Proceedings of Statistics Canada Symposium 2014
Beyond traditional survey taking: adapting to a changing world

# Measurement Error for Welfare Receipt and Its Impact on Fixed-Effects Models

Johannes Eggs[1]

## Abstract

Measurement error is one source of bias in statistical analysis. However, its possible implications are mostly ignored One class of models that can be especially affected by measurement error are fixed-effects models. By validating the survey response of five panel survey waves for welfare receipt with register data, the size and form of longitudinal measurement error can be determined. It is shown, that the measurement error for welfare receipt is serially correlated and non-differential. However, when estimating the coefficients of longitudinal fixed effect models of welfare receipt on subjective health for men and women, the coefficients are biased only for the male subpopulation.

Key Words: longitudinal data, measurement error, fixed effect models, welfare receipt

## 1. Introduction

The use of longitudinal panel data is popular in economics and the social sciences. Panel data like any survey data is in almost always affected by some forms of measurement errors. Measurement errors are deviations of the answers of respondents from the true values of the measures (Groves, 1991p. 2). When analyzing panel data, one has to keep in mind that the measurement error for a given variable is not necessarily time-invariant, but that it can also change over time.

The change of measurement error over time can especially bias parameters of longitudinal fixed-effects models (see Angrist and Pischke (2008) for discussion). A substantial attenuation of effect estimates due measurement error was found in previous studies (Freeman, 1984; Chowdhury & Nickell, 1985). Yet, an attenuation of effect estimates is not compulsory. The direction and size of the bias, caused by measurement error, depend on the particular model specifications.

Most research on the influence and extent of measurement error in surveys is conducted cross-sectionally and not longitudinally. Lack of research on the impact of longitudinal measurement error is related to the lack of longitudinal validation data, which is more difficult to acquire than validation data for one specific date. Some research was conducted for two subsequent panel waves (Bound & Krueger, 1991; Lynn et al., 2012).

In this work, validation data is available for a longer period. The extent and impact of measurement error can be evaluated for up to five panel waves. This study focuses on measurement error for unemployment benefit II (UB II), a type of welfare. The extent of underreporting for welfare receipt is known to be considerable in surveys (Kreuter et al, 2010; Czajka, 2013). The measurement error for UB II can be evaluated for the data of the German panel study “Labour market and social security” (PASS). The PASS data can be linked on individual level to register data that is provided by the German employment agency and which serves as the necessary source of validation data. Comparing the data entries for welfare receipt between the two sources, the measurement error can be determined for each respondents and its impact can be evaluated for data that covers a longer period of time than in previous research.

This paper focuses on the following research questions. (1) Are classic assumptions about the distributions and correlations of measurement error receipt met for the measurement error for UB II-receipt? In order to correct for the bias, a range of measurement error models have been introduced over time. Assumptions for these models are also discussed. (2) Whether and in which direction does measurement error for UB-II-receipt distort estimates for fixed effect models? For this purpose, analyses of the study by Eggs (2013) are recalculated with administrative information.

[1]Johannes Eggs, Institute for Employment Research, Regensburger Str. 104, 90478 Nürnberg, Germany

## 2. Data

The empirical analysis is based on data from the PASS panel study, a survey designed for research on the labour market and poverty in Germany (Trappmann et al, 2013). Data from the first five panel waves (2007–2011) is used. In the first wave, about 18 000 individuals in 13 000 households were interviewed. The total sample is a combination of two subsamples, one of which is drawn from the unemployment II registers of the Federal Agency of Employment, while the second is a general population sample. In each household, first an interview with the household target person is sought, followed by individual interviews with all members of the household aged 15+. In this study, UB-II-receipt is the variable of interest. UB II was introduced in 2005 as a part of the "Hartz" reforms, a major reform package of the social security system. UB II was supposed to be the new basic welfare scheme and as such supposed to provide the minimum resources necessary for an individual to meet his or her basic needs. UB-II-recipients are not necessarily unemployed or vice versa. E.g. if work-related earnings are not sufficient to provide the bare minimum, UB II can be claimed to bridge the resource gap. Information on UB II is collected in the household questionnaire.

For the purpose of this study, PASS has the advantage of providing a sufficient number of UB-II-recipients and the possibility to link survey reports and register information. In wave 1 PASS had household response rates of 28.7% for the recipient sample and 24.7% for the population sample (RR1 according to the definitions of AAPOR 2009). In each subsequent wave, refreshment samples were drawn. The refreshment samples consist of households that are first time recipients of UB II. Sizes of the refreshment samples vary around 1 000 households and 1 400 individuals.

The administrative data used to validate survey data is drawn from the IEB (Integrated Employment Biographies) data. The IEB contains longitudinal information on employment, unemployment benefits, and UB-II receipt. The linkage between PASS survey data and IEB data required the informed consent of respondents. Respondents in the population sample were linked by their name and address, gender and date of birth using error tolerant procedures based on Jaro (1989). Direct linkage with the register was possible for those individuals that were sampled from the registers.

