights at hand, it is straightforward to modify these adjustments. For instance, we experimented with industry-specific values

<sup>12</sup> We omit subscripts for spell-duration, working-time status and year in Eq. (3) to save notation.

<sup>13</sup> That is, we now choose  $\alpha_i = p_i + 1/r$ , where  $r$  is the number of reports for the specific hours value.



instead of the averages outlined in Table 1, but this did not yield considerable improvements. Our proposed correction also does not include an overtime adjustment, which might be desirable in some research contexts. In Appendix A.3, we suggest one simple way to achieve this.

#### 4 Quality of the correction procedure

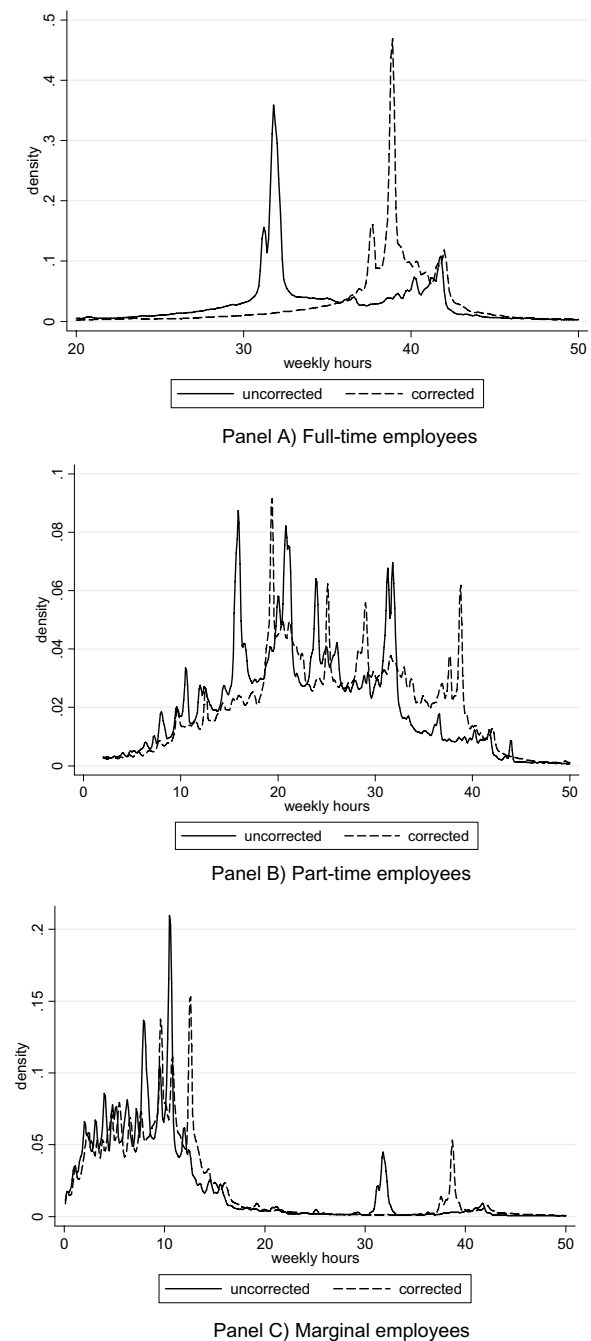
In this section, we describe the distribution of weekly working hours in the BeH before and after the correction for different reporting schemes, and compare it to the distribution from the German Structural Earnings Survey (SES, Statistisches Bundesamt 2016). Figure 3 shows kernel density estimates for working hours in the BeH for the year 2014, separately for full-time, part-time and marginal part-time employment in panels (A) to (C). Additionally, Table 2 displays the respective means of the distributions. As one special case of full-time work, we add the mean of working hours of apprentices in an extra row of the table.

For full-time workers, the raw weekly hours distribution shows a high density somewhere around 31 to 32 h. These values are implausibly small for full-time workers and reflect the full-time reference value. After correction, one large peak occurs at values around 38 to 39 h. Our correction procedure increases the means from 35 to 39 h of work per week for both regular full-time workers and apprentices, who are typically classified as working full-time, alike.

Not surprisingly, working hours per week are far more disperse for part-time than for full-time work. Before correction, the average weekly working hours of part-time workers in 2014 in the BeH was 21. This value increases to 24 h after correction while the whole distribution shifts to the right. The general pattern of the distribution remains largely intact before versus after correction.

For marginal part-time workers, the weekly working hours distribution is concentrated around values of seven to ten hours. The distribution also shows one spike somewhat above 30 h, which is likely the result of employers misreporting the full-time reference value but not accounting for the appropriate fraction of the marginal job. Therefore, caution should be taken with analyses involving working hours of marginal part-time workers. After correction, the distribution is shifted to the right with the mean increasing from 7.5 to 8.2 h of work per week.

