Center for Quantitative Economics

We Might Both Be Wrong - Reconciliation of Survey and Administrative Earnings Measurements

Stella Martin [†]

110/2025

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

wissen.leben
Universität Münster

We Might Both Be Wrong - Reconciliation of Survey and Administrative Earnings Measurements

Stella Martin

Abstract

This paper investigates measurement error using a German linked survey-administrative dataset. By contributing to a small branch of the literature that allows the possibility of erroneous earnings from both sides of a linked survey-administrative dataset, this study provides a replication of two previously proposed versions of a representation of different combinations of (correct and incorrect) survey and administrative earnings in a finite mixture model. Additionally, the paper investigates the robustness of the results with respect to the empirical choice of what tolerance for the differences between to earnings measures should be accepted to consider them close enough to represent the true earnings. While the estimated distributions of true earnings and error types are relatively robust to the choice of the accepted tolerances, substantial differences in the estimated latent class probabilities are elicited, which highlights the importance of robustness checks for the choice of a tolerance level with alternative values above and below the one chosen in any future study.

1 Introduction

While it is commonly acknowledged that self-reported earnings from surveys are prone to bias and erroneous measurement, the literature on measurement error in earnings remains divided on whether administrative data sources can be a remedy for errors-in-variable or are potentially error-prone themselves. The literature on measurement error of labor earnings can thus be divided into two main branches. The first contains studies which assume that administrative records provide error-free measures of earnings, and in linked data sets, the administrative measurement can either validate or falsify the self-reported survey measurement of earnings. The term *validation studies* has been coined for investigations adopting this assumption. Seminal studies in this branch include Bound and Krueger (1991); Bound et al. (1994) and Pischke (1995). These studies use linked data that consists of a survey component and some form of administrative record - employer-kept payroll records, social security or tax records represent the the most commonly linked validation sources. The second –newer and smaller– branch of the literature does not use the assumption of administrative records providing accurate measures of true earnings. Initial studies, pioneered by Kapteyn and Ypma (2007), focused on the probability of error introduced through the linkage process that matches administrative records to the survey respondents and their self-reported earnings. This approach was recently extended to include the probability of inherent measurement error in administrative records, even in the case of correctly conducted matches (Jenkins and Rios-Avila, 2023b). What unites the studies regardless of their assumption on the types of error in administrative earnings (only linkage error versus linkage error and genuine mismeasurement) is their representation of earnings and the probability of measurement error in (finite) mixture models, also known as latent class models.

The present paper contributes to the second branch of the literature and is the first to investigate of measurement error in earnings, allowing for error in both components of a linked survey-administrative dataset from Germany. For German data, there are only two investigations of measurement error so far. Both Valet et al. (2018) and Gauly et al. (2019) are validation studies, so that there is no analysis for German data that allows the possibility of error in administrative data as well. Secondly, this study replicates of the analysis that Kapteyn and Ypma (2007) performed using a Swedish sample of older individuals (ages 50 and up). The present study can supplement their analysis by providing estimates from a sample that incorporates a broader range of ages, as the German data used in the present study does not impose any age restrictions for the adult respondents. The seminal innovation of the Kapteyn and Ypma (2007) model was its incorporation of the possibility of linkage error in the administrative data. We also provide a replication of the model extension of Jenkins and Rios-Avila (2023b), in which linkage error and genuine mismeasurement in administrative earnings are allowed, which has not been replicated previously. This paper also contributes to the field by providing empirical evidence on the consequences of a decision that has to be made in order to estimate any of these models that frame earnings observations as a finite mixture model: The tolerance with which two earnings measurements are deemed sufficiently close to be considered equal, or at least sufficiently close to each other, is a necessary assumption for identifying the group of the mixture that contains the error-free earning pairs. This paper investigates the robustness of results in the face of different tolerance assumptions and whether estimation results are dependent on the choice of accepted tolerance.

In Section 2, the literature that deals specifically with error sources in earnings on both sides of linked data is presented to provide an overview. The finite mixture model estimated in this paper using ML estimation is described in Section 3. A brief description of the linked survey-administrative data used in the estimation is provided in Section 4, before estimation results are presented in section 5. Section 6 concludes and provides an outlook on further research.

2 Literature

Existing literature on the field of measurement error in earnings can be divided in terms of their approach to conceptualization of the measurement error. The field is united in assuming that there is potential for erroneous earnings reports in surveys beyond classical error, which is incorporated in standard models as an additive, independent error term with constant variance. The discerning factor lies in the attitude studies express toward matched administrative data (from company payroll or government agencies) and the validity of the earnings information they report. The first branch of the literature uses administrative earnings as a source of validation of the earnings measured in a survey. The most common validation sources are governmental administrative records, ranging from income tax to social security records. All studies which assume that administrative earnings are error-free, and hence represent true earnings, are referred to as *validation studies*. The second branch of the measurement error literature considers there to be at least some degree of error in administrative earnings measurements. In essence, these studies consider the administrative earnings to be mainly a second, potentially erroneous measurement from a different source. These studies are referred to as *second-wave studies* (Jenkins and Rios-Avila, 2023b).

The initial study which not only acknowledges the inherent error potential of both income measurements, but incorporates it in their econometric model, is from Kapteyn and Ypma (2007). They can be considered the founders of the second wave of studies by incorporating the possibility of errors in both measurements of earnings. Beyond earnings, their Swedish subsample of the Survey on Health, Ageing and Retirement in Europe (SHARE) linked to the Swedish registers also allows for considering measurement error in pensions and the amount of taxes paid. For the administrative side, they consider earnings (and pensions and tax) information to be either correct or incorrect, due to linkage error from matching an incorrect person's record to a survey observation. They do not, however, incorporate inherent mismeasurement of earnings in records on the administrative side. Another investigation of the same linked Swedish data as in Kapteyn and Ypma (2007) is conducted by Meijer et al. (2012). They further investigate the proposed model and show that in the presence of even a small probability of linkage error in administrative records, the survey data delivers more reliable estimates, where the reliability of a data source is measured as the normalized squared correlation of true earnings and the observed earnings measure.

The full Kapteyn and Ypma model, although laying the foundation for the literature branch considering the faultiness of earnings reports from either source in linked data, is not without criticism. Bee and Rothbaum (2019) hypothesize that including the probability of misreported earnings in administrative data would lower the size their estimates on linkage error. They imply that in the Kapteyn and Ypma (2007) model, in the absence of a possible misspecification error in administrative earnings, estimates on linkage error in the administrative data capture misspecification as well, hence inflating the estimate on linkage error. They hypothesize that including the probability of erroneous administrative earnings measurement, beyond linkage error, would reduce the estimates of mis-linkage. In estimating their bivariate (finite) mixture model, Kapteyn and Ypma (2007) define the first distribution as the one which contains correct observations, i.e. representing the true earnings. A pair of administrative and survey earnings is labelled as part of that group if their absolute difference is tolerable, where the tolerance is defined as 1000 Swedish Crowns per year. By determining this tolerance, they assume that 14.8% of their sample provides error-free earnings. Thus, these observations are defined as part of the first group, from which the distribution of true earnings is estimated. Jenkins and Rios-Avila (2020) criticize that 14.8% is too large a share to assume to be free of error. They propose a different formulation for the tolerance, where the absolute difference in logarithmized earnings is used to re-estimate the Kapteyn and Ypma full model and replicate it with different, mostly smaller, tolerances. Using UK linked data, they find that robustness issues arise for smaller tolerances.

Jenkins and Rios-Avila (2023b) address the criticism from Bee and Rothbaum (2019) and extend the model initially formulated by **?** to include the possibility of (correctly linked) administrative earnings records being subject to inherent misreporting. In essence, administrative earnings records are reported to the authorities for official purposes such as tax collection, unemployment insurance or retirement calculation by some non-official source, often by the employer. This extension increases the number of latent classes for the mixture model from six, in Kapteyn and Ypma (2007), to a total of nine. See section 3 for a detailed description of the model. As in the estimation of Kapteyn and Ypma (2007), an assumption about the tolerance applied to determine which observations are considered to provide (almost) correct earnings has to be made. While Jenkins and Rios-Avila (2023b) continue to use the tolerance definition introduced in their replication of the original Kapteyn and Ypma model, they fail to provide sufficient robustness checks of their tolerance assumption for the extension of the original model that they themselves propose.

Abowd and Stinson (2013) present an approach to dealing with the possibly erroneous nature of both measurements of earnings, but it does not directly compare to other studies in the field, as it does not include a concept of (unobserved) true earnings. The two measurements earnings provided by their US dataset (survey component: Survey of Income an Program Participation by the US Census Bureau, administrative part: Detailed Earnings Record from the Social Security Administration) are weighted using a weighting vector of length two. Different variations of the weighting vector are used for estimation, essentially depicting an individual's earnings as a weighted mean of two measurements from different sources. Consequently, this approach only allows for the weighted earnings measure to lie within the bounds of the two measurements. This does not account for the possibility that, in light of two erroneous measurements, the true earnings, although not explicitly conceptualized, might be larger or smaller than both measurements.

Bingley and Martinello (2017) investigate measurement error in income and schooling for the estimation of ordinary least squares (OLS) and instrumental variables (IV) regression using the Danish subsample of the Survey of Health, Ageing and Retirement (SHARE) and a corresponding linkage to Danish administrative registers. As mentioned above, SHARE only collects data for the population from 50 years of age, hence resulting in a bottom-censored sample with respect to age, like Kapteyn and Ypma (2007). Bingley and Martinello (2017) argue that (implicitly) assuming classical measurement error with additive, independent error and constant variance is convenient, since instrumental variable estimator are robust to classical measurement error in explanatory variables. In allowing measurement error in their supposed validation measurement of income, they find that the negative correlation of reported (gross annual) income and measurement error, which has been established as a central feature by previous validation studies like Bound and Krueger (1991) and Bound et al. (1994), does not occur in their data. They conclude that previous findings of the pattern called mean-reversion might be the result of erroneous validation data.

