Center for
Quantitative
Economics

# Measurement Error in Earnings

Stella Martin [†], Kevin Stabenow [†], Mark Trede [†]

108/2024

[†] Department of Economics, University of Münster, Germany

wissen.leben
Universität Münster

# Measurement Error in Earnings

Stella Martin*, Kevin Stabenow and Mark Trede

**Abstract**

This paper investigates the statistical properties of measurement error in earnings with a linked panel comprising a survey and administrative information from pension records. We can replicate central properties from previous literature such as mean reversion and extend insights into longitudinal features with our decade-long panel. Central correlates in the decomposition of measurement error include gender, features related to the individual labor market biography and individual positions in the income distribution, where under-/overreporting of earnings is especially prevalent above/below the median.

*Keywords: measurement error, earnings, survey data, administrative data, record linkage*

## 1 Introduction

Correct and unbiased knowledge of the distribution of earnings is fundamental to economic research and social policy. Measures of earnings inequality and poverty like the Gini coefficient or the at-risk-of-poverty rate rely strongly on the correctly recorded income or earnings distribution. While issues of sample composition and selectivity can be addressed by using weights, incorrect earnings data distort the estimation and result in misleading inferences. For instance, the estimation of the at-risk-of-poverty-rate is very sensitive, even to small distributional changes, especially when the pattern of measurement errors varies along the earnings distribution (Gottschalk and Huynh, 2010).

A small body of literature revolves around measurement errors in earnings information including the seminal studies on measurement error in earnings by Bound and Krueger (1991), Bound et al. (1994) and Pischke (1995). The main features that existing research has outlined are: mean reversion, i.e. a negative correlation of measurement error with true earnings, and a strong positive serial autocorrelation of the measurement error. However, the longitudinal structure was limited, and research following up on the topic

***Corresponding author: Stella Martin**, Institute for Econometrics and Economic Statistics, University of Münster, Am Stadtgraben 9, 48143 Münster, Germany, E-Mail: stella.martin@wiwi.uni-muenster.de
**Kevin Stabenow**, Institute for Econometrics and Economic Statistics, University of Münster
**Mark Trede**, Institute for Econometrics and Economic Statistics, University of Münster

from the 2000s onwards, has a strong focus on US data, where the validation source for earnings were records on employee earnings held by the Social Security Administration, implying coverage bias. Social security and tax institutions in the US differ considerably from their European counterparts, making a direct comparison of measurement error in self-reported earnings difficult.

Like earlier studies, we utilize two distinct sources of income information, a survey measurement of self-reported earnings that is complemented with a secondary observation from an administrative source, which is considered the validation observation. We use a novel dataset that contains a direct linkage of a large German panel with the respondents' records kept at the German statutory pension insurance. The latter record is used to assess the survey measurement of earnings.

We add to the literature in two ways. First, we extend and substantiate previous findings on the properties of measurement error in earnings reports on the basis of a considerably larger – and longer – panel. In line with previous studies, we consider the administrative earnings measure to be the true earnings. This is the first study to employ a direct linkage of survey reports of earnings to the employment biographies from insuree accounts of the mandatory federal pension insurance. The source of administrative data in the only previous study about earnings measurement error in Germany is the Federal Employment Agency (Gauly et al., 2020). Second, we report trajectories of measurement errors, i.e. their longitudinal features. Our analysis is based on a sample that incorporates a full decade of survey and administrative earnings information of up to 4118 individuals per wave.

The remainder of this paper is structured as follows. After surveying the previous literature in the field (Section 2), we describe the novel data and discuss how self-reported (survey) earnings relate to the validation income that we observe in administrative records (Section 3). We discuss the statistical properties of the measurement error in Section 4.1 and determine explanatory variables of the measurement error in Section 4.2 before turning to the estimation of a Mincer-type model to discuss the consequences of measurement error in earnings. Section 5 concludes.

## 2 Literature

The risk of survey data in general being prone to biases has been widely established. Research in that field has shown that several variables, among them personal characteristics and those pertaining to the survey design and setting, may influence the response behavior and thus ultimately data quality.

Psychological research has delved into the underlying cognitive processes to explain why information in surveys might be misreported. Tourangeau et al. (2010) attribute misreports to issues of interpretation and understanding of the question and its underlying concepts, problems in retrieving and judging the relevant information as well as its placement in time, and lastly, difficulties in formulating a response in the format which the survey requires. More specifically, considering misreports of income, Angel et al. (2019) name the following four sources of incorrect information by survey respondents, irrespective of whether the biased reporting takes place consciously or subconsciously: social desirability, socio-demographic characteristics of the respondent, specifics of the survey design, and learning effects.

One of the central findings throughout the literature on measurement errors of earnings is mean-reversion. Survey respondents whose true earnings are at the lower end of the distribution tend to overreport their earnings, whereas high earners systematically underreport their earnings (Angel et al., 2019). One reason for mean-reversion is in connotation with the social desirability bias, implying that patterns of over and underreporting towards the middle ground may be due to socially desirable and undesirable behavior, attitudes and characteristics (Bound et al., 2001; Kim and Tamborini, 2014). With respect to the properties of measurement error in surveys, evaluations of the accuracy and validity of the data mostly fall into the category of validation studies, since they typically compare respondents' self-reports from a survey to some validation measure. The sources of this validation measure range from complementary survey data, usually stemming from the employer, to administrative records, like those maintained by tax or social security authorities. To assess the validity of one measure, the other measure is assumed to provide an accurate and valid measurement of true earnings. An overview of the earlier (pre-2000s) validation studies can be found in Bound et al. (2001).

Bound and Krueger (1991) use two waves from subsequent years of the Current Population Survey (CPS), matched to payroll information reported to the Social Security Administration by the employer. They describe the properties of measurement error in reported earnings for a sample of male salaried workers and find that measurement error is negatively correlated with true earnings (-0.4). Measurement error accounts for a quarter of the overall variation in log-earnings. The measurement error is positively serially autocorrelated (0.4) and weakly correlated with covariates that are commonly included in earnings regressions, which is of concern, since it might cause error-in-variables bias. Bound et al. (1994) and Pischke (1995) use the so-called PSID validation study, a subsample of workers from one mid-western firm who completed the US-based Panel Study on Income Dynamics survey. The self-reported data was complemented with the firm's payroll records serving as validation observations. The sample offers six years of payroll records (1981-1986) and two waves of the worker survey (1982 and 1986). The measurement error is negatively correlated with the true earnings in 1982 and 1986. Its variance amounts to 15-25% of the variance of true log-earnings, and the 1982 and 1986 measurement error are positively correlated (0.094). They conceptualize true earnings as resulting from a random walk component and a pure-noise transitory component. The measurement error is decomposed into three components and consists of (i) an individual time-invariant component unrelated to earnings, (ii) a component correlated to a transitory earnings shock, and (iii) a pure-noise classical random error.

The only study on measurement error in Germany so far has been conducted by Gauly et al. (2020). They used the German subsample of the Programme for the International Assessment of Adult Competencies (PIAAC) and linked it to administrative records from the German Integrated Employment Biographies. They find that the difference in average monthly earnings between the two data sources amounts to 70 Euros. The absolute value of the measurement error is significantly negatively correlated with age, years of education and training, work experience, as well as literacy and numeracy. In an exemplary estimation of the Mincer equation, separately using the two data sources for the dependent variable (monthly earnings), the administrative data source yielded larger coefficient estimates and standard errors. All differences between the two estimations (except for the coefficient on age) were Chow-tested and significant at the 5% level.

Kapteyn and Ypma (2007) use a sample of the Swedish register data LINDA (Longitudinal Individual Data) matched to survey data from the Swedish subsample of the Survey on Health, Ageing and Retirement (SHARE). This sample is, however, not representative of the Swedish population, since the subjects of SHARE are older residents. While pensioners and their pensions are included – which is unique compared to the rest of the literature – the structure of earnings from employment likely does not adequately represent the composition of the full Swedish labor force. In assuming administrative earnings information to be true, they are able to confirm the mean-reversion property of the measurement error in earnings.

# 3 Data

We use a direct record linkage of the German Socio-economic Panel (SOEP)[1] which is a survey dataset, and the SOEP-RV.VSKT2020 which provides administrative data from insuree accounts of the German statutory pension insurance.

## 3.1 Survey Data

The survey data that we use is the SOEP. Since its first wave in 1984, it has been conducted annually. It is the largest and longest panel study in Germany. Currently, around 30,000 individuals in 15,000 households take part. The micro-data provides information on individual, household and family levels, covering a wide range of fields such as demography, labor market, household finance, education, health, political attitudes and well-being. Its richness, representativeness and scope has made it a staple in economic and social science research, with a large variety of applications.

[1]SOEP version 37, see (Goebel et al., 2019) for a detailed description.

Ideally, each person above 16 years of age in a sampled household completes a questionnaire on themselves every year ("personal questionnaire"). One person, usually the household head, completes an additional questionnaire that focuses on household-level information. Within the personal questionnaire for adults that is conducted yearly, individuals are asked:

> *What did you earn from your work last month? Please state both: gross income, which means income before deduction of taxes and social security and net income, which means income after deduction of taxes, social security, and unemployment and health insurance* (SOEP, 2022).

Respondents can then provide their gross and/or net income in the currency of that time (Euro from 2002, Deutsche Mark for West Germany before that, Mark of the German Democratic Republic for East Germany until June 30, 1990). Additionally, the interviewer instructions contain information for the interviewer to be relayed to the respondent for further clarification of the initial survey question:

> *If you received extra income such as vacation pay or back pay, please do not include this. Please do include overtime pay. If you are self-employed: Please estimate your monthly income before and after taxes* (SOEP, 2022).

We use this information[2] as it provides the most directly measured earnings information. SOEP provides other options, but these differ in terms of the reference period and the directness of measurement. Alternatives such as the generated gross labor earnings consist of the same measurement that we use, but are enriched with imputed earnings.[3] Because inferring gross earnings from net earnings requires knowledge of tax-specific individual circumstances that are not explicitly collected in the survey, we abstain from using imputed earnings. On a more general level, as we aim to evaluate the erroneous measures provided by respondents of their own earnings, we refrain from using earnings information that is "contaminated" by imputations as well as information from third parties. For this reason we eliminate all observations that stem from an interview with a proxy rather than the respondents themselves.

### 3.2 Administrative Data

The pension records are provided by the German Pension Insurance. In the German three-pillar pension system, old age provision is taken care of by the statutory pension insurance, occupational pensions and private provision. The pillars are not equally strong. The statutory pension system is the most important source of old-age income in

[2] Stored in the variable `plc0013`.

[3] For those who provide only net income in the interview process, gross earnings are imputed by the research data center (variable `pglabgro`). If item non-response is present across net and gross earnings, net earnings are imputed first, using past individual earnings information and cross-sectional trend data for labor earnings. For a more detailed description see Frick and Grabka (2005). If no individual earnings history has been recorded so far, imputation is carried out regression-based, using Mincer covariates and considering net earnings (Deutsches Institut für Wirtschaftsforschung, 2022).

Germany. The insuree population covers all dependently employed and those receiving social transfers or providing care work. Civil servants are not covered and entrepreneurs can opt in, but rarely do so as they would have to cover both the employer and the employee contributions. Those legally required to contribute, pay a share of their gross earnings, currently at 9%. Employee contributions are matched by the employer.

The pension records are individual micro-data that consist of (i) a fixed demographic and pension-relevant part that does not vary over time, and (ii) a longitudinal component. Accounts are held for everyone above 15 years of age and who has ever been in a pension-relevant labor market state. The standard time unit for time-varying data is a month. For each individual and each month, information is available on up to five labor market states and the amount of pension points awarded for each of these states. The number of pension-relevant states that an individual can be observed in within one month is limited to five for data security reasons. However, only 0.33% of episodes in the raw data are censored due to that limit. The reported statuses in a given month are prioritized in descending order by the amount of pension points awarded for them, so that the first is the one with the highest daily gross earnings. Besides dependent employment, other recorded statuses include caregiving, child-rearing, inability to work due to illness, unemployment and receipt of other social transfers, and military or civil service.