Table 3 contrasts means of working hours per week from the BeH (after correction) and the SES, again in 2014. The SES, which is a large mandatory survey among employers, serves as a benchmark as it is typically considered to provide the most comprehensive and reliable information on working time in Germany.



**Fig. 3** Kernel density estimates of the distribution of weekly working hours. The figure shows kernel density estimates of working hours per week by working-time status [full-time in panel (A), part-time in panel (B) and marginal part-time in panel (C)] in the BeH in 2014. Source: IAB Beschäftigtenhistorik (BeH) V10.06.00-202012; working hours from DGUV notifications

Specifically, for the comparisons, we use paid working hours without paid overtime from the SES. To make the sample in the BEH as similar as possible to that in the official statistics, activities of households

**Table 2** Uncorrected and corrected average working hours in the BeH

	Uncorrected	Corrected
Full-time	34.8	39.1
Part-time	21.2	23.9
Marginal part-time	7.5	8.3
Apprentices	35	39.3

The table reports uncorrected and corrected (for days of annual and sick leave) average working hours in the BeH for 2014, separately for full-time workers, part-time workers, marginally employed workers and apprentices

Source: IAB Beschäftigtenhistorik (BeH) V10.06.00-202012; working hours from DGUV notifications, own calculations

and extra-territorial organizations (T, U according to NACE Rev.2) are excluded. To reduce the effect of outliers, full-time work excludes hours below 30 and above 50, part-time excludes hours below 3 and above 38 and marginal part-time excludes hours above 18. We provide these means by location of the establishment (East/West Germany), working-time status and gender.

Reassuringly, the means of weekly working hours hardly deviate between the BeH and the SES after our correction. This is the case for most of the subgroups considered. Deviations exceeding 0.6 h, in absolute terms, are only observed for part-time workers in East Germany. The BeH thus appears to underestimate the

weekly working hours slightly for both men and women working part-time in East Germany.

In addition to the means, Table 4 shows the standard deviation of weekly working hours in both the BeH (after correction) and the SES, again in 2014 and by location, working-time status and gender. Overall, standard deviations also do not deviate by much between the two data sets. Standard deviations exceeding 0.6, in absolute terms, arise only for women working full-time and for apprentices. Even after correction, the BeH still shows a somewhat larger variation in working hours for these groups than the SES.

As a final comparison, Table 5 presents the deciles of the weekly working hours distributions in the BeH (after correction) and the SES for 2014, by working-time status. Deciles below the median do usually not deviate by more than 0.6 h per week. Sole exceptions are the fourth decile of part-time workers and the first decile of apprentices. For part-time, the fourth decile is larger by 1.2 h in the BeH than in the SES while the first decile for apprentices is lower by 0.9 h. In the upper half of the distributions, the deviations are somewhat more pronounced. However, no decile in any subgroup deviates by more than 1.8 h in absolute terms, showing a generally good fit between the working hours in the BeH and the SES data after correction.

**Table 3** Average weekly hours according to official statistics and the BeH (after correction) in 2014

	All		East Germany		West Germany	
	BeH	SES	BeH	SES	BeH	SES
Full-time						
All	39.1	39.1	39.4	39.9	39.1	39
Men	39.2	39.2	39.5	40	39.1	39.1
Women	39.1	39	39.3	39.7	39	38.9
Part-time						
All	23.9	23.9	26.9	27.9	23.4	23.2
Men	23.9	23.8	25.6	26.4	23.6	23.3
Women	23.9	23.9	27.2	28.3	23.3	23.2
Marginal part-time						
All	8.3	7.8	8.2	8.1	8.3	7.8
Men	8.2	7.6	8.1	8	8.2	7.6
Women	8.4	7.9	8.3	8.1	8.4	7.9
Apprentices						
All	39.3	39	39.6	39.6	39.3	39
Men	39.2	39	39.5	39.7	39.2	38.9
Women	39.4	39.1	39.6	39.5	39.4	39

The table compares average working hours per week in 2014 according to the Structure of Earnings Survey (SES) from the German Statistical Office with those in the BeH after correction. To make the sample in the BeH as similar as possible to that in the official statistics, activities of households and extra-territorial organizations (T, U according to NACE Rev.2) are excluded. Working hours in the BeH have been adjusted to account for days of annual and sick leave, but exclude overtime adjustment. To reduce the effect of outliers, full-time excludes hours below 30 and above 50, part-time excludes hours below 3 and above 38 and marginal excludes hours above 18