Administrative data is not necessarily free of error or of better quality than survey data. However, the administrative data for UB-II-receipt is suited to analyse measurement error. For UB-II-receipt, register information is of high quality as it is directly produced by the software that administers benefit claims and payments (Köhler & Thomsen, 2009). Information on UB-II-receipt can be extracted for the same date, the date of the interview, from both data sources. Also, the same construct is measured in both data sources (whether any benefits were collected on the date of the interview). Hence, differences between the two data sources can be defined as measurement error.

The analysis is based on 12 169 respondents with 36 909 observations. On average, respondents participate three times in the survey. Individuals that participated only once were excluded, since they do not provide longitudinal information. In order to study the effect of measurement error for UB-II-receipt on the longitudinal association of UB-II-receipt with health, individuals are kept in the analysis sample, that are in the working age range (17<Age<66) and are employed or unemployed. For 10 458 respondents with 32 019 observations, the survey information can be linked with register information. According to Beste (2011), the use of the data for the linked subgroup can lead to minor selection bias in comparison to the complete sample. Stata 13.1 is used for all statistical analyses.

## 3. Measurement error for welfare receipt

Measurement error occurs, if the survey response of an individual deviates from the underlying true score. The deviation can happen during the answering process to a given question. For each question the survey response of an individual is the product of a retrieval process and an editing process. First, the individual has to retrieve the information from memory. The quality of the retrieval is based on the cognitive ability of the individual and the salience of the event for the individual. In a second step, the retrieved information is edited by the respondent. The direction of the editing process depends on social norms. Respondents systematically overreport social-desirable items and underreport social-undesirable items (Tourangeau et al., 2000).

In this study, the focus is on the measurement error for UB-II-receipt. The error is determined by comparing the survey information for UB-II-receipt with the register information at the time of the each interview on individual level.

Measurement error for UB-II-receipt has been the topic of previous studies using the same data sources. These previous studies found a substantial underreporting of UB-II-receipt (Kreuter et al., 2010) and a significant decline of the degree of underreporting over time (Jäckle et. al., 2012). Bruckmeier et al. (2014) showed that the measurement error for UB II is correlated with a range of socio-economic indicators and that respondents, which are closer to the labour market, are more likely to underreport.

In a statistical setting, one speaks of classical measurement error, if the measurement error is uncorrelated with the true score and has an expected value of zero (Carroll et al., 2006). Classical measurement error can cause an attenuation of effect estimates towards the zero. Hence, if the variable of interest is known to be affected by measurement error, it is convenient to assume classical measurement error as the estimator can be claimed as being more conservative. However, both conditions for classical measurement error are violated for UB II receipt using the data of this study. The measurement error is correlated (r=-0.27) with the true value. As underreporting (5.2 %) is on average across all panel waves more common than overreporting (2.2%), the expected mean of the error cannot be equal to zero.

Additional assumptions for longitudinal measurement error models can also state the absence of serial correlation and stable probabilities for the error over time (Freeman, 1984). Analyzing the study, a considerable degree of serial correlation over time (r=0.33) is found for the measurement error for UB II. This means, that individuals that misreport in one wave are also more likely to misreport in the next wave. The probabilities to observe measurement error are unequal over time, as the proportion of underreporting decreases from 7.4% in the first wave to 2.2% in the fifth wave. The extent of over-reporting on the other hand does not change over time and remains at around 2%. This means, that the overall data quality increases over subsequent panel waves. In this paragraph, some properties of the measurement error for UB-II- receipt were discussed. Having tested some common assumptions for the measurement error for UB-II-receipt, it is found that it violates the classical assumptions and is time-dependent.

## 4. Measurement Error and Fixed-Effect Models

In the following section, it is analyzed whether the measurement error distorts model coefficients of a longitudinal fixed-effects model. The bias will be exemplarily evaluated for a model analyzing the association between unemployment, UB-II-receipt and a score for subjective health. The fixed effect model (1) can be stated as,

$$y_{ij} = \alpha + \beta_1 U_{ij} + \beta_2 UB\ II_{ij} + \boldsymbol{\beta}\,\boldsymbol{X}_{ij} + \mu_i + \varepsilon_{ij} \qquad (1)$$

where $y_{ij}$ is a score for self-rated health for person $i$ at time $j$. $U$ is the binary indicator for the unemployment status. *UB II* is the binary indicator for the welfare (unemployment benefit II) receipt. **X** is a vector of additional control variables (age, the logarithm of the household income, the relationship status and dummy variables for each panel wave). $\alpha$ is the intercept. $\mu_i$ are person-specific fixed unknown parameters that are eliminated by a within transformation. The model was introduced and discussed in a previous study by Eggs (2013). The results of the previous study showed negative associations for unemployment on subjective health. The results for UB-II-receipt were inconclusive and showed substantial differences between the results for men and women that could not be explained. Negative effects were found for men, but not for women. Thus, the evaluation of the bias will also be conducted separately for men and women.