Since the data are process-generated and hence not explicitly collected for research purposes, limitations apply for secondary usage such as research. The earnings are not nominally recorded by the pension insurance, but in terms of "pension points", which are later added up and assigned a value in order to calculate a pay-out amount in retirement. One pension point equals the mean annual earnings in a given year. There is a cap on the maximum insurable yearly earnings that is set annually in advance via the legislative process. Consequently, earnings in the pension dataset are censored at the contribution ceiling. An overview of the median and maximum insurable income over time is provided in Table A1.

### 3.3 Linking SOEP and Pension Records

The first step in linking pension-account information to an individual's survey data is securing their informed consent. Respondents who have participated in at least one interview were asked to give consent to linking their pension account data to their survey data. They received an explanatory informational letter jointly signed by the vice-director of the SOEP, the president of the German Pension Insurance and the department head in charge of the study at the survey institute. The letter gives an overview of the type of information to be linked, explains the precautions with respect to data security and how anonymity is guaranteed, as well as how consent can be revoked. Attached to the informational letter is a short additional questionnaire to be filled out with the information necessary to identify the individual's pension accounts. The respondents are asked to provide their first and family name as well as (at least the first letter of) their family name at birth and their date and place of birth. Additionally, they can provide their insurance number.

For each consenting respondent, an (artificial) identification number ("SOEP-RV-ID") is created for the linked scientific use file. This way, both data providers, i.e. the SOEP and the Pension Insurance, have the advantage of being able to provide their data without having to transfer it from site to site and leaving the final merging to the researcher. On the side of the survey, the final data release is simply extended by the SOEP-RV-ID and remains otherwise unchanged. On the other side, the pension insurance uses the personal data and insurance number to identify the corresponding account. Since the pension insurance is organized in a federal system, the accounts are kept at the regional insurance entities. Using the unified system of account numbers, the federal research data center requests the data from the accounts of consenting respondents. Owing to its secondary nature, the data can only be extracted from the records and be prepared for distribution if there is no open application for a pension or rehabilitative measures that the pension insurance administration is working on. The accessible records are handed over to the research data center for preparation and then released to the scientific community in anonymized form, with the account numbers removed and just the SOEP-RV-ID left to merge the administrative and survey components of the linked data set.

Survey respondents were asked to give consent to linking their data in interviews starting 2018. In case of a decline, respondents were asked to reconsider in the two subsequent interviews. After three refusals, they were permanently removed from the pool of respondents to be presented with the linkage request. Up to now, 29,452 adults in the SOEP population have been asked for linkage consent. 52.6% consented and a total of 49.2% of cases could be linked successfully. Due to the nature of obtaining consent and the difference in time between the onset of the survey (1984) and the implementation of the linkage (from 2018 onward), there is a substantial difference in the composition of the entire SOEP population and the linked sample. Regardless of when they entered the survey, the earliest year in which respondents were asked to consent to the linkage was 2018. Anyone who had dropped out before 2018 (e.g. due to unwillingness, emigration or death) cannot be part of the linked sample. As a result, few of those with a decade-long history of participating in the survey are in the linked sample.

### 3.4 Analysis Sample

We restrict the sample to non-censored observations with linked pension information, as there would otherwise not be two earnings measurements to elicit a potential measurement error. Owing to the sample selection due to the late onset of obtaining consent, and our research question requiring both survey participation and a corresponding pension record for earnings in a given period, the main body of observations is from 2010 onward. Relatively and absolutely speaking, there are very few observations each year before 2010, which is why we restrict our analysis to those years in which exploiting the longitudinal structure does not rely only on a small and highly selective set of observations.

The true monthly earnings of individual $i$ at time $t$ are denoted by $y^*_{i,t}$. As we restrict the sample to individuals with information from both data sources, we observe the survey earnings ($y^s_{i,t}$) and the administrative earnings ($y^a_{i,t}$). While the true measurement error is given by

$$m^*_{i,t} = y^s_{i,t} - y^*_{i,t}, \tag{1}$$

the measurement error in our analysis is

$$m_{i,t} = y^s_{i,t} - y^a_{i,t} \tag{2}$$

as we use administrative earnings as validation measures and thus implicitly assume that administrative earnings are an accurate measurement of true earnings. Throughout the empirical investigation, we refer to the measurement error both in absolute and relative terms. The absolute measurement error is the difference between survey and administrative earnings as denoted by Equation 2, while the term "relative measurement error" refers to the share of the measurement error of (assumed) true earnings:

$$m^{rel}_{i,t} = m_{i,t}/y^a_{i,t}. \tag{3}$$

In terms of quality control of the matching procedure, we require sex and birth year to match in the survey and administrative sources. Additionally, we can only include observations where (i) respondents provided a measurement of their earnings for the month preceding the interview and (ii) their employment was a dependent one that required social security contributions, so that the survey observation can be complemented with an administrative earnings observation. We require the survey to have been answered by the respondents themselves rather than a proxy. Moreover, observations with censored administrative and survey earnings data are eliminated. As the censoring in administrative data would yield biased measures of the measurement error, censored administrative earnings data has to be excluded from the analysis. In order to treat both data sources symmetrically, the survey earnings are also only observed until the cap.[4]

As per the pension insurance's standard procedure, earnings are reported by employers, usually once a year. If the employer report is for a period longer than one month, the pension insurance splits the earnings into the months by their exact length in days. That procedure leads to a jitter in monthly earnings, while employers usually pay out the same income (a twelfth of the annual salary) regardless of the length of a month or the number of workdays it contains. To avoid this, we compute earnings on a daily basis to circumvent the imprecision that would arise if we simply used the monthly earnings.

To enable comparisons over the course of ten years (2010 - 2020), we inflation-adjust the measurement error in order to express it in real terms. We use the consumer price

[4] In fact, the empirical cumulative distribution function of the administrative earnings not only jumps at the cap of maximum insurable income, but also just below this threshold. This stylized fact was first described in Bönke et al. (2015).

index time series as provided with the SOEP cross-national equivalence file. The data are based on information provided by the Federal Statistical Office. The base year for the consumer price index is 2015 (Grabka, 2022).

The final matched sample that we use in the empirical part of the paper consists of 33,463 pairs of survey and administrative earnings data from 6,153 individuals (in 4,880 households) over a span of eleven years (2010-2020)[5]. We provide key characteristics of the final matched sample exemplary for the survey year 2018 in Table 1[6]. For reasons of comparison, we also present these summary statistics for two closely related samples. One is the full SOEP, that is, the full initial survey. The second is the fully linked SOEP-VSKT containing the subset of consenting members from the initial survey. The final linked sample consists of about two thirds of the initial SOEP-VSKT individuals, mostly due to the above-mentioned restrictions and missing information among the covariates. Of the three (sub-)samples, the final sample has the largest share of females, which seems to be mostly a product of a higher likelihood of women consenting. With respect to age, the average in the final sample is a little lower than in the linkage data set, yet closer to the initial survey mean age of respondents. For years in education and reported earnings, the survey-level means are the highest and the means in the final analysis sample are even smaller than in the full linked sample.

Table 1: Characteristics of SOEP 37, SOEP-RV.VKST2020 and Final Sample in 2018

| Year | | SOEP 37 | SOEP-VSKT | Final Sample |
|---|---|---|---|---|
| 2018 | Female (%) | 48.16 | 49.78 | 49.46 |
| | Age (mean) | 42.45 | 43.16 | 42.41 |
| | Years of education (mean) | 12.79 | 12.97 | 12.51 |
| | Tenure with firm (mean) | 10.44 | 10.45 | 10.85 |
| | Survey earnings (mean) | 2300.52 | 2125.34 | 2148.05 |
| | $N$ | 13101 | 2774 | 1379 |

# 4 Results

In this section, we present the results of our empirical analysis of the measurement error. We first elaborate on the statistical properties of measurement error. In particular, we examine features like mean reversion that have previously been established, and contribute new insights into the longitudinal features of measurement error. Secondly, we decompose the determinants of measurement error using regressions before turning towards a simple, illustrative example to outline the consequences of relying on either data source.

[5] A detailed overview with all cleaning steps performed on the sample can be found in Table A5.

[6] A full table with summary statistics for all eleven waves is given in Table A4.

### 4.1 Statistical Properties

In Table 2 we present basic summary statistics of the price-adjusted measurement error (in 2015 Euros). We observe a negative mean for each year between 2010 and 2020, ranging from $-112.06$ Euros in 2020 to $-244.91$ Euros in 2019. The standard deviation of the price-adjusted measurement error varies between 390.48 Euros (in 2010) and 523.74 Euros (in 2019).

Table 2: Summary Statistics of Measurement Error, Survey Earnings (SOEP), Administrative Earnings (VSKT)

| | | Measurement Error | | SOEP | | VSKT | |
|---|---|---|---|---|---|---|---|
| Year | $N$ | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| 2010 | 1379 | -166.24 | 387.32 | 2148.05 | 949.62 | 2314.29 | 1055.75 |
| 2011 | 2062 | -158.80 | 383.54 | 2174.86 | 980.33 | 2333.66 | 1085.33 |
| 2012 | 2501 | -183.28 | 397.47 | 2268.62 | 1020.15 | 2451.90 | 1135.51 |
| 2013 | 2637 | -181.01 | 425.54 | 2332.08 | 1045.92 | 2513.09 | 1169.26 |
| 2014 | 2704 | -171.42 | 436.42 | 2432.29 | 1079.10 | 2603.71 | 1218.35 |
| 2015 | 2800 | -213.76 | 450.72 | 2494.80 | 1112.38 | 2708.55 | 1269.34 |
| 2016 | 2939 | -190.62 | 458.21 | 2539.25 | 1160.91 | 2729.87 | 1304.83 |
| 2017 | 4026 | -199.30 | 472.92 | 2648.70 | 1188.00 | 2848.00 | 1339.18 |
| 2018 | 4283 | -225.87 | 505.05 | 2740.55 | 1222.60 | 2966.42 | 1387.24 |
| 2019 | 4069 | -244.91 | 521.81 | 2866.01 | 1278.53 | 3110.92 | 1453.06 |
| 2020 | 4063 | -112.06 | 505.16 | 3059.09 | 1391.09 | 3171.15 | 1504.80 |

Looking at the histograms of measurement error in Figure 1, we see, for each survey year, an asymmetric distribution that does not even closely resemble a normal distribution. An asymmetric Laplace distribution might fit the data well, but formal goodness-of-fit tests always reject this hypothesis. Regardless of the exact distribution, it is obvious that the assumption of Gaussian white noise is untenable for measurement error.