Sources: IAB Beschäftigtenhistorik (BeH) V10.06.00-202012; working hours from DGUV notifications, Statistisches Bundesamt (2016), own calculations

**Table 4** Standard deviation of weekly hours according to official statistics and the BeH (after correction) in 2014

	All		East Germany		West Germany	
	BeH	SES	BeH	SES	BeH	SES
Full-time						
All	3	2.4	2.9	2.1	3	2.4
Men	3.1	2.6	2.9	2.3	3.1	2.7
Women	2.8	1.8	2.8	1.6	2.8	1.8
Part-time						
All	7.8	7.6	7.3	7.2	7.7	7.5
Men	8.6	8.6	8.3	8.8	8.6	8.5
Women	7.5	7.4	7	6.7	7.5	7.2
Marginal part-time						
All	4	3.8	4.4	4.7	4	3.7
Men	4.1	3.9	4.4	4.7	4	3.8
Women	4	3.8	4.4	4.7	3.9	3.7
Apprentices						
All	2.7	2	2.7	1.5	2.7	1.9
Men	2.8	1.9	2.7	1.2	2.8	2
Women	2.6	1.9	2.7	1.7	2.6	1.9

The table compares standard deviations of working hours per week in 2014 according to the Structure of Earnings Survey (SES) from the German Statistical Office with those in the BeH after correction. To make the sample in the BEH as similar as possible to that in the official statistics, activities of households and extra-territorial organizations (T, U according to NACE Rev.2) are excluded. Working hours in the BeH have been adjusted to account for days of annual and sick leave, but exclude overtime adjustment. To reduce the effect of outliers, full-time excludes hours below 30 and above 50, part-time excludes hours below 3 and above 38 and marginal excludes hours above 18

Sources: IAB Beschäftigtenhistorik (BeH) V10.06.00-202012; working hours from DGUV notifications, Statistisches Bundesamt (2016), own calculations

**Table 5** Distribution of weekly hours according to official statistics and the BeH (after correction) in 2014

Percentile	Full-time		Part-time		Marginal part-time		Apprentices	
	BeH	SES	BeH	SES	BeH	SES	BeH	SES
10	35.2	35.4	13.3	13.6	2.7	2.6	36.1	37
20	37.5	38	17.4	18	4.4	4.1	38	38
30	38.5	38.5	19.5	20	5.7	5.3	38.6	38.5
40	38.9	39	21.4	20.2	7.3	6.9	38.9	39
50 (Median)	38.9	39.8	23.8	24.2	8.8	8.8	38.9	39.8
60	39.4	40	26.6	27.3	9.7	9.2	39.4	40
70	40.3	40	29	29.9	10.7	10	40.5	40
80	41.3	40	31.4	30	11.8	10.8	41.6	40
90	42.3	40.5	34	33.8	13.5	12.2	42.3	40.5

The table compares the distribution of working hours per week in 2014 according to the Structure of Earnings Survey (SES) from the German Statistical Office with those in the BeH after correction. To make the sample in the BeH as similar as possible to that in the official statistics, activities of households and extra-territorial organizations (T, U according to NACE Rev.2) are excluded. Working hours in the BeH have been adjusted to account for days of annual and sick leave, but exclude overtime adjustment. To reduce the effect of outliers, full-time excludes hours below 30 and above 50, part-time excludes hours below 3 and above 38 and marginal excludes hours above 18

Sources: IAB Beschäftigtenhistorik (BeH) V10.06.00-202012; working hours from DGUV notifications, Statistisches Bundesamt (2016), own calculations

## 5 Limitations of the correction procedure

### 5.1 Suitability depending on research question

We argue that our correction procedure provides a useful way of unifying the working hours information contained in the data module “German Social Accident Insurance”.

The corrected hours variable thus appears well suited for different types of exercises, including:

- Drawing distributions of (contractual) working hours in levels.

- Comparing average working hours or average hourly wages between establishments or regions.
- Estimating treatment effects based on the position in the working hours or hourly wage distribution at different levels of analysis, like the individual level, establishment level, or the regional level (as in evaluation studies of the minimum wage; see Dustmann et al. 2022).
- Analysing changes of working hours or hourly wages over time at the regional level.
- Analysing changes of working hours or hourly wages over time at the individual level, if the individual is changing jobs or employers.

Nevertheless, there are also situations in which the benefits of our approach are less clear. This is particularly the case for any kind of comparison of working hours or hourly wages within the same establishment. If all of an establishment's notifications strictly follow one reporting scheme, then our correction introduces unnecessary bias. If an establishment does not report clearly, in turn, then our correction might be beneficial. Also note that, by assuming that the full-time worker reference value uniformly represents a standard work week of 39 h, no industry-specific variation in working hours can be identified for this reporting scheme.

