

ORIGINAL ARTICLE

Open Access



# Measurement error in longitudinal earnings data: evidence from Germany

Achim Schmillen<sup>1\*</sup>, Matthias Umkehrer<sup>2</sup> and Till von Wachter<sup>3,4,5,6</sup>

## Abstract

We present evidence on the extent of measurement error in German longitudinal earnings data. Qualitatively, we confirm the main result of the international literature: longitudinal earnings data are relatively reliable in a cross section but much less so in first differences. Quantitatively, in the cross section our findings are very similar to those of Bound and Krueger (*J Labor Econ* 9:1–24, 1991) and Pischke (*J Bus Econ Stat* 13:305–314, 1995) for the United States while we find even stronger evidence that first-differencing exacerbates measurement error problems. We also show that measurement error in our survey data is not “classical” as it is negatively correlated with administrative earnings and positively autocorrelated over an extended period of time. Additionally, we estimate a model of measurement error stemming from underreporting of transitory earnings shocks in combination with a white-noise component and make a number of methodological contributions. Our results are robust to the use of two different linked survey-administrative data sets and various other sensitivity checks.

**Keywords** Measurement error, Transitory and permanent earnings, Earnings dynamics, Linked survey-administrative data

**JEL Classification** C18, C33, J30

## 1 Introduction

Over the last three decades, administrative data sets have been “transforming the analysis of economic policy” (Friedman 2010, p. 2). Card et al. (2010, p. 1) highlight that as compared to survey data “[a]dministrative data offer much larger sample sizes and have far fewer problems with attrition, non-response, and measurement error.” In one of the pioneering and most influential examinations of measurement error in survey data, Bound and Krueger (1991) (hereafter: BK) focus on the prevalence and properties of measurement error in

longitudinal earnings information. Their analysis compares potentially mismeasured Current Population Survey (CPS) data to administrative Social Security payroll tax records, which BK assume to be free of measurement error. According to the findings by BK, in a cross section survey-based earnings data are relatively reliable. At the same time, reliability ratios are much lower for specifications that rely on survey-based earnings data in first differences. In addition, measurement errors are not “classical” in the sense of being identically and independently distributed and uncorrelated with administrative earnings. Instead, measurement errors are serially correlated over 2 years and negatively correlated with administrative earnings (or “mean-reverting” according to BK’s parlance).

In the first part of this paper, we replicate and extend the exercise by BK with the help of German data that link the administrative “ADIAB” social security records with the “PASS” (“Panel Study Labour Market and Social Security”/“Panel Arbeitsmarkt und soziale Sicherung”)

\*Correspondence:

Achim Schmillen  
aschmillen@worldbank.org

<sup>1</sup> The World Bank, 1818 H Street NW, Washington, DC 20433, USA

<sup>2</sup> Institute for Employment Research, Nuremberg, Germany

<sup>3</sup> University of California, Los Angeles, USA

<sup>4</sup> National Bureau of Economic Research, Cambridge, USA

<sup>5</sup> Centre for Economic Policy Research, London, UK

<sup>6</sup> IZA - Institute of Labor Economics, Bonn, Germany



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

household survey. In order to compare our findings for Germany directly with those already available for the United States, we make the same assumptions as BK. In particular, we assume that administrative ADIAB earnings represent “true earnings” and hereafter designate them as such. In fact, because the underlying administrative data are used to compute social security contributions, the ADIAB earnings information is considered highly reliable. While the data set used by BK only spans two time periods, we are able to exploit cross-sectional and longitudinal variation in earnings over the first four waves of the PASS survey. Our data are also more recent than those of BK. In addition, the provision of internationally comparable evidence on the extent of measurement error in German longitudinal earnings data makes it possible to put the conclusions by BK into an international perspective.

Qualitatively, we confirm many of the main results of BK. In particular, we confirm that in a cross section survey-based earnings data are relatively reliable but that their reliability tends to be much lower when the data are specified in first differences. Quantitatively, in the cross section our findings for Germany are very similar to those of BK for the United States while we find even stronger evidence that first-differencing exacerbates measurement error problems. As yardstick for the reliability of survey-based earnings data, BK consider the “reliability ratio” defined as the ratio of the covariance between mismeasured earnings and true earnings to the variance of mismeasured earnings. A value for this ratio of 1 would indicate that the covariance between mismeasured earnings and true earnings equals the variance of mismeasured earnings. This would imply perfect reliability. Conversely, a value of 0 would imply a complete lack of reliability. Allowing for mean-reverting measurement error, BK estimate that in a cross section, the reliability ratio is 0.97 to 1.02 for men and 0.93 to 0.96 for women. They also estimate that this ratio falls to between 0.78 and 0.86 when the data are specified in first differences. Also allowing for mean-reverting measurement error, we find that in a cross section the reliability ratio is 0.93 to 0.95 for men and 0.90 to 1.00 for women. In first differences, it falls to 0.11 to 0.47 for men and 0.25 to 0.34 for women.

Our other noteworthy findings include (a) that measurement error in our survey data is not classical but mean-reverting, with a high degree of autocorrelation and a strong negative correlation with true earnings, and (b) that the mismeasurement of earnings leads to little bias when survey-based earnings are on the left-hand side of a typical Mincer-type earnings regression. Both these findings are again in line with conclusions reached by BK. Going beyond the exercise by BK, we analyze both first- and higher-order autocorrelations

of measurement error. Our results strongly suggest that measurement errors in our survey-based earnings data are positively correlated over an extended period of time. This leads us to conjecture that at least a sizable fraction of the autocorrelation in the measurement error is not actually due to a simple autoregressive process but either to a person fixed effect or to a more complex time series process.

In the second part of this paper, we use our data to estimate the dynamic model of measurement error by Pischke (1995). Pischke (1995) explains the measurement error in earnings data from the Panel Study of Income Dynamics Validation Study (PSIDVS)—which combines earnings information from the Panel Study of Income Dynamics (PSID) questionnaire with information from payroll records at a specific, anonymous firm which he assumes to be free of measurement error—by an individual fixed effect, the misreporting of transitory earnings shocks and a white-noise component. This simple model fits our data surprisingly well. Again under the assumption that administrative earnings information represents true earnings, we find mean-reverting measurement error and, like Pischke (1995), that individuals underreport the transitory component of earnings. Pischke (1995, p. 309) explains that the underreporting of transitory earnings changes is entirely plausible given that “[o]bviously, it is the changes in permanent earnings that are related to the more important events in people’s lives.” Also in line with Pischke (1995), we find that underreporting of transitory earnings leads to downwardly biased estimates of the variance of earnings growth while the white-noise component induces upward bias. In our case, the upward bias more than offsets the downward bias, implying that earnings growth observed in survey data appears to be more spread out than it actually is. This is in contrast to Pischke (1995), who finds upward and downward biases of similar magnitude. As a further contribution, we document an upward bias in estimates of the variance of transitory earnings and a downward bias in estimates of the variance of permanent earnings due to measurement error in the survey data. Thus, earnings inequality is actually more persistent than suggested by survey data.

Our results are robust to various sensitivity checks and arguably exhibit a high degree of external validity as demonstrated through the use of a second linked survey-administrative data set that combines the administrative social security records with the “WeLL” (“Continuing Education and Lifelong Learning”/“Weiterbildung und Lebenslanges Lernen”) household survey. While the PASS survey oversamples poorer households, the WeLL survey includes a disproportionately large number of high earners. Reassuringly, our results are qualitatively robust across the two surveys.

We also make three methodological contributions: (a) we extend the simple model of measurement error suggested by BK as well as the dynamic measurement error model by Pischke (1995) from the two-period to the four-period case, (b) we introduce the methodology for taking account of top-coded administrative earnings information developed by Card et al. (2013) to the measurement error literature and (c) we develop a procedure for merging the correct administrative records to the different waves of the PASS and WeLL surveys.<sup>1</sup> These methodological contributions will help researchers to further investigate the prevalence and properties of measurement error in longitudinal earnings data and tap the full potential of German linked survey-administrative data.

Our study is relevant for three distinct strands of literature. First, it is relevant for the literature that links administrative and survey data to investigate the extent of measurement error in longitudinal earnings data. This literature is partly surveyed in Bound et al. (2001) and Meyer and Mittag (2021) and almost exclusively focuses on the United States. Together with BK and Pischke (1995), pioneering studies include those by Duncan and Hill (1985) and Bound et al. (1994) which both rely on the PSIDVS. In accordance with the findings by BK, Duncan and Hill (1985) show that in levels the reliability ratio of earnings in the PSID data exceeds 80 percent while Bound et al. (1994) document mean reversion and a positive autocorrelation in measurement error. Other noteworthy studies that link American administrative and survey data sets to investigate the extent of measurement error in earnings data include those by Mellow and Sider (1983), Rodgers et al. (1993), Bollinger (1998) and Stinson (2002).<sup>2</sup>

In addition, our results on measurement error in longitudinal earnings data complement the emerging literature that compares the quality of German administrative and survey-based earnings information in the cross section. Notable contributions to this literature include Oberski et al. (2017), Antoni et al. (2019), Valet et al. (2019), Gaulty et al. (2020) and Stüber et al. (2023). One recurrent conclusion of this literature is that while there is a certain tendency for average earnings to be larger in administrative records than in surveys, in the cross section average earnings differ only relatively little between the two types of data sources.

Third and finally, our study is relevant for the literature that investigates whether findings regarding labor market characteristics and impacts of economic policy depend on using either administrative or survey-based earnings data (either in levels or with regard to dynamics). For instance, Gideon et al. (2017), Abowd and McKinney (2017) and Kopczuk et al. (2010) demonstrate that using administrative instead of survey-based earnings data can alter well-established and policy-relevant findings regarding topics as diverse as the extent of race-based wage discrimination, the role of firms in explaining earnings inequality and the degree of earnings mobility over the life cycle. Two recent studies on the effects of the introduction of a nationwide minimum wage in Germany neatly illustrate the advantages of administrative data, but also that these advantages come at a cost. Dustmann et al. (2022) rely on administrative data and convincingly argue that as compared to exploiting survey data this approach improves the measurement of earnings and the precision of estimates. In contrast, Caliendo et al. (2023) follow the more traditional route and employ survey data. While the authors acknowledge that there may be measurement error in survey-based earnings information, they also highlight that some other relevant variables such as working hours are only partially and imprecisely captured in German administrative data sources. Indeed, Dustmann et al. (2022) need to rely on certain assumptions and an imputation procedure to expand their analysis from daily earnings to hourly wages.

The remainder of this paper is structured as follows: the PASS and ADIAB data, our linkage and sampling procedure and summary statistics are presented in the next section. This is followed by a description of our static measurement error model in Sect. 3. In Sect. 4, we characterize the measurement error in the PASS earnings data and the induced bias in survey-based earnings regressions. Section 5 contains the findings of our main empirical analysis using the static model of measurement error, including sensitivity checks. Section 6 outlines our application of the dynamic measurement error model of Pischke (1995), presents relevant estimates and discusses biases in estimates of earnings processes when using error-ridden data. Section 7 concludes.

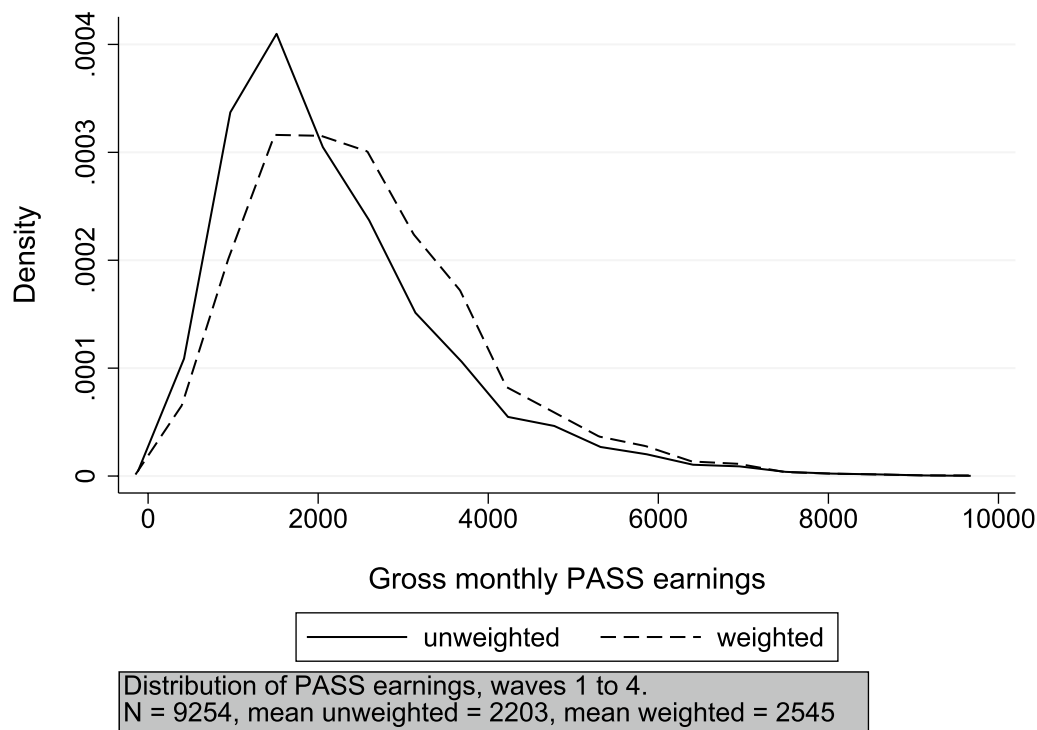
## 2 Data, sampling and summary statistics

We base our investigation on the PASS household survey linked with ADIAB administrative social security records prepared and maintained by the Institute for Employment Research (IAB) in Nuremberg, Germany.<sup>3</sup> PASS is

<sup>1</sup> This procedure also converts daily net earnings from WeLL to daily gross earnings using an empirical model of the German income tax system.

<sup>2</sup> The relevant literature for countries other than the United States is relatively limited. It includes Kristensen and Westergaard-Nielsen (2007) who rely on a matched sample of survey and administrative longitudinal data covering all sectors in the Danish economy. They find that in the Danish data measurement error in earnings is much larger than reported in studies for the United States.

<sup>3</sup> Cf. Trappmann et al. (2013) for an introduction to the PASS survey and Antoni and Bethmann (2019) for a detailed descriptions of standardized linked PASS-ADIAB data that differ slightly from the linked survey-administrative data used here.



**Fig. 1** Distribution of earnings in PASS estimation sample, weighted and unweighted

a household panel survey with a focus on poverty and receipt of non-contributory, means-tested unemployment benefits (“Arbeitslosengeld II”). The survey covers approximately 10,000 households and is carried out annually. Households receiving unemployment benefits are oversampled to allow for a detailed analysis of the dynamics of benefits receipt. As unemployment benefit recipients are comparatively more likely to transition into and out of lower-paying jobs, average earnings in PASS are lower than in the German labor market as a whole (cf. Fig. 1 which captures both the actual earnings distribution of PASS for our estimation sample as defined below and the counterfactual PASS earnings distribution approximating the German labor market using probability weights, i.e., where the weight of a respondent is equivalent to the reciprocal value of their inclusion probability).

PASS comprises information on, for example, individuals’ socio-demographic characteristics and subjective well-being. In our study, we incorporate the first four waves of the survey, collected in 2006/2007, 2007/2008, 2008/2009 and 2010, respectively. The PASS survey takes great care to collect data that is reliable, robust and comparable across households and across time. Given the declining prevalence of landline phones in Germany there is a mix of CATI and CAPI interviews. Based on respondents’ preferences, interviews are conducted

either in German or in one of three other languages (English, Russian and Turkish). Moreover, there are detailed interviewer training, outreach, engagement, follow-up and quality assurance processes. While there is a survey module regarding the overall household situation that is directed at one knowledgeable household member, the information on individual household members’ characteristics used here is directly solicited from these individual members.<sup>4</sup>

The version of the ADIAB data used here contains the universe of all individuals who were employed subject to social security contributions, received unemployment benefits or were registered as job seekers in the Federal Republic of Germany at least once between 1975 and 2010. All individuals with at least one spell of “marginal” employment, i.e., employment not covered by social security, in 1999 or later are also included. For most individuals in the data set, information on the majority of their labor market biography is available. Only spells of employment not covered by social security—like those as civil servants or family workers—and spells of

<sup>4</sup> Relevant questions are directed at all household members aged 15 to 64. For the first four waves of the PASS survey used here it was not recorded whether individuals were alone when they responded to the survey. Slotwinski and Roth (2020) show that social norms might contribute to the misreporting of income. Conceivably, this misreporting could be exacerbated by the presence of other household members.

self-employment are not covered. All in all, the ADIAB data cover more than 80 percent of Germany's total workforce. They encompass detailed longitudinal information on employment status, earnings, socio-demographic and firm characteristics to the exact day. Because Germany's social security agencies use the underlying administrative data to compute social security contributions and unemployment benefits, the earnings information in the ADIAB data is considered highly reliable.