3 Methods

This section presents the model of Jenkins and Rios-Avila (2023b). It is an extended version of the model initially formulated by Kapteyn and Ypma (2007). The full model of the latter (and any restriction imposed there) is included in the formulation of the model and can be viewed as a special case of the extended model.

The logarithm of true, unobserved earnings is defined as ξ_i for individual i. It is assumed to be normally distributed according to $\xi_i \sim N\left(\mu_\xi, \sigma_\xi^2\right)$, where μ and σ represent the mean and standard deviation of the distribution. True earnings cannot be observed directly. Instead, two measurements of earnings are made for each individual. One stems from a survey and the other one from linked administrative records.

Survey earnings of an individual i are denoted as s_i. Equation 1 summarizes the three possible types of observation that make up the survey measurement of earnings. They can be correctly measured (type S1), which means they represent the true earnings ξ_i with probability π_s. Or, in cases of erroneous measurement, which take place with probability $1 - \pi_s$, the survey measurement of earnings consists of mean-reverting error with ρ_s representing the correlation of error and true earnings, and error term η_i (type S2). Additional contamination error, also referred to as reference period error, denoted ω_i, affects (erroneous) observations with probability π_ω. In the initial formulation of the model by Kapteyn and Ypma (2007), ω was considered to represent contamination. However, little explanation is provided beyond that it is supposed to capture erroneous reports resulting, for example, from misreporting yearly earnings as monthly or vice versa, omitting a job, or reporting employment that did not last throughout the entire year when providing annual earnings. Ultimately, contamination is used to differentiate between conceptual mismeasurement and other types of errors (due to social desirability, lack of adequate recall, rounding etc.). The type of contaminated observation to which type S3 refers was later framed as reference period error in the survey-reported measure of earnings by Jenkins and Rios-Avila (2023b).

$$s_i = \begin{cases} \xi_i & \text{with prob. } \pi_s & \text{(type S1)} \\ \xi_i + \rho_s(\xi_i - \mu_\xi) + \eta_i & \text{with prob. } (1-\pi_s)(1-\pi_\omega) & \text{(type S2)} \\ \xi_i + \rho_s(\xi_i - \mu_\xi) + \eta_i + \omega_i & \text{with prob. } (1-\pi_s)\pi_\omega & \text{(type S3)} \end{cases} \quad (1)$$

Earnings from administrative records for an individual are denoted as r_i. The administrative earnings measure can be free of error (with probability π_υ) and linked correctly (with probability π_r), thus equalling true earnings (type R1). Alternatively, administrative earnings can contain an erroneous measurement of earnings inherent to the record (type R2), where υ_i denotes that error and ρ_r is the correlation analogously defined to ρ_s above. The presence of linkage error is covered by observations of type R3 (with probability $1 - \pi_r$). A linkage error in administrative data is the result of earnings not belonging to the survey respondent, but to the records of another person from the record base, who might not necessarily belong to the survey sample. Equation 2 summarizes the mixture of three types for the administrative earnings measurement:

$$r_i = \begin{cases} \xi_i & \text{with prob. } \pi_r\pi_\upsilon & \text{(type R1)} \\ \xi_i + \rho_r(\xi_i - \mu_\xi) + \upsilon_i & \text{with prob. } \pi_r(1-\pi_\upsilon) & \text{(type R2)} \\ \zeta_i & \text{with prob. } (1-\pi_r)\pi_\upsilon & \text{(type R3)} \end{cases} \quad (2)$$

In summary, each observation (s_i, r_i) consists of two earnings measurements - one from a survey and one from administrative records. As each measurement can potentially be of three types (S1-S3 or R1-R3); there is thus a total of nine possible combinations $j = 1, \cdots, 9$ of administrative and survey measurement from which an observation can arise. These nine combinations of observation types make up the latent classes in the mixture model of earnings. The latent class probabilities are denoted as π_j and the density of earnings in each class as $f_j(r_i, s_i)$. Refer to Table 1 for an overview of the nine classes.

Some assumptions of normality are made to facilitate fitting the maximum likelihood estimation of the model. Distributionally, it is assumed that ζ_i, η_i

and υ_i are each independently normally distributed, while true earnings, ξ_i, and contamination from reference period error, ω_i, are bivariate normal, so that distributional features can be summarized as follows:

$$\begin{aligned} \begin{pmatrix} \xi_i \\ \omega_i \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu_\xi \\ \mu_\omega \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & \rho_\omega \sigma_\xi \sigma_\omega \\ \rho_\omega \sigma_\xi \sigma_\omega & \sigma_\omega^2 \end{pmatrix} \right), \\ \zeta_i \sim N\left(\mu_\zeta, \sigma_\zeta^2\right), \quad \eta_i &\sim N\left(\mu_\eta, \sigma_\eta^2\right), \quad \text{and} \quad \nu_i \sim N\left(\mu_\nu, \sigma_\nu^2\right), \end{aligned} \tag{3}$$

where μ denotes the mean, and σ the standard deviation of a distribution. ρ_ω is the correlation of true earnings and contamination from reference period error.

The fitting of the finite mixture model is done by maximum likelihood. The log-likelihood function of the model is maximized based on assumptions made on the first class, i.e. the group of correctly measured information, where $s_i = r_i = \xi_i$, i.e. the pair of observations is of type (R1, S1). In practice, it seems unlikely that the exact measurement will be made in both earnings measurements, especially if survey questionnaires restrict responses to integers. Empirically, a range of tolerance needs to be defined, within which two measurements of earnings are close enough to be considered approximately equal. Kapteyn and Ypma (2007) define a tolerance of 1000 SEK (Swedish Crowns) per year, equivalent to around 90 Euros at an exchange rate of 1 Euro = 11 SEK, for which observations are defined to be sufficiently close to each other. Observations fulfilling this criterion are referred to as "completely labelled" (Redner and Walker, 1984). Under this assumption, membership of the first class consisting of observations from type (R1, S1) is known and fully observable, rather than latent. The parameters of the (normal) distribution of true earnings ξ_i are thus identified through the first group of observations. The log-likelihood function of the mixture can thus be expressed as

$$\log \mathcal{L}_i(\theta, \Pi) = \sum_{i=1}^{n_1} \pi_1 \log\{f_1(\xi_i \mid \theta)\} + \sum_{i=n_1+1}^{N} \log\left(\sum_{j=2}^{9} \pi_j f_j(r_i, s_i \mid \theta)\right), \tag{4}$$

where $i = 1, \cdots, n_1$ are the completely labelled observations and $i = (n_1 + 1), \cdots, N$ are the remaining ones that belong to one of the remaining eight genuinely latent classes. θ is the vector of parameters for each latent class distribution and Π contains the probabilities $\pi_s, \pi_r, \pi_\omega, \pi_\upsilon$, which ultimately make up the class probabilities π_j as displayed in Table 1 (Jenkins and Rios-Avila, 2023a).

A share of 14.8% of the observations of yearly earnings in Kapteyn and Ypma (2007) fall within their definition of "close-enough" observations to represent true earnings. Jenkins and Rios-Avila (2023b) argue that 14.8% is a fairly large share of observations that are considered reasonably correct. Accordingly, they instead introduce the threshold of $|r_i - s_i| < 0.005$.[1] The importance of reference period error for UK data, as used by Jenkins and

[1] While stating in the published paper that this inequality is defined as $|r_i - s_i| < 0.005$,

Rios-Avila (2023b), might lie in the different culture regarding pay frequencies. Among employees in the UK, it is much more common to be paid weekly (53 paychecks) or fortnightly (26 paychecks per year), which renders reference period error more important than for other labor markets, where the share of monthly payrolls would be higher. In aligning with the tolerance definition of Jenkins and Rios-Avila, 2.55% of our observations have earnings measurements that are sufficiently close. This share is even smaller than the 3.4% reported by Jenkins and Rios-Avila (2023b).

With respect to the identifiability of the model, Kapteyn and Ypma (2007) refer to (a preprint of) Meijer and Ypma (2008), who provide proof of identification for a mixture of two normal distributions and argue that their model is merely a generalization of the one used in the proof. Yakowitz and Spragins (1968) provide a more general investigation into the identifiability and find that finite mixtures are identifiable if they are mixtures of (multivariate) Gaussian distributions, which is true for the given model. An additional fact that contributes to convenience in the estimation of the model is its dependence on only four probabilities to determine a total of nine latent classes.

the replication material that they provide as part of their introductory remarks indicates that they instead defined $|r_i - s_i| \leq 0.005$. The present study aligns itself with the latter. In order to ensure comparable results, this paper uses the formulation that sets ≤ 0.05 as an acceptable difference between the logarithmized earnings measures.