The correction procedure that we propose here, however, is likely not helpful for analysing:

- Changes of working hours or hourly wages of an individual worker, if the worker is keeping the job, or
- Changes of working hours or hourly wages measured at the establishment level.

In both of these cases using the uncorrected working hours information is likely superior (but maybe still problematic).

### 5.2 Selectivity

One important point to note before working with the corrected working hours variable is that our procedure only corrects available information to make it more consistent, but does not try to impute missing information on working hours. As already noted in Sects. 2.3 and 3.1, a significant share of working hours is missing for various reasons. This means that biases due to selectivity in missing values will not be addressed by our procedure. To gain a better understanding of the selectivity of missing values in the working hours variable, Appendix A.4 presents tabulations for several key characteristics for spells with and without missing hours information. It shows that missing hours information is more prevalent,

e.g., for marginal workers, very large establishments, and establishments in the agricultural and the public sectors.

### 5.3 Other caveats

While our corrected working hours dataset includes the years 2010 to 2014, experimentation with the version outlined in Appendix A.2 that we used in Dustmann et al. (2022) led us to believe that the quality of the 2010 working hours data is somewhat worse than for the other years, at least for the kind of analysis carried out there. This is why we decided against using 2010 data in that publication. We do not generally advise against using the 2010 corrected working hours data, but want to remind users of the data to be especially careful in case they do.

In our experience, precision of results can be considerably sharpened by treating outliers in the working hours variable differently, depending on working-time status. For example, the restrictions used in Sect. 4 (excluding full-time weekly working hours below 30 and above 50, part-time weekly working hours below 3 and above 38 and marginal weekly working hours above 18) work quite well in our view. However, we note that such restrictions come at a cost, as they might introduce additional selection bias or measurement error, respectively.

## 6 Availability

The corrected working hours variable will be available to external researchers for the coming SIAB Version 7521, which will be released in 2023. The IAB-FDZ will assess whether it can also be made available for selected other IAB-FDZ standard data products in that year. Since the DGUV data series is discontinued, the corrected working hours variable will not become part of the standard portfolio in future data updates.

The data structure is as follows:

Main data

persnr	Individual ID (dataset specific).
spell	Observation counter per person.
hours_orig	Hours in reporting period, original notification.
hours_week	Hours per week, corrected.
hours_full	Hours in reporting period, corrected.

Supplementary probabilities data

persnr	Individual ID (dataset specific).
spell	Observation counter per person.
prob_0	Probability notification is 'contractual'.
prob_1	Probability notification is 'actual'.
prob_2	Probability notification is 'fulltime reference value'.

## 7 Conclusion

Usually, the information on working hours in the German administrative labour market data is restricted to working full-time or part-time. For the years 2010 to 2014, however, it is possible to merge working hours reported by employers to the German Social Accident Insurances at the job-level. However, this raw variable "hours worked" has some weaknesses, in particular that employers were able to choose between different reporting schemes. These could be actual hours worked, contractual hours worked, or fulltime-worker benchmarks. Furthermore, it is not obvious from the data which reporting scheme was actually used. In this report, we propose a correction procedure for creating a harmonised measure of contractual working time from these supplementary working hours information. However, one should be cautious when using it in other contexts, such as analysing changes of average hours worked measured at the establishment level over time.

## Appendix

### A.1 Example for determining the full-time worker reference value

The full-time worker reference value enters the social security notification process with a two-year delay, i.e. for notifications concerning the year 2011, the reference value calculated on data from the year 2009 must be used as a basis.

Even though this delay is understandable due to the availability of data and the reporting deadlines of the establishments, the full-time worker reference value does not reflect the level of the current calendar year, but rather the holiday and sickness patterns as well as the working day effects delayed by two years. Another point

of criticism concerns the calculation of the reference value by the DGUV (cf. also Lehner and Ruppert 2009 and Appendix Table 6). When calculating the full-time worker reference value, Saturdays, Sundays and public holidays are first deducted from the calendar days of a year and then sick days and annual leave are also considered. The resulting working days are then multiplied by daily working hours based on statistics from the Federal Statistical Office and rounded to an annual full-worker reference value. However, the adjustment of working days for sick leave erroneously does not take into account that the days of incapacity to work in the sick leave statistics refer to calendar days and not to the potential working days of a calendar year. As a result, weekends and public holidays that are included in the sick days are deducted twice in the calculation. Thus, the full-worker reference value significantly underestimates the hours actually worked.