In the ADIAB data, precise earnings information is not recorded beyond Germany's "contribution assessment ceiling", the maximum level of earnings that are subject to contributions to the country's various social security programs. To address the resulting top-coding of earnings information for 1.6 percent of observations, we rely on the imputation procedure developed by Card et al. (2013). This procedure uses separate Tobit regressions for different calendar years and for East and West Germany with a series of imputation variables that can be calculated when an establishment's entire workforce and a time series of earnings are observed.<sup>5</sup> Importantly, as reported in Sect. 5.3 all our results are robust to instead excluding observations with top-coded ADIAB earnings or simply keeping these observations with the original earnings information.

For the purpose of our study, we disregard spells of unemployment and inactivity and focus on individuals' spells in employment only (which will include time spent on annual or sick leave as long as the employment spell is not interrupted). Labor market biographies from the PASS and ADIAB data are linked using individuals' social security numbers. Linkage is only possible for those survey participants who explicitly permitted a match of their survey data to administrative records (depending on the PASS wave, this is the case for 80 to 87 percent of participants). A number of steps are needed to derive comparable information on earnings. PASS refers to gross earnings in the *last month* before the survey. In the first wave the respondents were to report the earnings from their *main job*, in the other waves they were to report the *total amount* of earnings. The administrative data, in turn, encompass all jobs of a person and include the sum of gross earnings over a *reporting period* of up to a year from a *given job*. There are also differences in the PASS and ADIAB data regarding whether one-time payments are captured in the definition of earnings and regarding the coverage of non-standard forms of employment.

<sup>5</sup> The following regressors are included in the relevant Tobit regressions: (a) an intercept, (b) a gender dummy, (c) 16 dummy variables for age brackets, (d) a second order polynomial of establishment size, (e) a dummy variable for single-employee establishments, (f) a dummy variable for establishments with more than ten employees, (g) a dummy variable for whether a worker has only one earnings observation and (h) a range of variables reflecting firm quality and worker productivity.

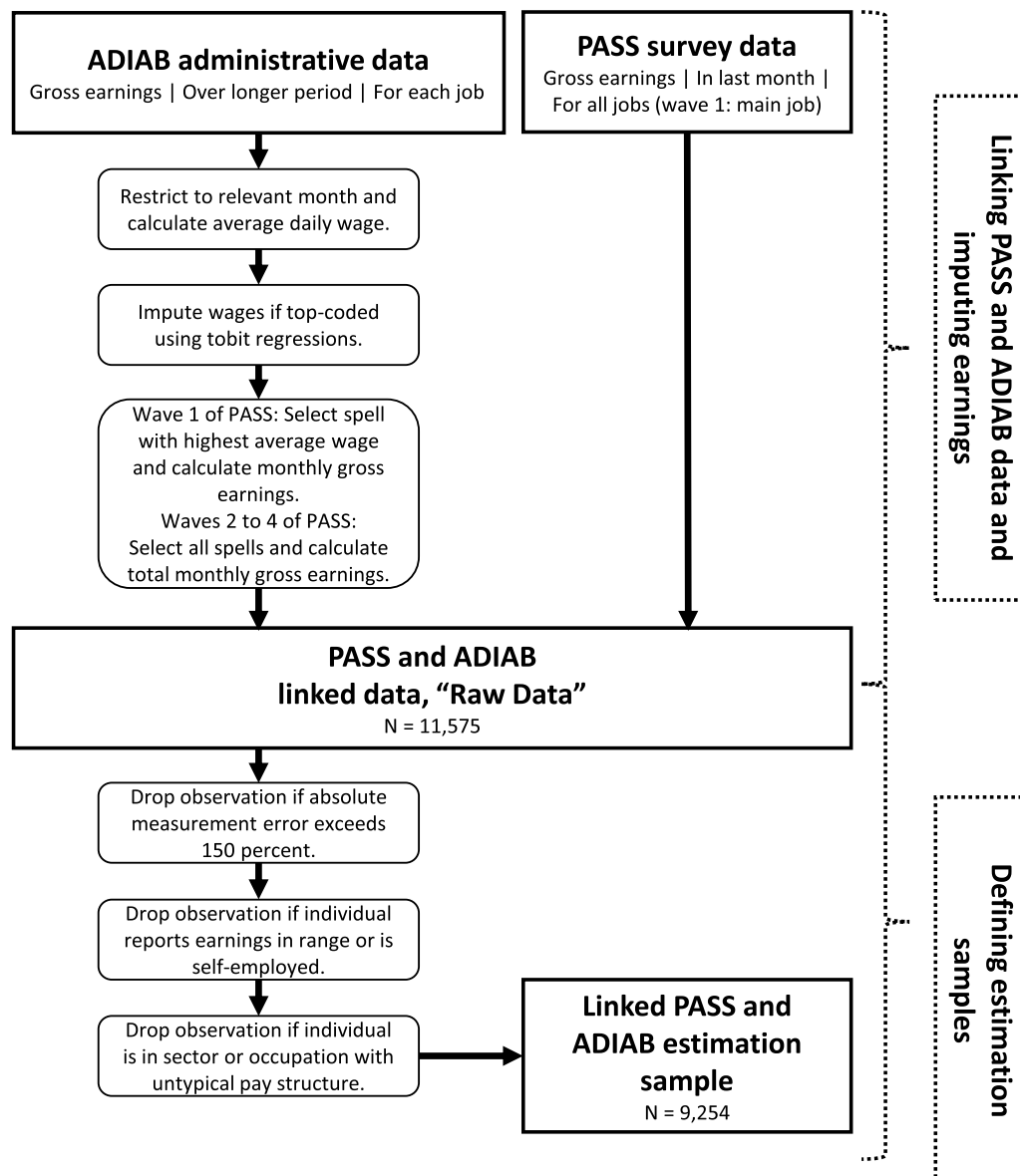
These differences are the subject of robustness checks in Sect. 5.3.<sup>6</sup>

The process of linking PASS with ADIAB data and defining the estimation sample through data cleansing and taking account of the top-coding of the ADIAB earnings data is visualized in simplified form in Fig. 2. First, we identify the ADIAB spells covering the month the respective PASS survey waves refer to. Next, for this month we calculate average daily gross earnings in the ADIAB data. In case of multiple earnings records within the month, for wave one of PASS we select the job with the highest average daily earnings as the main job while for waves two to four we combine earnings from all spells. Finally, we calculate monthly gross earnings by multiplying average daily gross earnings with  $(7/12 * 31 + 4/12 * 30 + 1/12 * 28.75)$ . In practical terms, we are able to successfully link 11,575 spells with information on earnings from both ADIAB and PASS. We label these 11,575 spells our "raw" data.

As the literature emphasizes the potentially substantial consequences of mismatch between different data sources, we go to great lengths to assure that the earnings information gathered from the PASS and ADIAB data are in fact comparable. Starting with our raw data, we implement four discrete data cleansing steps. First, we exclude data points where earnings information from PASS is unlikely to be directly comparable to earnings as recorded in the ADIAB data. This is evident for 179 cases with a measurement error that in absolute terms exceeds 150 percent. Second, we exclude 851 cases where respondents are only willing to give their earnings in terms of a broad range instead of a precise number. Third, we drop 158 cases where respondents indicate that they are self-employed at least once in addition to or instead of being in dependent employment. Fourth, following BK we exclude some occupations and sectors with untypical pay structures. For this reason, all observations for workers in agriculture and engaged as coachmen, managers, artists, performers, clerks, hairdressers, innkeepers, waiters, cleaners, housekeepers, cab drivers, barkeepers or homeworkers are dropped. The fourth step affects

<sup>6</sup> The precise question in the first wave of PASS that we rely on was: "How high was your employment income from your main job last month? If you had special payments in the last month, such as Christmas bonus or back payments, do not include them. However, do count the pay for overtime. Please indicate your gross income first, i.e. your income before deduction of taxes and social security contributions." In the other waves the question was: "What was your employment income last month? In the case of multiple jobs please calculate the total sum. Please indicate your gross income first, i.e. your income before the deduction of taxes and social security contributions. If you had special payments last month, such as Christmas bonus or back payments, do not include these. However, do include any pay for overtime." While respondents were not explicitly encouraged to use payroll or bank statements or other supporting documentation to help them recall their earnings information, they were also not advised against doing so.





**Fig. 2** Linking PASS and ADIAB data and defining the estimation sample. The figure visualizes the process of linking the PASS survey data and ADIAB administrative data and defining the estimation sample. Relevant samples are the raw data without the data cleansing steps implemented for selecting the estimation sample ("raw data") and the full estimation sample with top-coded earnings information replaced by imputed earnings

1404 observations in our raw PASS data. Altogether, data cleansing cuts the estimation sample from 11,575 to 9254 observations.

Table 1 contains separate summary statistics for men and women for (a) the raw linked survey-administrative data, (b) the estimation sample and (c) the estimation sample reduced to a strongly balanced panel (which is required for some of our subsequent analyses). The table focuses on the following individual characteristics: education (whether an individual holds a school leaving

certificate from an academic high school, also called "Abitur"), citizenship (German passport or not), marital status, geographic location (East or West Germany), monthly gross earnings according to both the survey and the administrative information, age and weeks in employment in the respective year.

Table 1 makes it possible to compare differences in terms of observable individual characteristics between the raw data and the estimation sample. Almost all of the individuals considered here are German citizens.

**Table 1** Summary statistics by gender and sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
	Raw data	Sample	Balanced	Raw data	Sample	Balanced
"Abitur"	0.22 (0.01)	0.21 (0.01)	0.26 (0.01)	0.22 (0.01)	0.24 (0.01)	0.23 (0.01)
German	0.95 (0.00)	0.95 (0.00)	0.98 (0.00)	0.96 (0.00)	0.97 (0.00)	0.99 (0.00)
Not married	0.38 (0.01)	0.39 (0.01)	0.27 (0.01)	0.48 (0.01)	0.47 (0.01)	0.37 (0.01)
East German	0.26 (0.01)	0.25 (0.01)	0.21 (0.01)	0.28 (0.01)	0.28 (0.01)	0.26 (0.01)
Survey earnings	2,700.39 (28.64)	2,678.45 (22.90)	3,223.72 (45.60)	1,622.45 (12.72)	1,708.25 (14.40)	1,974.68 (32.97)
Admin earnings	2,773.63 (21.59)	2,838.41 (23.07)	3,514.98 (51.42)	1,742.72 (13.90)	1,857.61 (15.61)	2,149.54 (34.23)
Age	41.38 (0.14)	41.11 (0.16)	44.31 (0.28)	42.14 (0.13)	41.91 (0.15)	44.56 (0.26)
Weeks employed	47.11 (0.14)	47.25 (0.15)	49.56 (0.24)	47.39 (0.13)	47.63 (0.15)	49.93 (0.22)
N	5,721	4,720	1,016	5,854	4,534	1,056

The table shows characteristics of men and women for the raw data without the data cleansing steps implemented for selecting the estimation sample ("raw data"), for the estimation samples ("sample") and for a strongly balanced panel ("balanced"); robust standard errors are in parentheses

On average, sampled individuals are employed for 47 to 48 weeks per year (with a minimum of less than 1 week of employment and a maximum of 52 weeks, denoting continuous employment throughout the respective year). This means they are quite closely attached to the labor market. A comparison across genders shows that women tend to be comparatively less likely to be married and to receive lower earnings. Table 1 also shows that apart from minor exceptions like the earnings information for women the different data cleansing steps leave the mean values of almost all observable characteristics practically unchanged. Thus, data cleansing does not appear to induce a significant amount of selectivity. In contrast, reducing the sample to a strongly balanced panel has more marked implications. Not surprisingly, individuals in the strongly balanced panel have higher average earnings both according to the survey and administrative data. On average, they also have a stronger labor market attachment and are older and more likely to be married, among other notable differences. We return to this issue in Sect. 5 below and show that despite the selectivity of the strongly balanced panel our results are robust to either using this sample or the entire estimation sample.

Table 1 also provides a first comparison between earnings as recorded in the administrative data and reported in the PASS survey. The mean average monthly gross earnings differ little between administrative records and survey

responses, though there is a slight tendency for average administrative earnings to be larger than survey earnings. In the following sections, we analyze the distribution of measurement error in survey-based earnings information in greater detail.

### 3 A static model of measurement error

For our simple static model of measurement error, we follow BK and assume that log gross monthly earnings recorded in the PASS survey  $x$  in wave  $t = \{1, \dots, 4\}$  are composed of a true earnings component in logs  $y$  and an error term  $m$ :

$$x_t = y_t + m_t. \quad (1)$$

Again following BK, we assume that the log gross monthly earnings recorded in the administrative ADIAB data equal the true earnings  $y$  (after adjustment for the top-coding). In addition, we assume that measurement error  $m$  is either classical or "general" in nature. As already mentioned, classical measurement error assumes that  $m$  is identically and independently distributed, with  $E[m_t] = \text{cov}[y_t, m_s] = 0 \forall t, s$ . General measurement error also assumes  $E[m_t] = 0$  but allows for  $\text{cov}[y_t, m_s] \neq 0$ . Below, we compute variance-covariance matrices for  $y$  and  $m$ . Similar to BK, we find a negative correlation between  $y_t$  and  $m_s$  for  $t = s$  (and in weaker form also for  $t \neq s$ ). The negative correlation between true earnings

and measurement error leads us to reject the assumption of classical measurement error. Given that higher earnings tend to be underreported and lower earnings overreported, we follow BK and refer to the general measurement error as mean-reverting measurement error.

As already highlighted in Sect. 1, a suitable statistic for the reliability of the earnings data in the PASS survey is given by the ratio of the covariance between mismeasured earnings and true earnings to the variance of mismeasured earnings. In the case of general measurement error, the resulting reliability ratio  $\lambda^L$  for log earnings in levels is given by

$$\begin{aligned}\lambda_t^L &= \text{cov}[y_t + m_t, y_t] / \text{var}[y_t + m_t] \\ &= (\sigma_{y_t}^2 + \sigma_{y_t m_t}) / (\sigma_{y_t}^2 + \sigma_{m_t}^2 + 2\sigma_{y_t m_t}).\end{aligned}\quad (2)$$

A value for  $\lambda^L$  of 1 would indicate that the covariance between mismeasured earnings and true earnings equals the variance of mismeasured earnings. This would imply perfect reliability. Conversely, a value of 0 would imply a complete lack of reliability.

Correspondingly, the reliability ratio for log earnings in first differences  $\lambda^D$  can be computed as follows:

$$\begin{aligned}\lambda_t^D &= \text{cov}[\Delta(y_t + m_t), \Delta y_t] / \text{var}[\Delta(y_t + m_t)] \\ &= (\sigma_{y_t}^2 + \sigma_{y_{t-1}}^2 - 2\sigma_{y_t, y_{t-1}} + \sigma_{y_t, m_t} - \sigma_{y_{t-1}, m_t} - \sigma_{y_t, m_{t-1}} + \sigma_{y_{t-1}, m_{t-1}}) / \\ &[\sigma_{y_t}^2 + \sigma_{y_{t-1}}^2 + \sigma_{m_t}^2 + \sigma_{m_{t-1}}^2 + 2(\sigma_{y_t, m_t} + \sigma_{y_{t-1}, m_{t-1}} - \sigma_{y_t, y_{t-1}} - \sigma_{m_t, m_{t-1}} - \sigma_{y_t, m_{t-1}} - \sigma_{y_{t-1}, m_t})].\end{aligned}\quad (3)$$

If we allow for mean-reverting measurement error and replace top-coded earnings with imputed earnings based on the procedure described above, we can estimate the reliability ratios in levels and first differences by regressing the ADIAB earnings variable on the earnings information from PASS:

$$y_t = c_t^L + \lambda_t^L x_t + u_t^L, \quad (4)$$

or

$$\Delta y_t = c_t^D + \lambda_t^D \Delta x_t + u_t^D, \quad (5)$$

respectively.

Alternatively, under the assumption of both classical and general measurement error reliability ratios can be estimated by first estimating the variance-covariance matrix of  $y$  and  $m$  and then inserting the relevant estimates of variances and covariances into Eqs. (2) and (3). The variance-covariance matrix of  $y$  and  $m$  is also interesting in itself.

#### 4 Measurement error and induced bias in survey-based earnings regressions

Figure 3 presents two plots of the probability density of measurement error  $m$ , calculated by subtracting log earnings as captured in the PASS survey  $x$  from log earnings

recorded in the administrative ADIAB data  $y$ . The figure pools all four waves of the survey. Panel (a) shows the probability density of  $m$  for men, Panel (b) for women. Both panels also depict the density of a hypothetical normal distribution, vertical lines for zero measurement error and a number of summary statistics. Among the summary statistics, we care most about the mean measurement error as its sign will indicate whether on average earnings are under- or overreported while its absolute value is a key indicator of the typical extent of measurement error.