Table 1: Overview of Latent Classes in Extended Model, Table from Jenkins and Rios-Avila (2023b)

j	Observation type	Latent class probability, π_j	Class densities $f_j(r_i, s_i)$
1	R1, S1	$\pi_1 = \pi_r\pi_\upsilon\pi_s$	$N\left(\begin{pmatrix}\mu_\xi \\ \mu_\xi\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & \sigma_\xi^2 \\ \sigma_\xi^2 & \sigma_\xi^2\end{pmatrix}\right)$
2	R1, S2	$\pi_2 = \pi_r\pi_\upsilon(1-\pi_s)(1-\pi_\omega)$	$N\left(\begin{pmatrix}\mu_\xi \\ \mu_\xi+\mu_\eta\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & (1+\rho_s)\sigma_\xi^2 \\ (1+\rho_s)\sigma_\xi^2 & (1+\rho_s)^2\sigma_\xi^2+\sigma_\eta^2\end{pmatrix}\right)$
3	R1, S3	$\pi_3 = \pi_r\pi_\upsilon(1-\pi_s)\pi_\omega$	$N\left(\begin{pmatrix}\mu_\xi \\ \mu_\xi+\mu_\eta+\mu_\omega\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & (1+\rho_s)\sigma_\xi^2+\rho_\omega\sigma_\xi\sigma_\omega \\ (1+\rho_s)\sigma_\xi^2+\rho_\omega\sigma_\xi\sigma_\omega & (1+\rho_s)^2\sigma_\xi^2+\sigma_\eta^2+\sigma_\omega^2+2\rho_\omega\sigma_\xi\sigma_\omega\end{pmatrix}\right)$
4	R2. S1	$\pi_4 = \pi_r(1-\pi_\upsilon)\pi_s$	$N\left(\begin{pmatrix}\mu_\xi+\mu_\upsilon \\ \mu_\xi\end{pmatrix}, \begin{pmatrix}(1+\rho_r)^2\sigma_\xi^2+\sigma_\upsilon^2 & (1+\rho_r)\sigma_\xi^2 \\ (1+\rho_r)\sigma_\xi^2 & \sigma_\xi^2\end{pmatrix}\right)$
5	R2, S2	$\pi_5 = \pi_r(1-\pi_\upsilon)(1-\pi_s)(1-\pi_\omega)$	$N\left(\begin{pmatrix}\mu_\xi+\mu_\upsilon \\ \mu_\xi+\mu_\eta\end{pmatrix}, \begin{pmatrix}(1+\rho_r)^2\sigma_\xi^2+\sigma_\upsilon^2 & (1+\rho_r)(1+\rho_s)\sigma_\xi^2 \\ (1+\rho_r)(1+\rho_s)\sigma_\xi^2 & (1+\rho_s)^2\sigma_\xi^2+\sigma_\eta^2\end{pmatrix}\right)$
6	R2, S3	$\pi_6 = \pi_r(1-\pi_\upsilon)(1-\pi_s)\pi_\omega$	$N\left(\begin{pmatrix}\mu_\xi+\mu_\upsilon \\ \mu_\xi+\mu_\eta+\mu_\omega\end{pmatrix}, \begin{pmatrix}(1+\rho_r)^2\sigma_\xi^2+\sigma_\upsilon^2 & (1+\rho_r)(1+\rho_s)\sigma_\xi^2 +(1+\rho_r)\rho_\omega\sigma_\xi\sigma_\omega \\ (1+\rho_r)(1+\rho_s)\sigma_\xi^2 +(1+\rho_r)\rho_\omega\sigma_\xi\sigma_\omega & (1+\rho_s)^2\sigma_\xi^2+\sigma_\eta^2 +\sigma_\omega^2+2\rho_\omega\sigma_\xi\sigma_\omega\end{pmatrix}\right)$
7	R3, S1	$\pi_7 = (1-\pi_r)\pi_s$	$N\left(\begin{pmatrix}\mu_\zeta \\ \mu_\xi\end{pmatrix}, \begin{pmatrix}\sigma_\zeta^2, 0 \\ 0, \sigma_\xi^2\end{pmatrix}\right)$
8	R3, S2	$\pi_8 = (1-\pi_r)(1-\pi_s)(1-\pi_\omega)$	$N\left(\begin{pmatrix}\mu_\zeta \\ \mu_\xi+\mu_\eta\end{pmatrix}, \begin{pmatrix}\sigma_\zeta^2 & 0 \\ 0 & (1+\rho_s)^2\sigma_\xi^2+\sigma_\eta^2\end{pmatrix}\right)$
9	R3, S3	$\pi_9 = (1-\pi_r)(1-\pi_s)\pi_\omega$	$N\left(\begin{pmatrix}\mu_\zeta \\ \mu_\xi+\mu_\eta+\mu_\omega\end{pmatrix}, \begin{pmatrix}\sigma_\zeta^2 & 0 \\ 0 & (1+\rho_s)^2\sigma_\xi^2+\sigma_\eta^2+\sigma_\omega^2+2\rho_\omega\sigma_\xi\sigma_\omega\end{pmatrix}\right)$

4 Data

For the empirical analysis in this paper, the linkage of the German Socioeconomic Panel (SOEP) with the insuree records held at the German statutory pension insurance (VSKT) is used. The latest available version of the SOEP data with corresponding linked records is version 37, and the most recent survey year is 2020. A detailed description is provided by Goebel et al. (2019). The administrative component of the data set used in this paper is called SOEP-RV.VSKT2020 (source: FDZ-RV). It includes the insuree account records held at the German statutory pension insurance. The labor market biography it provides includes the earnings records, which are available up to 2020.

The Socioeconomic Panel is a yearly survey. Its first interviews were conducted in 1984 and have always included a question on earnings from labor. Each respondent above 16 who completes a personal questionnaire is asked for their personal income from work in the month prior to the interview with the following question, which is then stored as `plc0013`:

> *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).

To maintain a harmonized definition of earnings over time, the interviewer instruction manual clarifies what to include in the answer:

> *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).

This is the most direct measurement of earnings provided by the survey. Alternative variables either include imputations or provide a measurement of earnings which refers to a different time frame, mostly annual earnings. The difference in time frames, which has been criticized by the measurement literature (Jenkins and Rios-Avila, 2023b) and implicitly requires information about the number of salary payments a respondent receives (besides the regular monthly salary this could include vacation or Christmas pay, bonuses etc.). The number of salary payments can only be identified through further survey-based information, which has a lower response rate than the earnings question. It would furthermore be necessary to either make assumptions about the number of paychecks for respondents with lacking data or to discard observations with uncertainty about the number of salary payments they received in a year.

The administrative component of the data contains a set of time-invariant variables and the earnings history stored in the insuree account at the German statutory pension insurance on a monthly basis. Anyone who is above

15 years of age and has been in a retirement-relevant state has an account with the German statutory pension insurance. Up to five parallel states (employment, child-rearing, care-provision, unemployment etc.) can be observed for each month as well as the corresponding gross earnings (for employment). These earnings are recorded in pension points rounded to three decimal points. At the end of each year, one pension point is awarded for the median earnings of the insuree base in that year. The pension insurance contribution rate is a share of gross earnings in Germany, half of which is borne by the employee and the other half by the employer. Employers deduct the amount to be paid by the employee from gross earnings and match the contribution. Employers are required to submit the necessary information to the social security system according to the legal ordinance called "Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung (Datenerfassungs- und -übermittlungsverordnung - DEÜV)". The pension insurance, together with other social security institutions in Germany conducts internal plausibility checks of the submitted information (Deutsche Rentenbersicherung, 2024). However, there is no guarantee for earnings to be transmitted without error. Employees receive yearly account information in the form of an informational mailing, but records can only be changed when insurees participate in a reconciliation of their pension account, which aims to clarify the insurance history including earnings. Insurees have to provide proof (payslips, tax statements etc.) that the record does not reflect their earnings before changes are made. The fact that the process of account clarification exists, can be interpreted as an implicit acknowledgement of the error potential of the pension records. However, it could take years from the point in time when an employer submits erroneous information to a final correction. The account reconciliation is generally voluntary and only has to be done when submitting a pension claim. Although errors in earnings records might be corrected towards the later stages of an individual's employment biography, it seems reasonable to relax the assumption of error-free administrative earnings measurements.

The pension points recorded in the insuree's record varies from month to month, which need not be a result of changing gross earnings, but of the pension insurance calculating calendar-daily earnings and multiplying them by the number of days in a month. This leads to a jitter in the data if the monthly pension points are transformed to nominal earnings, as the length of a month can be between 28 and 31 days. Thus, both earnings measures are transformed to calendar-daily earnings measured in pension points. One pension point is awarded for annual earnings of the amount of the median earnings in that calendar year and set by law[2]. After the difference in earnings is calculated in pension points, they are re-transferred to monthly earnings in Euro for reasons of interpretability, which are then price-adjusted with the consumer-price index from the SOEP's cross-national equivalence file.

The linkage of survey respondents to their pension records takes place in

[2]The German Pension Code details the procedure in book 6, §69, paragraph 2.

several steps. When the survey is conducted, respondents are asked to give consent for their pension records to be linked to their survey data. They then complete a short questionnaire in which they provide data required to generate their social security number, i.e. their full name, family name at birth, location and date of birth and, if they can and choose to, their insuree or pensioner number. At the end of the questionnaire they sign it in order to legally document their consent to the linkage. The form is displayed in Figure 1. It is collected by the survey institute, which also generates a person-identifier for the linkage and then reports the data from the linkage consent form to the pension insurance, and the interview data with the linkage identifier to the SOEP data department. The pension insurance then verifies the insuree or pensioner number and at which regional office an insuree's account is maintained. Each regional office receives a request to deliver a pseudonymized copy of the account data to the headquarters, where the data is anonymized and only the linkage identifier is kept alongside the record data before the final data product, i.e. the administrative component of the linkage, is released. For a complete description of the production process for the SOEP-VSKT data, please refer to the data security concept provided by Deutsche Rentenversicherung (nd).

It is important to note that there is a cap on the insurable income. The value is set annually, and only earnings below it are subject to social security contributions paid by employee and employer. Any earnings in excess of the cap are not insured. Table 2 provides an overview of the cap in the insurable income per year, as well as the maximum attainable earnings points this implies. Since earnings beyond the cap are not recorded beyond the value of the cap, this means that earnings information from the administrative records are top-censored. Also, there have been reports of inexplicable bunching just below the maximum insurable earnings (Bönke et al., 2015), which is the reason why earnings at or above 98% of each year's maximum insurable earnings are excluded from the sample.