### A.2 Alternative correction procedure used in Dustmann et al. (2022)

The correction procedure in this report is an alternative to the one used in Dustmann et al. (2022). A brief description of that procedure (called "alternative version" from here on) can be found in Data Appendix A.1 of that paper. The main differences between the alternative version and the one presented here are:

1. The alternative version only distinguishes between actual and contractual working hours and does not treat the full-time reference value as a separate group in the algorithm. It is treated as part of the group classified as reporting actual working hours.
2. Instead of the probabilistic approach presented in step 1 above (see Sect. 3.2), the alternative version classifies establishments according to a simpler heuristic. All establishments that report at least 90% of

**Table 6** Example for determining the full-time worker reference value

	Full-time worker reference value (2011) DGUV	Corrected version
Calendar days	365	365
Saturdays/Sundays	105	105
Public holidays	8.5	8.5
Potential working days	251.5	251.5
Annual leave and other release times	30.7	30.7
Sick leave in calendar days	16.6	
Sick leave in working days		9.5
Working days (1)	204.2	209.4
Weekly working hours (2)	39	39
Yearly working hours in full time [Weekly working hours/5 * Working days]	1592.8	1633.3
Results in rounded full-time worker reference value	1590	1630

Calculations according to the calculation scheme of the German Social Accident Insurance for the year 2011.

Source: DGUV, (2014), Lehner and Ruppert, (2009) and own calculations.

their full-time workers as working less than 35 h per week are classified as safely reporting actual working hours. All establishments that report at least 90% of their full-time workers as working more than 35 h per week are classified as safely reporting contractual working hours. Assuming that an establishment uses the same notification variant for all its employees, this classification is also transferred to all part-time and marginal part-time workers in the establishment.

3. For establishments not classified already according to (2), it uses a step somewhat similar to step 2 above (see Sect. 3.3). For each specific value of reported working hours, it computes the likelihood that the employer reported actual or contractual working hours, based only on the sample of establishments classified as ‘safe’ in (2). It then randomly classifies observations as reporting actual versus contractual working hours according to the estimated relative likelihood (in case they were not classified in (2) already).

The alternative version is available as a separate data file for interested researchers.

### A.3 Additional adjustment for overtime

Some empirical research questions require that the working hours variable includes overtime work. Unfortunately, the working hours data from the module “German Social Accident Insurance” does not allow us to identify whether a reported number of working hours contains overtime or not. We can expect a notification to exclude overtime if employers report contractual hours or the full-time worker reference value. In case of actual hours reporting, however, overtime hours should be included. In this report, we decided to leave the decision whether and how to adjust overtime to the researcher. Our corrected data therefore do not include an additional overtime correction by default. Appendix Table 7 reports average overtime hours per month by worker group for the years 2010 to 2014, based on SOEP data. The German SES also reports overtime statistics, but only every four years (see Statistisches Bundesamt 2016). These or more elaborate overtime adjustments can be added to the corrected working hours data module using the combined probability measures  $p^{act}$ ,  $p^{con}$  and  $p^{ref}$  available as a data add-on.

### A.4 Selectivity of missing values

This appendix provides information on the share of missing values in the working hours variable reported to the DGUV by employers for several key worker characteristics. It is important to note that there are several peculiarities of the reporting scheme that lead to pronounced spikes in missing values (see also the discussion in

**Table 7** Suggested monthly overtime adjustment for corrected working hours variable

	Full-time	Part-time	Marginal
2010	5.79	2.40	1.11
2011	6.76	2.43	0.89
2012	6.03	2.79	0.82
2013	5.36	2.37	0.85
2014	5.35	2.43	0.84

The table reports a suggested overtime adjustment value per month, separately for full-time workers, part-time workers and marginally employed workers, based on SOEP data. For an adjustment on a weekly basis, divide by 4.33.

Source: SOEPv33.1; own calculations. See Schupp et al. (2017) for further information on the SOEP.

Sect. 2.3).<sup>14</sup> As a result, there is some selectivity of missing values in the working hours variable with respect to worker and establishment characteristics.

To provide an overview of this selectivity, Appendix Table 8 depicts contrasts calculated after running a probit regression model with an indicator variable marking missing information in working hours as the dependent variable and several dummies for worker and establishment characteristics as regressors. We use the SIAB as our data base for the analysis to present results for a data product that is available to the international research community through the IAB-FDZ.<sup>15</sup>

With an overall rate of missing values of 17 percent,<sup>16</sup> the coefficients show the difference to this grand mean for the selected criteria. In several cases, missings in working hours are more common when the variable in question shows a missing value, too. For example, missings are 8 percentage points more likely for missing nationality status, 11 percentage points more likely for missing employment status and 16 percentage points more likely for missing industry classification. This seems plausible, since missing values in other characteristics point to establishments with low reporting standards. There are some other notable differences, however. Information on hours is missing more often for university graduates and workers with higher secondary education (plus 2 percentage points), marginal workers (plus 2 percentage points) and workers in very large establishments with more than 1000 workers (plus 6 percentage points). When looking at industry sections, major discrepancies show up as expected. Missings are much more likely in agriculture (A: plus 67 percentage points), for private households (T: plus

<sup>14</sup> The following list is not exhaustive, but focuses on aspects likely relevant for empirical labour market research.