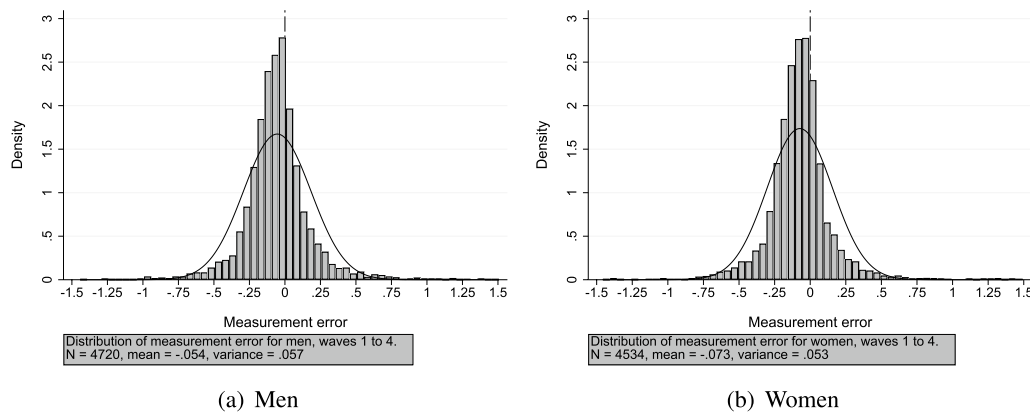
Both panels of Fig. 3 show plots that are unimodal and largely symmetric, with maybe a slight positive skew. Although both distributions are more or less bell-shaped, the tails appear thinner than one would expect with a normal distribution. The histograms indicate that in the PASS survey measurement error has a slightly negative mean. This is confirmed by the summary statistics given at the bottom of the panels. Mean measurement error is  $-5.4$  percent for men and  $-7.3$  percent for women, indicating that both men and women have a slight tendency to underreport their earnings. This result is somewhat in contrast to findings by BK who report that women but not men have such a tendency.

At 5.3 percent for women and 5.7 percent for men, the magnitude of the error variance is substantial for both genders and does not notably differ across genders. The error variances documented here are of a similar order of magnitude to those found by BK. In the CPS data used by BK, the error variance is about 11.4 percent for men and 5.1 percent for women.

Somewhat in contrast to what BK find for the United States, in our linked survey-administrative data set there are very few observations with measurement error at or very close to zero. In PASS only 13.3 percent of observations for men and 13.7 percent of observations for women are associated with an absolute measurement error in log earnings that is smaller or equal to 2.5 percent. 44.4 percent of observations show absolute measurement error not exceeding ten percent, and deviations of more than 30 but less than 150 percent occur in about 13.7 percent of cases. In contrast, according to BK more than 40 percent of CPS respondents report earnings that are within 2.5 percent of their true earnings.

The comparatively small fraction of observations in our sample where measurement error is non-existent or negligible is likely due to differences in the definition of earnings in the PASS and the ADIAB data. Apparently, this causes some residual uncertainty or fuzziness even after the implementation of our thorough procedure to derive comparable information on earnings from the survey-based and administrative data sources.





**Fig. 3** Distribution of measurement error by gender. The figure shows the distribution of measurement error for men and women in the PASS survey; top-coded earnings have been replaced by imputed earnings; measurement error  $m_t$  is defined as the log difference in earnings between survey-based earnings and administrative earnings; also depicted is the density of a normal distribution as are vertical lines for zero measurement error

In Table 2, we turn our attention to the potential biases measurement error might inflict on the estimation of economic relationships. As mentioned by BK, an error-ridden measure of earnings as a left-hand-side variable will produce biased results if and only if the measurement error is correlated with a right-hand-side regressor. Based on this reasoning, Table 2 presents results of regressing measurement error  $m$  on a variety of socio-demographic variables typically found in Mincer-type earnings regressions. For comparison purposes, the table also depicts results from a second set of regressions where log earnings according to the PASS survey data  $x$  are regressed on the same socio-demographic variables. These more standard though simplified Mincer-type earnings regressions serve as a useful benchmark—under the assumption that researchers without access to data on true earnings would want to use  $x$  as their earnings measure.

Columns (1) and (2) of Table 2 show that there is little correlation between some of the socio-demographic variables and the measurement error. For both genders, this is for instance the case for the variables capturing whether an individual holds a school leaving certificate from an academic high school or has German citizenship. However, some other socio-demographic variables exhibit a statistically significant correlation with measurement error either for men or for women or for both genders. For instance, the regression results indicate that earnings of men from East Germany tend to be overreported in the PASS survey while earnings for both men and women with stronger labor market attachment as measured by weeks employed are underreported.

BK point out that the magnitude of the coefficients reported in regressions akin to those in columns (1) and

(2) of Table 2 equal the magnitude of the bias expected for the relevant socio-demographic variables when the log of earnings is the dependent variable of a regression. Although some of the coefficients in each of the two regressions of measurement error are statistically

**Table 2** Individual characteristics associated with measurement error and survey-based earnings by gender

Regression of	(1) Measurement error ( $m_t$ )		(3) Survey-based earnings ( $x_t$ )	
	Men	Women	Men	Women
"Abitur"	0.0024 (0.0086)	-0.0004 (0.0081)	0.3381 (0.0195)	0.2719 (0.0206)
German	-0.0053 (0.0177)	0.0246 (0.0188)	0.3345 (0.04)	0.1646 (0.0475)
Not married	0.0058 (0.0079)	-0.0057 (0.0072)	-0.2174 (0.0163)	0.0522 (0.0172)
East German	0.0230 (0.0080)	0.0146 (0.0074)	-0.3705 (0.0167)	-0.0691 (0.0178)
Age	0.0001 (0.0004)	0.0006 (0.0004)	0.0146 (0.0009)	0.0099 (0.0009)
Log weeks employed	-0.0739 (0.0139)	-0.0571 (0.0155)	0.1823 (0.0255)	0.2126 (0.0291)
Intercept	0.2202 (0.0582)	0.0969 (0.0622)	6.2207 (0.1093)	5.8334 (0.1279)
F	6.57	3.94	333.72	60.49
R <sup>2</sup>	0.013	0.008	0.326	0.084
N	4720	4534	4720	4534

The table shows OLS regressions of measurement error and survey-based earnings; top-coded earnings have been replaced by imputed earnings; survey-based earnings  $x_t$  are defined as log earnings in the PASS survey data and measurement error  $m_t$  as the log differences in earnings between survey-based earnings and administrative earnings; robust standard errors are in parentheses

**Table 3** Variance-covariance matrix by gender

Panel (a): Men (n=254)								
	$y_1$	$y_2$	$y_3$	$y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$y_1$	<b>0.254</b>	0.922	0.923	0.905	-0.226	-0.177	-0.180	-0.164
$y_2$	0.219	<b>0.223</b>	0.962	0.934	-0.127	-0.242	-0.165	-0.193
$y_3$	0.219	0.213	<b>0.221</b>	0.963	-0.130	-0.199	-0.286	-0.232
$y_4$	0.225	0.217	0.223	<b>0.242</b>	-0.096	-0.166	-0.236	-0.307
$m_1$	-0.023	-0.012	-0.012	-0.009	<b>0.040</b>	0.435	0.310	0.159
$m_2$	-0.016	-0.020	-0.017	-0.015	0.016	<b>0.032</b>	0.534	0.349
$m_3$	-0.018	-0.016	-0.027	-0.023	0.012	0.019	<b>0.041</b>	0.420
$m_4$	-0.018	-0.020	-0.024	-0.033	0.007	0.014	0.019	<b>0.048</b>
Panel(b): Women (n=264)								
	$y_1$	$y_2$	$y_3$	$y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$y_1$	<b>0.262</b>	0.936	0.870	0.843	-0.187	-0.131	-0.092	-0.170
$y_2$	0.237	<b>0.244</b>	0.923	0.905	-0.209	-0.209	-0.121	-0.204
$y_3$	0.224	0.230	<b>0.254</b>	0.958	-0.199	-0.162	-0.197	-0.210
$y_4$	0.222	0.230	0.248	<b>0.263</b>	-0.256	-0.199	-0.179	-0.311
$m_1$	-0.020	-0.021	-0.021	-0.027	<b>0.043</b>	0.309	0.188	0.313
$m_2$	-0.015	-0.022	-0.018	-0.022	0.014	<b>0.047</b>	0.304	0.267
$m_3$	-0.010	-0.012	-0.020	-0.019	0.008	0.014	<b>0.043</b>	0.364
$m_4$	-0.014	-0.017	-0.018	-0.027	0.011	0.010	0.012	<b>0.028</b>

The table shows variance-covariance matrices of true earnings and measurement error for men and women; covariances are at and below the diagonal, correlations above the diagonal; top-coded earnings have been replaced by imputed earnings; measurement error  $m_t$  is defined as the log differences in earnings between survey-based PASS earnings and administrative ADIAB earnings and true earnings  $y_t$  as log earnings in the administrative ADIAB data; strongly balanced panel; variances in bold

significant, the estimated bias for each variable is generally small compared to its typical effect in an earnings equation. This is evident from a comparison of the coefficients in columns (1) and (2) with those in columns (3) and (4). For instance, as mentioned earnings of East German men tend to be overreported in the PASS survey. But the relevant coefficient is less than ten percent of that of the “East German dummy” in a simplified Mincer-type earnings regressions for men using PASS survey data.

At 1.3 percent for men and 0.8 percent for women, the coefficients of determination  $R^2$  implied by the regressions of measurement error on socio-demographic variables are substantially smaller than the coefficients from the comparable regressions of survey-based earnings (which amount to 32.6 percent and 8.4 percent for men and women, respectively). Together with the estimated bias for each variable being generally small compared to its typical effect in an earnings equation, and confirming what BK find for the United States, this suggests that measurement error leads to little bias when earnings information from the PASS data is on the left-hand side of a regression.

## 5 Reliability of the data

### 5.1 Variance-covariance matrices of true earnings and measurement error

In Table 3, we estimate variance-covariance matrices and correlation coefficients of true earnings  $y$  and measurement error  $m$  separately for men and women. The table gives an indication of the magnitude and covariance structure of measurement error over the first four waves of the PASS survey and allows us to assess the reliability of the survey-based earnings data. The imputation procedure of top-coded administrative earnings developed by Card et al. (2013) makes it possible to derive the variance-covariance matrix of  $y$  and  $m$  directly from the data. Using the formula  $\text{cor}[X, Y] = \text{cov}[X, Y] / (\text{sd}[X]\text{sd}[Y])$  it is then straightforward to calculate the correlation coefficients from the variances and covariances. Covariances are depicted at and below the diagonal, correlations above the diagonal and the two samples are reduced to strongly balanced panels.

Four noteworthy results emerge from Table 3. First, true earnings exhibit a high degree of first- and higher-order autocorrelation. First-order autocorrelation

coefficients of true earnings  $\text{cor}[y_1, y_2]$ ,  $\text{cor}[y_2, y_3]$  and  $\text{cor}[y_3, y_4]$  are 0.92, 0.96 and 0.96 for men and 0.94, 0.92 and 0.96 for women, respectively. Higher-order autocorrelation coefficients are pronounced as well. This points at a high degree of earnings persistence in Germany.

Second, Table 3 shows large negative correlations between measurement error and true earnings for both men and women. The correlation coefficients  $\text{cor}[y_1, m_1]$ ,  $\text{cor}[y_2, m_2]$ ,  $\text{cor}[y_3, m_3]$  and  $\text{cor}[y_4, m_4]$  for men amount to  $-0.23$ ,  $-0.24$ ,  $-0.29$  and  $-0.31$  in waves 1, 2, 3 and 4, respectively. For women, they are  $-0.19$ ,  $-0.21$ ,  $-0.20$  and  $-0.31$ . We also document sizable negative correlations between measurement error and true earnings in different years that hardly abate as the distance between years increases. The strong negative correlations between measurement error and true earnings imply that measurement error is not classical as this assumption would require all these correlations to be zero. Instead, higher earnings tend to be underreported and lower earnings overreported, making measurement errors in PASS earnings data mean-reverting. This was also documented by BK for men—but not necessarily for women.<sup>7</sup>

Third, our results strongly suggest that measurement errors in the earnings information available in PASS are positively correlated from one year to the next and even over an extended period of time. BK document a first-order autocorrelation coefficient for  $m$  of 0.40 for men and 0.10 for women. We find similar and even somewhat more pronounced first-order autocorrelation coefficients  $\text{cor}[m_1, m_2]$ ,  $\text{cor}[m_2, m_3]$  and  $\text{cor}[m_3, m_4]$  of 0.44, 0.53 and 0.42 for men and 0.31, 0.30 and 0.36 for women, respectively. This implies that a positive measurement error in 1 year is relatively likely to be followed by another positive measurement error in the next (and similarly for negative measurement errors).

Fourth, going beyond the exercise by BK, our extended panel makes it possible to investigate higher order autocorrelations of  $m$ . At 0.31 and 0.35 for men and 0.19 and 0.27 for women, second-order autocorrelation coefficients  $\text{cor}[m_1, m_3]$  and  $\text{cor}[m_2, m_4]$  are only slightly lower than first-order autocorrelation coefficients. At 0.16 for men and 0.31 for women, even third-order autocorrelation coefficients  $\text{cor}[m_1, m_4]$  are pronounced. Apparently, a sizable fraction of the autocorrelation in the measurement error is not actually due to an autoregressive process but rather caused by either individual-specific time-invariant characteristics (i.e., person fixed effects) or other, more complex time series processes.

## 5.2 Reliability of PASS earnings data

Table 4 summarizes reliability ratios for PASS earnings data. These have been derived by first estimating the variance-covariance matrix of  $y$  and  $m$  and then inserting the relevant estimates into Eqs. (2) and (3). Reliability ratios are presented separately for men and women as well as for the pooled sample. The table also differentiates between the first four waves of PASS, between assuming classical measurement error (Panel (a)) and allowing for mean-reverting measurement error (Panel (b)), and between the cross section and first differences.

In a cross section our longitudinal earnings data are relatively reliable: Under the assumption of classical measurement error, the cross-sectional reliability of PASS earnings data is around 0.85 for both genders. This is not too far removed from the value of 1 which would imply perfect reliability. If we allow for mean-reverting measurement error—our preferred specification—this further increases the estimates of cross-sectional reliability. As noted by BK, this is exactly as one would expect in the case of  $\sigma_m^2 < \sigma_y^2$  which is what we find in our data (as discussed above). In the mean-reverting specification, cross-sectional reliability ratios for male PASS respondents range from 0.93 to 0.95 and for female respondents from 0.90 to 1.00. For the combined sample, they range from 0.93 to 0.98. These estimates are similar to those by BK and other researchers using data sets for the United States. In particular, BK report that in a cross section, the reliability ratio is 0.82 to 1.02 for men and 0.92 to 0.96 for women.

Similar to what is reported by BK for the CPS, measurement error is significantly exacerbated when considering earnings changes. In fact, the reliability ratios for PASS earnings data in first differences are even lower than the comparable statistics for the CPS documented by BK. Their estimates generally range from 0.65 to 0.85. In contrast, when we allow for mean-reverting measurement error we document point estimates for reliability ratios in first differences that range from 0.18 to 0.38. Under the assumption of classical measurement error, estimates range from 0.29 to 0.41. In contrast to what is found by BK, in our case allowing for mean-reverting measurement error does not increase but in fact decreases the reliability ratio in first differences.

As mentioned above, an alternative procedure to estimate the reliability ratios in levels and first differences is to regress the ADIAB earnings variable on PASS earnings information. Results from this procedure are summarized in Table 5. The table again estimates reliability ratios separately by gender and survey wave. All estimates depicted in the table allow for mean-reverting measurement error. The main value-add of the alternative procedure is that it allows us to exploit the full extent of our estimation

<sup>7</sup> The finding of mean-reverting measurement error led BK to emphasize empirical results based on a sample where top-coded earnings are imputed. In this study, we follow a similar approach.

**Table 4** Reliability ratios inferred from variance-covariance matrix by gender and type of measurement error

Wave	(1) All	(2) Men	(3) Women
<b>Panel (a): Classical measurement error</b>			
1	0.888	0.864	0.858
2	0.882	0.874	0.838
3	0.878	0.845	0.856
4	0.893	0.835	0.905
2-1	0.407	0.484	0.345
3-2	0.366	0.333	0.382
4-3	0.289	0.255	0.327
<b>Panel (b): Mean-reverting measurement error</b>			
1	0.940	0.931	0.912
2	0.928	0.945	0.900
3	0.930	0.935	0.913
4	0.984	0.934	0.996
2-1	0.376	0.474	0.322
3-2	0.290	0.095	0.342
4-3	0.182	0.110	0.249
n	518	254	264

The table reports reliability ratios  $\lambda$  derived from the variance-covariance matrix of true earnings  $y$  and measurement error  $m$ ; top-coded earnings have been replaced by imputed earnings; Panel (a) imposes classical measurement error assumptions while Panel (b) allows for mean-reverting measurement error; in either case, serial correlation in  $m$  is allowed; strongly balanced panel

sample without the need to construct a balanced panel. In addition, the alternative procedure also makes analytical standard errors readily available.

Results from the regressions summarized in Table 5 are qualitatively and for the most part also quantitatively very similar to those from the more indirect procedure depicted in Table 4. In fact, reliability ratios both in levels and in first differences estimated from the full samples tend to be somewhat larger than the estimates from the strongly balanced panels. In addition, Table 5 makes it clear that all reliability ratios are statistically significantly different from zero. This means that even in the case of earnings data in first differences, an analysis relying purely on PASS earnings data would be able to capture some meaningful information. Nevertheless, while Table 4 and Table 5 leave us confident to encourage analyses using PASS earnings data in levels, more caution might be in order when analyzing earnings data in first differences.