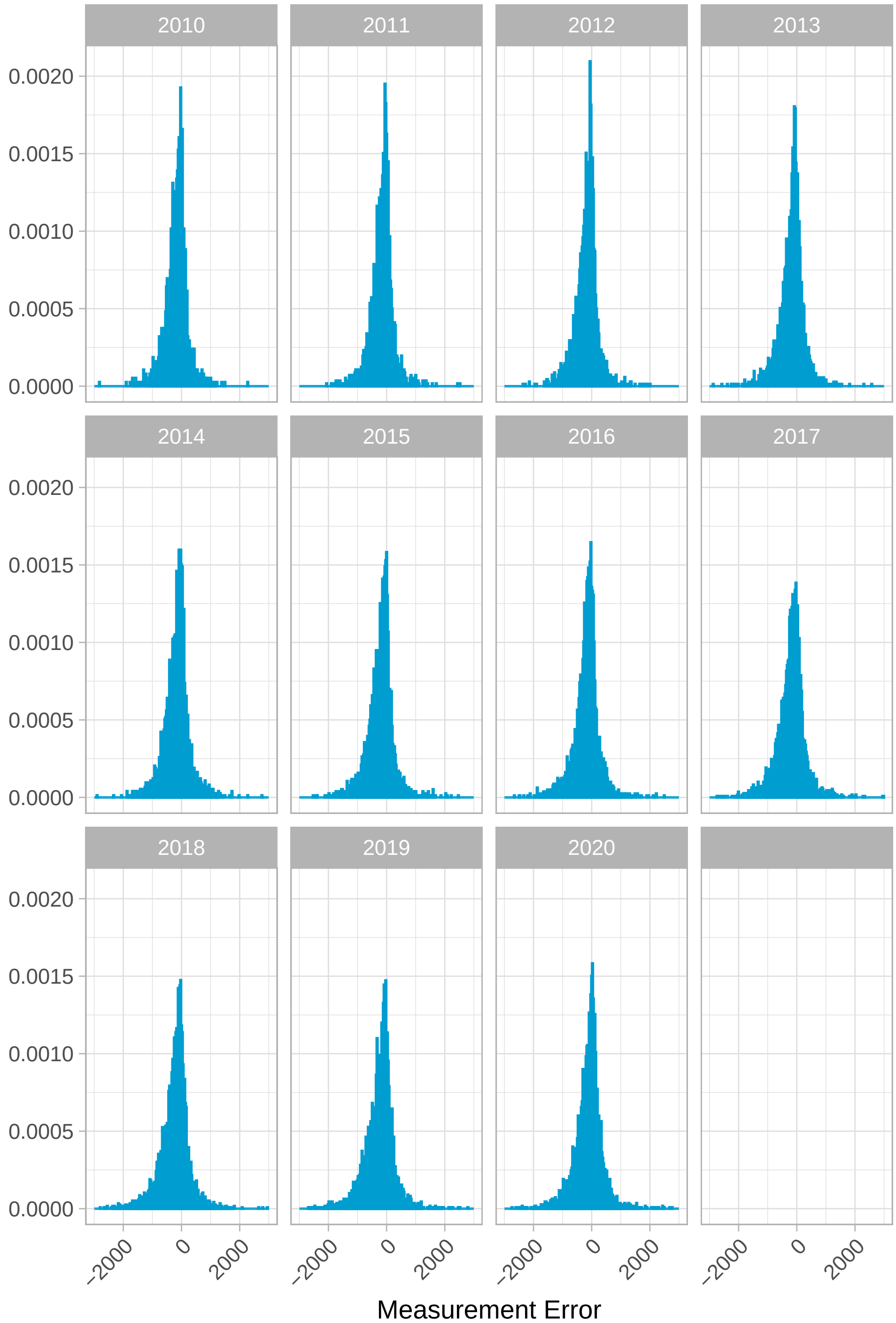


Figure 1: Histogram of the Price-adjusted Measurement Error for the Years 2010 to 2020, $m_{i,t} \in [-3000, 3000]$. The full range histogram can be found in Figure A5 in the Appendix.

To align ourselves with previous literature and to provide a descriptive basis for the estimation of the mincer equation that is based on the logarithmized earnings, we depict the relationship between true (log) earnings and measurement error for 2010 - 2020 in Figure 2. In order to visually account for the fact that there is considerable variation in the number of observations made in each year, we use semi-transparent dots. The more of them are more concentrated in one area, the darker that area appears. Although it is more strongly pronounced in the later years, roughly the same pattern emerges for all years in the sample: There is a negative relationship between true logarithmized earnings and measurement error. The higher the true earnings, the more observations of negative measurement error exist, which is the result of survey respondents underreporting their earnings.

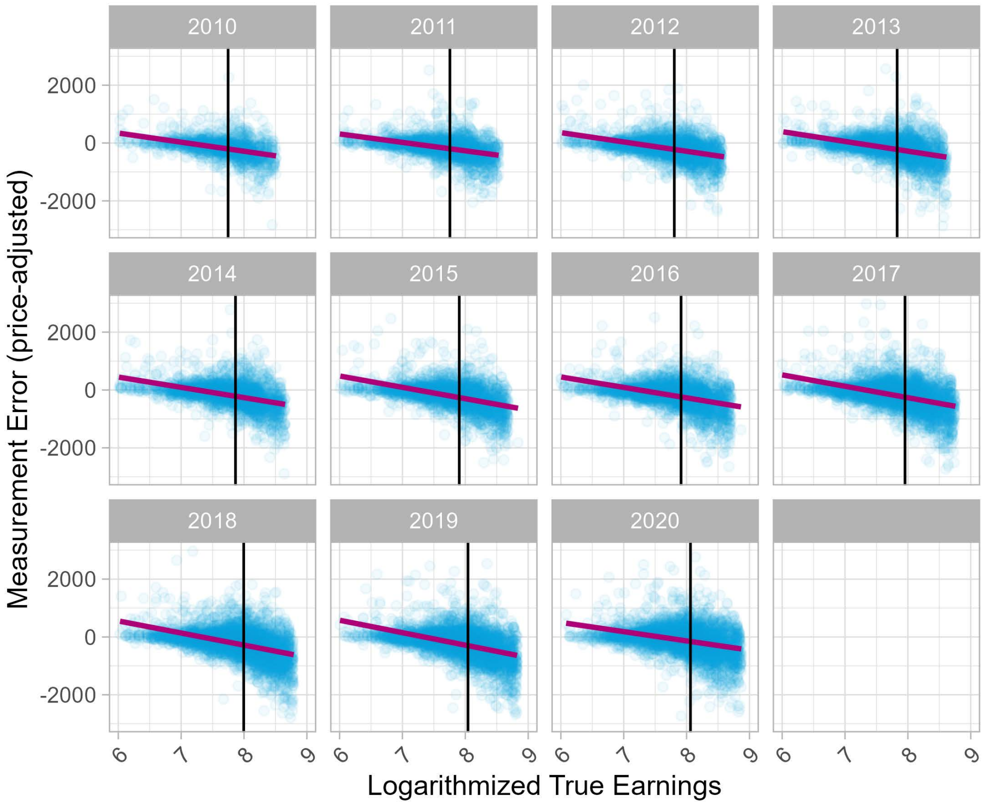


Figure 2: Scatterplot of the Relationship between True (log) Earnings and Price-adjusted Measurement Error for the Years 2010 to 2020. Log earnings restricted to interval [6, 9]. A full range graph can be found in Figure A6 in the Appendix. Black vertical line represents mean log administrative earning for each year.

One central result from the previous literature is the mean-reversion property of measurement error. We also see this pattern in our sample. The measurement error is negatively correlated with true earnings in all waves of our sample, see Table 3. The correlation coefficients are significantly different from zero in all years.

Table 3: Mean Reversion Property of the Measurement Error

| Year | $N$ | $\rho_t$ |
|---|---|---|
| 2010 | 1379 | -0.44*** |
| 2011 | 2062 | -0.44*** |
| 2012 | 2501 | -0.45*** |
| 2013 | 2637 | -0.46*** |
| 2014 | 2704 | -0.48*** |
| 2015 | 2800 | -0.50*** |
| 2016 | 2939 | -0.47*** |
| 2017 | 4026 | -0.48*** |
| 2018 | 4283 | -0.49*** |
| 2019 | 4069 | -0.49*** |
| 2020 | 4063 | -0.38*** |

$^{***}p < 0.001$

Positive autocorrelation, which is another finding from the literature, can also be confirmed as present in our sample. In contrast to previous studies, due to our rich dataset, we are able to examine the empirical autocorrelation function for its longitudinal features in up to ten lags. For all analyses regarding autocorrelative features of measurement error, we restrict the sample to a balanced panel. The resulting panel consists of 450 individuals. For each we have 11 years of earnings (measurement error) data.

Table 4 presents the lower triangular of the Pearson correlation matrix and the upper triangular of the Spearman correlation matrix. Bearing in mind that the measurement error is asymmetrically distributed with a large range, the correlation coefficient of Spearman seems more suitable. However, we find very similar results regarding the autocorrelation for both kinds of correlation coefficient (Pearson and Spearman). The values vary between 0.22 and 0.55 (0.28 and 0.58) for the Pearson (Spearman) correlation and point to a long lasting time dependence in the measurement error. In the samples used by Pischke (1995) and Bound and Krueger (1991), the autocorrelation in measurement error decreases with the lag length. The decrease in the autocorrelation coefficient that we find is much smaller. Even at a lag of 5 or 10 years, the correlation coefficients are remarkably high.

Table 4: Autocorrelation functions of the price-adjusted measurement error. The lower triangular depicts the Pearson autocorrelation, the upper triangular the Spearman autocorrelation. Observation years are given as column and row names.

| | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2010 | | 0.43 | 0.39 | 0.40 | 0.44 | 0.40 | 0.33 | 0.32 | 0.35 | 0.33 | 0.25 |
| 2011 | 0.38 | | 0.48 | 0.52 | 0.47 | 0.42 | 0.36 | 0.35 | 0.29 | 0.34 | 0.24 |
| 2012 | 0.35 | 0.47 | | 0.51 | 0.48 | 0.46 | 0.38 | 0.46 | 0.39 | 0.42 | 0.30 |
| 2013 | 0.36 | 0.55 | 0.52 | | 0.58 | 0.50 | 0.46 | 0.46 | 0.40 | 0.42 | 0.33 |
| 2014 | 0.44 | 0.49 | 0.40 | 0.55 | | 0.55 | 0.53 | 0.52 | 0.49 | 0.50 | 0.38 |
| 2015 | 0.38 | 0.42 | 0.42 | 0.50 | 0.55 | | 0.49 | 0.55 | 0.43 | 0.51 | 0.40 |
| 2016 | 0.28 | 0.35 | 0.28 | 0.41 | 0.46 | 0.44 | | 0.58 | 0.48 | 0.50 | 0.34 |
| 2017 | 0.34 | 0.38 | 0.43 | 0.50 | 0.52 | 0.58 | 0.55 | | 0.52 | 0.54 | 0.45 |
| 2018 | 0.33 | 0.30 | 0.33 | 0.39 | 0.46 | 0.45 | 0.43 | 0.52 | | 0.54 | 0.45 |
| 2019 | 0.36 | 0.36 | 0.40 | 0.43 | 0.51 | 0.53 | 0.51 | 0.57 | 0.53 | | 0.53 |
| 2020 | 0.22 | 0.24 | 0.24 | 0.28 | 0.33 | 0.38 | 0.29 | 0.42 | 0.38 | 0.51 | |

We fit an ARMA$(p,q)$ model to the measurement errors using minimum distance estimation,

$$m_{i,t} = \sum_{j=1}^{p} \varphi_j m_{i,t-j} + \sum_{j=1}^{q} \vartheta_j \epsilon_{i,t-j} + \epsilon_{i,t}. \tag{4}$$

The minimum distance estimator of the ARMA coefficients minimizes the weighted sum of quadratic distances between elements of the vectorized lower triangular empirical autocorrelation matrix, $\hat{\rho}$, and the vectorized lower triangular theoretical autocorrelation matrix $g(\theta)$ of an ARMA$(p,q)$ process (Cameron and Trivedi, 2005),

$$Q_N(\theta) = (\hat{\rho} - g(\theta))' W_N (\hat{\rho} - g(\theta)). \tag{5}$$

The vector $\hat{\rho} = [\hat{\rho}_{2010,2010}; \hat{\rho}_{2010,2011}; \ldots; \hat{\rho}_{2019,2020}; \hat{\rho}_{2020,2020}]'$ consists of the estimated correlation coefficients

$$\hat{\rho}_{t_1,t_2} = \frac{\sum_{i=1}^{N} (m_{i,t_1} - \overline{m}_{t_1})(m_{i,t_2} - \overline{m}_{t_2})}{\sqrt{\sum_{i=1}^{N} (m_{i,t_1} - \overline{m}_{t_1})^2} \sqrt{\sum_{i=1}^{N} (m_{i,t_2} - \overline{m}_{t_2})^2}} \tag{6}$$

for $t_1 \in \{2010, \ldots, 2020\}$, $t_2 \in \{2010, \ldots, 2020\}$ and $t_2 \leq t_1$.

We allow differing degrees of AR and MA terms with $p \in \{1,2\}$, $q \in \{1,2\}$ in the minimization. The weighting matrix $W_N$ is set to the optimal weighting matrix as described in Andrews (1999). The term $Q_N(\theta)$ from Equation 5 is numerically minimized. The starting values for the numerical optimization of the ARMA$(1,1)$ are zeros. The distance-minimizing estimates for the ARMA$(1,1)$ coefficients are used as the starting values for the numerical optimization of the ARMA$(1,2)$, ARMA$(2,1)$ and ARMA$(2,2)$.