<sup>15</sup> For general information on the SIAB, see vom Berge et al. (2021a, b).

<sup>16</sup> This differs slightly from the 16 percent mentioned in the main text due to sampling variation.

**Table 8** Deviations of probabilities for missing working hours by categories

	Contrast	Std. err.	z	P > z
Year				
2010	- 0.006	0.000	- 24.210	0.000
2011	0.000	0.000	1.430	0.152
2012	- 0.002	0.000	- 6.600	0.000
2013	0.001	0.000	3.590	0.000
2014	0.006	0.000	25.380	0.000
Gender				
Male	0.005	0.000	33.470	0.000
Female	- 0.005	0.000	- 33.470	0.000
Nationality				
German	0.000	0.000	7.040	0.000
Foreign	- 0.003	0.000	- 8.290	0.000
Missing	0.077	0.006	13.220	0.000
Qualification				
Lower secondary w/o vocational qual	- 0.004	0.000	- 13.070	0.000
Lower secondary with vocational qual	- 0.004	0.000	- 35.290	0.000
Entrance qual. for FH or Uni. w/o voc. qual	0.021	0.001	33.160	0.000
Entrance qual. for FH or Uni. with voc. qual	0.000	0.000	0.500	0.617
University of Applied Sciences (FH)	- 0.004	0.001	- 4.820	0.000
University	0.021	0.000	54.650	0.000
Missing	- 0.000	0.001	- 0.190	0.851
Place of residence				
East	- 0.009	0.000	- 36.780	0.000
West	0.002	0.000	33.260	0.000
Missing	0.040	0.002	17.630	0.000
Employment status				
Fulltime	- 0.011	0.000	- 79.620	0.000
Parttime	0.003	0.000	11.210	0.000
Marginal	0.017	0.000	69.890	0.000
Missing	0.111	0.002	53.480	0.000
Industry				
A: agriculture, forestry and fishing	0.672	0.002	420.690	0.000
B: Mining and Quarrying	- 0.099	0.002	- 43.790	0.000
C: Manufacturing	- 0.096	0.001	- 172.940	0.000
D: Electricity, gas, steam and air conditioning supply	- 0.077	0.002	- 46.050	0.000
E: Water suppl.; sewerage, waste man. and remed. act	0.080	0.002	34.830	0.000
F: Construction	- 0.060	0.001	- 76.000	0.000
G: Wholes. and ret. trade; rep. of motor vehic. ...	- 0.079	0.001	- 131.680	0.000
H: Transportation and storage	- 0.066	0.001	- 89.680	0.000
I: Accommodation and food service activities	- 0.049	0.001	- 64.000	0.000
J: Information and communication	- 0.085	0.001	- 106.010	0.000
K: Financial and insurance activities	0.003	0.001	2.130	0.033
L: Real estate activities	- 0.020	0.001	- 14.460	0.000
M: Professional, scientific and technical activities	- 0.069	0.001	- 93.770	0.000
N: Administrative and support service activities	- 0.054	0.001	- 78.030	0.000
O: Public admin. and defence; compulsory social sec	0.580	0.001	498.190	0.000
P: Education	0.289	0.001	217.020	0.000
Q: Human health and social work activities	- 0.015	0.001	- 19.430	0.000
R: Arts, entertainment and recreation	0.084	0.002	54.860	0.000

**Table 8** (continued)

	Contrast	Std. err.	z	P > z
S: Other service activities	− 0.021	0.001	− 21.620	0.000
T: Activities of households as employers, ...	0.764	0.001	634.200	0.000
U: Activities of extraterritorial organisations and bodies	0.742	0.005	150.250	0.000
Missing	0.159	0.023	6.990	0.000
Establishment size				
1–9	0.016	0.000	38.580	0.000
10–49	− 0.015	0.000	− 41.760	0.000
50–99	− 0.027	0.000	− 61.990	0.000
100–499	− 0.012	0.000	− 31.970	0.000
500–999	0.018	0.001	29.810	0.000
1000+	0.061	0.001	110.210	0.000
Missing	− 0.027	0.011	− 2.360	0.019

The table shows margin contrasts from the observation-weighted grand mean from a probit model with a dummy indicator variable for missing working hours information as the dependent variable

Source: SIAB\_7519\_v1 with added working hours module, own calculations. See vom Berge et al. (2021a; b) for further details on the SIAB

76 percentage points) and sections with a high share of public sector employers (O/P/U: between plus 29 and plus 74 percentage points).