### 5.3 Robustness and sensitivity checks

Our main results on the reliability of the PASS earnings data are robust to a wide range of alternative specifications. This is demonstrated through a series of sensitivity checks summarized in Table 6 that estimate reliability ratios  $\lambda$  derived from variance-covariance matrices of true earnings  $y$  and measurement error  $m$ . For conciseness, all estimates allow for mean-reverting measurement error and are pooled for men and women (but depicted separately for earnings information in levels and first differences and for each of the four waves of PASS).

As described above, our preferred “baseline” estimates of  $\lambda$  from Table 4 are derived using an estimation sample where top-coded earnings are imputed using Tobit regressions and various data cleansing steps are implemented to minimize mismatch between different data sources. Columns (1) to (3) of Table 6 deal with the issue of top-coding in the administrative data in alternative ways while in columns (4) to (7) we check if our results

**Table 5** Reliability ratios inferred from OLS regressions by gender

Wave	(1) All	(2) Men	(3) Women
1	0.955 (0.008) [2,543]	0.950 (0.012) [1,388]	0.961 (0.012) [1,155]
2	0.952 (0.011) [2,105]	0.953 (0.016) [1,094]	0.955 (0.018) [1,011]
3	0.952 (0.011) [2,283]	0.944 (0.019) [1,120]	0.968 (0.016) [1,163]
4	0.993 (0.012) [2,323]	0.973 (0.019) [1,118]	1.008 (0.016) [1,205]
2-1	0.370 (0.053) [1,197]	0.435 (0.083) [630]	0.308 (0.056) [567]
3-2	0.314 (0.051) [1,283]	0.217 (0.068) [640]	0.362 (0.068) [643]
4-3	0.361 (0.060) [1,337]	0.211 (0.081) [642]	0.459 (0.078) [695]

The table reports reliability ratios  $\lambda$  given by slope-coefficient estimates from OLS regressions of ADIAB earnings on PASS earnings and an intercept, or their first differences; top-coded earnings have been replaced by imputed earnings; estimates allow for mean-reverting measurement error; robust standard errors are in parentheses and sample sizes in brackets

are robust to alternative delineations of the estimation sample. Finally, column (8) deals with differences in the definition of earnings in the PASS and ADIAB data.

In column (1) of Table 6, all observations with top-coded earnings are excluded. This exclusion leads to a loss of external validity but not necessarily internal validity.<sup>8</sup> Estimates of reliability ratios are somewhat lower than the baseline reliability ratios with regard to earnings in levels and somewhat higher when earnings are specified in first differences. Overall, differences between baseline estimates and those of column (1) are not particularly pronounced. In column (2), top-coded earnings information is not replaced by imputed values but kept in its original form. In levels, the resulting—biased—estimates of  $\lambda$

<sup>8</sup> BK highlight that under the assumption of classical measurement error excluding top-coded earnings information does not even lead to a nonrandom sample of  $m$  nor to sample selection bias in the estimates of  $\lambda^L$  and  $\lambda^D$ . This is because under classical measurement error, sampling on  $y$  falling below the social security contribution assessment ceiling is equivalent to sampling on an orthogonal variable. However, if we allow for mean-reverting measurement error sampling on  $y$  falling below the social security contribution assessment ceiling will result in a nonrandom sample.

tend to be lower than the baseline estimates. In first differences, time-invariant effects are eliminated and differences between baseline estimates and those of column (2) are small. In column (3), top-coded earnings information is replaced through an alternative imputation procedure proposed by BK. Instead of directly imputing top-coded earnings, BK impute the variance-covariance matrices of earnings and measurement error using a maximum likelihood procedure. Column (3) yields reliability ratios that are very similar to those of our baseline estimates.<sup>9</sup>

In column (4) of Table 6, we do not exclude outliers (defined as observations with absolute measurement error in excess of 150 percent) from the observation sample, in column (5) we do not exclude observations where survey earnings were reported as a range instead of a precise number and in column (6) we do not exclude observations for individuals working in sectors or occupations with untypical pay structures, such as agriculture. In column (7), we restrict our sample to standard employment spells only by excluding spells of student interns, apprentices or those in marginal employment (i.e., with earnings below the social security contribution threshold). The underlying motivation is that the first wave of PASS includes earnings information for apprentices and student interns but excludes those in marginal employment while the ensuing waves include the marginally employed but exclude apprentices and students—whereas in the ADIAB data all these groups in non-standard employment are consistently included. Finally, in column (8) we account for the fact that one-time payments are always included in the ADIAB earnings data but not in the standard PASS earnings variable by adding the value of one-time payments (which is collected separately) to the standard PASS earnings variable.<sup>10</sup> Although the point estimates for reliability ratios in levels or first differences differ slightly across the different specifications of columns (4) to (8), the general picture and key findings stay very robust.

In addition to the sensitivity checks of Table 6, the appendix replicates all estimation results of this paper using a second linked survey-administrative data set that combines the WeLL household panel with the ADIAB administrative records. The main objective of this replication exercise is to investigate the external validity of our

<sup>9</sup> In one case the estimate for  $\lambda^L$  is greater than 1, which would be implausible. This is likely a rounding error.

<sup>10</sup> Depending on the survey wave, between 29 and 33 percent of respondents in our estimation sample note that they received a one-time payment in the preceding year. While relatively frequent, at six percent of total annual earnings the average amount of one-time payments is comparatively small. This robustness exercise comes with the caveat that PASS records one-time payments for the entire year and earnings information on a monthly basis. For comparability, we divide all one-time payments by 12 before adding them to the PASS earnings variable.



**Table 6** Reliability ratios inferred from variance-covariance matrix by survey and sample restriction

Sample restriction	(1) Top-coded earnings excluded	(2) Top-coded earnings included without adjustments	(3) Top-coded earnings replaced by imputed earnings (BK's procedure)	(4) Outliers not excluded from sample	(5) Imprecise earnings not excluded from sample	(6) Occupations with unique structure not excluded from sample	(7) Standard employment spells only	(8) One-time payments included in PASS earnings
1	0.918	0.908	0.977	0.900	0.921	0.949	0.936	0.913
2	0.909	0.887	0.962	0.905	0.911	0.940	0.941	0.934
3	0.898	0.894	0.958	0.892	0.921	0.932	0.940	0.937
4	0.962	0.931	1.006	0.969	0.975	0.990	0.995	0.971
2-1	0.384	0.377	0.384	0.331	0.292	0.386	0.241	0.385
3-2	0.303	0.289	0.297	0.212	0.268	0.275	0.290	0.334
4-3	0.192	0.179	0.195	0.151	0.166	0.182	0.141	0.189
n	452	518	518	525	607	574	51*	518

The table reports estimates of reliability ratios  $\lambda$  derived from the variance-covariance matrix of true earnings  $y$  and measurement error  $m$  that allow for mean-reverting measurement error; serial correlation in  $m$  is allowed; strongly balanced panel; \* indicates that a small number of observations is dropped in column (7) as compared to the baseline specifications but we are unable to report the precise number due to data protection regulations

results across the earnings distribution, given that PASS primarily encompasses low-income households whereas WeLL focuses on employees of a random but stratified sample of all German establishments. As compared to the labor market as a whole, WeLL respondents are positively selected in terms of their labor market attachment and also earnings (with average gross monthly earnings of 3,126 Euro as compared to 2,203 Euro for the PASS estimation sample and 2,545 Euro for the PASS sample weighted to approximate the German labor market). As documented in detail in the appendix, qualitatively and for the most part also quantitatively this replication exercise leads to very consistent results.

## 6 Measurement error and earnings dynamics

### 6.1 A dynamic model of measurement error

In this section, we assess how measurement error biases parameter estimates in a dynamic model of earnings. Building on Pischke (1995), we assume that true earnings  $y_t$  are composed of a permanent component  $z_t$  with a random walk on  $\epsilon_t$  and a transitory component  $\eta_t$ :

$$y_t = z_t + \eta_t; \quad z_t = z_{t-1} + \epsilon_t. \quad (6)$$

Pischke (1995, p. 308–309) notes that “[i]n many applications it is useful to decompose earnings in to a permanent and transitory component” as “[d]ifferent events are associated with these components. Promotions or job loss in a high-paying industry lead to permanent changes in earnings, and overtime and temporary layoffs induce temporary variations.”

A specification of the earnings process defined by Eq. (6) in first differences illustrates that changes in earnings follow a first-order moving average process:

$$\Delta y_t = \epsilon_t + \eta_t - \eta_{t-1}. \quad (7)$$

Pischke (1995) observes only two periods of survey-based earnings data. He therefore has to make a number of relatively specific assumptions about the measurement error process and the structure of the variances of permanent and transitory earnings shocks. For reasons of comparability, we impose the exact same assumptions. More specifically, measurement error is assumed to be composed of three additive components, (a) the transitory earnings shock  $\eta_t$  with “transmission” factor  $\alpha$ , (b) an individual fixed effect  $\mu$  that is uncorrelated with earnings and (c) a white noise error term  $\xi_t$ :

$$m_t = \alpha \eta_t + \mu + \xi_t. \quad (8)$$

Due to the individual fixed effect the model assumes a time-invariant autocorrelation of measurement error. If  $\alpha$  is smaller than zero then this implies mean-reverting measurement error and underreporting of transitory earnings shocks. In this simple model, measurement error is assumed to be unrelated to permanent earnings shocks. Again following Pischke (1995), we additionally assume that the variance of permanent earnings shocks  $\text{var}[\epsilon_t]$  is constant over time while the variance of transitory earnings shocks  $\text{var}[\eta_t]$  may vary:

$$\begin{aligned} \text{var}[\epsilon_t] &= \sigma_\epsilon^2; & \text{var}[\eta_t] &= \sigma_{\eta t}^2; & \text{cov}[\epsilon_t, \eta_t] &= 0; \\ \text{var}[\xi_t] &= \sigma_{\xi t}^2; & \text{var}[\mu] &= \sigma_\mu^2. \end{aligned} \quad (9)$$

The resulting model for earnings dynamics and measurement error implies the following moment conditions:

**Table 7** GMM estimates of earnings dynamics and measurement error

Estimates of earnings dynamics					
$\hat{\sigma}_{\eta_1}^2$	$\hat{\sigma}_{\eta_2}^2$	$\hat{\sigma}_{\eta_3}^2$	$\hat{\sigma}_{\eta_4}^2$	$\hat{\sigma}_{\epsilon}^2$	
0.0114	0.0083	0.0052	0.0054	0.0133	
(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	
Estimates of measurement error process					
$\hat{\sigma}_{\xi_1}^2$	$\hat{\sigma}_{\xi_2}^2$	$\hat{\sigma}_{\xi_3}^2$	$\hat{\sigma}_{\xi_4}^2$	$\hat{\sigma}_{\mu}^2$	$\hat{\alpha}$
0.0232	0.0230	0.0264	0.0221	0.0128	-0.6974
(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0099)

The table reports GMM estimates of earnings dynamics and measurement error; top-coded earnings have been replaced by imputed earnings; standard errors are in parentheses

$$\begin{aligned}
\text{var}[\Delta y_t] &= \sigma_{\epsilon}^2 + \sigma_{\eta_t}^2 + \sigma_{\eta_{t-1}}^2; \\
\text{cov}[\Delta y_t, \Delta y_{t-1}] &= -\sigma_{\eta_{t-1}}^2; \\
\text{var}[m_t] &= \alpha^2 \sigma_{\eta_t}^2 + \sigma_{\xi_t}^2 + \sigma_{\mu}^2; \\
\text{cov}[\Delta y_t, m_t] &= \alpha \sigma_{\eta_t}^2; \\
\text{cov}[\Delta y_t, m_{t-1}] &= -\alpha \sigma_{\eta_{t-1}}^2; \\
\text{cov}[m_t, m_{t-j}] &= \sigma_{\mu}^2.
\end{aligned} \tag{10}$$

All other covariances are assumed to be zero.

## 6.2 Estimation results

We identify the parameters of the dynamic model of measurement error and earnings as described in Eqs. (7) and (8) by fitting the moment conditions of Eq. (10) to the sample covariance matrix of earnings changes and measurement error using Generalized Method of Moments (GMM). For this purpose, we rely on the same two linked survey-administrative data sets also used for the static model of measurement error and earnings and compute estimation results separately for the PASS survey data (as described here) and for WeLL (cf. the appendix).

The earnings process of Eq. (7) can be decomposed into a permanent and a transitory component. The relevant estimates in Table 7 imply that according to PASS earnings data, between 40 percent and 56 percent of the variation in earnings changes is due to the permanent component. For instance, for the second wave of PASS the variation in earnings changes amounts to  $\text{var}[\Delta y_2] = \hat{\sigma}_{\eta_1}^2 + \hat{\sigma}_{\eta_2}^2 + \hat{\sigma}_{\epsilon}^2 = 0.0114 + 0.0083 + 0.0133 = 0.0330$  and the variation in the permanent component to  $\hat{\sigma}_{\epsilon}^2 = 0.0133$ . In turn,  $0.0133/0.0330 = 0.4030$  indicates that for the second wave of PASS 40 percent of the variation in earnings changes is due to the permanent component. Over all four waves of PASS, the importance of the permanent component for explaining the variation in earnings changes is rather sizable. It is also much higher than what is found by Pischke (1995) who only attributes between ten and 25 percent of

the variation in earnings changes to the permanent component. According to these results, compared to the United States earnings changes in Germany are relatively more likely to result from events like promotions or job loss in a high-paying industry that lead to permanent changes in earnings rather than occurrences like overtime and temporary layoffs that induce only temporary variations.

The estimates of the measurement error process of Eq. (8) make it possible to account for the contributions of white noise, individual fixed effects and underreporting of transitory earnings. According to our estimates, the variance of the white-noise component is quite large, between 0.0221 and 0.0264. At 0.0128, the variance of the individual fixed effect is sizable as well. In contrast, the variance of underreporting is only between 0.0052 and 0.0114. The contribution of underreporting is further moderated by the transmission factor  $\alpha$  at  $-0.6974$ . Overall, the white-noise component accounts for between 56 percent to 63 percent of the total variance of the measurement error, individual fixed effects for between 31 and 34 percent and the underreporting of transitory earnings for between six and 13 percent.<sup>11</sup> This means that according to our model and data both random fluctuations and individual-specific time-invariant characteristics account for most of the total variance of the measurement error while the underreporting of transitory earnings is much less important. While Pischke (1995) also finds that underreporting plays a relatively small role in explaining the total variance, he attributes more than 80 percent of this variance to the white-noise component.

In turn, the underreporting of transitory earnings depends on the extent of transitory earnings shocks and the transmission factor  $\alpha$ . The negative estimate of  $\alpha$  confirms the finding of mean-reverting measurement error already documented in Sect. 5 and implies that individuals underreport transitory earnings by 70 percent. Pischke (1995) also finds mean-reverting measurement error and underreporting of transitory earnings. As already mentioned, Pischke (1995) regards it as very plausible that people underreport changes in transitory earnings as such changes tend to be related to relatively minor events.<sup>12</sup>

<sup>11</sup> For instance, for the first period the contribution of the underreporting of transitory earnings to the overall variance of measurement error equals  $[\alpha^2 * \sigma_{\eta_1}^2] / [\alpha^2 * \sigma_{\eta_1}^2 + \sigma_{\xi_1}^2 + \sigma_{\mu}^2] = [(-0.6974)^2 * 0.0114] / [(-0.6974)^2 * 0.0114 + 0.0232 + 0.0128] = 0.1339$ .

<sup>12</sup> As a robustness check, we added information on one-time payments to the standard PASS earnings variable and then repeated the decomposition of the earnings process of Eq. (7) into a permanent and a transitory component. Results from this robustness check are not reported here but available upon request. They show that the results of Table 7 are not driven in any noteworthy way by the omission of one-time payments in the standard PASS earnings variable.

**Table 8** Estimated and fitted variance-covariance matrices

Panel (a): Estimated variance-covariance matrix (n = 518)							
	$\Delta y_2$	$\Delta y_3$	$\Delta y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$\Delta y_2$	0.0357						
$\Delta y_3$	-0.0054	0.0280					
$\Delta y_4$	0.0003	-0.0039	0.0198				
$m_1$	0.0045	0.0002	-0.0019	0.0416			
$m_2$	-0.0063	0.0041	-0.0012	0.0147	0.0399		
$m_3$	-0.0002	-0.0098	0.0025	0.0102	0.0165	0.0418	
$m_4$	-0.0019	-0.0023	-0.0090	0.0089	0.0114	0.0152	0.0376
Panel (b): Fitted variance-covariance matrix (n = 518)							
	$\Delta y_2$	$\Delta y_3$	$\Delta y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$\Delta y_2$	0.0331						
$\Delta y_3$	-0.0083	0.0268					
$\Delta y_4$	0	-0.0052	0.0239				
$m_1$	0.0080	0	0	0.0416			
$m_2$	-0.0058	0.0058	0	0.0128	0.0399		
$m_3$	0	-0.0036	0.0036	0.0128	0.0128	0.0417	
$m_4$	0	0	-0.0038	0.0128	0.0128	0.0128	0.0376

Panel (a) reports the variance-covariance matrix estimated from the data; Panel (b) depicts the fitted variance-covariance matrix from the model; top-coded earnings have been replaced by imputed earnings; entries marked "0" are constrained to be exactly zero

In order to investigate the fit and potential shortcomings of the dynamic model of measurement error, Table 8 uses the PASS data to report both the actual covariance matrix of earnings dynamics and measurement error observed in the data and a fitted covariance matrix that accounts for the constraints imposed by our model. Overall, our simple model fits the data very well: Estimated and fitted values of variances and first-order autocovariances of earnings changes and measurement error are all very similar. In the model, higher-order autocovariances of measurement error are assumed to be identical to first-order autocovariances of measurement error while higher-order autocovariances of earnings changes are assumed to be zero. Both assumptions seem to be relatively good approximations of the data-generating process.