The estimated model reported below is selected on the basis of the information criteria GMM-BIC and GMM-AIC described by Andrews (1999). The selected model is an ARMA$(2, 2)$ with the estimated parameters shown in Table 5. Table 6 gives the values of the empirical autocorrelation matrix on the lower triangular, and the estimated theoretical autocorrelation matrix on the upper triangular of the matrix. Both matrices are depicted as a heatmap in Figure 3, the grey tiles correspond to those on the diagonal.

Table 5: Minimum Distance Estimates for the Fitted ARMA(2,2) Model

| $j$ | $\varphi_j$ | $\vartheta_j$ |
|---|---|---|
| 1 | 0.3981 | -0.1493 |
| 2 | 0.4908 | -0.2902 |

Table 6: The lower triangular depicts the empirical Pearson autocorrelation, the upper triangular shows the estimated autocorrelation from the estimated ARMA(2, 2) model. Observation years are given as column and row names.

| | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2010 | | 0.51 | 0.51 | 0.46 | 0.43 | 0.40 | 0.37 | 0.34 | 0.32 | 0.29 | 0.27 |
| 2011 | 0.38 | | 0.51 | 0.51 | 0.46 | 0.43 | 0.40 | 0.37 | 0.34 | 0.32 | 0.29 |
| 2012 | 0.35 | 0.47 | | 0.51 | 0.51 | 0.46 | 0.43 | 0.40 | 0.37 | 0.34 | 0.32 |
| 2013 | 0.36 | 0.55 | 0.52 | | 0.51 | 0.51 | 0.46 | 0.43 | 0.40 | 0.37 | 0.34 |
| 2014 | 0.44 | 0.49 | 0.40 | 0.55 | | 0.51 | 0.51 | 0.46 | 0.43 | 0.40 | 0.37 |
| 2015 | 0.38 | 0.42 | 0.42 | 0.50 | 0.55 | | 0.51 | 0.51 | 0.46 | 0.43 | 0.40 |
| 2016 | 0.28 | 0.35 | 0.28 | 0.41 | 0.46 | 0.44 | | 0.51 | 0.51 | 0.46 | 0.43 |
| 2017 | 0.34 | 0.38 | 0.43 | 0.50 | 0.52 | 0.58 | 0.55 | | 0.51 | 0.51 | 0.46 |
| 2018 | 0.33 | 0.30 | 0.33 | 0.39 | 0.46 | 0.45 | 0.43 | 0.52 | | 0.51 | 0.51 |
| 2019 | 0.36 | 0.36 | 0.40 | 0.43 | 0.51 | 0.53 | 0.51 | 0.57 | 0.53 | | 0.51 |
| 2020 | 0.22 | 0.24 | 0.24 | 0.28 | 0.33 | 0.38 | 0.29 | 0.42 | 0.38 | 0.51 | |

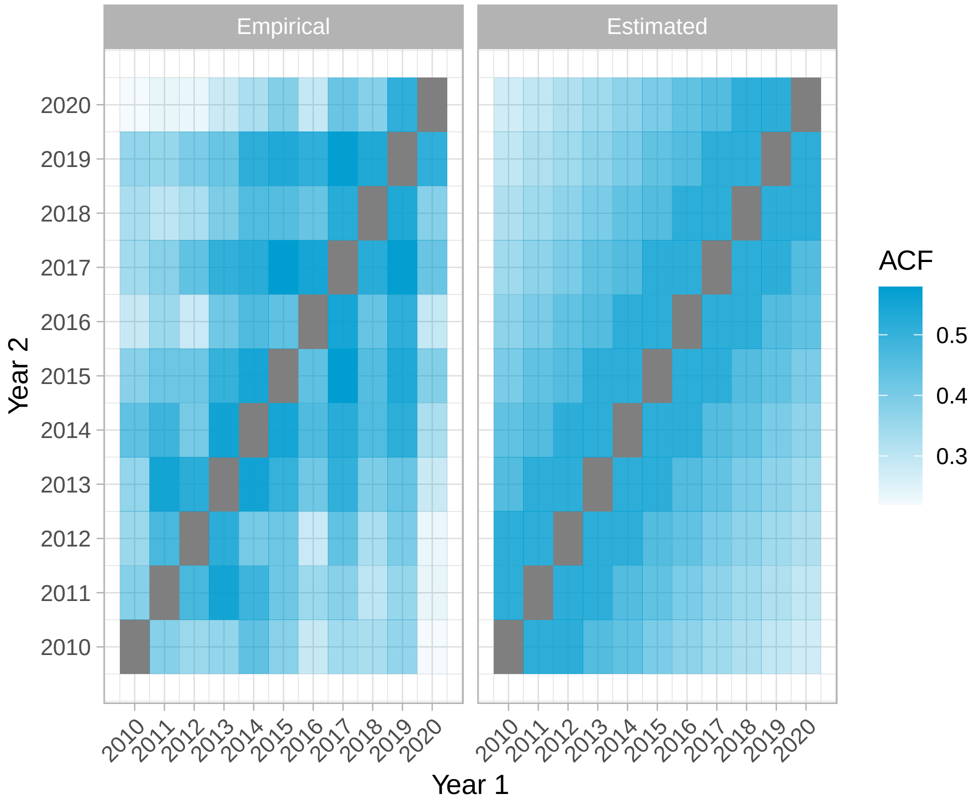


Figure 3: Empirical Autocorrelaton Matrix and Estimated Theoretical Autocorrelation Matrix of the Measurement Error.

### 4.2 Determinants of Measurement Error

To determine which characteristics are associated with the measurement error, we regress the absolute magnitude of measurement error (cf. Equation 2) and the relative measurement error, i.e. the ratio of (absolute) measurement error and administrative income on the covariates (cf. Equation 3).

**Individual demographic characteristics:** gender, age, migration background[7], region of birth (East or West Germany)

**Household characteristics:** household size, number of children

**Partnership information:** married, in a cohabitating relationship

**Labor market related characteristics:** years of education and training, occupational qualification, work experience, unemployment experience, employment in the civil sector, tenure with current employer

**Earnings distribution:** position (quantile) of individual earnings in the overall true income distribution of a given year[8].

The entire sample is used to estimate fixed-effect regressions with year-fixed effects. Individual effects cannot be incorporated here, as invariant characteristics such as gender[9] would be captured within those individual effects. The results are presented in Table 7.

Women seem to produce self-reports of their earnings with significantly smaller measurement error. While behavioral research has shown that women are generally less likely to produce overreports (Roth and Slotwinski, 2020; Bertrand et al., 2015), this observation in our sample might also be a result of sample selection. While the sample consists of an almost equal mix of males and females, we can only elicit measurement error in earnings for those who have earnings that are subject to social security contributions. Those women who voluntarily select themselves into the labor force (and into employment above the social security requirement) might have a greater level of awareness of their personal finances than those not in employment. In support of this notion, both the household size as well as being married and cohabitating with a partner are highly negatively correlated with measurement error. For reasons of comparison, an indicator for living in a cohabitating relationship while not being married was included. Unlike cohabitating partners, married couples are eligible to benefit from income splitting for tax purposes, and hence, might be more aware of their earnings situation than non-married

[7] We distinguish between "direct migration background" and "indirect migration background". An individual born abroad who migrated to Germany has a "direct migration background". An individual born and raised in Germany whose parents migrated to Germany has an "indirect migration background".

[8] The quantile is measured as a value between 0 and 1.

[9] Gender could technically vary within the SOEP, although we do not observe that for the individuals in our sample.

Table 7: Determinants of Absolute and Relative Measurement Error (Time Fixed Effects Regression)

| | Absolute | Relative |
|---|---|---|
| Female (Dummy) | -146.14*** | -9.36*** |
| Age | 1.98 | -0.17 |
| Age squared | -0.02 | 0.00 |
| Direct Migration Background (Dummy) | -10.43 | 0.45 |
| Indirect Migration Background (Dummy) | 18.65 | -0.82 |
| East-Germany (Dummy) | 13.54* | 1.18 |
| Household Size | -15.28*** | -0.27 |
| Number of Children | -19.83 | -1.87 |
| Cohabitation (Dummy) | 33.87*** | 0.01 |
| Married (Dummy) | -27.03*** | -2.82*** |
| Years of Education | 13.64*** | 0.76*** |
| Untrained Worker (Dummy) | -28.22*** | -0.90 |
| Work Experience (in Years) | 2.14*** | 0.26*** |
| Years of Unemployment | -2.59* | -0.44** |
| Public Service (Dummy) | -67.14*** | -2.17** |
| Tenure with the Firm (in Years) | -1.29*** | -0.09** |
| Quantile of Income | -232.75*** | -115.24*** |
| Squared Quantile of Income | -610.08*** | 76.63*** |
| | | |
| $R^2$ (adj.) | 0.24 | 0.05 |
| $N$ | 33463 | 33463 |

${}^{*}p < 0.05, {}^{**} p < 0.01, {}^{***} p < 0.001$

couples who are not eligible for such a tax break.

Comparing the estimation of absolute and relative measurement error as reported in Table 7, an individual's labor market biography seems to produce the most significant correlates besides a respondent's position in the earnings distribution and their gender. Further, employment in public service has a large negative impact on the magnitude of measurement error, which might be a result of the comprehensive coverage of public service jobs by collective wage bargaining agreements and the contractually fixed working hours, resulting in little to no variation of pay from month to month, which could contribute to a more accurate recollection of respondents in a survey.

Strikingly, measurement error is strongly negatively correlated with an individual's position in the income distribution, in terms of the decomposition of both relative and absolute measurement error. For the squared term, the estimates are of opposing magnitude, because, for the relative measurement error, the estimated coefficient on the squared quantile of income is 76.63, while it is $-610.08$ in the decomposition of the absolute measurement error. The high significance level in both regressions suggests a non-linear relationship between an individual's position in the income distribution and their measurement error. To elaborate on this further, we refrain from employing an individual's quantile membership in the distribution of true earnings in a (quasi-)continuous manner, and include quantile membership in a discrete fashion. Previous research has developed and supported the hypothesis that mean-reversion is the result of a social desirability bias in reporting one's labor earnings, in the sense that reports are biased towards the mean. That implies that reports from individuals with true earnings from the lower end of the income distribution tend to overreport their earnings, whereas those from the upper end seem to systematically underreport their earnings. Thus, we substitute the variable *Quantile of Income* and its square with the income decile of each respondent and re-estimate the model with the fifth decile as the reference category. This way, we can further disentangle reporting patterns across the deciles. The results are reported in Table 8, where the second column presents estimates for regressions on measurement error and the third on measurement error relative to the true earnings. Both specifications contain the full set of controls as well as time-fixed effects.

Table 8: Time Fixed Effects Regression of (Relative) Measurement Error on Income Deciles and Other Controls

| | Absolute | Relative |
|---|---|---|
| Earnings Decile (ref.: 5th) | | |
| 1st decile | 286.7639*** | 36.3741*** |
| 2nd decile | 152.7298*** | 7.5459*** |
| 3rd decile | 101.8178*** | 4.4905*** |
| 4th decile | 53.2192*** | 2.0261 |
| 6th decile | -74.6815*** | -2.427* |
| 7th decile | -126.7446*** | -3.5104** |
| 8th decile | -214.4795*** | -5.3426*** |
| 9th decile | -378.7501*** | -8.5764*** |
| 10th decile | -593.9213*** | -11.4551*** |
| | | |
| Other controls | | |
| Demographics | yes | yes |
| Household characteristics | yes | yes |
| Partnership information | yes | yes |
| Labor market charcateristics | yes | yes |
| | | |
| $R^2$ (adj.) | 0.24 | 0.05 |
| $N$ | 33463 | 33463 |

$^{*}p < 0.05,^{**}p < 0.01,^{***}p < 0.001$

The sign of coefficients estimated on the first to fourth deciles in the distribution of true earnings is positive. This implies that individuals from the lower half tend to overreport their earnings, while individuals with earnings above the 5th decile tend to underreport. Most estimates are highly significant at $p < 0.001$. In relation to the base category, i.e. the fifth decile in the distribution of true incomes, earnings further towards either end of the earnings distribution, are associated with a larger measurement error. The same overall pattern also holds for relative measurement error, except for the significance pattern. With the relative share of measurement error, membership in the median-adjacent deciles (i.e. fourth and sixth) is not associated with significant coefficient estimates.