Our final sample, with which we estimate the finite mixture model, only consists of complete observations. A complete observation is an individual's earnings report for which a corresponding administrative earnings record is available. If one of the two measurements of earnings is missing, that person's observation is discarded for the given period. Furthermore, any individual who reports being self-employed is eliminated, as they are not mandatorily covered and selection into coverage might take place that could also influence the measurement error. Observations for which either earnings measure is reported to be zero are discarded in line with Jenkins and Rios-Avila (2023b). Overall, the sample is made up of 33,690 complete observations and restricted to the years 2010 - 2020, from which the majority of observations stem. For the empirical analysis, all observations are pooled. In pooling a panel, several observation from different points in time for one individual are treated as if they were independent observations, even though observations for the same entity are likely correlated with each other and heteroskedasticity will likely be present. We

Leben in Deutschland

SOEP

Einwilligung zur Zusammenführung von Befragungsdaten der Studie „Leben in Deutschland“ mit den Kontendaten der Rentenversicherungsträger

Ich erkläre mich damit einverstanden, dass infas meine unten angegebenen Daten an die Deutsche Rentenversicherung Bund übermittelt, damit mein Rentenversicherungsträger meinen Versicherungsverlauf und die Höhe meiner Rente nach heutigem Stand ermitteln kann. Dies sind Angaben zum beitragspflichtigen Entgelt, Zeiten der Erwerbstätigkeit oder Erwerbslosigkeit (z.B. auf Grund von Arbeitslosigkeit) sowie Kindererziehungszeiten und Zeiten nicht erwerbsmäßiger Pflege. Das DIW Berlin und andere wissenschaftliche Nutzende der SOEP-Daten dürfen die von der Rentenversicherung ermittelten und anonymisierten Versicherungsdaten mit meinen Antworten aus dem von infas durchgeführten Interview zusammenführen und im Rahmen wissenschaftlicher Forschung auswerten. Die anonymisierten Daten lassen keinen direkten Bezug zu meiner Person zu. Ich habe die Information für Teilnehmende der Studie „Leben in Deutschland“ zur Kenntnis genommen. Ich behalte mir vor, jederzeit mein Einverständnis wieder zurückzuziehen. Hierzu genügt eine formlose Nachricht (E-Mail: LiD@infas.de; Tel.: 0800/66 77 876).

Familienname   Vorname

Geburtsname (Anfangsbuchstabe ist ausreichend)   Geburtsort

Geburtsdatum

Sie helfen uns bei der Ermittlung der Daten sehr, wenn Sie hier noch Ihre **Versicherungsnummer der gesetzlichen Rentenversicherung/Sozialversicherung** eintragen. Sie finden diese Nummer u.a. in Ihrem Sozialversicherungsausweis, bei Arbeitnehmern in der jährlichen Meldung Ihres Arbeitgebers an die Rentenversicherung (eine Kopie geht an Sie) oder neuerdings in den jährlichen Meldungen der Rentenversicherung, in den monatlichen Gehaltsabrechnungen oder den Rentenbescheiden bei Rentenbeziehern.

Personen, die noch keine Rente beziehen:

Versicherungsnummer

Rentner/innen:

PAN   Versicherungsnummer   BX/ZA

Sozialversicherungsnummer ist nicht bekannt ☐

Ich habe keine Sozialversicherungsnummer ☐

Bitte geben Sie das ausgefüllte Formular auch dann an uns zurück, wenn Sie Ihre Sozialversicherungsnummer nicht zur Hand haben.

Ort, Datum   Unterschrift

Vielen Dank!

Lfd.Nr.:

infas

infas Institut für angewandte Sozialwissenschaft GmbH

Postfach 240101
53154 Bonn
Tel. 0800/66 77 876
LiD@infas.de
www.leben-in-deutschland.de

Deutsche Rentenversicherung Bund

Deutsche Rentenversicherung Bund

10704 Berlin
www.fdz-rv.de

DIW

Deutsches Institut für Wirtschaftsforschung
Sozio-oekonomisches Panel (SOEP)

Mohrenstr. 58
10117 Berlin
www.diw.de/soep

7707/2023

Figure 1: Consent Form for Linkage of Pension Records

Table 2: 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 Euro	Max. insurable earnings in Euro	Max. number of pension points
2020	39,167	82,800	2.1140
2019	39,301	80,400	2.0457
2018	38,212	78,000	2.0412
2017	37,077	76,200	2.0552
2016	36,187	74,400	2.0560
2015	35,363	72,600	2.0530
2014	34,514	71,400	2.0687
2013	33,659	69,600	2.0678
2012	33,002	67,200	2.0362
2011	32,100	66,000	2.0561
2010	31,144	66,000	2.1192

follow the literature in the field, e.g. Kapteyn and Ypma (2007) and Jenkins and Rios-Avila (2023b), in computing cluster-robust standard errors, where the cluster variable is a respondent's household.

Since the present study replicates two previous studies, insuring harmonization is important in order to maintain a high level of comparability. For the estimation of their analysis, Jenkins and Rios-Avila (2023b) and Kapteyn and Ypma (2007) report models both with and without covariates. Their sets of covariates, however, differ from each other. Additionally, it is outlined that extensive covariate specifications of the finite mixture models may cause fitting problems and complicate achieving convergence (Jenkins and Rios-Avila, 2023b). Due to missing data, using different sets of covariates would complicate a comparison of model estimates across different covariate specifications. Most importantly, it is concluded that while increasing the overall model fit (Jenkins and Rios-Avila, 2023b), the inclusion of covariates does not substantially change the estimated results and yields conclusions that are qualitatively similar (Kapteyn and Ypma, 2007). The inclusion of covariates is therefore omitted in this work.

5 Results

5.1 Kapteyn and Ypma (2007) Replication

I replicate the analysis of Kapteyn and Ypma (2007) using our sample of linked German survey-administrative data consisting of 33690 cases from 6264 individuals. Since they also estimate restricted versions of their model, they refer to it as the *full model*. Since the model of Jenkins and Rios-Avila (2023b), as presented in Section 3, is an extension of the original Kapteyn and Ypma model, it can be restricted to represent the Kapteyn and Ypma model as a special case. While allowing for all three types of survey earnings measurement, administrative measurements of earnings can only be of type R1 or R3. This results in a total number of six latent classes: (R1, S1), (R1, S2), (R1, S3), (R2, S1), (R2, S2) and (R2, S3). The corresponding restrictions on parameters that lead to the original formulation of the full model by Kapteyn and Ypma (2007) are as follows. As there is assumed to be no error beyond pure linkage error in administrative earnings data in the original model, the probability of accurate measurement of earnings, given a record is correctly linked, is a certainty and thus $\pi_v = 1$. In turn, this implies $\mu_v = 0$ and $\sigma_v = 0$. Since there is no erroneous administrative earnings measure in this framework, there is also no ρ_r.

In order to make replication results as comparable as possible to the original study of Kapteyn and Ypma (2007), the tolerance used to define earnings measurements as close enough to each other to be considered part of class 1 is retained and converted with respect to the implied share of the sample. It was 1,000 Swedish Crowns for yearly earnings in the original study, which includes 14.8% of their sample. I transfer that share to a tolerance of 50 Euros in the data, within which 14.21% of the sample observations fall, which is my conversion of the original tolerance in Kapteyn and Ypma (2007). I complement the replication with other thresholds, i.e. the acceptable absolute difference of the two (non-logarithmized) earnings measurements is varied, and estimation results are reported for tolerances below (5, 10, 25, 33 Euros) and above (75, 100 Euros) the converted tolerance. Results for tolerances of the kind used in Jenkins and Rios-Avila (2023b) are also reported, where the absolute difference of logarithmized earnings is not to exceed 0.005. Estimates for this type of tolerance definition are complemented with re-estimations of the model for tolerances below (0.001, 0.0025) and above (0.0075, 0.01, 0.0015, 0.02) that. The shares of our sample that are assumed correct for each tolerance are presented in Table 3. With respect to the choice of the tolerance, it has to be kept in mind that for relatively large tolerances, the difference between and administrative and survey measurement of earnings can be smaller than the accepted tolerance, even if erroneous measurement is present. Ultimately, this would result in observations being (wrongly) classified as (R1,S1), even though they are actually part of another class that contains mismeasured earnings.

Estimates for the full Kapteyn and Ypma (2007) model are displayed in Tables 4 and 6, where the former presents the results for the tolerances ap-

Table 3: Sample Shares Within Varied Tolerances

Tolerance type	$\mid$ admin. earnings$_i$ − survey earnings$_i$ $\mid \leq$ ...						
Tolerance value	5	10	25	33	50	75	100
Share of obs. in %	1.46	2.83	7.13	9.44	14.21	20.55	26.32
Tolerance type	$\mid r_i - s_i \mid \leq$ ...						
Tolerance value	0.001	0.0025	0.005	0.0075	0.01	0.015	0.02
Share of obs. in %	0.49	1.39	2.82	4.21	5.68	8.61	11.6

plied to the first class that refer to differences in administrative and survey earnings measurements as in the original publication, and the latter contains estimation results for the tolerances that are defined as the difference in logarithmized earnings as in Jenkins and Rios-Avila (2023b).

With regard to the parameters of the distribution of true earnings (ξ) in the estimations of the full model with tolerance definitions according to Kapteyn and Ypma (2007), as reported in Table 4, it can be seen that threshold size does not seem to impact mean and variance estimates. The estimate for the mean of true earnings is the same across all tolerances except for the 100 Euro tolerance, for which it is slightly larger. The standard deviation of true earnings also seems mostly insensitive to the chosen tolerance, as it only varies at the fourth decimal digit, again except for the 100 Euro tolerance, for which the estimate differs from the others at the third and fourth decimal digit. For ζ, which denotes the earnings from a mismatched record, the estimates for the mean and standard deviation of the distribution decrease with increasing tolerance amounts, with the exception of the estimates for a tolerance of 100 Euros, which are larger than that for the previous tolerance of 75 Euros. Regarding η, denoting the measurement error in survey earnings reports, the observed patterns of estimates for mean and standard deviation across different thresholds are similar, in that for the highest threshold, there is a substantial deviation from what is estimated for the other thresholds. As the tolerance amount increases, the estimates for the mean decrease incrementally. The estimates for σ_η increase with the tolerance, except for the largest tolerance, for which the standard deviation drops by a third from 0.1246 for a tolerance of 75 Euros to 0.0473 for a tolerance of 100. However, all changes only affect the decimals of the estimates. ρ_s, which is a measure of mean reversion, indicates the presence of mean reversion in the data, but as with other estimates, the largest tolerance breaks the pattern of the estimate which decreases with increasing tolerance. While up to a tolerance of 75 Euros all estimates range between -0.0718 and -0.0827 and are estimated to be highly significant with $p < 0.001$, the estimated mean reversion for a tolerance of 100 Euros increases to -0.0062 and is only significant at $p < 0.05$.