The model also does a good job in fitting contemporary and lagged covariances between earnings changes and measurement error. Both in terms of signs and magnitudes, all relevant covariances allowed to be non-zero are fitted very well. The relevant constraints also appear broadly appropriate although some of the covariances constrained to zero in the fitted model—such as  $\text{cov}[\Delta y_2, m_4]$ ,  $\text{cov}[\Delta y_3, m_4]$  and  $\text{cov}[\Delta y_4, m_1]$ —are estimated to be about 0.002 in absolute terms in the data.

### 6.3 Implications for the estimation of earnings dynamics

Reconsider the simple model in Eq. (6) and assume that we wanted to estimate the variances of the innovation to

the permanent component  $\epsilon_t$  and to the transitory component  $\eta_t$  but that we only had access to PASS survey data. In other words, assume data were available not on true earnings  $y_t$  but only on reported earnings  $x_t$ . Then, instead of the process for changes in true earnings of Eq. (7), we would need to rely on the following process for changes in reported earnings:

$$\Delta x_t = \epsilon_t + (1 + \alpha)(\eta_t - \eta_{t-1}) + \xi_t - \xi_{t-1}. \quad (6.6)$$

In this case, the corresponding moment conditions would be:

$$\begin{aligned} \text{var}[\Delta x_t] &= \sigma_\epsilon^2 + (1 + \alpha)^2(\sigma_\eta^2 + \sigma_{\eta_{t-1}}^2) + \sigma_{\xi_t}^2 + \sigma_{\xi_{t-1}}^2; \\ \text{cov}[\Delta x_t, \Delta x_{t-1}] &= -(1 + \alpha)\sigma_{\eta_{t-1}}^2 - \sigma_{\xi_{t-1}}^2. \end{aligned} \quad (6.7)$$

A comparison of Eqs. (10) and (12) gives an indication of the degree to which measurement error biases survey-based estimates of the parameters in a dynamic model of earnings. Estimates of the variance of true earnings changes using Eq. (12) are confounded by two bias terms of opposite sign. On the one hand, the variance is underestimated because of the underreporting of transitory earnings,  $((1 + \alpha)^2 - 1)(\sigma_\eta^2 + \sigma_{\eta_{t-1}}^2)$  with  $\alpha < 0$ . On the other hand, the white-noise term in measurement error  $\xi_t$  results in the overestimation of the variance of true earnings changes due to the additional positive variance component,  $\sigma_{\xi_t}^2 + \sigma_{\xi_{t-1}}^2$ . Two bias terms of opposite sign also arise for the first-order autocovariances of true earnings changes.

**Table 9** Bias in estimates of variance of earnings growth

Wave	Underreporting of transitory earnings	White-noise error	Total bias	var[ $\Delta y_t$ ]	Reliability ratios
2	-0.0180	0.0461	0.0282	0.0331	0.5401
3	-0.0123	0.0494	0.0371	0.0268	0.4193
4	-0.0096	0.0485	0.0389	0.0239	0.3805

The table summarizes the biases in estimates of the variance of earnings growth that result from the underreporting of transitory earnings and white noise error for the PASS survey; top-coded earnings have been replaced by imputed earnings

**Table 10** Bias in estimates of first-order autocovariance of earnings growth

Wave	Underreporting of transitory earnings	White-noiseerror	Total bias	cov[ $\Delta y_t, \Delta y_{t-1}$ ]	Reliability ratios
2	0.0080	-0.0232	-0.0152		
3	0.0058	-0.0230	-0.0171	-0.0083	0.3274
4	0.0036	-0.0264	-0.0228	-0.0052	0.1844

The table summarizes the biases in estimates of the first-order autocovariance of earnings growth that result from the underreporting of transitory earnings and white noise error for the PASS survey; top-coded earnings have been replaced by imputed earnings

Tables 9 and 10 use the GMM estimates from Table 7 to quantify the biases that result from the underreporting of transitory earnings and white noise error. The tables also depict the total biases in estimates of the variance and first autocovariance of earnings growth and, with reference to the fitted variance-covariance matrices from Table 8, the variances of true earnings changes  $\text{var}[\Delta y_t]$ . Finally, they list the reliability ratios, i.e., the variance of true earnings changes to the variance of reported earnings changes  $\lambda_t = \text{var}[\Delta y_t]/\text{var}[\Delta x_t]$ .

Table 9 focuses on the bias in estimates of the variance of earnings growth. Its first column reports the bias due to underreporting of transitory earnings, the second column gives the bias due to the white-noise component in the measurement error and the third the total bias.<sup>13</sup> As is evident from these first three columns, in the PASS data the biases due to underreporting of transitory earnings and white noise error are both substantial. In accordance with findings by Pischke (1995), the bias due to white noise error consistently has the opposite sign as the bias due to the underreporting of transitory earnings. In absolute terms, the bias due to white noise error is larger than the bias due to the underreporting of transitory earnings. Therefore, the two biases do not cancel each other out. Instead, relying entirely on the PASS data leads to a substantial overestimation of the overall variance of earnings

growth. This is different to what Pischke (1995) finds for the PSID.

At between 0.02 and 0.03, the values for the variance of true earnings changes  $\text{var}[\Delta y_t]$  listed in the fourth column of Table 9 are also slightly larger than the comparable values reported by Pischke (1995), which range from 0.01 to 0.03. This should probably come as no surprise as our linked survey-administrative data span the entire German economy, whereas Pischke's (1995) data cover just one single plant. Reliability ratios reported in the fifth column of Table 9 are between 0.38 and 0.54. This is somewhat lower as what is documented by Pischke (1995) who finds a reliability ratio of 0.79 for the first year of his survey and a ratio of 0.54 for the second one.

Table 10 repeats the exercise underlying Table 9 but for  $\text{cov}[\Delta y_t, \Delta y_{t-1}]$  instead of  $\text{var}[\Delta y_t]$ . The first three columns of Table 10 confirm for  $\text{cov}[\Delta y_t, \Delta y_{t-1}]$  that the biases due to underreporting of transitory earnings and white noise error are of opposite signs and rather substantial. As in the case of  $\text{var}[\Delta y_t]$ , the bias due to white noise error is consistently larger than the bias due to the underreporting of transitory earnings (in absolute terms). As a consequence, the overall bias is negative and substantial.

At below 0.01 in absolute terms, the values for  $\text{cov}[\Delta y_t, \Delta y_{t-1}]$  reported in Table 10 are relatively small. They are also of a similar order of magnitude to comparable values reported by Pischke (1995) who also documents negative values at or below 0.01 in absolute terms. As a result of the relatively substantial biases and the relatively modest values for  $\text{cov}[\Delta y_t, \Delta y_{t-1}]$ , reliability ratios for the estimated first-order autocovariance of earnings

<sup>13</sup> Pischke (1995) is only able to estimate the variance of the white noise error in two specific years. Therefore, he has to assume that the variance of the white noise component is the same in two adjacent periods,  $\sigma_{\xi_t}^2 = \sigma_{\xi_{t-1}}^2$ . In contrast, we are able to relax this simplifying assumption.

**Table 11** Implied estimates of the innovation variances of earnings growth

Wave	$\text{var}[\Delta x_t]$	$\text{cov}[\Delta x_t, \Delta x_{t-1}]$	$\text{var}[\eta_t]^{\text{biased}}$	$\text{var}[\epsilon_t]^{\text{biased}}$
2	0.0613	-0.0266	0.0266	0.0080
3	0.0639	-0.0255	0.0255	0.0129
4	0.0628	-0.0280	0.0280	0.0068

The table depicts estimates of the variance and first-order autocovariance of earnings changes as well as estimates of the (biased) variance of the transitory and permanent component of earnings innovations obtained purely from survey-based data; top-coded earnings have been replaced by imputed earnings

growth tend to be rather low, ranging from 0.18 to 0.33, which is lower than the reliability ratios for the estimated variance of earnings growth. They are also lower than those documented by Pischke (1995) who finds a ratio of 0.49 for the first year of his survey and a ratio of 0.40 for the second one.

Once again following Pischke (1995), Table 11 presents results for the estimation of the variance and first-order autocovariance of earnings changes under the assumption that only PASS survey data potentially contaminated by measurement error are available. In addition to that, the table also depicts results for a decomposition of the variance of earnings changes into a permanent component  $\eta_t$  and a transitory component  $\epsilon_t$ .

Estimates of the variance and first-order autocovariance of earnings changes ( $\text{var}[\Delta x_t]$  and  $\text{cov}[\Delta x_t, \Delta x_{t-1}]$ , respectively) can be derived from Eq. (12) and are presented in the first and second column of Table 11. The table's third and fourth columns present the estimates for the variance of the transitory and permanent component of earnings innovations obtained purely from the survey-based data ( $\text{var}[\eta_t]^{\text{biased}}$  and  $\text{var}[\epsilon_t]^{\text{biased}}$ ). To differentiate these variances from the relevant variances obtained from estimating the true process of earnings dynamics, they are labelled by the superscript "biased". The biased variance of the transitory component  $\text{var}[\eta_t]^{\text{biased}}$  is identified from  $\text{cov}[\Delta x_t, \Delta x_{t-1}]$ . Given an estimate for  $\text{var}[\eta_t]^{\text{biased}}$ , the variance of the biased permanent component  $\text{var}[\epsilon_t]^{\text{biased}}$  can be identified from  $\text{var}[\Delta x_t]$ :

$$\begin{aligned} \text{var}[\eta_t^{\text{biased}}] &= -\text{cov}[\Delta x_t, \Delta x_{t-1}]; \\ \text{var}[\epsilon_t^{\text{biased}}] &= \text{var}[\Delta x_t] - 2\text{var}[\eta_t^{\text{biased}}]. \end{aligned} \quad (13)$$

It is illuminating to compare the biased estimates for the variance of the transitory and permanent component of earnings innovations obtained purely from the survey-based data as depicted in Table 11 with the true estimates for the same parameters as summarized in Table 7. This comparison gives an indication whether according to the estimates that rely purely on survey-based earnings the

variances of the transitory and permanent component of earnings innovations are over- or underestimated in the PASS survey.

According to the purely survey-based earnings estimates depicted in Table 11, the estimates of  $\text{var}[\eta_t]^{\text{biased}}$  range from 0.026 to 0.028. According to the GMM estimates of earnings dynamics and measurement error summarized in Table 7 and already discussed above, the estimates of  $\text{var}[\eta_t]$  lie between 0.005 and 0.011. Thus, survey-based earnings data generally overstate the variance of the transitory component of earnings innovations. The opposite is generally the case with regard to the permanent component. According to survey-based earnings data, the estimates of  $\text{var}[\epsilon_t]^{\text{biased}}$  are 0.007, 0.008 and 0.013. In contrast, the true estimate of the same parameter  $\text{var}[\epsilon_t] = \hat{\sigma}_\epsilon^2$  is 0.013. Thus, similar to what is also found by Pischke (1995), if inferences are made on the basis of survey data with measurement error the transitory component in earnings innovations will be relatively overestimated as compared to the permanent component.<sup>14</sup>

Somewhat in contrast to the relatively encouraging findings by Pischke (1995), our dynamic model of measurement error and earnings presents at best mixed news for the estimation of earnings dynamics with survey data. On the one hand, the negative correlation of measurement error with transitory earnings attenuates the role of white-noise measurement error. On the other hand, reliability ratios are generally low and variance estimates of earnings growth from survey data overstate the true variance. In addition, a decomposition of survey-based earnings data into a permanent and a transitory component overstates the variance of the transitory component of earnings innovations. In our reading, these results suggest caution about using the PASS survey to analyze earnings dynamics.

## 7 Conclusions

In this paper, we replicate and extend the analyses of the extent of measurement error in longitudinal earnings data by Bound and Krueger (1991) (BK) and Pischke (1995) with German linked administrative and survey data. As compared to BK and Pischke (1995), our findings provide more recent evidence on measurement error in survey data. They also provide first evidence from Germany that makes it possible to put the conclusions by BK and Pischke (1995) into an international perspective. We also make the following methodological contributions:

<sup>14</sup> While Pischke (1995) also documents that survey-based earnings data generally overstate the variance of the transitory component of earnings innovations, he finds very similar values for  $\text{var}[\epsilon_t]^{\text{biased}}$  and  $\text{var}[\epsilon_t]$ .



(a) we extend the simple model of measurement error suggested by BK from the two-period to the four-period case, (b) we introduce the methodology for taking account of top-coded administrative earnings information developed by Card et al. (2013) to the measurement error literature and (c) we develop a procedure for merging the correct administrative records to the different waves of the PASS survey (and in a robustness exercise also to the WeLL survey).

Qualitatively, we confirm many of the main results of BK and Pischke (1995). Most crucially, we confirm that in a cross section survey-based earnings data are relatively reliable but that their reliability tends to be much lower when the data are specified in first differences. Other findings include (a) that measurement error in our survey data is not classical but can be more adequately described as mean-reverting (in the sense that measurement errors are serially correlated over 2 years and negatively correlated with true earnings), (b) that the mismeasurement of earnings leads to little bias when survey-based earnings are on the left-hand side of a typical Mincer-type earnings regression, (c) that measurement error in survey-based earnings data is positively correlated from one year to the next and even over an extended period of time, (d) that survey respondents underreport innovations to the transitory component of earnings (i.e., those resulting from transitory events like overtime and temporary layoffs) and (e) that variance estimates of earnings growth from error-ridden data overstate the true variance. Our results are robust to various sensitivity checks and arguably exhibit a high degree of external validity as demonstrated through the use of a second linked survey-administrative data set with a focus on a different part of the earnings distribution.

In order to compare our findings for Germany directly with those already available for the United States, our models impose the same assumptions as BK and Pischke (1995). This approach comes with certain drawbacks. First, BK and Pischke (1995) observe only two periods of survey-based earnings data and therefore need to make relatively specific assumptions regarding measurement error processes and other model features. With data spanning more than two periods, it would in principle be possible to relax some of these assumptions. Second, BK

and Pischke (1995) assume that administrative earnings information is free of measurement error and represents true earnings. This is a commonly made assumption. However, it has been challenged by a number of studies, including Kapteyn and Ypma (2007), Abowd and Stinson (2013) and Oberski et al. (2017) who assess the relative reliability of administrative and survey-based earnings data and argue that while reliability is generally higher for administrative data (at least for full-time workers) neither data source represents true earnings. More specifically, Abowd and Stinson (2013) argue that administrative earnings might not always represent true earnings due to filing or processing errors. While the earnings information in the ADIAB data is considered highly reliable, such errors can arguably never be completely ruled out.<sup>15</sup> The extension of our analysis to models with more general measurement error processes or that allow for measurement error in both survey and administrative data is left for further research.

Further research may also be useful to determine whether our results are generalizable beyond the period covered (2006 to 2010). Germany was hit by the Global Financial Crisis in 2009, which might affect results for the last two periods of our analysis. While our results appear robust across the entire observation period, follow-up research could potentially more explicitly analyze measurement error in earnings data in crisis as compared to non-crisis periods. Finally, while we confirm many of the main results of BK and Pischke (1995) there are also some noteworthy disparities, for instance regarding the role of individual fixed effects in explaining measurement error. An analysis of possible reasons underlying these disparities (e.g. cultural or institutional differences between the United States and Germany) would constitute another promising avenue for follow-up research.

In spite of these caveats and possibilities for extensions, our results leave us confident to encourage empirical analyses using German survey-based earnings data in levels but somewhat more cautious about using the same data in first differences. They also open the door for more research on understanding the precise nature and properties of measurement error in longitudinal earnings data, on the quality of German administrative and survey-based earnings information and on whether findings regarding labor market characteristics and impacts of economic policy depend on using either administrative or survey-based earnings data.

<sup>15</sup> Other studies that challenge the assumption that administrative earnings information is free of measurement error include Meijer et al. (2012) and Jenkins and Rios-Avila (2023) which propose generalized versions of the model by Kapteyn and Ypma (2007) and apply them to Swedish and British data, respectively. They show that administrative earnings data perform relatively poorly if there is even a small probability of a mismatch in the process of linking data from different sources. Bollinger et al. (2018) and Hyslop and Townsend (2020) also introduce models that allow for measurement error in both administrative and survey-based earnings.