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12651-023-00331-0>.

**Additional file 1:** Reproduction package.

## Acknowledgements

We would like to thank Jennifer Vallé for the provision of her unpublished IAB project paper "Möglichkeiten und Grenzen der Erhebung von Arbeitszeiten der Beschäftigungsstatistik der Bundesagentur für Arbeit" (Possibilities and limitations of collecting working hours of the employment statistics of the Federal Employment Agency), which provided us with valuable suggestions and information for this report. We also thank Dana Müller for helpful comments.

## Author contributions

PVB and MU have developed the correction procedure. PVB has conducted the analyses and created the graphs and tables. All three authors are responsible for discussing the findings and writing the manuscript. PVB is the corresponding author. All authors read and approved the final manuscript.

## Funding

We gratefully acknowledge financial support from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, grant numbers BE 6283/5-1) in an early phase of the project. The funding did not influence the design of the study, analysis, and interpretation of data.

## Availability of data and materials

For our analyses, we use administrative data of the Institute for Employment Research (IAB) [IAB Beschäftigtenhistorik (BeH) V10.06.00-202012]. The data are social security data with administrative origin which are processed and kept by IAB, Regensburger Str. 104, D-90478 Nürnberg, iab@iab.de, phone: + 49 911 1790, according to the German Social Code III. There are certain legal restrictions due to the protection of data privacy. The data contain sensitive information and therefore are subject to the confidentiality regulations of the German Social Code (Book I, Sect. 35, Paragraph 1). The raw data, computer programs,

and results have been archived by IAB in accordance with good scientific practice. Computer programs and results can be found in the Reproduction package (Additional file 1). If you wish to access the full data for replication purposes, please contact Philipp vom Berge (philipp.vom-berge@iab.de). Please visit <https://www.iab.de/en/daten/replikationen.aspx>. For information on availability of further data, please refer to Sect. 6.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

Received: 21 September 2022 Accepted: 1 January 2023

Published online: 11 March 2023

## References

- Deutsche Gesetzliche Unfallversicherung (DGUV): DGUV-Statistiken für die Praxis 2014. Aktuelle Zahlen und Zeitreihen aus der Deutschen Gesetzlichen Unfallversicherung. <https://publikationen.dguv.de/zahlen-fakten/ueberblick/3076/dguv-statistiken-fuer-die-praxis-2014> (2014). Accessed 29 Nov 2022
- Deutsche Gesetzliche Unfallversicherung (DGUV): In good hands. Your German Social Accident Insurance. <https://publikationen.dguv.de/uebergreifende-themen/2865/in-good-hands-your-german-social-accident-insurance> (2016). Accessed 29 Nov 2022
- Deutsche Gesetzliche Unfallversicherung (DGUV): Lohnnachweis digital. Informationen und wichtige Termine zum UV-Meldeverfahren der Unfallkassen und Berufsgenossenschaften. <https://publikationen.dguv.de/uebergreifende-themen/3383/lohnachweis-digital-informationen-und-wichtige-termine-zum-uv-meldeverfahren-der-unfallkassen-und> (2018). Accessed 29 Nov 2022
- Deutsche Rentenversicherung (DRV): Daten zur Unfallversicherung: Angabe zu Arbeitsstunden in DEÜV-Meldungen. In: Summa summarum 6, 12–14. [https://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Fachliteratur\\_Kommentare\\_Gesetzestexte/summa\\_summarum/e\\_paper\\_und\\_broschueren/2008/2008\\_6.pdf?\\_\\_blob=publicationFile&v=1](https://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Fachliteratur_Kommentare_Gesetzestexte/summa_summarum/e_paper_und_broschueren/2008/2008_6.pdf?__blob=publicationFile&v=1) (2008). Accessed 29 Nov 2022
- Deutsche Rentenversicherung (DRV): Besprechung der Spitzenorganisationen der Sozialversicherungen zu Fragen des gemeinsamen Meldeverfahrens am 8.6. und 9.6.2011 in Berlin. Besprechungsergebnisse, Anlage 1. <https://>