Because a considerable difference in reporting behavior across the distribution of incomes becomes obvious, we split the sample into under- and overreports and re-ran the regressions. The adjusted $R^2$ is substantially lower for over- than underreports, implying that the model is of much more use in explaining underreports than overreports.

Table 9: Determinants of Absolute and Relative Measurement Error by Over- and Underreporting (Time Fixed Effects Regression)

| | Underreport | Overreport |
|---|---|---|
| Female (Dummy) | -66.95*** | -118.48*** |
| Age | 0.55 | 2.85 |
| Age squared | -0.00 | -0.04 |
| Direct Migration Background (Dummy) | -10.18 | 5.79 |
| Indirect Migration Background (Dummy) | 8.15 | 14.93 |
| East-Germany (Dummy) | 13.41** | -29.46*** |
| Household Size | -14.88*** | 1.38 |
| Number of Children | -1.18 | -19.21 |
| Cohabitation (Dummy) | 17.60* | 14.27 |
| Married (Dummy) | -6.49 | -34.27*** |
| Years of Education | 11.20*** | 6.86*** |
| Untrained Worker (Dummy) | -46.51*** | 1.97 |
| Work experience | 1.95*** | 1.50* |
| Years of Unemployment | 1.12 | -5.32*** |
| Public Service (Dummy) | -39.39*** | 32.49*** |
| Tenure with the Firm | -1.20*** | 0.82 |
| Quantile of Income | 87.08** | -262.66*** |
| Quantile of Income Squared | -802.85*** | 211.55*** |
| $R^2$ (adj.) | 0.31 | 0.03 |
| $N$ | 23360 | 10103 |

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Coefficient estimates that are highly significant usually are so among both over- and underreporters, with household size being a notable exception. For underreporters, a larger household is associated with a significantly smaller measurement error, whereas no such pattern emerges for overreporters, reflected by an insignificant coefficient for household size. Similarly, work experience and tenure with the current employer are highly significantly associated with measurement error among underreporters, but not for those who overreport their income. The opposite pattern emerges for unemployment experience. A longer experience of unemployment in overreporters is significantly negatively associated with measurement error, but the same does not hold for underreporters. The adjusted $R^2$ that we estimate as a measure of the goodness-of-fit indicates that while underreports can be explained comparably well (adjusted $R^2$ of 0.31), the coefficient of determination for the same model is about a tenth (adjusted $R^2$ of 0.03) in the subsample of overreports.

### 4.3 Estimation of Mincer Equation

As one of the fundamental models in labor economics, we estimate the Mincer equation with both administrative and survey information to illustrate the difference in the estimates obtained from both income observations. Introduced by Mincer (1958), it is a widely used workhorse model rooted in human capital theory that specifies labor earnings as a function of education, work experience and its square. Unlike common specifications of the Mincer equation, Gauly et al. (2020) use non-logarithmized earnings and do not report estimating an intercept. However, there are justifications from a statistical as well as economic theory perspective to use a logarithmized earnings variable when estimating Mincer-type earnings equations. Besides allowing for the interpretation of coefficient estimates as elasticities, economic theory suggests that education and work experience have a multiplicative (rather than additive) impact on earning, which would result in an additional year of education or work experience increasing earnings by a certain percentage rather than a fixed amount. We align ourselves with Gauly et al. (2020) in using a simple specification of the Mincer equation, including only the number of years spent in education or training and the work experience to estimate monthly labor earnings as regressors. We also estimate the simple specification of the Mincer equation without further control variables, but deviate from Gauly et al. (2020) by including an intercept, separately for each year, and using the logarithmized price-adjusted survey and administrative earnings as the dependent variables. For the sake of completeness and comparability with the literature, the results for the estimation of the non-logarithmized earnings specification without intercept in alignment with Gauly et al. (2020) are reported in Table A3.

The final formulation of the Mincer equation that we estimate for a given year $t$ is

$$\ln(y_{i,t}) = \ln(w_0) + \rho s_i + \beta_1 x_i + \beta_2 x_i^2 + \epsilon, \tag{7}$$

where $w_0$ represents the earnings for an individual with no education and no experience, $s$ is their years of education and training, and $x$ the work experience of an individual

$i$. The coefficient $\rho$ represents returns to education and $\beta_1, \beta_2$ the returns to experience. We estimate the Mincer equation separately for administrative earnings, $y^a_{i,t}$, and survey earnings, $y^s_{i,t}$.

The results of the estimation of the classical Mincer equation for the entire period from 2010 to 2020 are reported in Table A2. It is noteworthy that since earnings are logarithmized, we exclude earnings of zero. The estimation results for the 2018 subsample are presented in Table 10. 2018 is the largest wave of data ($N = 4282$) and is also one of the more recent years that is not impacted by any COVID-19-related changes to the labor market.

Table 10: Mincer Equation, Year = 2018

| | Year = 2018 | | |
|---|---|---|---|
| | SOEP | VSKT | Difference |
| | $\ln\left(y^s_{i,t}\right)$ | $\ln\left(y^a_{i,t}\right)$ | |
| DV = logarithmized monthly earnings | | | |
| Intercept | 6.7594*** | 6.7088*** | 0.0507 |
| | (0.0416) | (0.0460) | |
| Years of education | 0.0614*** | 0.0665*** | -0.0051 |
| | (0.0029) | (0.0031) | |
| Work experience | 0.0191*** | 0.0242*** | -0.0051* |
| | (0.0024) | (0.0028) | |
| Work experience sq. | -0.0002** | -0.0002*** | 0.0001 |
| | (0.0001) | (0.0001) | |
| $R^2$ (adj.) | 0.1522 | 0.1549 | |
| $N$ | 4282 | 4282 | |

Robust standard errors in parentheses.
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

For the 2018 subsample, the estimations with the two logarithmized earnings measures yields similar results. Both time spent in education and training, as well as work experience, are highly significantly associated with log earnings. The differences between the administrative-based and the survey-based earnings estimation are small (-0.0051 for the coefficients estimated on education years, -0.0051 for work experience and 0.0001 for squared work experience). They also vary in different directions, i.e. the estimate of the intercept is larger for the survey-based estimation, while those on education years and work experience are larger in the estimation based on administrative earnings. The overall model fit is similar, in that around 15% of the variation in earnings are explained for either source of the earnings measure, indicating that other (unobserved) factors play a role in determining earnings, but using either earnings measure does not severely distort the estimation results of the classical Mincer equation in a cross-sectional setting, and the impact on the goodness of fit is minimal.

Considering all waves of data available, the use of survey earnings information elicits similar results to when administrative earnings are employed (cf. Table A2), confirming the notion from the 2018 estimation above. Estimated coefficients have the same signs and generally exhibit the same level of significance. The estimated coefficients on the intercept are always highly significant ($p < 0.001$). They are also always slightly larger when using administrative earnings. This does not hold for the coefficient on work experience, which is not consistently larger with survey earnings.

Using potentially erroneous survey earnings instead of administrative earnings information with this sample would result in an estimated absolute difference in earnings of 235.44 Euros (nominal, 2018 level), which amounts to 8.45% of of the predicted survey earnings. An overview of the absolute and relative deviation in predicted earnings, using the average years of schooling and average work experience in the respective year's subsample, is provided in Table 11.

Table 11: (Relative) Prediction Error Given Average Characteristics

| Year | Error | % of SOEP earnings |
|---|---|---|
| 2010 | -195.41 | 8.75 |
| 2011 | -184.68 | 8.26 |
| 2012 | -216.95 | 9.35 |
| 2013 | -192.18 | 8.06 |
| 2014 | -193.89 | 7.82 |
| 2015 | -240.00 | 9.52 |
| 2016 | -202.92 | 7.89 |
| 2017 | -200.85 | 7.51 |
| 2018 | -235.44 | 8.45 |
| 2019 | -255.43 | 8.87 |
| 2020 | -112.24 | 3.64 |

With the exception of 2020, using survey earnings instead of administrative earnings results in a 7-9% underestimation of earnings. The absolute error is consistently negative, i.e. administrative earnings records are higher than self-reported earnings from the survey. The smaller misspecification obtained for 2020 might be a result of COVID 19 and its impact on the labor market, as well as on the data collection within the survey or heightened respondent awareness of their earnings in extraordinary circumstances.

## 4.4 Gini Coefficient and ARP

Highly aggregated measures of inequality like the Gini coefficient or the at-risk-of-poverty rate (ARP) rely on correct reports of earnings information. To illustrate the effect of measurement error in survey earnings on the yearly calculated inequality measures, we determined these measures from the individual monthly earnings as reported by the German Pension Insurance and the individuals themselves. The results are stated in Table 12 and depicted in Figure 4. For all years of observation, the inequality is underestimated when using the individual survey earnings. The difference between the Gini coefficient calculated from the administrative earnings, $Gini^a$, and based on the survey earnings, $Gini^s$, is roughly constant at around 0.01.

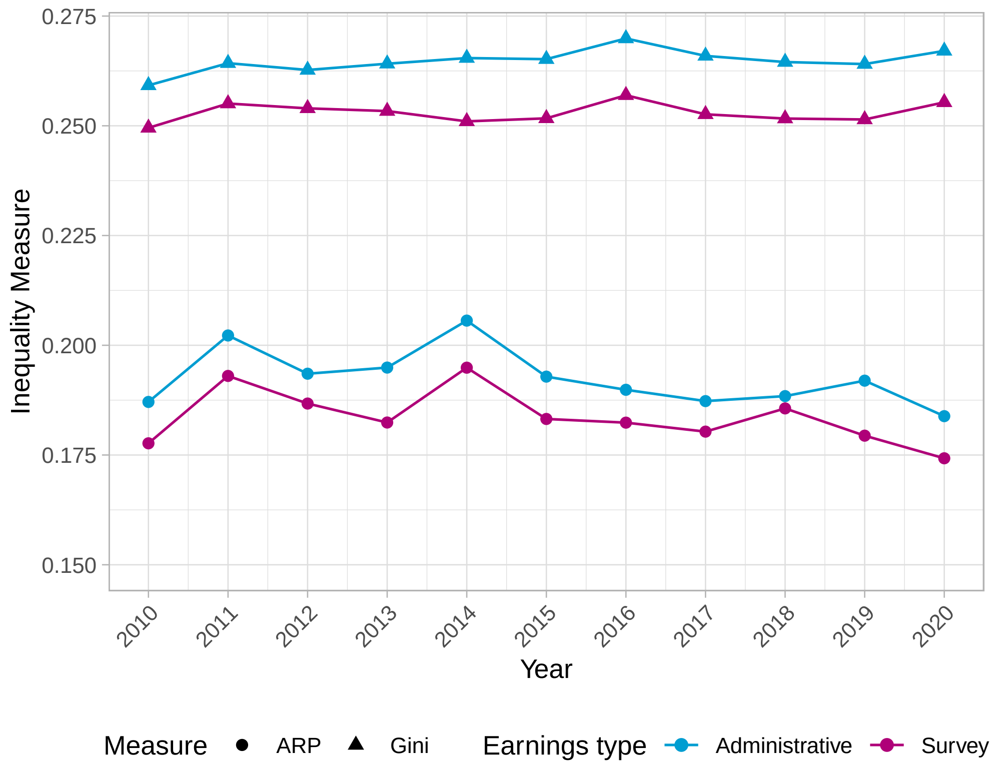


Figure 4: Gini Coefficient and At-Risk-of-Poverty-Rate calculated on survey and administrative earnings.