## Appendix: Measurement error in WeLL earnings data

### Data, sampling and summary statistics

This appendix replicates all estimation results of this paper using a second linked survey-administrative data set that combines the “WeLL” (“Continuing Education and Lifelong Learning”/“Weiterbildung und Lebenslanges Lernen”) household panel with the ADIAB administrative data. The main objective of this replication exercise is to investigate the external validity of our results across the earnings distribution, given that PASS focuses primarily on low-income households. To collect the WeLL survey data, a two-stage procedure was implemented. First, a sample of 149 establishments was drawn from the IAB Establishment Panel, which in turn is a random but stratified sample of all German establishments.<sup>16</sup> Second, a random but stratified sample was drawn from all employees in these 149 establishments. WeLL is a panel that covers both East and West Germany. The survey was implemented during four consecutive years, from 2007 to 2010, resulting in four waves. Individuals are tracked even if they leave the establishment initially sampled for WeLL. Survey instruments cover information on individuals’ socio-demographic characteristics, labor market biographies, training experience and other topics.<sup>17</sup>

The basic process for linking the WeLL and ADIAB data is the same as the one for linking the PASS and ADIAB data. Therefore, this appendix focuses on describing some additional complications specific to WeLL. For instance, deriving comparable information on earnings from the WeLL and the ADIAB data is particularly challenging because WeLL survey respondents were asked to report the *net* earnings in the *current month* of their *main job*. To address this challenge, we rely on a three-step procedure to merge the correct ADIAB records to the WeLL survey. First, we identify the ADIAB spells covering the month the respective WeLL survey wave refers to and for this month calculate average daily gross earnings in the ADIAB data. Second, we calculate monthly gross earnings by multiplying average daily gross earnings

with  $(7/12 * 31 + 4/12 * 30 + 1/12 * 28.75)$ . Third, we convert monthly net earnings in WeLL to monthly gross earnings using an empirical model of the German income tax system applicable to labor earnings. This appendix provides more details on the third step of the procedure, that is, the conversion of monthly net earnings to monthly gross earnings.

As detailed for example in Beninger (2010), Germany has an income tax system administered at the federal level and regulated through the Personal Income Tax Law (“Einkommensteuergesetz”). Income tax covers labor earnings together with other income such as investment income. In addition to income tax in the narrow sense of the word, any attempt to convert net earnings to gross earnings also needs to be able to account for some additional taxes or pseudo-taxes on labor earnings, such as (a) the “supplemental solidarity tax for the reconstruction of East Germany” (“Solidaritätszuschlag”) mandatory for all income tax payers, (b) the “church tax” (“Kirchensteuer”) directly collected by the taxation authorities for registered Roman Catholics and Protestants and members of some other congregations and (c) compulsory social security contributions levied up to the relevant contribution assessment ceiling.

In the German income tax system, the effective tax rate and tax burden depend on a number of distinct variables on the individual and household level. More specifically, these are (a) the amount of taxable income, (b) the household composition including the number of dependent children in the household, (c) the age of household members, (d) the location (for the period considered here special rules apply to East Germany as a whole and the states of Saxony in East Germany and Bavaria and Baden-Wuerttemberg in West Germany), (e) church membership and (f) household members’ “tax classes”.

Tax classes are a central feature of the German income tax system. There are six different classes depending on the household composition and tax payers’ choices. In simplified form, tax class I covers households with one adult and no dependent children, tax class II households with one adult and at least one dependent child, and tax classes III to V married couples. Tax class VI is relatively rarely used, for instance in the case of parallel employment spells. If only one spouse in a married couple works, tax class III applies. However, if both spouses are active on the labor market, married couples can choose either to be both taxed under tax class IV or to have a combination of tax classes III and V. In the first system, both spouses pay taxes as if they were the only adult with labor income in the household (which usually results in an overpayment of taxes and a refund at the end of the year). In the second system, the household member with the highest earnings pays income tax as if he/she was the

<sup>16</sup> For the IAB Establishment Panel, the universe of all establishments is stratified according to establishment size, industry and federal state. For WeLL, all establishments in the IAB Establishment Panel are stratified according to establishment size (100 to 199 employees, 200 to 499 employees and 500 to 1,999 employees), industry (services and manufacturing) and location (East and West Germany).

<sup>17</sup> Cf. Knerr et al. (2012) for more information on the WeLL survey while Schmucker et al. (2014) provide a detailed description of standardized linked WeLL-ADIAB data. Similar to the linked PASS-ADIAB data, the standardized WeLL-ADIAB data differ slightly from the linked survey-administrative data used here. For instance, in contrast to the standardized WeLL-ADIAB version we can make use of precise information on individuals’ earnings.

only earner in the couple while the other spouse pays the difference between the total tax burden and the amount paid by the main earner. Usually, the combination of tax class III and V results in a lower tax burden if there are large differences in earnings between spouses whereas having both spouses taxed under tax class IV is beneficial if their earnings are very similar.

Most of the information required for a conversion of net to gross earnings is present in our linked survey-administrative data. Importantly, the WeLL data make it possible to use information on the household composition and other variables that are not present in the administrative ADIAB data. Conversely, a conversion of net earnings to gross earnings (or *vice versa*) purely based on German administrative labor market data would have been more cumbersome and/or dependent on relatively restrictive assumptions.<sup>18</sup>

Nevertheless, some assumptions are necessary. In particular, neither the WeLL nor the ADIAB data contain direct data on tax classes. While in many cases the appropriate tax class can be inferred with high certainty from information on the household context available in WeLL, it is not as straightforward to know whether a specific married couple with both spouses active on the labor market chooses to have both partners taxed under tax class IV or instead through a combination of tax classes III and V. In these instances, we assume that the spouse earning more than half of the total household labor income always chooses tax class III, and the other one always tax class V.

Besides, we do not have data on church membership and some information is not gathered in all survey wave. Since during the observation period the majority of Germany's population belonged to one of the relevant churches, we simply assume that all workers pay church taxes. In addition, we assume that tax-relevant individual characteristics do not change between survey waves. As a consequence, if some information is not gathered (or missing) in any particular survey wave it can be imputed from information available in another wave.

In order to practically implement the conversion of net earnings to gross earnings, we rely on the detailed program flow charts ("Programmablaufpläne") for this conversion published each year by the German Federal Ministry of Finance. The flow charts are meant to assist employers' efforts to correctly and automatically deduct income tax payments from workers' gross earnings and transfer them to the tax authorities. They are publicly available on the official website [www.bundesfinanzministerium.de](http://www.bundesfinanzministerium.de) while the privately-run website [www.permentier.de](http://www.permentier.de) contains useful pseudo code that makes it straightforward to implement the procedure described in the flow charts in various different programming languages. Here, we use information published on both websites to translate the flow charts for the years 2007 to 2010 into Stata code. As the flow charts describe the conversion of gross earnings to net earnings while we instead want to convert net earnings to gross earnings we implement a simple iterative algorithm. For each observation, this algorithm (a) calculates a plausible value for gross earnings as a starting point, (b) plugs this value into our representation of the flow chart, (c) adjusts the value of gross earnings based on the value of net earnings calculated by the flow chart and (d) repeats steps (b) and (c) until the difference between the value of net earnings calculated by the flow chart and actual net earnings becomes negligible.<sup>19</sup>

Using the procedure to convert net earnings to gross earnings described in this appendix, we are able to successfully link 17,407 spells with information on earnings from the administrative ADIAB data with corresponding survey-based information on earnings from WeLL. Then, the same four discrete data cleansing steps as with PASS—excluding data points where earnings information from WeLL is unlikely to be directly comparable to earnings as recorded in the ADIAB data, excluding cases where respondents are only willing to give their earnings in terms of a broad range instead, excluding cases where respondents indicate that they are self-employed in addition to or instead of being in dependent employment

(See figure on next page.)

**Fig. 4** Program flow charts for converting gross earnings to net earnings. The figure shows the program flow chart ("Programmablaufplan") for converting gross earnings to net earnings in 2010 as published by the German Federal Ministry of Finance

<sup>18</sup> Cf. Junge (2017) for a discussion of how to convert gross earnings to net earnings purely with German administrative labor market data and a comparison of different ways to perform this conversion. More specific suggestions are made by Gunselmann (2014). Cf. also Reichert (2014) for a further application.

<sup>19</sup> Cf. Fig. 4 for the flow chart for the year 2010 as published on the website [www.bundesfinanzministerium.de](http://www.bundesfinanzministerium.de). For most years, the German Federal Ministry of Finance also publishes an annex to the flow charts with tables that make it possible to verify that the relevant procedures have been implemented in a correct fashion. On a spot-check basis, we use these tables to verify that our translation of the program flow charts into Stata is valid.

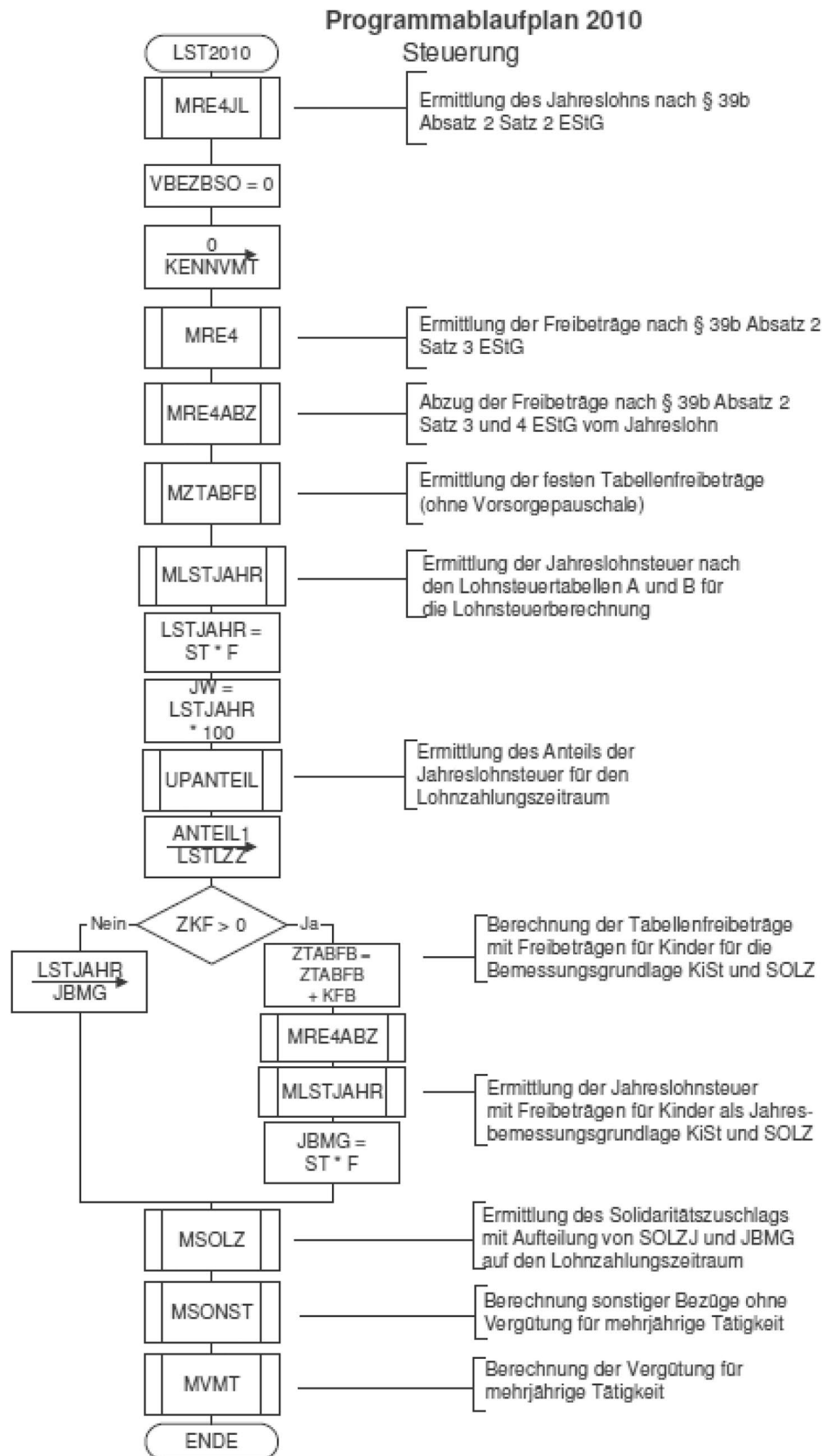
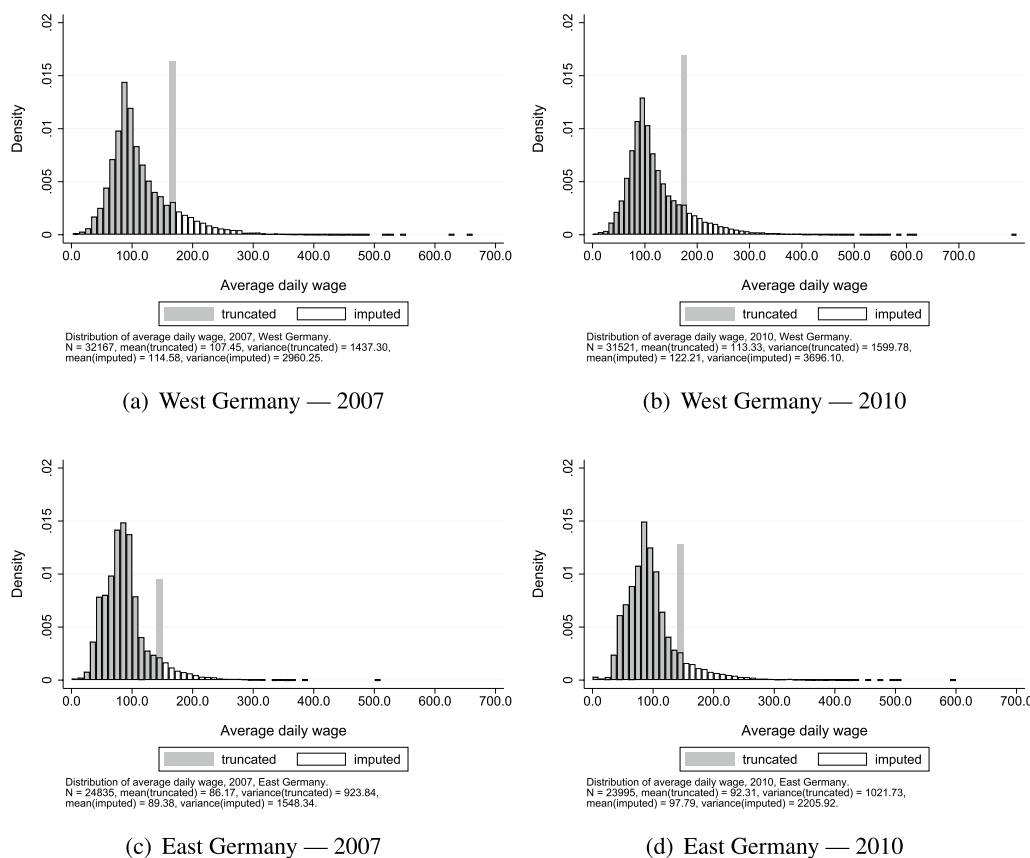


Fig. 4 (See legend on previous page.)



**Fig. 5** Distribution of daily earnings by location and year according to administrative data. The figure shows the distribution of daily earnings in East and West Germany for 2007 and 2010 for both the full estimation sample linking WeLL and ADIAB data with the original—top-coded—earnings information and for the estimation sample with top-coded earnings information replaced by imputed earnings; all information is based on administrative ADIAB data

and excluding occupations and sectors with untypical pay structures—reduce the size of the estimation sample from 17,407 to 14,220 observations.

The top-coding of administrative earnings information presents particular challenges in the case of the linked WeLL-ADIAB data. For these data, top-coding is present for 1,899 out of 14,220 observations or 13.4 percent of the entire sample. In contrast, due to the oversampling of low-income earners in PASS, in the PASS estimation sample top-coding is much less prevalent. Given the prevailing gender wage gap in Germany, top-coding is more widespread for men than for women. In the WeLL estimation sample it is present for 18.9 percent of observations for men and 3.9 percent of observations for women.

Formally, top-coding can be described as follows: Instead of observing a worker’s true average daily earnings  $D^*$ , in the ADIAB data we observe  $D$  and it holds that  $D = D^*$  if and only if  $D < C_{t,r}$  while  $D = C_{t,r}$  if  $D \geq C_{t,r}$ . In this formal setting,  $C_{t,r}$  denotes the social security contribution assessment ceiling which varies by

year  $t$  (2007 to 2010) and location  $r$  (East Germany and West Germany).