- [www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Fachliteratur\\_Kommentare\\_Gesetzestexte/summa\\_summarum/besprechungsergebnisse/meldeverfahren/2011/2011\\_06\\_09\\_08\\_download.zip?\\_\\_blob=publicationFile&v=1](http://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Fachliteratur_Kommentare_Gesetzestexte/summa_summarum/besprechungsergebnisse/meldeverfahren/2011/2011_06_09_08_download.zip?__blob=publicationFile&v=1) (2011). Accessed 29 Nov 2022
- Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., vom Berge, P.: Reallocation effects of the minimum wage. *Q J Econ* **137**(1), 267–328 (2022). <https://doi.org/10.1093/qje/qjab028>
- Fitzenberger, B., Seidlitz, A.: The 2011 break in the part-time indicator and the evolution of wage inequality in Germany. *J Labour Mark Res* (2020). <https://doi.org/10.1186/s12651-019-0265-0>
- Frodermann, C., Müller, D., Abraham, M.: Determinanten des Wiedereinstiegs von Müttern in den Arbeitsmarkt in Vollzeit oder Teilzeit. *KZfSS Kölner Zeitschrift Soziologie Sozialpsychologie* **65**(4), 645–668 (2013). <https://doi.org/10.1007/s11577-013-0236-3>
- Höller, E.: Ungewollte Nebenwirkungen Folgen des Übergangs der Betriebsprüfung auf die Rentenversicherung. *DGUV-Forum*. 4, 14–19. [https://forum.dguv.de/issues/archiv/2010/03\\_DGUV\\_Forum\\_4\\_2010.pdf](https://forum.dguv.de/issues/archiv/2010/03_DGUV_Forum_4_2010.pdf) (2010). Accessed 29 Nov 2022
- Institut für Arbeitsmarkt- und Berufsforschung (IAB): IAB-working time measurement concept—Data on the development of working hours and their components. [https://doku.iab.de/arbeitsmarktdaten/AZ-Komponenten\\_en.xlsx](https://doku.iab.de/arbeitsmarktdaten/AZ-Komponenten_en.xlsx) (2022). Accessed 29 Nov 2022
- Lehner, N., Ruppert, L.: Die neue gesetzliche Unfallversicherung: Grundlagen und Praxistipps für die Entgeltabrechnung, Lösungsbeispiele für die Anwendung in der Praxis; [die Reform der gesetzlichen Unfallversicherung; Tipps und Umsetzungshilfen für die Lohnabrechnung]. Hüthig Jehle Rehm, Heidelberg (2009)
- Ludsteck, J., Thomsen, U.: Imputation of the working time information for the employment register data. *FDZ-Methodenreport*, 01/2016 (en), (2016)
- Schupp, J., Goebel, J., Kroh, M., Schröder, C., Bartels, C., Erhardt, K., Fedorets, A., Franken, A., Giesselmann, M., Grabka, M., Krause, P., Kröger, H., Kühne, S., Metzinger, M., Nebelin, J., Richter, D., Schacht, D., Schmelzer, P., Schmitt, C., Schnitzlein, D., Siegers R., Wenzig, K.: Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984–2016 (2017). <https://doi.org/10.5684/soep.v33>
- Statistisches Bundesamt: Verdienststrukturerhebung 2014-Niveau, Verteilung und Zusammensetzung der Verdienste und der Arbeitszeiten abhängiger Beschäftigungsverhältnisse-Ergebnisse für Deutschland, Fachserie 16 Heft 1, Wiesbaden (2016)
- Vom Berge, P., Frodermann, C., Graf, T., Grießemer, S., Kaimer, S., Köhler, M., Lehnert, C., Oertel, M., Schmucker, A., Schneider, A., Seth, S.: Weakly anonymous version of the sample of integrated labour market biographies (SIAB)—Version 7519 v1. research data centre of the federal employment agency (BA) at the institute for employment research (IAB) (2021a). <https://doi.org/10.5164/IAB.SIAB7519.de.en.v1>
- Vom Berge, P., Frodermann, C., Schmucker, A., Seth, S.: Sample of integrated labour market biographies (SIAB) 1975–2019. *FDZ Datenreport*, 01/2021b (en), (2021b). <https://doi.org/10.5164/IAB.FDZD.2101.en.v1>
- Wanger, S., Weigand, R., Zapf, I.: Measuring hours worked in Germany contents, data and methodological essentials of the IAB working time measurement concept. *J. Labour Mark. Res.* **49**(3), 213–238 (2016). <https://doi.org/10.1007/s12651-016-0206-0>
- Wirtschafts- und Sozialwissenschaftliches Institut (WSI): WSI-Arbeitszeitkalender 2014 - Tarifdaten aus 25 Wirtschaftszweigen, Elemente qualitativer Tarifpolitik Nr. 78, Düsseldorf. [https://www.boeckler.de/pdf/p\\_ta\\_elemente\\_78\\_2014.pdf](https://www.boeckler.de/pdf/p_ta_elemente_78_2014.pdf) (2014). Accessed 29 Nov 2022

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)