A similar pattern applies for the at-risk-of-poverty-rate. Estimates based on the survey earnings information, $\text{ARP}^s$, are too low in comparison to the estimates calculated on the validation earnings, $\text{ARP}^a$. However, one has to bare in mind that the ARP is usually calculated based on equivalized household net income. In our dataset, we have information on the household structure of which an individual is part. Even so, not all individuals in the household consented to the linkage of the survey data to the administrative data, or were even asked. Individuals under the age of 18 are not asked to give consent, but might nevertheless have monthly labor market income. Moreover, the tax burden of an individual depends on household decisions like the marital status

Table 12: Gini Coefficient and At-Risk-of-Poverty Rate Calculated on Survey and Administrative Earnings

| Year | N | Gini$^a$ | Gini$^s$ | ARP$^a$ | ARP$^s$ |
|---|---|---|---|---|---|
| 2010 | 1379 | 0.2592 | 0.2495 | 0.1871 | 0.1777 |
| 2011 | 2062 | 0.2643 | 0.2551 | 0.2022 | 0.1930 |
| 2012 | 2501 | 0.2627 | 0.2540 | 0.1935 | 0.1867 |
| 2013 | 2637 | 0.2641 | 0.2534 | 0.1949 | 0.1824 |
| 2014 | 2704 | 0.2654 | 0.2510 | 0.2056 | 0.1949 |
| 2015 | 2800 | 0.2652 | 0.2517 | 0.1929 | 0.1832 |
| 2016 | 2939 | 0.2699 | 0.2570 | 0.1899 | 0.1824 |
| 2017 | 4026 | 0.2659 | 0.2526 | 0.1873 | 0.1803 |
| 2018 | 4283 | 0.2645 | 0.2516 | 0.1884 | 0.1856 |
| 2019 | 4069 | 0.2641 | 0.2514 | 0.1919 | 0.1794 |
| 2020 | 4063 | 0.2671 | 0.2553 | 0.1839 | 0.1743 |

of spouses and tax deductions for children. Thus, an accurate calculation of net income or net equivalent income is not possible given the available information.

# 5 Conclusion

We use the linkage dataset of the German Socio-economic Panel and the administrative pension records. Earnings information plays a crucial role in economic, demographic and social science research, whether as an explanatory variable or as a dependent one. In using the twice validated administrative record of earnings, and assuming them to be an accurate representation of true earnings, we evaluate the accuracy of self-reported monthly earnings from the SOEP survey. Estimating a simple specification of the Mincer equation, we show that survey earnings rather than administrative earnings produce an underestimation of, on average, around 8% of monthly earnings in a Mincer equation.

We confirm the main findings on measurement errors in the literature. The measurement error is non-normal and skewed as well as mean-reverting, i.e. it is negatively correlated with true earnings. Regarding the longitudinal properties, we find that autocorrelation is present and long-lasting. Unlike previous studies, we are able to consider lag orders of up to ten, and observe a pattern of slowly declining autocorrelation over time. The main driver of measurement error, both in absolute terms and relative to the level of earnings, seems to be an individual's position in the income distribution. Labor market characteristics are also shown to be of some importance, while age and household-specific characteristics are not.

In this study, we consider a sample of employed individuals, hence non-labor income is entirely ignored. No inferences can be made on measurement error from further types of income like capital gains or even social transfers. Patterns of social desirability could be discussed more comprehensively with regard to social norms like the male breadwinner norm if income from social transfers is considered as well.

For policy makers, our results suggest a need for greater caution with respect to certain demographic groups. Females who participate in the labor force seem to produce more accurate measurements of their earnings, while males' reports are commonly an over/understatement of their gross earnings, rendering descriptions of inequality potentially inaccurate.

## References


Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, 67(3):543–564.

Angel, S., Disslbacher, F., Humer, S., and Schnetzer, M. (2019). What did you really earn last year?: Explaining measurement error in survey income data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(4):1411–1437.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.

Bound, J., Brown, C., Duncan, G., and Rodgers, W. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12(3):345–368.

Bound, J., Brown, C., and Mathiowetz, N. (2001). *Measurement Error in Survey Data*, volume 5, pages 3705–3843. Elsevier, New York.

Bound, J. and Krueger, A. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, 9(1):1–24.

Bönke, T., Corneo, G., and Lüthen, H. (2015). Lifetime earnings inequality in Germany. *Journal of Labor Economics*, 33(1):171–208.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications.* Cambridge University Press.

Deutsches Institut für Wirtschaftsforschung (2022). SOEP-Core v37-PGEN: Person-related status and generated variables. Technical report, SOEP Survey Papers.

Frick, J. R. and Grabka, M. M. (2005). Item nonresponse on income questions in panel surveys: Incidence, imputation and the impact on inequality and mobility. *Allgemeines Statistisches Archiv (AStA)*, 89:49–61.

Gauly, B., Daikeler, J., Gummer, T., and Rammstedt, B. (2020). What's my wage again? Comparing survey and administrative data to validate earning measures. *International Journal of Social Research Methodology*, 23(2):215–228.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Journal of Economics and Statistics*, 239(2):345–360.

Gottschalk, P. and Huynh, M. (2010). Are earnings inequality and mobility overstated? the impact of nonclassical measurement error. *The Review of Economics and Statistics*, 92(2):302–315.

Grabka, M. M. (2022). SOEP-Core v37-Codebook for the $PEQUIV File 1984-2020: CNEF variables with extended income information for the SOEP. Technical report, SOEP Survey Papers.

Kapteyn, A. and Ypma, J. Y. (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economics*, 25(3):513–551.

Kim, C. and Tamborini, C. (2014). Response error in earnings. *Sociological Methods Research*, 43(1):39–72.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4):281–302.

Pischke, J.-S. (1995). Individual income, incomplete information, and aggregate consumption. *Econometrica*, 63(4):805.

Roth, A. and Slotwinski, M. (2020). Gender norms and income misreporting within households. *ZEW-Centre for European Economic Research Discussion Paper*, 20-001.

SOEP (2022). Soep-Core – 2020: Person (CAPI, mit Verweis auf Variablen). In *SOEP Survey Papers*, number 1135 in Series A – Survey Instruments (Erhebungsinstrumente). Berlin: DIW Berlin/SOEP.

Tourangeau, R., Rips, L., and Rasinski, K. (2010). *The Psychology of Survey Response*, pages 315–346. Cambridge University Press, Bingley.

# A Appendix

## A.1 Median income and insurance cap

Table A1: Relevant annual thresholds from pension law. Median annual earnings per year in Euros from SGB VI Appendix 1, maximum insurable annual earnings per year in Euros from SGB VI, Appendix 2. Maximum number of pension points derives from the ratio of maximum insurable annual earnings to average annual income, rounded to four decimal points.

| Year | Median earnings in Euros | Max. insurable earnings in Euros | Max. number of pension points |
|---|---:|---:|---:|
| 2020 | 39167 | 82800 | 2.1140 |
| 2019 | 39301 | 80400 | 2.0457 |
| 2018 | 38212 | 78000 | 2.0412 |
| 2017 | 37077 | 76200 | 2.0552 |
| 2016 | 36187 | 74400 | 2.0560 |
| 2015 | 35363 | 72600 | 2.0530 |
| 2014 | 34514 | 71400 | 2.0687 |
| 2013 | 33659 | 69600 | 2.0678 |
| 2012 | 33002 | 67200 | 2.0362 |
| 2011 | 32100 | 66000 | 2.0561 |
| 2010 | 31144 | 66000 | 2.1192 |

## A.2 Mincer equation, year-wise

Table A2: Year-wise Mincer Equations and Test of Differences

| Year | | Intercept | Education | Work Exp. | Sq. Work Exp. | $R^2$ (adj.) | $N$ |
|---|---|---|---|---|---|---|---|
| 2020 | SOEP | 6.7524*** | 0.0698*** | 0.0168*** | -0.0001 | 0.1836 | 4062 |
| | | (0.0411) | (0.0028) | (0.0024) | (0.0001) | | |
| | VSKT | 6.6864*** | 0.0745*** | 0.0196*** | -0.0001* | 0.1903 | 4062 |
| | | (0.0435) | (0.0030) | (0.0026) | (0.0001) | | |
| | Diff. | 0.0660 | -0.0047 | -0.0028 | 0.0000 | | |
| 2019 | SOEP | 6.8038*** | 0.0625*** | 0.0160*** | -0.0001 | 0.1603 | 4068 |
| | | (0.0420) | (0.0029) | (0.0024) | (0.0001) | | |
| | VSKT | 6.7775*** | 0.0667*** | 0.0197*** | -0.0001* | 0.1672 | 4068 |
| | | (0.0446) | (0.0031) | (0.0026) | (0.0001) | | |
| | Diff. | 0.0263 | -0.0041 | -0.0037 | 0.0001 | | |
| 2018 | SOEP | 6.7594*** | 0.0614*** | 0.0191*** | -0.0002** | 0.1522 | 4282 |
| | | (0.0416) | (0.0029) | (0.0024) | (0.0001) | | |
| | VSKT | 6.7088*** | 0.0665*** | 0.0242*** | -0.0002*** | 0.1549 | 4282 |
| | | (0.0460) | (0.0031) | (0.0028) | (0.0001) | | |
| | Diff. | 0.0507 | -0.0051 | -0.0051 | 0.0001 | | |
| 2017 | SOEP | 6.7267*** | 0.0610*** | 0.0185*** | -0.0001* | 0.1497 | 4025 |
| | | (0.0435) | (0.0030) | (0.0025) | (0.0001) | | |
| | VSKT | 6.6693*** | 0.0671*** | 0.0216*** | -0.0002* | 0.1573 | 4025 |
| | | (0.0472) | (0.0032) | (0.0028) | (0.0001) | | |
| | Diff. | 0.0574 | -0.0061 | -0.0031 | 0.0000 | | |
| 2016 | SOEP | 6.6334*** | 0.0634*** | 0.0204*** | -0.0001* | 0.1632 | 2939 |
| | | (0.0509) | (0.0035) | (0.0029) | (0.0001) | | |
| | VSKT | 6.5915*** | 0.0681*** | 0.0243*** | -0.0002** | 0.1642 | 2939 |
| | | (0.0550) | (0.0038) | (0.0032) | (0.0001) | | |
| | Diff. | 0.0419 | -0.0046 | -0.0039 | 0.0001 | | |
| 2015 | SOEP | 6.6398*** | 0.0624*** | 0.0194*** | -0.0001 | 0.1569 | 2800 |
| | | (0.0550) | (0.0037) | (0.0030) | (0.0001) | | |
| | VSKT | 6.5970*** | 0.0667*** | 0.0254*** | -0.0002** | 0.161 | 2800 |
| | | (0.0562) | (0.0038) | (0.0033) | (0.0001) | | |
| | Diff. | 0.0429 | -0.0043 | -0.0060 | 0.0001 | | |
| 2014 | SOEP | 6.5901*** | 0.0629*** | 0.0221*** | -0.0002** | 0.154 | 2702 |
| | | (0.0555) | (0.0037) | (0.0029) | (0.0001) | | |
| | VSKT | 6.5067*** | 0.0691*** | 0.0274*** | -0.0003*** | 0.153 | 2702 |
| | | (0.0614) | (0.0041) | (0.0032) | (0.0001) | | |
| | Diff. | 0.0834 | -0.0062 | -0.0052 | 0.0001 | | |

continued from previous page

| Year | | Intercept | Education | Work Exp. | Sq. Work Exp. | $R^2$ (adj.) | $N$ |
|---|---|---|---|---|---|---|---|
| 2013 | SOEP | 6.5074*** | 0.0637*** | 0.0253*** | -0.0003*** | 0.159 | 2635 |
| | | (0.0583) | (0.0039) | (0.0031) | (0.0001) | | |
| | VSKT | 6.4808*** | 0.0681*** | 0.0284*** | -0.0003*** | 0.1541 | 2635 |
| | | (0.0622) | (0.0041) | (0.0034) | (0.0001) | | |
| | Diff. | 0.0266 | -0.0043 | -0.0031 | 0.0001 | | |
| 2012 | SOEP | 6.4767*** | 0.0643*** | 0.0251*** | -0.0003*** | 0.147 | 2500 |
| | | (0.0596) | (0.0039) | (0.0032) | (0.0001) | | |
| | VSKT | 6.4316*** | 0.0684*** | 0.0320*** | -0.0004*** | 0.1455 | 2500 |
| | | (0.0642) | (0.0041) | (0.0036) | (0.0001) | | |
| | Diff. | 0.0451 | -0.0041 | -0.0068 | 0.0001 | | |
| 2011 | SOEP | 6.4391*** | 0.0628*** | 0.0281*** | -0.0004*** | 0.1391 | 2062 |
| | | (0.0668) | (0.0045) | (0.0035) | (0.0001) | | |
| | VSKT | 6.4200*** | 0.0652*** | 0.0331*** | -0.0005*** | 0.1356 | 2062 |
| | | (0.0725) | (0.0049) | (0.0039) | (0.0001) | | |
| | Diff. | 0.0191 | -0.0024 | -0.0050 | 0.0001 | | |
| 2010 | SOEP | 6.4736*** | 0.0575*** | 0.0314*** | -0.0004*** | 0.124 | 1378 |
| | | (0.0829) | (0.0055) | (0.0046) | (0.0001) | | |
| | VSKT | 6.4001*** | 0.0632*** | 0.0383*** | -0.0006*** | 0.1296 | 1378 |
| | | (0.0896) | (0.0058) | (0.0051) | (0.0001) | | |
| | Diff. | 0.0734 | -0.0056 | -0.0069 | 0.0001 | | |