To impute top-coded earnings in our linked WeLL and ADIAB data, we implement the two step-procedure proposed by Gartner (2005). First, for the sample of non-top-coded earnings we run a number of Tobit regressions of log earnings on the following regressors: (a) an intercept, (b) a gender dummy, (c) a dummy for German nationality, (d) a third order polynomial of age, (e) four dummies for the occupational status (specialist workers, master craftsmen, white-collar workers, blue-collar workers), (f) six dummies for the education status (no school-leaving qualification, primary/lower secondary school qualification, school leaving certificate from an academic high school (“Abitur”) without vocational training degree, Abitur with vocational training degree, technical college qualification, university degree), (g) nine dummies for the occupational status on the one-digit level, (h) 16 dummies for the federal state of the individual’s workplace, (i) ten dummies for his or her sector of employment and (j)



**Table 12** Summary statistics by survey data set, gender and sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	WeLL				PASS			
	Men		Women		Men		Women	
	Raw data	Sample	Raw data	Sample	Raw data	Sample	Raw data	Sample
"Abitur"	0.26 (0.00)	0.24 (0.00)	0.29 (0.01)	0.29 (0.01)	0.22 (0.01)	0.21 (0.01)	0.22 (0.01)	0.24 (0.01)
German	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.95 (0.00)	0.95 (0.00)	0.96 (0.00)	0.97 (0.00)
Not married	0.26 (0.00)	0.27 (0.01)	0.33 (0.01)	0.34 (0.01)	0.38 (0.01)	0.39 (0.01)	0.48 (0.01)	0.47 (0.01)
East German	0.38 (0.01)	0.38 (0.01)	0.40 (0.01)	0.41 (0.01)	0.26 (0.01)	0.25 (0.01)	0.28 (0.01)	0.28 (0.01)
Survey earnings	3681.08 (21.99)	3517.53 (18.66)	2426.64 (16.22)	2466.94 (17.18)	2700.39 (28.64)	2678.45 (22.90)	1622.45 (12.72)	1708.25 (14.40)
Admin earnings	3885.47 (18.14)	3816.06 (19.27)	2443.09 (15.99)	2517.10 (16.68)	2773.63 (21.59)	2838.41 (23.07)	1742.72 (13.90)	1857.61 (15.61)
Age	45.94 (0.10)	45.85 (0.11)	44.72 (0.12)	44.44 (0.14)	41.38 (0.14)	41.11 (0.16)	42.14 (0.13)	41.91 (0.15)
Weeks employed	48.88 (0.09)	48.97 (0.09)	48.85 (0.11)	49.04 (0.12)	47.11 (0.14)	47.25 (0.15)	47.39 (0.13)	47.63 (0.15)
N	10,800	8927	6607	5293	5721	4720	5854	4534

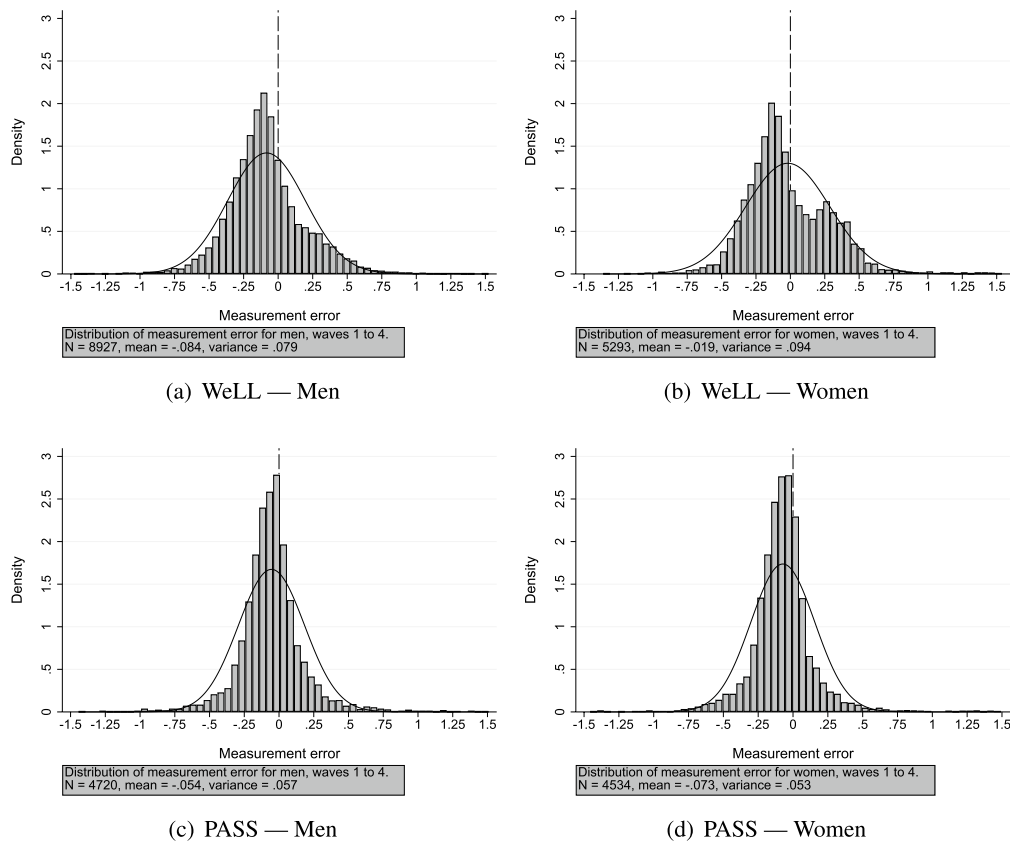
The table shows characteristics of men and women for the raw data without the data cleansing steps implemented for selecting the estimation sample ("raw data") and for the estimation samples ("sample"); robust standard errors are in parentheses

four dummies for the size of the employing establishment (one, two to 199, 200 to 999, 1000 or more employees). These Tobit regressions are run separately by calendar year and for East and West Germany. Given that our sample runs from 2007 to 2010 and that we have data for both East and West Germany for each year, we run a total of eight distinct Tobit regressions. Second, we use the fitted values from the Tobit regressions together with estimates of the residual variance to impute earnings in the case of top-coding.

Figure 5 shows four histograms for the distribution of average daily earnings for our linked WeLL and ADIAB data. Panel (a) shows the earnings distribution for West Germany in 2007, Panel (b) for West Germany in 2010, Panel (c) for East Germany in 2007 and Panel (d) for East Germany in 2010. Given that top-coding is much less prevalent for the linked PASS-ADIAB data as compared to the linked WeLL-IAB, similar histograms for the PASS-ADIAB data are not displayed. However, they are available upon request as are similar histograms for 2008 and 2009. The four panels show the distributions of daily earnings for both the full estimation sample with the original—top-coded—earnings information and for the estimation sample with top-coded earnings information replaced by imputed earnings. All earnings information is based on administrative ADIAB data.

As is evident from Fig. 5, across the two time periods and two locations the histograms look qualitatively relatively similar, though average earnings are somewhat higher in West Germany as compared to East Germany and in 2010 as compared to 2007. Notably, all distributions of daily earnings for the full estimation sample with the top-coded earnings information show very marked spikes at or around the social security contribution assessment ceilings. In contrast, the distributions for the estimation sample with top-coded earnings information replaced by imputed earnings show no such spikes. Instead, the respective distributions are generally much smoother and suggest that the imputation procedure results in plausible distributions of daily earnings.

Table 12 replicates Table 1 but adds information on WeLL, in addition to PASS. A comparison of the two estimation samples shows that the average WeLL respondent is comparatively older, better educated, higher-earning and more likely to be married. At about 40 percent, the percentage among WeLL respondents residing in East Germany is also significantly higher than among PASS respondents and among the German population as a whole. This can be explained by the sampling procedure for WeLL which in its first step oversamples East German establishments to make separate analyses for East and West Germany possible.



**Fig. 6** Distribution of measurement error by survey and gender The figure shows the distribution of measurement error for men and women in both the WeLL and the PASS surveys; top-coded earnings have been replaced by imputed earnings; measurement error  $m_i$  is defined as the log difference in earnings between survey-based WeLL or PASS earnings and administrative ADIAB earnings; also depicted is the density of a normal distribution as are vertical lines for zero measurement error

Table 12 confirms that mean average monthly gross earnings differ only relatively little between administrative records and survey responses but that there is a certain tendency for average administrative earnings to be larger than average survey earnings. Qualitatively similar results are also reported by Antoni et al. (2019), Valet et al. (2019) and Gauly et al. (2020) who link the ADIAB data with yet three more household data sets, the “National Educational Panel Study”/“Nationales

Bildungspanel”, the “Legitimation of Inequality Over the Life-Span (LINOS)” survey and the German sample of the OECD’s “Programme for the International Assessment of Adult Competencies (PIAAC),” respectively. For instance, according to Valet et al. (2019) respondents to the LINOS survey on average underreported their gross earnings by about 5 percent.

**Table 13** Individual characteristics associated with measurement error and survey-based earnings by data set and gender

Regression of	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	WeLL				PASS			
	Measurement error ( $m_t$ )		Survey-based earnings ( $x_t$ )		Measurement error ( $m_t$ )		Survey-based earnings ( $x_t$ )	
	Men	Women	Men	Women	Men	Women	Men	Women
"Abitur"	0.0097 (0.0075)	-0.0003 (0.0091)	0.3444 (0.0104)	0.229 (0.017)	0.0025 (0.0086)	-0.0004 (0.0081)	0.3380 (0.0195)	0.2719 (0.0206)
German	-0.0047 (0.0164)	-0.0235 (0.0294)	0.1177 (0.03)	-0.0289 (0.0541)	-0.0054 (0.0177)	0.0246 (0.0188)	0.3348 (0.04)	0.1646 (0.0475)
Not married	0.0326 (0.0069)	-0.1326 (0.0085)	-0.101 (0.0102)	0.125 (0.0147)	0.0059 (0.0079)	-0.0057 (0.0072)	-0.2174 (0.0163)	0.0522 (0.0172)
East German	0.0232 (0.0063)	0.0217 (0.0085)	-0.2945 (0.0085)	0.1496 (0.0144)	0.0230 (0.0080)	0.0146 (0.0074)	-0.3705 (0.0167)	-0.0691 (0.0178)
Age	-0.0007 (0.0003)	-0.0022 (0.0004)	0.0082 (0.0005)	0.0065 (0.0008)	0.0001 (0.0004)	0.0006 (0.0004)	0.0146 (0.0009)	0.0099 (0.0009)
Log weeks employed	-0.0374 (0.0149)	-0.0404 (0.0215)	0.0854 (0.0153)	0.1206 (0.0268)	-0.0736 (0.0139)	-0.0571 (0.0155)	0.1818 (0.0255)	0.2126 (0.0291)
Intercept	0.078 (0.0604)	0.2918 (0.0926)	7.2993 (0.07)	6.7817 (0.1228)	0.2192 (0.0582)	0.0969 (0.0622)	6.2223 (0.1093)	5.8334 (0.1279)
F	11.62	42.33	519.62	63.15	6.54	3.94	333.63	60.49
$R^2$	0.008	0.041	0.254	0.073	0.013	0.008	0.326	0.084
N	8927	5293	8927	5293	4720	4534	4720	4534

The table shows OLS regressions of measurement error and survey-based earnings; top-coded earnings have been replaced by imputed earnings; survey-based earnings  $x_t$  are defined as log earnings in the WeLL or PASS survey data and measurement error  $m_t$  as the log differences in earnings between survey-based WeLL/PASS earnings and administrative ADIAB earnings; robust standard errors are in parentheses

### Measurement error, induced bias in survey-based earnings regressions and reliability of the data

Figure 6 replicates Fig. 3 but presents four plots of the probability density of  $m$ , calculated by subtracting log earnings as captured in both the WeLL or PASS surveys ( $x$ ) from log earnings recorded in the administrative ADIAB data ( $y$ ). The figure pools all four waves of each survey. Panel (a) shows the probability density of  $m$  for men from WeLL, Panel (b) for women from WeLL, Panel (c) for men from PASS and Panel (d) for women from PASS. The figure shows that the distributions of measurement errors in WeLL and PASS look generally similar, but that error variances are somewhat larger in the WeLL as compared to the PASS survey. Again similar to what we document for PASS, with a mean measurement error of  $-0.084$  for men and  $-0.019$  for women both male

and female survey respondents have a slight tendency to underreport their earnings. Gender differences in the extent that earnings are underreported are more pronounced in WeLL than in PASS.

Table 13 replicates Table 2 but presents results separately for the WeLL and the PASS surveys. Again, although some of the coefficients in the regressions of measurement error are statistically significant and all coefficients put together are also often jointly statistically significant, the estimated bias for each variable is generally small compared to its typical effect in an earnings equation. The only case in which the parameter estimate in the regression of measurement error is not only statistically significant but also sizable is the dummy for not being married in the WeLL survey. In this case, the corresponding estimate in the survey-based

**Table 14** Variance-covariance matrix by gender—WeLL

Panel (a): Men (n=1003)								
	$y_1$	$y_2$	$y_3$	$y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$y_1$	<b>0.169</b>	0.931	0.899	0.888	-0.323	-0.183	-0.222	-0.158
$y_2$	0.153	<b>0.159</b>	0.925	0.906	-0.204	-0.269	-0.238	-0.151
$y_3$	0.157	0.157	<b>0.181</b>	0.913	-0.194	-0.187	-0.362	-0.158
$y_4$	0.152	0.151	0.162	<b>0.174</b>	-0.183	-0.159	-0.223	-0.293
$m_1$	-0.037	-0.023	-0.023	-0.021	<b>0.079</b>	0.555	0.435	0.419
$m_2$	-0.019	-0.028	-0.021	-0.017	0.040	<b>0.067</b>	0.594	0.460
$m_3$	-0.025	-0.026	-0.042	-0.025	0.033	0.042	<b>0.075</b>	0.510
$m_4$	-0.017	-0.015	-0.017	-0.031	0.030	0.030	0.036	<b>0.066</b>

Panel (b): Women (n=494)								
	$y_1$	$y_2$	$y_3$	$y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$y_1$	<b>0.301</b>	0.953	0.928	0.889	-0.287	-0.235	-0.236	-0.236
$y_2$	0.281	<b>0.288</b>	0.968	0.923	-0.256	-0.290	-0.273	-0.248
$y_3$	0.275	0.281	<b>0.292</b>	0.948	-0.262	-0.265	-0.307	-0.275
$y_4$	0.264	0.268	0.277	<b>0.293</b>	-0.252	-0.274	-0.287	-0.320
$m_1$	-0.048	-0.042	-0.043	-0.042	<b>0.094</b>	0.481	0.517	0.444
$m_2$	-0.039	-0.047	-0.043	-0.045	0.044	<b>0.091</b>	0.648	0.565
$m_3$	-0.036	-0.041	-0.046	-0.043	0.044	0.054	<b>0.077</b>	0.696
$m_4$	-0.032	-0.033	-0.036	-0.042	0.033	0.042	0.047	<b>0.060</b>

The table shows variance-covariance matrices of true earnings and measurement error for men and women; covariances are at and below the diagonal, correlations above the diagonal; top-coded earnings have been replaced by imputed earnings; measurement error  $m_t$  is defined as the log differences in earnings between survey-based WeLL earnings and administrative ADIAB earnings and true earnings  $y_t$  as log earnings in the administrative ADIAB data; strongly balanced panel; variances in bold

**Table 15** Reliability ratios inferred from variance-covariance matrix by survey, gender and type of measurement error

Wave	(1) WeLL			(4) PASS		
	All	Men	Women	All	Men	Women
Panel (a): Classical measurement error						
1	0.755	0.683	0.762	0.888	0.864	0.858
2	0.766	0.704	0.761	0.882	0.874	0.838
3	0.774	0.708	0.791	0.878	0.845	0.856
4	0.802	0.726	0.831	0.893	0.835	0.905
2-1	0.245	0.259	0.223	0.407	0.484	0.345
3-2	0.290	0.314	0.239	0.366	0.333	0.382
4-3	0.339	0.311	0.419	0.289	0.255	0.327
Panel (b): Mean-reverting measurement error						
1	0.858	0.761	0.846	0.940	0.931	0.912
2	0.859	0.770	0.846	0.924	0.945	0.900
3	0.888	0.810	0.887	0.930	0.935	0.913
4	0.896	0.806	0.935	0.984	0.934	0.996
2-1	0.076	0.001	0.142	0.376	0.474	0.322
3-2	0.115	0.080	0.160	0.290	0.095	0.342
4-3	0.167	0.008	0.392	0.182	0.110	0.249
n	1497	1003	494	518	254	264

The table reports reliability ratios  $\lambda$  derived from the variance-covariance matrix of true earnings  $y$  and measurement error  $m$ ; top-coded earnings have been replaced by imputed earnings; Panel (a) imposes classical measurement error assumptions while Panel (b) allows for mean-reverting measurement error; in either case, serial correlation in  $m$  is allowed; strongly balanced panel