The probability of correctly reported survey earnings π_s naturally increases with the tolerance amount, since allowing for a larger tolerance implies that a larger share of observations are assumed to be correct. The estimated probability of contamination in the survey earnings data decreases with an increasing tolerance until the amount of 75 Euros. For a tolerance of 100 Euros, the estimate more than triples from 18.07% to 63.5%. The probability of a correct match for the administrative earnings measurement is, on the other hand, estimated to be roughly the same size across all tolerances. Differences in the decimals of the estimates for π_r can be observed, but remain small overall.

Table 4: Parameter estimates, Kapteyn and Ypma (2007) Full Model, Varied Tolerances for non-logarithmized Earnings Differences

	(1) 5	(2) 10	(3) 25	(4) 33	(5) 50	(6) 75	(7) 100
μ_ξ	7.7895***	7.7895***	7.7895***	7.7895***	7.7895***	7.7895***	7.7910***
	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)
σ_ξ	0.5569***	0.5569***	0.5568***	0.5567***	0.5565***	0.5566***	0.5537***
	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0049)
μ_ζ	6.7167***	6.7046***	6.6730***	6.6577***	6.6312***	6.6135***	6.8602***
	(0.0657)	(0.0659)	(0.0665)	(0.0668)	(0.0676)	(0.0704)	(0.0541)
σ_ζ	0.8718***	0.8714***	0.8699***	0.8689***	0.8664***	0.8590***	0.8761***
	(0.0355)	(0.0359)	(0.0369)	(0.0374)	(0.0384)	(0.0398)	(0.0296)
μ_ω	-0.0364***	-0.0366***	-0.0364***	-0.0370***	-0.0365**	-0.0313*	0.0411***
	(0.0065)	(0.0069)	(0.0082)	(0.0090)	(0.0111)	(0.0155)	(0.0036)
σ_ω	0.2349***	0.2393***	0.2518***	0.2587***	0.2720***	0.2846***	0.2015***
	(0.0110)	(0.0115)	(0.0132)	(0.0141)	(0.0168)	(0.0229)	(0.0026)
μ_η	-0.0576***	-0.0584***	-0.0611***	-0.0624***	-0.0655***	-0.0706***	-0.1173***
	(0.0014)	(0.0015)	(0.0016)	(0.0017)	(0.0019)	(0.0026)	(0.0014)
σ_η	0.0992***	0.1014***	0.1075***	0.1109***	0.1174***	0.1246***	0.0473***
	(0.0023)	(0.0023)	(0.0025)	(0.0025)	(0.0028)	(0.0037)	(0.0009)
ρ_s	-0.0718***	-0.0729***	-0.0765***	-0.0784***	-0.0811***	-0.0827***	-0.0062*
	(0.0025)	(0.0026)	(0.0030)	(0.0032)	(0.0038)	(0.0051)	(0.0029)
π_s	0.0148***	0.0289***	0.0725***	0.0960***	0.1444***	0.2088***	0.2697***
	(0.0007)	(0.0010)	(0.0017)	(0.0020)	(0.0025)	(0.0031)	(0.0036)
π_ω	0.2521***	0.2433***	0.2216***	0.2105***	0.1923***	0.1807***	0.6350***
	(0.0226)	(0.0226)	(0.0229)	(0.0230)	(0.0244)	(0.0311)	(0.0097)
π_r	0.9803***	0.9807***	0.9816***	0.9821***	0.9827***	0.9833***	0.9728***
	(0.0018)	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0015)	(0.0019)
log L.	-16663.8813	-18945.6558	-24750.6421	-27398.4663	-32175.4469	-37396.6867	-40352.9450
AIC	33351.7625	37915.3116	49525.2842	54820.9326	64374.8939	74817.3734	80729.8900
BIC	33452.8620	38016.4110	49626.3837	54922.0321	64475.9934	74918.4729	80830.9895
No. cases	33690	33690	33690	33690	33690	33690	33690

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The resulting set of estimated latent class probabilities is presented in Table 5. Each probability of the six latent classes is a function of the error probabilities π_s, π_r and π_ω according to Equations 1 and 2. An exact specification of how to the latent class probabilities are computed is given in the third column of Table 1 for the extended Jenkins and Rios-Avila (2023b) model. Recall that since the Kapteyn and Ypma (2007) model does not include erroneous measurement of earnings in administrative records beyond linkage error, there is no R2-type observation, so that there are only six latent class probabilities for this model.

As mentioned previously, Jenkins and Rios-Avila (2020), in their replication of the Kapteyn and Ypma (2007) full model, employ a tolerance definition

Table 5: Latent Class Probabilities for Kapteyn and Ypma (2007) Full Model, Varied Tolerances for First Class: |admin. earnings − survey earnings| $\leq \cdots$

	5	10	25	33	50	75	100
π_1	0.0146***	0.0283***	0.0712***	0.0942***	0.1419***	0.2053***	0.2623***
	(0.0007)	(0.0010)	(0.0016)	(0.0019)	(0.0025)	(0.0031)	(0.0035)
π_2	0.7223***	0.7207***	0.7087***	0.7009***	0.6792***	0.6374***	0.2593***
	(0.0223)	(0.0220)	(0.0213)	(0.0208)	(0.0208)	(0.0246)	(0.0073)
π_3	0.2435***	0.2317***	0.2017***	0.1869***	0.1617***	0.1406***	0.4511***
	(0.0217)	(0.0214)	(0.0208)	(0.0204)	(0.0204)	(0.0242)	(0.0070)
π_4	0.0003***	0.0006***	0.0013***	0.0017***	0.0025***	0.0035***	0.0073***
	(0.0000)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0005)
π_5	0.0145***	0.0142***	0.0133***	0.0128***	0.0119***	0.0109***	0.0073***
	(0.0012)	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0010)	(0.0005)
π_6	0.0049***	0.0046***	0.0038***	0.0034***	0.0028***	0.0024***	0.0126***
	(0.0007)	(0.0007)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0009)

Robust standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

which differs from that in the original formulation, as it refers to the absolute difference in logarithmized earnings r_i and s_i. Also, all but one of tolerance amounts they use for their replication imply a much smaller share of observations to be considered part of group 1. Results for the re-estimation of the original Kapteyn and Ypma (2007) full model with seven different tolerance amounts, formulated based on logarithmized earnings, are displayed in Table 6. Although the replication of the full model by Jenkins and Rios-Avila (2020) includes a set of eight different thresholds, Jenkins and Rios-Avila (2023b) only report estimates for one threshold ($|r_i - s_i| \leq 0.005$) use when reporting their extended model. The present study contains estimates of the full model for a set of seven tolerances closely related to those used in Jenkins and Rios-Avila (2020).

Table 6 presents the estimation results for the full model using tolerances as in the replication by Jenkins and Rios-Avila (2020). Considering the mean and standard deviation of true earnings ξ, there is virtually no change in the estimates for the investigated tolerances up to 0.0015. Estimates for μ_ξ are either 7.7895, 7.7894 or 7.7893. The 0.02 tolerance value is the only one for which the estimate is 7.7891, which is a very small deviation from the other estimates. The estimates for the standard deviation $_\xi$ are the same for all four lower tolerances and increase slightly for tolerances from 0.01. For ζ, the estimates for the mean range from 6.6933 to 6.7269, becoming smaller as the tolerance amount increases. The estimates for σ_ζ are all close to each other, but also increase slightly for the four largest tolerances. They only differ at the third and fourth decimal, ranging from 0.8722 to 0.8756. As for the contamination ω, the observable pattern is also a slight increase over the estimates with growing tolerance. However, the last tolerance comes with a substantially larger increase and that estimate lacks significance. A similar pattern

Table 6: Parameter estimates, Kapteyn and Ypma (2007) Full Model, Varied Tolerances $|r_i - s_i| \leq \cdots$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	0.001	0.0025	0.005	0.0075	0.01	0.0015	0.02
μ_ξ	7.7895***	7.7895***	7.7895***	7.7894***	7.7894***	7.7893***	7.7891***
	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)
σ_ξ	0.5570***	0.5570***	0.5570***	0.5570***	0.5571***	0.5572***	0.5574***
	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
μ_ζ	6.7269***	6.7213***	6.7131***	6.7065***	6.7001***	6.6934***	6.6933***
	(0.0655)	(0.0658)	(0.0663)	(0.0668)	(0.0675)	(0.0696)	(0.0745)
σ_ζ	0.8722***	0.8722***	0.8722***	0.8723***	0.8725***	0.8735***	0.8756***
	(0.0353)	(0.0355)	(0.0357)	(0.0360)	(0.0363)	(0.0367)	(0.0370)
μ_ω	-0.0356***	-0.0345***	-0.0329***	-0.0309***	-0.0286***	-0.0227**	-0.0156
	(0.0062)	(0.0064)	(0.0067)	(0.0070)	(0.0074)	(0.0082)	(0.0095)
σ_ω	0.2314***	0.2332***	0.2359***	0.2381***	0.2402***	0.2421***	0.2413***
	(0.0106)	(0.0109)	(0.0114)	(0.0120)	(0.0126)	(0.0144)	(0.0176)
μ_η	-0.0573***	-0.0582***	-0.0596***	-0.0611***	-0.0627***	-0.0661***	-0.0699***
	(0.0014)	(0.0014)	(0.0014)	(0.0015)	(0.0015)	(0.0016)	(0.0019)
σ_η	0.0974***	0.0984***	0.0999***	0.1011***	0.1024***	0.1041***	0.1052***
	(0.0023)	(0.0023)	(0.0024)	(0.0026)	(0.0027)	(0.0033)	(0.0044)
ρ_s	-0.0709***	-0.0716***	-0.0730***	-0.0742***	-0.0754***	-0.0778***	-0.0802***
	(0.0024)	(0.0024)	(0.0025)	(0.0025)	(0.0026)	(0.0027)	(0.0029)
π_s	0.0050***	0.0142***	0.0287***	0.0428***	0.0579***	0.0876***	0.1180***
	(0.0004)	(0.0007)	(0.0010)	(0.0012)	(0.0014)	(0.0017)	(0.0021)
π_ω	0.2602***	0.2575***	0.2537***	0.2515***	0.2497***	0.2526***	0.2628***
	(0.0227)	(0.0231)	(0.0239)	(0.0249)	(0.0262)	(0.0310)	(0.0412)
π_r	0.9800***	0.9802***	0.9805***	0.9807***	0.9810***	0.9813***	0.9815***
	(0.0018)	(0.0018)	(0.0018)	(0.0017)	(0.0017)	(0.0017)	(0.0018)
log. L.	-14766.6251	-16505.9610	-18855.5248	-20859.8865	-22818.3620	-26273.5297	-29396.8849
AIC	29557.2502	33035.9221	37735.0495	41743.7729	45660.7240	52571.0593	58817.7698
BIC	29658.3497	33137.0216	37836.1490	41844.8724	45761.8234	52672.1588	58918.8693
No. cases	33690	33690	33690	33690	33690	33690	33690