Robust standard errors in parentheses.
$^{*}p < 0.05,^{**} p < 0.01,^{***} p < 0.001$

## A.3 Mincer equation analogously to Gauly et al. (2020)

Gauly et al. (2020) estimate a Mincer equation using nominal labor earnings as the dependent variable. For reasons of comparison, we estimate an identical specification and report the results below in addition to the original Mincer-formulation that entails a logarithmized earnings variable, which is reported in the main text (cf. Section 4.3). Table A3 provides an overview of the estimation results for the eleven sample years (2010 - 2020, descending order in table).

Table A3: Year-wise Mincer Equations and Test of Differences

| Year | | Education | Work. Exp. | Work. Exp. (sq.) | $R^2$ (adj.) | $N$ |
|---|---|---|---|---|---|---|
| 2020 | SOEP | 189.54*** | 38.49*** | -0.08 | 0.86 | 4063 |
| | | (3.88) | (5.76) | (0.14) | | |
| | VSKT | 196.55*** | 38.72*** | -0.02 | 0.86 | 4063 |
| | | (4.24) | (6.27) | (0.16) | | |
| | Diff. | -7.01 | -0.23 | -0.06 | | |

continued from previous page

| Year | | Education | Work. Exp. | Work. Exp. (sq.) | $R^2$ (adj.) | $N$ |
|---|---|---|---|---|---|---|
| 2019 | SOEP | 177.29*** | 37.33*** | -0.08 | 0.86 | 4069 |
| | | (3.73) | (5.49) | (0.14) | | |
| | VSKT | 188.93*** | 43.89*** | -0.11 | 0.86 | 4069 |
| | | (4.22) | (6.27) | (0.16) | | |
| | Diff. | -11.64 | -6.56 | 0.03 | | |
| 2018 | SOEP | 165.17*** | 46.92*** | -0.38** | 0.86 | 4283 |
| | | (3.40) | (5.07) | (0.13) | | |
| | VSKT | 177.80*** | 51.60*** | -0.40** | 0.85 | 4283 |
| | | (3.88) | (5.81) | (0.15) | | |
| | Diff. | -12.64 | -4.68 | 0.02 | | |
| 2017 | SOEP | 160.00*** | 41.61*** | -0.23 | 0.86 | 4026 |
| | | (3.45) | (5.25) | (0.13) | | |
| | VSKT | 170.98*** | 44.48*** | -0.19 | 0.85 | 4026 |
| | | (3.87) | (5.94) | (0.15) | | |
| | Diff. | -10.98 | -2.86 | -0.04 | | |
| 2016 | SOEP | 153.28*** | 41.63*** | -0.24 | 0.86 | 2939 |
| | | (3.97) | (5.90) | (0.15) | | |
| | VSKT | 162.94*** | 47.22*** | -0.30 | 0.85 | 2939 |
| | | (4.39) | (6.56) | (0.17) | | |
| | Diff. | -9.67 | -5.59 | 0.05 | | |
| 2015 | SOEP | 150.12*** | 40.49*** | -0.22 | 0.86 | 2800 |
| | | (3.89) | (5.88) | (0.15) | | |
| | VSKT | 158.48*** | 51.13*** | -0.39* | 0.85 | 2800 |
| | | (4.34) | (6.65) | (0.17) | | |
| | Diff. | -8.36 | -10.64 | 0.17 | | |
| 2014 | SOEP | 142.91*** | 46.30*** | -0.41** | 0.86 | 2704 |
| | | (3.76) | (5.59) | (0.14) | | |
| | VSKT | 148.40*** | 55.71*** | -0.54*** | 0.85 | 2704 |
| | | (4.06) | (6.15) | (0.16) | | |
| | Diff. | -5.48 | -9.42 | 0.14 | | |
| 2013 | SOEP | 134.81*** | 47.05*** | -0.41** | 0.86 | 2637 |
| | | (3.70) | (5.52) | (0.14) | | |
| | VSKT | 144.76*** | 52.04*** | -0.48** | 0.85 | 2637 |
| | | (4.10) | (6.14) | (0.16) | | |
| | Diff. | -9.95 | -4.99 | 0.06 | | |
| 2012 | SOEP | 131.92*** | 45.76*** | -0.43** | 0.86 | 2501 |
| | | (3.72) | (5.62) | (0.15) | | |
| | VSKT | 137.92*** | 58.38*** | -0.69*** | 0.85 | 2501 |
| | | (4.07) | (6.23) | (0.16) | | |
| | Diff. | -6 | -12.62 | 0.26 | | |

continued from previous page

| Year | | Education | Work. Exp. | Work. Exp. (sq.) | $R^2$ (adj.) | $N$ |
|---|---|---|---|---|---|---|
| 2011 | SOEP | 125.13*** | 47.83*** | -0.54*** | 0.86 | 2062 |
| | | (4.01) | (6.03) | (0.16) | | |
| | VSKT | 130.27*** | 58.46*** | -0.75*** | 0.85 | 2062 |
| | | (4.44) | (6.65) | (0.17) | | |
| | Diff. | -5.14 | -10.63 | 0.22 | | |
| 2010 | SOEP | 116.76*** | 58.02*** | -0.81*** | 0.86 | 1379 |
| | | (5.07) | (7.79) | (0.21) | | |
| | VSKT | 122.89*** | 68.90*** | -1.04*** | 0.85 | 1379 |
| | | (5.48) | (8.38) | (0.22) | | |
| | Diff. | -6.13 | -10.88 | 0.23 | | |

Robust standard errors in parentheses.
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

## A.4 Summary Statistics for All Waves

Table A4: Yearly Characteristics of SOEP 37, SOEP-RV.VKST2020 and Final Sample

| Year | | SOEP 37 | SOEP-VSKT | Final Sample |
|---|---|---|---|---|
| 2010 | Female (%) | 48.16 | 49.78 | 49.46 |
| | Age (mean) | 42.45 | 43.16 | 42.41 |
| | Years of education (mean) | 12.79 | 12.97 | 12.51 |
| | Tenure with firm (mean) | 10.44 | 10.45 | 10.85 |
| | Survey earnings (mean) | 2300.52 | 2125.34 | 2148.05 |
| | $N$ | 13101 | 2774 | 1379 |
| 2011 | Female (%) | 50.44 | 52.45 | 52.13 |
| | Age (mean) | 42.96 | 43.72 | 42.97 |
| | Years of education (mean) | 12.82 | 12.87 | 12.47 |
| | Tenure with firm (mean) | 10.55 | 10.70 | 10.78 |
| | Survey earnings (mean) | 2374.66 | 2205.97 | 2174.86 |
| | $N$ | 14953 | 3878 | 2062 |
| 2012 | Female (%) | 51.03 | 52.80 | 53.10 |
| | Age (mean) | 43.38 | 44.16 | 43.28 |
| | Years of education (mean) | 12.78 | 12.86 | 12.46 |
| | Tenure with firm (mean) | 10.55 | 10.64 | 10.76 |
| | Survey earnings (mean) | 2425.26 | 2289.23 | 2268.62 |
| | $N$ | 15048 | 4477 | 2501 |
| 2013 | Female (%) | 50.92 | 53.21 | 53.77 |
| | Age (mean) | 42.78 | 44.39 | 43.76 |
| | Years of education (mean) | 12.54 | 12.83 | 12.43 |
| | Tenure with firm (mean) | 9.71 | 10.50 | 10.69 |
| | Survey earnings (mean) | 2412.50 | 2348.69 | 2332.08 |
| | $N$ | 17031 | 4751 | 2637 |
| 2014 | Female (%) | 51.11 | 53.21 | 54.55 |
| | Age (mean) | 43.52 | 44.79 | 44.05 |
| | Years of education (mean) | 12.63 | 12.84 | 12.44 |
| | Tenure with firm (mean) | 10.13 | 10.71 | 10.85 |
| | Survey earnings (mean) | 2537.80 | 2487.33 | 2432.29 |
| | $N$ | 15029 | 4787 | 2704 |
| 2015 | Female (%) | 50.63 | 53.25 | 54.93 |
| | Age (mean) | 43.43 | 44.97 | 44.34 |
| | Years of education (mean) | 12.58 | 12.83 | 12.44 |
| | Tenure with firm (mean) | 9.82 | 10.66 | 10.71 |
| | Survey earnings (mean) | 2621.10 | 2519.62 | 2494.80 |
| | $N$ | 14570 | 4909 | 2800 |

continued from previous page

| Year | | SOEP 37 | SOEP-VSKT | Final Sample |
|---|---|---|---|---|
| 2016 | Female (%) | 50.79 | 53.57 | 54.71 |
| | Age (mean) | 43.74 | 45.23 | 44.67 |
| | Years of education (mean) | 12.55 | 12.80 | 12.39 |
| | Tenure with firm (mean) | 9.79 | 10.66 | 10.76 |
| | Survey earnings (mean) | 2639.51 | 2584.41 | 2539.25 |
| | $N$ | 13787 | 4995 | 2939 |
| 2017 | Female (%) | 49.87 | 52.99 | 54.55 |
| | Age (mean) | 43.88 | 45.16 | 44.69 |
| | Years of education (mean) | 12.59 | 12.83 | 12.45 |
| | Tenure with firm (mean) | 9.99 | 10.84 | 10.96 |
| | Survey earnings (mean) | 2707.45 | 2681.75 | 2648.70 |
| | $N$ | 15734 | 6786 | 4026 |
| 2018 | Female (%) | 50.00 | 52.72 | 54.45 |
| | Age (mean) | 43.99 | 45.28 | 44.81 |
| | Years of education (mean) | 12.58 | 12.83 | 12.47 |
| | Tenure with firm (mean) | 9.88 | 10.73 | 10.74 |
| | Survey earnings (mean) | 2803.29 | 2782.80 | 2740.55 |
| | $N$ | 15132 | 7178 | 4283 |
| 2019 | Female (%) | 47.81 | 53.18 | 55.12 |
| | Age (mean) | 45.17 | 45.68 | 45.12 |
| | Years of education (mean) | 12.78 | 12.89 | 12.53 |
| | Tenure with firm (mean) | 10.80 | 10.83 | 10.79 |
| | Survey earnings (mean) | 3618.96 | 2892.42 | 2866.01 |
| | $N$ | 15757 | 6754 | 4069 |
| 2020 | Female (%) | 47.35 | 53.23 | 54.96 |
| | Age (mean) | 44.16 | 45.89 | 45.47 |
| | Years of education (mean) | 12.61 | 12.94 | 12.63 |
| | Tenure with firm (mean) | 9.94 | 10.92 | 11.03 |
| | Survey earnings (mean) | 3388.37 | 3023.33 | 3059.09 |
| | $N$ | 15690 | 6728 | 4063 |