**Table 16** GMM estimates of earnings dynamics and measurement error

<b>Panel (a): WeLL (n = 1497)</b>					
Estimates of earnings dynamics					
$\hat{\sigma}_{\eta_1}^2$	$\hat{\sigma}_{\eta_2}^2$	$\hat{\sigma}_{\eta_3}^2$	$\hat{\sigma}_{\eta_4}^2$	$\hat{\sigma}_{\epsilon}^2$	
0.0096	0.0053	0.0099	0.0101	0.0097	
(0.00001)	(0.00001)	(0.00000)	(0.00001)	(0.00002)	
Estimates of measurement error process					
$\hat{\sigma}_{\xi_1}^2$	$\hat{\sigma}_{\xi_2}^2$	$\hat{\sigma}_{\xi_3}^2$	$\hat{\sigma}_{\xi_4}^2$	$\hat{\sigma}_{\mu}^2$	$\hat{\alpha}$
0.0318	0.0290	0.0224	0.0095	0.0391	-1.2251
(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0005)
<b>Panel (b): PASS (n = 518)</b>					
Estimates of earnings dynamics					
$\hat{\sigma}_{\eta_1}^2$	$\hat{\sigma}_{\eta_2}^2$	$\hat{\sigma}_{\eta_3}^2$	$\hat{\sigma}_{\eta_4}^2$	$\hat{\sigma}_{\epsilon}^2$	
0.0114	0.0083	0.0052	0.0054	0.0133	
(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	
Estimates of measurement error process					
$\hat{\sigma}_{\xi_1}^2$	$\hat{\sigma}_{\xi_2}^2$	$\hat{\sigma}_{\xi_3}^2$	$\hat{\sigma}_{\xi_4}^2$	$\hat{\sigma}_{\mu}^2$	$\hat{\alpha}$
0.0232	0.0230	0.0264	0.0221	0.0128	-0.6974
(0.0001)	(0.0001)	(0.00004)	(0.00002)	(0.00003)	(0.0099)

The table reports GMM estimates of earnings dynamics and measurement error; top-coded earnings have been replaced by imputed earnings; standard errors are in parentheses

**Table 17** Estimated and fitted variance-covariance matrices—WeLL

<b>Panel (a): Estimated variance-covariance matrix (n = 1497)</b>							
	$\Delta y_2$	$\Delta y_3$	$\Delta y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$\Delta y_2$	0.0244						
$\Delta y_3$	-0.0050	0.0238					
$\Delta y_4$	-0.0012	-0.0094	0.0309				
$m_1$	0.0120	-0.0003	0.0016	0.0854			
$m_2$	-0.0079	0.0063	0.0018	0.0432	0.0761		
$m_3$	-0.0020	-0.0124	0.0121	0.0381	0.0471	0.0764	
$m_4$	0.0006	-0.0023	-0.0114	0.0318	0.0347	0.0400	0.0638
<b>Panel (b): Fitted variance-covariance matrix (n = 1497)</b>							
	$\Delta y_2$	$\Delta y_3$	$\Delta y_4$	$m_1$	$m_2$	$m_3$	$m_4$
$\Delta y_2$	0.0246						
$\Delta y_3$	-0.0053	0.0249					
$\Delta y_4$	0	-0.0099	0.0297				
$m_1$	0.0118	0	0	0.0854			
$m_2$	-0.0065	0.0065	0	0.0391	0.0761		
$m_3$	0	-0.0121	0.0121	0.0391	0.0391	0.0764	
$m_4$	0	0	-0.0124	0.0391	0.0391	0.0391	0.0638

Panel (a) reports the variance-covariance matrix estimated from the data; Table (b) depicts the fitted variance-covariance matrix from the model; top-coded earnings have been replaced by imputed earnings; entries marked "0" are constrained to be exactly zero



**Table 18** Bias in estimates of variance and first-order autocovariance of earnings growth—WeLL

Panel (a): Variance of earnings growth (n = 1497)					
Wave	Underreporting of transitory earnings	White-noise error	Total bias	var[ $\Delta y_t$ ]	Reliability ratios
2	-0.0142	0.0608	0.0466	0.0246	0.3454
3	-0.0144	0.0514	0.0370	0.0249	0.4020
4	-0.0190	0.0319	0.0128	0.0297	0.6984
Panel (b): First-order autocovariance of earnings growth (n = 1497)					
Wave	Underreporting of transitory earnings	White-noise error	Total bias	cov[ $\Delta y_t, \Delta y_{t-1}$ ]	Reliability ratios
2	0.0118	-0.0318	-0.0199		
3	0.0065	-0.0290	-0.0225	-0.0053	0.1904
4	0.0121	-0.0224	-0.0102	-0.0099	0.4917

The table summarizes the biases in estimates of the variance and first-order autocovariance of earnings growth that result from the underreporting of transitory earnings and white noise error for the WeLL survey; top-coded earnings have been replaced by imputed earnings

earnings regression appears substantially biased towards zero (under the assumption that the model is correctly specified, of course). Similarly, Gauly et al. (2020, p. 33) document with linked ADIAB-PIAAC data that due to nonclassical measurement error Mincer wage regressions with survey-based instead of true earnings data might lead to some biased results but that, overall, there are “more similarities than differences between the estimated coefficients.”

Replicating Table 3, Panel (a) of Table 14 reports the estimated variance-covariance matrix of true earnings  $y$  and measurement errors  $m$  for men from the WeLL survey. Panel (b) of the same table contains the corresponding estimates for women. As in Table 3, covariances are at and below the diagonal and correlations above the diagonal. The table confirms that true earnings exhibit a high degree of first- and higher-order autocorrelation, that there are large negative correlations between measurement error and true earnings, that measurement errors in the earnings information available in survey data are positively correlated from 1 year to the next and even over an extended period of time and that higher order autocorrelations of  $m$  tend to be pronounced.

Table 15 replicates Table 4 but summarizes reliability ratios for both WeLL and PASS earnings data. The table confirms that in a cross section longitudinal earnings data are relatively reliable. Under the assumption of classical measurement error, the reliability of WeLL earnings data in the cross section is around 0.7 or slightly higher for men and around 0.8 or slightly lower for women. If we allow for mean-reverting measurement error this further increases the estimates of cross-sectional reliability.

Also according to Table 15, our results for reliability ratios in first differences are even less encouraging for

WeLL than for PASS earnings data, especially for men. When we allow for mean-reverting measurement error we document some point estimates for reliability ratios in first differences that range from basically 0 to 0.39 for WeLL as compared to 0.10 to 0.47 for PASS. Under the assumption of classical measurement error, estimates range from 0.22 to 0.42 for WeLL and from 0.26 to 0.48 for PASS.

In tables not reported here but available upon request we also summarize results for WeLL for the alternative procedure to estimate the reliability ratios in levels and first differences that regresses the ADIAB earnings variable on survey-based earnings information depicted in Table 15 and replicate the sensitivity checks summarized in Table 6 for WeLL earnings data. Reassuringly, according also to WeLL data results from the regressions of the ADIAB earnings on survey-based earnings information are qualitatively and for the most part also quantitatively very similar to those from the more indirect procedure. Similarly, although the precise estimates for reliability ratios in levels or first differences differ slightly between the two data sets as well as across the different estimation samples, the general picture and key findings continue to stay very robust across various sensitivity checks that rely on WeLL data.

### Measurement error and earnings dynamics

Estimation results for the replication of the dynamic model of measurement error of Sect. 6 are summarized in Table 16. Panel (a) lists estimation results for the WeLL survey and Panel (b) for PASS. In general, most parameters are precisely estimated, appear sensible and are qualitatively similar for the WeLL and PASS surveys.

In order to investigate the robustness of results regarding the fit and potential shortcomings of the dynamic model of measurement error, Table 8 uses the PASS data to report both the actual covariance matrix of earnings dynamics and measurement error observed in the data and a fitted covariance matrix that accounts for the constraints imposed by our model. Table 17 does the same for the WeLL data. The table shows that our simple model of measurement error and earnings dynamics fits the data almost as well when we draw on the WeLL survey as compared to when we rely on PASS data.

Table 18 replicates Tables 9 and 10 using the GMM estimates of earnings dynamics and measurement error from Table 16 to quantify the biases that result from the underreporting of transitory earnings and white noise error for the WeLL survey. The table also depicts the total bias in estimates of the variance and first-order autocovariance of earnings growth. In addition to that, with reference to the fitted variance-covariance matrices from Table 17, it also lists  $\text{var}[\Delta y_t]$  and the reliability ratios  $\lambda_t = \text{var}[\Delta y_t]/\text{var}[\Delta x_t]$  using the WeLL survey. While for conciseness the relevant estimates from PASS are not reproduced in Table 18, a comparison of that table with Tables 9 and 10 readily confirms that our results are robust to using the WeLL instead of the PASS data.

Finally, in results not reported here but available upon request we also replicate Table 11 with the WeLL data. This replication exercises uses WeLL data to estimate the variance and first-order autocovariance of earnings changes under the assumption that only survey data that are potentially contaminated by measurement error are available. In addition to that, the replication exercise decomposes the variance of earnings changes into a permanent effect  $\eta_t$  and a transitory effect  $\epsilon_t$ . Once again, results for the WeLL and PASS surveys are generally consistent.

#### Acknowledgements

The authors thank participants at EALE, ESPE and SOLE conferences and a seminar at the Institute for Employment Research, as well as two anonymous referees and an associate editor (Britta Gehrke) for helpful comments and suggestions. Findings, interpretations and conclusions expressed in this paper do not necessarily represent the views of the World Bank, its affiliated organizations, its Executive Directors or the governments these represent. The usual disclaimer applies

#### Author contributions

TVW and AS conceptualized the study. MU conducted the analyses and created the graphs and tables, with contributions by AS. AS wrote the manuscript, with contributions by MU. All three authors discussed the findings, and read and approved the final manuscript.

#### Funding

No funding was received.

#### Availability of data and materials

For our analyses, we use administrative and survey data of the Institute for Employment Research (IAB). The social security data with administrative origin are processed and kept by IAB, Regensburger Str. 104, D-90,478 Nuremberg,

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. If you wish to access the full data for replication purposes, please contact Matthias Umkehrer (matthias.umkehrer@iab.de).

#### Declarations

##### Competing interests

The authors declare no conflict of interest.

Received: 28 July 2023 Accepted: 13 May 2024

Published online: 04 June 2024

#### References

- Abowd, J.M., McKinney, K.L.: Earnings inequality and the role of the firm. Paper presented at the Labor and Employment Relations Association (2017)
- Abowd, J.M., Stinson, M.H.: Estimating measurement error in annual job earnings: a comparison of survey and administrative data. *Rev. Econ. Stat.* **95**, 1451–1467 (2013)
- Antoni, M., Bethmann, A.: PASS-ADIAB: Linked survey and administrative data for research on unemployment and poverty. *Jahrbücher für Nationalökonomie und Statistik* **239**, 747–756 (2019)
- Antoni, M., Bela, D., Vicari, B.: Validating earnings in the German national educational panel study: determinants of measurement accuracy of survey questions on earnings. *Methods Data Anal.* **13**, 59–90 (2019)
- Beninger, D.: The perception of the income tax: evidence from Germany. Unpublished working paper, Centre for European Economic Research (2010)
- Bollinger, C.R.: Measurement error in the current population survey: a nonparametric look. *J. Labor Econ.* **16**, 576–594 (1998)
- Bollinger, C.R., Hirsch, B.T., Hokayem, C.M., Ziliak, J.P.: The good, the bad and the ugly: measurement error, non-response and administrative mismatch in the CPS. Unpublished paper, University of Kentucky (2018)
- Bound, J., Krueger, A.B.: The extent of measurement error in longitudinal earnings data: do two wrongs make a right? *J. Labor Econ.* **9**, 1–24 (1991)
- Bound, J., Brown, C., Duncan, G.J., Rodgers, W.L.: Evidence on the validity of cross-sectional and longitudinal labor market data. *J. Labor Econ.* **12**, 345–368 (1994)
- Bound, J., Brown C., Mathiowetz, N.: Measurement error in survey data. In: Heckman, J., Leamer, E. (eds.) *Handbook of Labor Economics*, vol. 5, pp. 3707–3843. Elsevier, Amsterdam (2001)
- Caliendo, M., Fedorets, A., Preuss, M., Schröder, C., Wittbrodt, L.: The short- and medium-term distributional effects of the German minimum wage reform. *Empir. Econ.* **64**, 1149–1175 (2023)
- Card, D., Chetty, R., Feldstein, M., Saez, E.: Expanding access to administrative data for research in the United States. National Science Foundation Directorate of Social, Behavioral, and Economic Sciences White Paper 112 (2010)
- Card, D., Heining, J., Kline, P.: Workplace heterogeneity and the rise of west german wage inequality. *Q. J. Econ.* **128**, 967–1015 (2013)
- Duncan, G.J., Hill, D.H.: An investigation of the extent and consequences of measurement error in labor-economic survey data. *J. Labor Econ.* **3**, 508–532 (1985)
- Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., vom Berge, P.: Reallocation effects of the minimum wage. *Q. J. Econ.* **137**, 267–328 (2022)
- Friedman, J.: The value of public administrative data in policy-relevant economic research. Paper presented at the American Economic Association Committee on Economic Statistics (2010)
- Gartner, H.: The imputation of wages above the contribution limit with the German IAB Employment Sample. FDZ Methodenreport 2005/2 (2005)

- Gauly, B., Daikeler, J., Gummer, T., Rammstedt, B.: What's my wage again? Comparing survey and administrative data to validate earning measures. *Int. J. Soc. Res. Methodol.* **23**, 215–228 (2020)
- Gideon, M., Heggeness, M.L., Murray-Close, M., Myers, S.L. Jr.: The influence of source data on understanding human capital characteristics and the black-white earnings gap. Paper presented at the Labor and Employment Relations Association (2017)
- Gunselmann, I.: Programmierbeispiel zur Umrechnung des Brutto- in ein Netto-Tagesentgelt für die Administrativen Daten des FDZ. *FDZ Methodenreport 2014/1* (2014)
- Hyslop, D., Townsend, W.: Earnings dynamics and measurement error in matched survey and administrative data. *J. Bus. Econ. Stat.* **38**, 457–469 (2020)
- Jenkins, S., Rios-Avila, F.: Reconciling reports: modelling employment earnings and measurement errors using linked survey and administrative data. *J. Royal Stat. Soc. Ser. A: Stat. Soc.* **186**, 110–136 (2023)
- Junge, H.: From gross to net wages in German administrative data sets. *DIW Data Documentation* 89 (2017)
- Kapteyn, A., Ypma, J.Y.: Measurement error and misclassification: a comparison of survey and administrative data. *J. Labor Econ.* **25**, 513–551 (2007)
- Knerr, P., Schröder, H., Aust, F., Gilberg, R.: Weiterbildung als Bestandteil Lebenslangen Lernens: WeLL-Erhebung 2010 – Methodenreport FDZ-Methodenreport 2012/7 (2012)
- Kopczuk, W., Saez, E., Song, J.: Earnings inequality and mobility in the United States: evidence from social security data since 1937. *Q. J. Econ.* **125**, 91–128 (2010)
- Kristensen, N., Westergaard-Nielsen, N.: A large-scale validation study of measurement errors in longitudinal survey data. *J. Econ. Soc. Meas.* **32**, 65–92 (2007)
- Meijer, E., Rohwedder, S., Wansbeek, T.: Measurement error in earnings data: using a mixture model approach to combine survey and register data. *J. Bus. Econ. Stat.* **30**, 191–201 (2012)
- Mellow, W., Sider, H.: Accuracy of response in labor market surveys: evidence and implications. *J. Labor Econ.* **1**, 331–344 (1983)
- Meyer, B.D., Mittag, N.: Combining administrative and survey data to improve income measurement. In: Chun, A.Y., Larsen, M.D., Durrant, G., Reiter, J.P. (eds.) *Administrative Records for Survey Methodology*, pp. 297–322. Wiley, Hoboken (2021)
- Oberski, D.L., Kirchner, A., Eckman, S., Kreuter, F.: Evaluating the quality of survey and administrative data with generalized multitrait-multimethod models. *J. Am. Stat. Assoc.* **112**, 1477–1489 (2017)
- Pischke, J.-S.: Measurement error and earnings dynamics: some estimates from the PSID validation study. *J. Bus. Econ. Stat.* **13**, 305–314 (1995)
- Reichert, A.: Nettoeinkommen in Deutschland – Ein Vergleich zwischen Befragungsdaten und administrativen Daten des IAB. *FDZ-Methodenreport 2014/13* (2014)
- Rodgers, W.L., Brown, C., Duncan, G.J.: Errors in survey reports of earnings, hours worked, and hourly wages. *J. Am. Stat. Assoc.* **88**, 1208–1218 (1993)
- Schmucker, A., Seth, S., Eberle, J.: WeLL-Befragungsdaten verknüpft mit Administrativen Daten des IAB. *FDZ-Datenreport 2014/2* (2014)
- Slotwinski, M., Roth, A.: Gender norms and income misreporting within households. *Centre for European Economic Research Discussion Paper* 20-001 (2020)
- Stinson, M.: Estimating measurement error in SIPP annual job earnings: a comparison of census survey and SSA administrative data. *U.S. Census Bureau Technical Report TP-2002-24* (2002)
- Stüber, H., Grabka, M., Schnitzlein, D.: A tale of two data sets: comparing German administrative and survey data using wage inequality as an example. *J. Labour Market Res.* **57**, 8 (2023)
- Trappmann, M., Beste, J., Bethmann, A., Müller, G.: The PASS panel survey after six waves. *J. Labour Market Res.* **46**, 275–281 (2013)
- Valet, P., Adriaans, J., Liebig, S.: Comparing survey data and administrative records on gross earnings: nonreporting, misreporting, interviewer presence and earnings inequality. *Qual. Quant.* **53**, 471–491 (2019)

## Publisher's Note

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