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

holds for η. The estimates for mean and standard deviation all have the same direction across all tolerance amounts. The estimates for the mean decrease as the tolerance increases, and for the standard deviation of η, the estimates increase as the tolerance amount rises. With respect to the measure of mean reversion of the (survey) measurement error, denoted as ρ_s, all estimates come from roughly the same "ballpark", ranging from -0.0709 to -0.0802, decreasing slightly with the increases in tolerance. The smaller estimates from model versions with higher tolerance are accompanied by slightly higher standard errors than the smaller ones.

The probability of correctly reported survey earnings, π_s varies substantially. It is lowest at 0.5% for a tolerated absolute difference in logarithmized earnings of 0.001, and the highest at 11.8% for the largest tolerance reported. The probability of contamination π_ω and the probability of a correct administrative match π_r are, on the other hand, both not increasing or decreasing with increasing tolerances throughout the entire range of tolerances, and remain within small ranges with no detectable pattern related to increasing tolerance amounts. π_ω ranges from 24.97% to 26.28% and π_r from 98.00% to 98.15%. The upper end of the range of estimates, however, always stems from the largest accepted tolerance, and the lower one from the model version with the lowest tolerance.

The two different threshold types used in the replication of the Kapteyn and Ypma (2007) full model are the absolute differences of (1) unlogarithmized and (2) logarithmized administrative and survey earnings measures. In comparing the two sets of estimations reported in Tables 4 and 6, it should be acknowledged that the share of observation that is assumed error-free should be as similar as possible, so as to enable a direct comparison. The share of the sample that is assumed to provide true earnings for each tolerance is tabulated in Table 3. The most directly comparable pair of tolerances is the threshold of 10 Euros per month (2.83% of observations) to the difference in logarithmized earnings of 0.005 (2.82% of observations). Comparing the pair confirms that they yield very similar results, which is to be expected given how similar the implied size of group 1 is for the two thresholds, and underlines the importance of bearing in mind that for small deviations in the assumed share of error-free measurements, the estimation will probably seem robust, while large differences in the assumed share may be more insightful about actual robustness of the model estimation.

Kapteyn and Ypma (2007) do not report goodness of fit statistics beyond the log likelihood for their estimation, so that a comparison of the present study's results with their results focuses on the estimated parameters of their full model estimation using no covariates for earnings rather than a question of model fit. Note that their sample is highly selective. The survey component of their linked data only includes individuals over 50 years of age and thus presents a possibly more homogeneous sample. More tightly clustered data makes it easier for the likelihood to be closer to zero, since fewer large variations have to be accounted for. All of the estimates in column (5) of Table 4 (referred to as the replication) have the same direction as in Kapteyn and Ypma (2007, Table C2) (referred to as the original estimation). With respect to the estimates of mean and standard deviation for true earnings ξ, Kapteyn and Ypma estimate a higher mean (12.283 versus 7.7895 in the present replication), but also a higher standard deviation (0.717 versus 0.5565). As for the mean, the currency in which earnings are measured, differs. While the original study uses earnings measured in Swedish Crowns, this replication is carried out with earnings measured in Euros. With respect to the standard deviations, it is important to note that the inequality in earnings is estimated to be higher for the Swedish data. This might be the result of the German (replication) data being top-censored by the yearly cap on insurable income (so-called *Beitragsbemessungsgrenze*), which cuts off any individuals with earnings above that contribution ceiling and thus artificially eliminates inequality from the observed data. A similar picture emerges for the mean of the distribution of ζ (9.187 versus 6.6312), while the estimate for ζ_σ is larger in our data (0.843 versus 0.8664). For η, the replication yields a smaller estimate for the mean, but a higher one for the standard deviation (μ: -0.044 versus -0.0655, σ: 0.217 versus 0.1176). For contamination ω, the replication yields a much larger estimate for μ_ω (-1.632 versus -0.0365 in the replication) and at the same time a much smaller estimate of the standard deviation σ_ω (3.801 vs 0.2720). With respect to mean reversion, both estimations establish the presence of mean-reverting measurement error. However, the replication reveals that it is present to a different extent that the original estimation (ρ_s: -0.113 vs -0.0811). The estimated probability of a correct match of administrative records is high and similar at 98.1% in the original and 98.21% in the replication, the estimate for the probability of contamination π_ω is also much lower in the original estimation than in the replication (5% versus 19.23%). However, the replication does yield a much smaller estimate for the probability of correctly reported survey incomes π_s (26.8% vs. 14.44%).

5.2 Jenkins and Rios-Avila (2023b) Full Model

Since the original Kapteyn and Ypma (2007) model was extended to contain erroneous measurement of earnings besides linkage error, it is crucial to investigate the consequences of different choices for the tolerance of *close enough* observation pairs to belong to group 1 (both observations report true earnings). Estimations by Jenkins and Rios-Avila (2023b) were carried out using $|r_i - s_i| \leq 0.005$, but no further investigation into that choice was made, other than reporting that repeating the analysis with a tolerance of 0.01 delivers conclusions that are robust. Only investigating the robustness with one larger tolerance, which is close to the originally used tolerance hardly seems sufficient for concluding robustness of the results with respect to the choice of tolerance amounts. A more thorough investigation of tolerances above and below the original choice is required to draw more valid conclusions about the robustness of the estimation with respect to the accepted tolerances and implied assumption of the share of accurate pairs of measurement. Therefore, this study investigates a set of seven tolerance choices for the Jenkins and Rios-Avila extended model, namely $\{0.001, 0.0025, 0.005, 0.0075, 0.01, 0.0015 \text{ and } 0.02\}$. As for the model specifications estimated, Jenkins and Rios-Avila (2023b) refer to their model, described in Section 3 of this paper, as the *extended* model, since it is an extended generalization of Kapteyn and Ypma (2007), which is referred to as the *full* model, even though it is a restricted special case of the extended model.

Jenkins and Rios-Avila (2023b) estimate two versions of their model. The first is the extended model as presented, while in the second version they restrict the correlation of true and contaminated earnings, ρ_ω to zero, i.e. disallowing a relationship between reference period error and true earnings, which they argue to be justified empirically with regard to of the insignificance of the estimate for that correlation. They compare the goodness of fit for the constrained and unconstrained extended model, and since the two are almost the same with regard to the value of the log likelihood, but the constrained version is associated with lower AIC and the BIC, they focus on the constrained version of the extended model. The summary statistics we estimate across all tolerances for the constrained and unconstrained extended model are presented in Table 7. The superiority of the constrained model cannot be confirmed for the present data. For the information criteria, for which lower values imply better model fits, the picture is not as clear-cut as in Jenkins and Rios-Avila (2023b). While they are smaller in the unconstrained model for most tolerances, this is not the case for the tolerances of 0.005 and 0.02. Notably, the former is the tolerance that Jenkins and Rios-Avila (2023b) chose. Since these results do not suggest undisputed superiority of the constrained version over the unconstrained model, we report results for the unconstrained version of the model.

Comparing the model fit for the full and extended model versions, it can be concluded that, for the present sample, the goodness of fit for the extended Jenkins and Rios-Avila (2023b) model (results displayed in Table 8) is superior to that for the corresponding full model replication of Kapteyn and Ypma (2007) for all tolerances. The log likelihood of the extended model is closer to zero, compared to their counterparts, for the same tolerances in the full model replication (cf. Table 6), indicating a better fit of the extended models. The same holds for the likelihood-based information criteria, AIC and BIC.