## A.5 Cleaning Steps

Table A5: Cleaning Steps, Number of Observations of Measurement Error, Number of Unique Individuals in the Sample

| Step | Description | Observations | Individuals |
|---|---|---|---|
| 0 | Initial sample | 130227 | 12054 |
| 1 | Matching sex information [SOEP (sex), VSKT (gevs)] | 130180 | 12051 |
| 2 | Matching birth year observation [SOEP (gebjahr), VSKT (gbjavs)] | 130147 | 12041 |
| 3 | No proxy interview [SOEP (pgmode)] | 130128 | 12041 |
| 4 | Retain observations from 2010 on [SOEP (syear)] | 95191 | 12041 |
| 5 | Only dependently employed individuals [SOEP (pgstib)] | 84738 | 11470 |
| 6 | Full-time or part-time employment [SOEP (pgemplst)] | 43942 | 7623 |
| 7 | Retain observations from February – November [SOEP (pgmonth)] | 43901 | 7621 |
| 8 | No self-employed [VSKT (status_1)] | 43899 | 7619 |
| 9 | No self-employed [SOEP (plb0568_v1)] | 43891 | 7617 |
| 10 | No self-employed [SOEP (plb0719)] | 43798 | 7599 |
| 11 | Drop high-income sample members [SOEP (psample)] | 42551 | 7405 |
| 12 | Admin. earnings below cap (West Germany) [VSKT] | 40386 | 7342 |
| 13 | Admin. earnings below cap (West Germany, Knappsch.) [VSKT] | 40384 | 7342 |
| 14 | Admin. earnings below cap (East Germany) [VSKT] | 40159 | 7340 |
| 15 | Admin earnings below cap (East Germany, Knappsch.) [VSKT] | 40159 | 7340 |
| 16 | Survey earnings below cap (West Germany) [SOEP] | 39714 | 7228 |
| 17 | Survey earnings below cap (West Germany, Knappsch.) [SOEP] | 39713 | 7227 |
| 18 | Survey earnings below cap (East Germany) [SOEP] | 39632 | 7214 |
| 19 | Survey earnings below cap (East Germany, Knappsch.) [SOEP] | 39632 | 7214 |
| 20 | No missing employment [VSKT (status_1)] | 36307 | 6517 |
| 21 | No missing survey income [SOEP (plc0013)] | 34775 | 6423 |
| 22 | No missing admin. earnings [VSKT] | 34775 | 6423 |
| 23 | No missing survey earnings [SOEP] | 34775 | 6423 |
| 24 | Observed ME [SOEP, VSKT] | 34427 | 6382 |
| 25 | Observed gender [SOEP (sex), VSKT (gevs)] | 34427 | 6382 |
| 26 | Observed age [SOEP (gebjahr), VSKT (gbjavs)] | 34427 | 6382 |

continued from previous page

| Step | Description | Observations | Individuals |
|---|---|---|---|
| 27 | Information on migration background (direct) [SOEP (migback)] | 34427 | 6382 |
| 28 | Information on migration background (indirect) [SOEP (migback)] | 34427 | 6382 |
| 29 | Information on sample region (East-Germany dummy) [SOEP (sampreg)] | 34427 | 6382 |
| 30 | Observed household size [SOEP (hhsize)] | 34427 | 6382 |
| 31 | Observed number of children [SOEP] | 34427 | 6382 |
| 32 | Observed relationship status [SOEP (parid)] | 34427 | 6382 |
| 33 | Observed marital status [SOEP (pgfamstd)] | 34385 | 6379 |
| 34 | Observed time of education [SOEP (pgbilzeit)] | 33707 | 6183 |
| 35 | Observed work experience [VSKT] | 33707 | 6183 |
| 36 | Observed experience with unemployment [SOEP (pgexpue)] | 33618 | 6162 |
| 37 | Observed civil service dummy [SOEP (pgoeffd)] | 33494 | 6158 |
| 38 | Observed tenure with the firm [SOEP (pgerwzeit)] | 33492 | 6157 |
| 39 | Observed observed occupational position [SOEP (pgstib)] | 33463 | 6153 |
| 40 | Observed assistance status of interview [SOEP (pgmode)] | 33463 | 6153 |

## A.6 Histogram of Price-Adjusted Measurement Error, Full Range

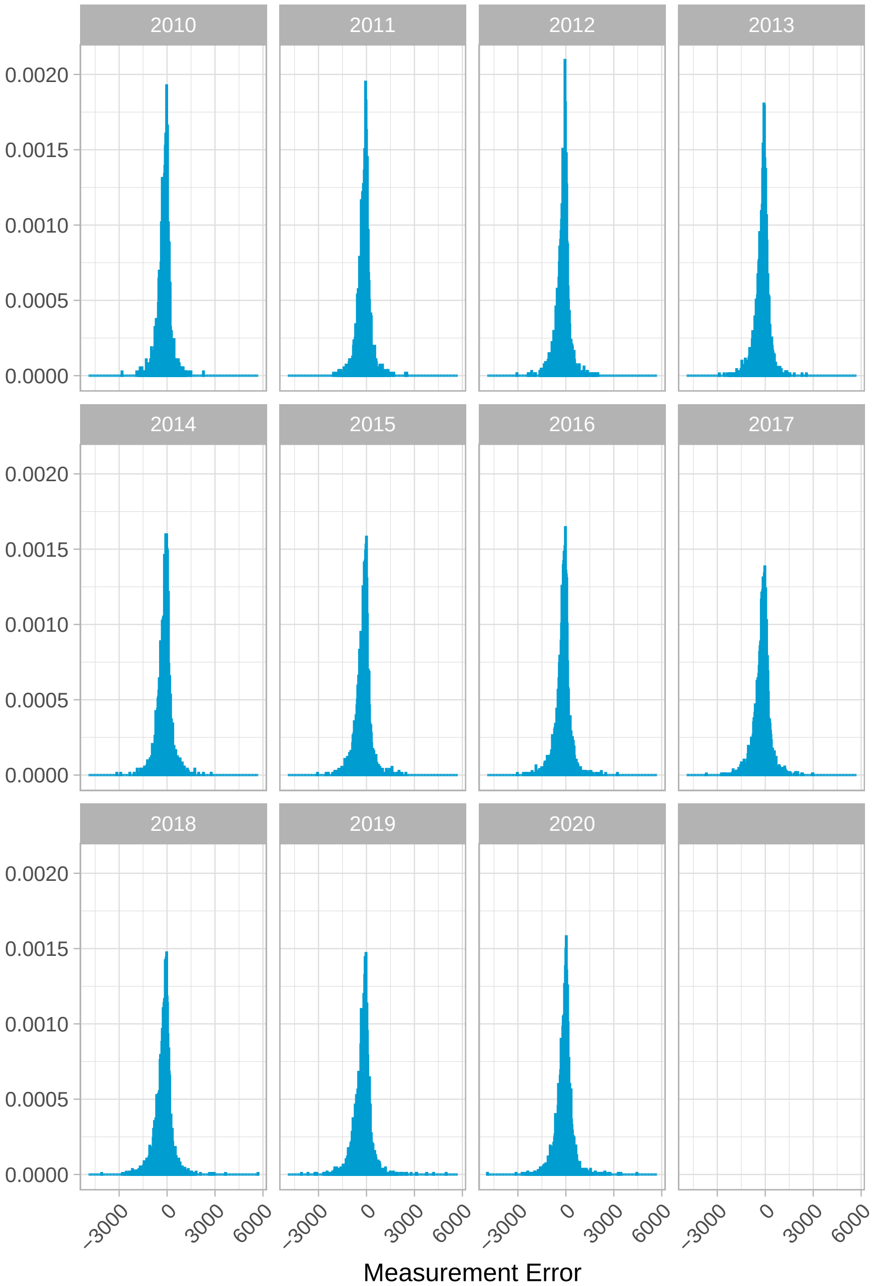


Figure A5: Histogram of the Price-adjusted Measurement Error for the Years 2010 to 2020.

## A.7 Scatterplot of Logarithmized True Earnings vs. Price-Adjusted Measurement Error, Full Range

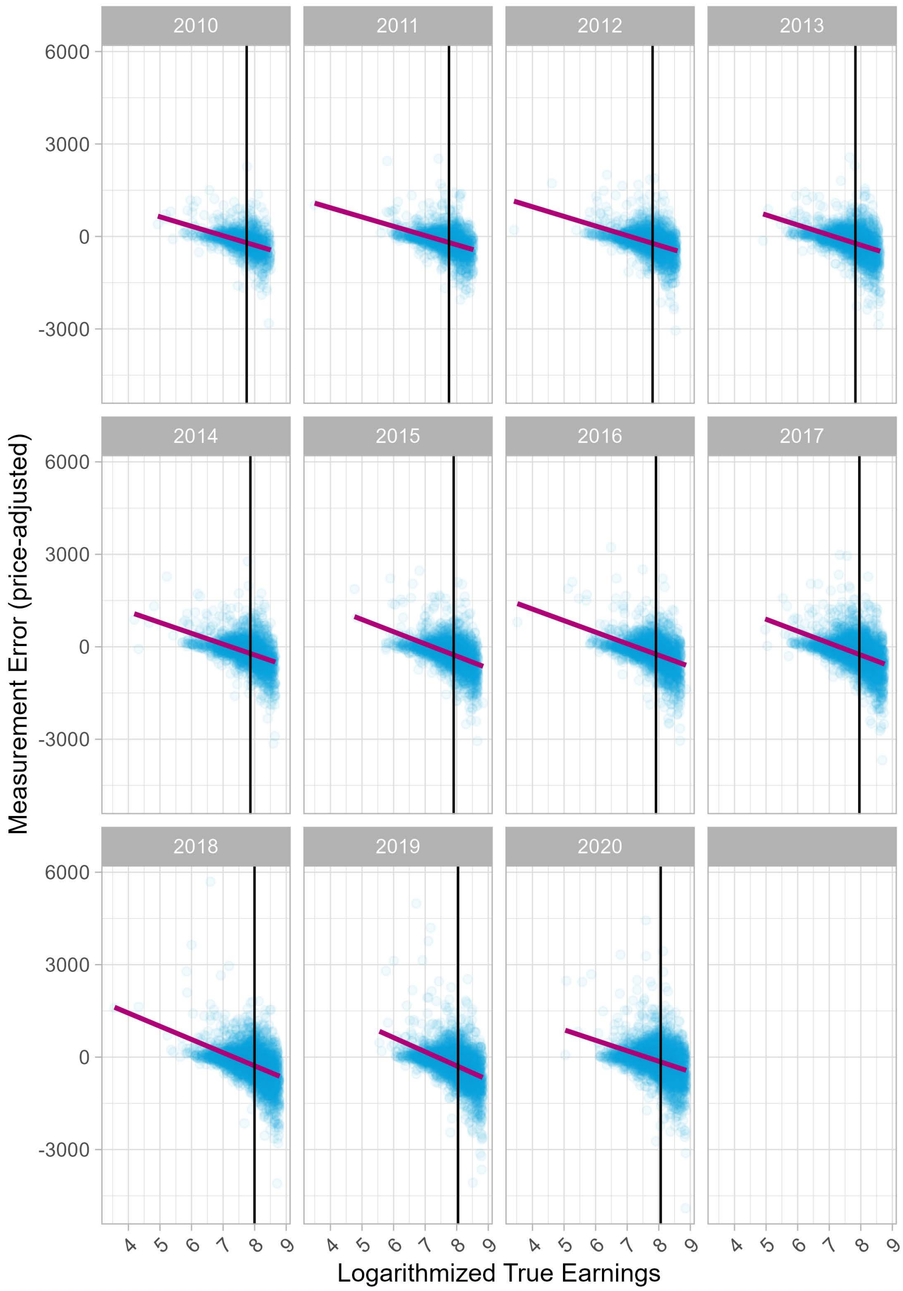


Figure A6: Scatterplot of the Price-adjusted Measurement Error for the Years 2010 to 2020. Unrestricted Range of Earnings. Black vertical line represents the mean log administrative earnings of each year.