The estimation results of the model parameters in the unconstrained extended model are reported in Table 8. Regarding the true earnings ξ, the estimated mean ranges from 7.7885 to 7.8079. There is no visible trend as the accepted tolerance increases; the extremes of the range of estimates for μ_ξ come from the two highest tolerances, 0.0015 for the upper limit and 0.02 for

Table 7: Goodness of fit statistics, Jenkins and Rios-Avila (2023b) Constrained and Unconstrained Extended Model, Varied Tolerances

	(1) 0.001	(2) 0.0025	(3) 0.005	(4) 0.0075	(5) 0.01	(6) 0.0015	(7) 0.02
log L. (constrained)	-14747	-16488	-18840	-20847	-22807	-26265	-29389
AIC (constrained)	29520	33002	37706	41719	45640	52555	58805
BIC (constrained)	29629	33111	37816	41829	45749	52665	58914
log L.	-14257	-15995	-18825	-20345	-22299	-25745	-29375
AIC	28549	32024	37685	40724	44633	51524	58784
BIC	28692	32168	37828	40867	44776	51667	58927
No. cases	33690	33690	33690	33690	33690	33690	33690

Table 8: Parameter estimates, Jenkins and Rios-Avila (2023b) Extended Model, Varied Tolerances

	(1) 0.001	(2) 0.0025	(3) 0.005	(4) 0.0075	(5) 0.01	(6) 0.0015	(7) 0.02
μ_ξ	7.8053***	7.8055***	7.7887***	7.8061***	7.8066***	7.8079***	7.7885***
	(0.0068)	(0.0068)	(0.0069)	(0.0068)	(0.0068)	(0.0069)	(0.0069)
σ_ξ	0.5427***	0.5427***	0.5589***	0.5426***	0.5424***	0.5417***	0.5590***
	(0.0048)	(0.0048)	(0.0052)	(0.0048)	(0.0048)	(0.0049)	(0.0053)
μ_ζ	6.8215***	6.8184***	6.7714***	6.8121***	6.8103***	6.8117***	6.7618***
	(0.0919)	(0.0921)	(0.0850)	(0.0924)	(0.0921)	(0.0910)	(0.1062)
σ_ζ	0.9982***	0.9983***	0.7804***	0.9974***	0.9955***	0.9871***	0.7841***
	(0.0488)	(0.0489)	(0.0400)	(0.0489)	(0.0488)	(0.0487)	(0.0422)
μ_ω	-0.1678***	-0.1689***	-0.0283***	-0.1709***	-0.1719***	-0.1711***	-0.0098
	(0.0134)	(0.0136)	(0.0070)	(0.0142)	(0.0145)	(0.0156)	(0.0122)
σ_ω	0.2141***	0.2162***	0.2325***	0.2221***	0.2250***	0.2301***	0.2339***
	(0.0141)	(0.0141)	(0.0111)	(0.0144)	(0.0146)	(0.0151)	(0.0213)
μ_η	-0.0620***	-0.0631***	-0.0602***	-0.0669***	-0.0692***	-0.0745***	-0.0711***
	(0.0017)	(0.0017)	(0.0014)	(0.0018)	(0.0020)	(0.0024)	(0.0026)
σ_η	0.0985***	0.0993***	0.0981***	0.1014***	0.1023***	0.1029***	0.1026***
	(0.0020)	(0.0020)	(0.0026)	(0.0021)	(0.0023)	(0.0029)	(0.0061)
μ_υ	-0.1345***	-0.1358***	-3.2926***	-0.1392***	-0.1393***	-0.1351***	-3.3352***
	(0.0190)	(0.0189)	(0.8303)	(0.0188)	(0.0188)	(0.0198)	(0.9039)
σ_υ	0.2154***	0.2145***	0.5522*	0.2106***	0.2075***	0.1987***	0.5401*
	(0.0115)	(0.0117)	(0.2145)	(0.0125)	(0.0129)	(0.0142)	(0.2290)
ρ_s	-0.0573***	-0.0576***	-0.0668***	-0.0589***	-0.0591***	-0.0589***	-0.0744***
	(0.0030)	(0.0030)	(0.0034)	(0.0033)	(0.0036)	(0.0043)	(0.0056)
ρ_r	0.1201***	0.1201***	-0.6578	0.1197***	0.1186***	0.1134***	-0.6733
	(0.0151)	(0.0149)	(0.3789)	(0.0146)	(0.0144)	(0.0144)	(0.4132)
ρ_ω	0.3249***	0.3282***	-0.1015**	0.3372***	0.3419***	0.3494***	-0.0783
	(0.0369)	(0.0365)	(0.0327)	(0.0358)	(0.0357)	(0.0363)	(0.0433)
π_s	0.0057***	0.0163***	0.0287***	0.0493***	0.0669***	0.1028***	0.1180***
	(0.0005)	(0.0008)	(0.0010)	(0.0017)	(0.0022)	(0.0037)	(0.0021)
π_ω	0.1204***	0.1185***	0.2696***	0.1143***	0.1123***	0.1106***	0.2877***
	(0.0144)	(0.0141)	(0.0258)	(0.0136)	(0.0134)	(0.0137)	(0.0591)
π_r	0.9876***	0.9876***	0.9819***	0.9876***	0.9875***	0.9871***	0.9822***
	(0.0016)	(0.0016)	(0.0018)	(0.0017)	(0.0017)	(0.0018)	(0.0020)
π_υ	0.8607***	0.8610***	0.9996***	0.8608***	0.8583***	0.8460***	0.9996***
	(0.0172)	(0.0173)	(0.0003)	(0.0181)	(0.0193)	(0.0256)	(0.0003)
log L.	-14257.4492	-15995.1534	-18825.3132	-20344.8069	-22299.4088	-25745.0527	-29375.0937
AIC	28548.8984	32024.3069	37684.6263	40723.6137	44632.8175	51524.1054	58784.1875
BIC	28692.1226	32167.5311	37827.8506	40866.8380	44776.0418	51667.3296	58927.4117
No. cases	33690	33690	33690	33690	33690	33690	33690

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the lower limit of the range. The estimated standard deviation of true earnings varies within an even smaller range (0.5417 to 0.5590), also without a detectable pattern besides the 0.005 and 0.02 tolerances yielding larger estimates that the other reported tolerances. The mean of logarithmized survey earnings is 7.7232 (with a standard deviation of 0.5386) and for logarithmized administrative earnings, the mean is 7.7776 (std. dev. 0.5783), so that the means are slightly smaller than the mean of true earnings we estimate. For the standard deviation, survey earnings seem to provide a more accurate representation of the estimated standard deviation of true earnings. This is contradiction of the findings in Jenkins and Rios-Avila (2023b). In their data, the true standard deviation is substantially overestimated, which we cannot confirm. Interestingly, the standard deviation of survey earnings is more accurate, i.e. closer to the estimated standard deviation of true earnings for most tolerances.

For the different errors ζ, ω, η and υ, the estimation of their means and standard deviations follows no trend in relation to the increasing tolerance. All estimates have the same direction and the$<$ vary only in the decimals. However, the model variant with the 0.005 tolerance produces estimates that break the pattern of in- or decreasing estimates with the tolerance amounts, although this does not hold for the goodness of fit statistics.

The estimated probability of correct linkages is high, and ranges from 98.19% to 98.75% with a very low estimated standard deviation, ranging from 0.0016 to 0.0020) that increases with the tolerance amount. This implies that the probability of linkage error, $1 - \hat{\pi}_r$ is estimated to be below 2%, which is about a third of what Jenkins and Rios-Avila (2023b) find, and a fifth of that reported by Bollinger et al. (2018) for a US sample, and half of what Kapteyn and Ypma (2007) find. It has been suggested that not explicitly allowing for misreported earnings in administrative records and only allowing linkage error inflates the estimates for the latter (Bee and Rothbaum, 2019). Indeed, the estimates for π_r reported for the full Kapteyn and Ypma (2007) model in Table 6 range from 98% to 98.15% and are lower than their counterparts in the extended model as reported in Table 8. Thus, the implied probability of linkage error of approximately 2% is estimated to be slightly higher in the full model than in the extended model. The magnitude, however, does not compare to the 4 percentage point difference found by Jenkins and Rios-Avila (2023b).

The probability of mismeasurement in survey data, $1 - \hat{\pi}_s$ is fairly high, especially compared to to the probability of mismeasurement of earnings in the administrative records, $1 - \hat{\pi}_\upsilon$. It is important to note, however, that the estimates for π_s, π_ω and π_υ vary substantially across different accepted tolerances, while in contrast, the probability of linkage-free administrative earnings remains roughly constant across varied tolerances. Table 9 presents the latent class probabilities that are computed from the estimated error probabilities according to column three of Table 1. Note that even though the first class is not latent, because it is defined by the accepted tolerance chosen, there is a

Table 9: Estimated Latent Class Probabilities for Jenkins and Rios-Avila (2023b) Extended Model, Varied Tolerances for First Class $|r_i - s_i| \leq \cdots$

	0.001	0.0025	0.005	0.0075	0.01	0.015	0.02
π_1	0.0056***	0.0161***	0.0282***	0.0487***	0.0660***	0.1015***	0.1159***
	(0.0004)	(0.0008)	(0.0010)	(0.0017)	(0.0021)	(0.0036)	(0.0021)
π_2	0.8637***	0.8564***	0.6966***	0.8315***	0.8180***	0.7877***	0.6171***
	(0.0145)	(0.0140)	(0.0250)	(0.0131)	(0.0128)	(0.0132)	(0.0519)
π_3	0.1182***	0.1151***	0.2571***	0.1073***	0.1035***	0.0979***	0.2493***
	(0.0141)	(0.0137)	(0.0244)	(0.0127)	(0.0123)	(0.0121)	(0.0509)
π_4	0.0001***	0.0002***	0.0005***	0.0006***	0.0008***	0.0013***	0.0021***
	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
π_5	0.0109***	0.0108***	0.0129***	0.0104***	0.0103***	0.0103***	0.0112***
	(0.0014)	(0.0014)	(0.0012)	(0.0014)	(0.0014)	(0.0014)	(0.0011)
π_6	0.0015***	0.0014***	0.0047***	0.0013***	0.0013***	0.0013***	0.0045***
	(0.0003)	(0.0003)	(0.0007)	(0.0003)	(0.0003)	(0.0003)	(0.0013)
π_7	0.0001***	0.0002***	0.0005***	0.0006***	0.0008***	0.0013***	0.0021***
	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
π_8	0.0109***	0.0108***	0.0129***	0.0104***	0.0103***	0.0103***	0.0112***
	(0.0014)	(0.0014)	(0.0012)	(0.0014)	(0.0014)	(0.0014)	(0.0011)
π_9	0.0015***	0.0014***	0.0047***	0.0013***	0.0013***	0.0013***	0.0045***
	(0.0003)	(0.0003)	(0.0007)	(0.0003)	(0.0003)	(0.0003)	(0.0013)

Robust standard errors in parentheses. $\hat{\pi}_1$ computed from $\hat{\pi}_r\hat{\pi}_v\hat{\pi}_s$.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

standard error reported for π_1, as it is computed as $\hat{\pi_1} = \hat{\pi_r}\hat{\pi_v}\hat{\pi_s}$ according to Table 1. The estimated error probabilities $\hat{\pi_j}$ exhibit a rather mixed picture. Two main patterns emerge among the estimates. The first is characterized by the estimates slightly increasing as the accepted tolerance rises, which affects $\hat{\pi_1}, \hat{\pi_4}$ and $\hat{\pi_7}$. The second pattern exhibits rather constant estimates, with the exception of those for the tolerance amounts of 0.005 and 0.02. For $\hat{\pi_3}, \hat{\pi_6}$ and $\hat{\pi_9}$ the estimates for the tolerances of 0.005 and 0.02 are two to four times than the other estimates, which only vary slightly. The same pattern, but with much smaller deviations of the estimates for the 0.005 and 0.02 tolerances from to the other tolerances, is observed for $\hat{\pi_5}$. In the case of $\hat{\pi_2}$ the estimates for 0.005 and 0.02 are considerably smaller, which also holds for the tolerance of 0.015, but to a lesser degree. $\hat{\pi_8}$ exhibits a less clear-cut picture and falls between the two patterns, as the estimates for tolerances 0.005 and 0.02 are slightly larger than for the rest of the tolerances, for which a consistent image emerges. Yet the differences are 0.2 and 0.1 percentage points, so that the overall range of estimates is narrow and it could also be concluded that the deviations from the estimates for the other tolerances are so minimal, that the overall pattern of estimates could be described as sufficiently invariant to consider estimates rather constant.

6 Conclusion

This paper investigates measurement error in earnings in a matched survey-administrative dataset. Previous literature using German data is rare and assumes that administrative data provide an accurate and error-free measurement of true earnings. This paper is the first to relax the assumption of error-free administrative records in German data. It provides a replication of prior seminal work by, among others, Kapteyn and Ypma (2007) and Jenkins and Rios-Avila (2023b) in the branch of the measurement error literature allowing both sides of a linked dataset to contain error. Also, this paper investigates the robustness of the empirical choice of acceptable tolerances for earnings measurements to be considered (as if) correct, in order to identify the first class, which is essentially an empirical judgement at the discretion of the researcher.

Overall, the choice of accepted tolerance is not as insignificant as has been previously implied by the lack of explicit robustness checks for the extended model. While the big picture remains the same and estimates follow the same direction as well, as mostly a similar significance pattern prevails across the bandwidth of tolerances estimated in this paper, which span from 0.49% of observations to 11.6% of the sample, small shifts in estimates can be impactful, such as when latent class probabilities are computed and multiple shifted estimates are multiplied. In summary, the most conservative choice regarding the accepted tolerance under which earnings measurements are considered *close enough*, which is ultimately an assumption about what share of the sample is considered to provide correct information, would be to use the smallest possible tolerance. With the present data, as the administrative records store earnings information in a different pseudo-currency, i.e. pension points, conversion might render a numerically identical measurement of earnings very unlikely. Another reason for using at least some degree of accepted tolerance can lie in the collection mode of survey earnings, which sometimes only allow integer amounts. While assumptions are necessary and sometimes unavoidable in empirical work, attention needs to be directed to the consequences of such assumptions regarding the end result. This work shows that while the *big picture* in terms of estimated directions and significance levels, as well as distributional features of the different errors is robust to the choice of tolerance, more detail-oriented questions, like that on the presence of mean-reverting errors or the magnitude of estimated error probabilities, require more careful consideration, as the estimates can vary substantially in their magnitudes. Readers of empirical research in the field should be cautioned by the observation that significance patterns seem to hold irrespective of the chosen tolerance. The significance of empirical findings should therefore not be accepted as an argument justifying the empirical judgements made in an analysis.

With respect to the robustness of the present analysis, several points should be carefully considered:

Firstly, representativeness is a potential issue of the analysis sample. Sakshaug and Kreuter (2012) come to the conclusion that consent and linkage bias are negligible in their linked data, as non-consent biases are present for few of their estimates and are small in magnitude compared to other sources of bias. However, the sampling of the survey is an ongoing issue. The sample of analysis is pooled, i.e. observations come from a range of ten years (2010 - 2020) and for some individuals, there is more than one observation, although there is never more than one of an individual per year. Neither cross-sectional nor longitudinal weights could entirely mitigate the consequences of pooling. Cluster-robust standard errors are computed to address some of the issues that arise from pooling. Further research could focus on the longitudinal dimension of the data, and formulate mixtures over several periods, which would ultimately result in a filtering problem that would enable inferences on the unobserved latent class probability at each point in time from the observed data.

Secondly, the administrative data used here has the disadvantage of a cap on social security contributions, which results in censored observations, i.e. observations beyond a yearly-set social security maximum are not observed. There are two possible remedies to this limitation. One could be imputing administrative earnings beyond the social security cap. The forthcoming work of Schluter et al. (2024) estimating optimal extreme values of a distribution using rank-size regression could function as the groundwork in imputing these unobservable values. To impute individual incomes beyond the censoring threshold would require an assumption about the rank of an individual's (censored) earnings, which could only come from the distribution of survey earnings and might be problematic considering mean-reverting properties of measurement error in earnings. So far, imputations have deliberately been avoided in the measurement error literature and observations containing imputation were consistently removed from the samples under study. Jenkins and Rios-Avila (2023b, p. 123) report having fitted their models including imputed and otherwise edited observation, finding hardly any change to the results using non-imputed data only. However, for their data, they do not report any top-censoring, which is present in this paper. Alternatively, the need for imputations could be avoided by a reformulation of the finite mixture model so as to account for the right-censored observation of administrative earnings as another way to address the censoring introduced by the social security contribution ceiling.

References

Abowd, J. M. & Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *The Review of Economics and Statistics*, *95*(5), 1451–1467, https://doi.org/10.1162/rest_a_00352.

Bee, A. & Rothbaum, J. (2019). *The administrative income statistics (AIS) project: Research on the use of administrative records to improve income and resource estimates.* Technical Report 2019-36, SEHSD Working Paper, US Census Bureau.

Bingley, P. & Martinello, A. (2017). Measurement error in income and schooling and the bias of linear estimators. *Journal of Labor Economics*, *35*(4), 1117–1148, https://doi.org/10.1086/692539.

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., & Ziliak, J. P. (2018). The good, the bad and the ugly: measurement error, non-response and administrative mismatch in the cps. Unpublished paper, University of Kentucky https://www.cemmap.ac.uk/wp-content/legacy/uploads/GoodBadUglyFull.pdf.

Bönke, T., Corneo, G., & Lüthen, H. (2015). Lifetime earnings inequality in germany. *Journal of Labor Economics*, *33*(1), 171–208, https://doi.org/https://doi.org/10.1086/677559.

Bound, J., Brown, C., Duncan, G. J., & Rodgers, W. L. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, *12*(3), 345–368, https://doi.org/10.1086/298348.

Bound, J. & Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, *9*(1), 1–24, https://doi.org/10.1086/298256.

Deutsche Rentenbersicherung (2024). Meldeverfahren nach der deÜv. https://www.deutsche-rentenversicherung.de/DRV/DE/Experten/Arbeitgeber-und-Steuerberater/Meldeverfahren_nach_DEUEV/Meldung-nach-DEUV.html. Accessed: 06/09/2024.

Deutsche Rentenversicherung (n.d.). *Datenschutzkonzept SOEP-RV*. Technical report, available on request at Research Data Center, Deutsche Rentenversicherung Bund.

Gauly, B., Daikeler, J., Gummer, T., & Rammstedt, B. (2019). What's my wage again? comparing survey and administrative data to validate earning measures. *International Journal of Social Research Methodology*, *23*(2), 215–228, https://doi.org/10.1080/13645579.2019.1657691.

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

Jenkins, S. P. & Rios-Avila, F. (2020). Modelling errors in survey and administrative data on employment earnings: Sensitivity to the fraction assumed to have error-free earnings. *Economics Letters*, *192*, 109253, https://doi.org/10.1016/j.econlet.2020.109253.

Jenkins, S. P. & Rios-Avila, F. (2023a). Finite mixture models for linked survey and administrative data: Estimation and postestimation. *The Stata Journal*, *23*(1), 53–85, https://doi.org/10.1177/1536867X231161976.

Jenkins, S. P. & Rios-Avila, F. (2023b). Reconciling reports: modelling employment earnings and measurement errors using linked survey and administrative data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *186*(1), 110–136, https://doi.org/10.1093/jrsssa/qnac003.

Kapteyn, A. & Ypma, J. Y. (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economics*, *25*(3), 513–551, https://doi.org/10.1086/513298.

Meijer, E., Rohwedder, S., & Wansbeek, T. (2012). Measurement error in earnings data: Using a mixture model approach to combine survey and register data. *Journal of Business and Economic Statistics*, *30*(2), 191–201, https://doi.org/10.1198/jbes.2011.08166.

Meijer, E. & Ypma, J. Y. (2008). A simple identification proof for a mixture of two univariate normal distributions. *Journal of Classification*, *25*(1), 113–123, https://doi.org/10.1007/s00357-008-9008-6.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the psid validation study. *Journal of Business & Economic Statistics*, *13*(3), 305–314, https://doi.org/10.1080/07350015.1995.10524604.

Redner, R. A. & Walker, H. F. (1984). Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, *26*(2), 195–239, https://doi.org/10.1137/1026034.

Sakshaug, J. W. & Kreuter, F. (2012). Assessing the magnitude of non-consent biases in linked survey and administrative data. *Survey Research Methods*, *Vol 6*, No 2 (2012), https://doi.org/10.18148/SRM/2012.V6I2.5094.

Schluter, C., Schröder, C., Retter, I., Beckmannshagen, M., & König, J. (2024). The beyondpareto command for optimal extreme value index estimation. *The Stata Journal*, (pp. forthcoming).

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.

Valet, P., Adriaans, J., & Liebig, S. (2018). Comparing survey data and administrative records on gross earnings: nonreporting, misreporting, interviewer presence and earnings inequality. *Quality & Quantity*, *53*(1), 471–491, https://doi.org/10.1007/s11135-018-0764-z.

Yakowitz, S. J. & Spragins, J. D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, *39*(1), 209–214, https://doi.org/10.1214/aoms/1177698520.