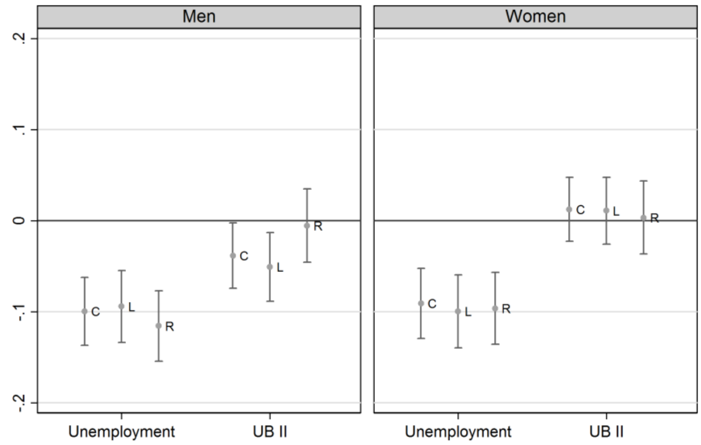
Fixed-effects models are sensitive to measurement error as they rely solely on transitions for the estimation of the coefficients. A relatively small amount of cross-sectional error may cause larger amounts of errors in transitions. For this data, the average degree of measurement error for UB II receipt is 7.4%. Yet, 20.5% of all reported transitions in or out UB II receipt are artefacts. This is also influenced by the decrease of underreporting over time. Every respondent, that stops to underreport his receipt from one wave to the next, causes an artificial transition in UB-II-receipt.

**Table 4.1**
**Multilevel multinomial logistic regressions: Average marginal effects (AME) of model covariates on over- and underreporting**

| | Underreporting | | Overreporting | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| | AME | AME | AME | AME |
| Self rated health | -0.001 | 0.001 | -0.000 | -0.0003* |
| Unemployment | -0.008*** | -0.007*** | 0.001*** | 0.001** |
| Log(HHIncome) | -0.015*** | -0.009*** | -0.001** | -0.001** |
| Age | -0.000*** | -0.000* | -0.000 | -0.000** |
| Partner (Yes) = 1 | -0.004* | -0.002 | 0.000 | -0.000 |
| Wave 1 | Ref. | | | |
| Wave 2 | -0.008** | -0.005* | 0.000 | -0.000 |
| Wave 3 | 0.001 | 0.000 | 0.001 | 0.000 |
| Wave 4 | -0.008** | -0.010*** | 0.001** | 0.000 |
| Wave 5 | -0.014*** | -0.011*** | 0.001* | 0.000 |
| Observations | 16100 | 15354 | 16100 | 15354 |
| *p<0.01,**p<0.001,***p<0.0001 | | | | |

If the measurement error for a dichotomous variable is uncorrelated with the independent variables, model coefficients might still be attenuated towards the zero. This is tested by estimating a multilevel multinomial logistic regression model as the measurement error for this variable can take three different expressions. No measurement error serves as base category. The results are presented table 4.1. The dependent variable health score is negatively associated with overreporting for men. There are associations between unemployment and most of the control variables and measurement error for men and women. Regarding the association between the dependent variable and the outcome, it might be that individuals with low health scores might misclassify other welfare benefits as UB II. The same reason might apply, why unemployed respondents are more likely to overreport. Unemployed are less likely to underreport. Employed recipients that are on the brink to be eligible to claim welfare are more likely to underreport than individuals that are certain recipients. Respondents that are more integrated in the labour market might be less inclined to report their UB-II-receipt. This would divulge that their work does not provide sufficient resources to make ends meet. Individuals with a higher household income are less likely to over- or underreport. Individuals with a higher income are less at risk to be welfare recipients and hence less likely to misreport. Older respondents are less likely to misreport. In later panel waves, individuals are less likely to underreport. The results show that variables, which are related with the risk to receive welfare benefits, are significantly associated with the respective measurement error. However, one has to keep in mind that the size of the associations is on average close to zero.

**Figure 4.1**
**Coefficients and 95% confidence intervals of linear fixed effect models on self-rated health**




C=complete sample; L=linked sample using survey entries; R=linked sample using register entries
Adjusted for age, log(HH-income), partner, panel waves
Source: PASS/IEB data waves 1-5

In the previous paragraph it was shown that the measurement error for UB-II-receipt is associated with the model variables. Hence, it cannot be predicted, whether or in which direction the error will bias the coefficients of the linear fixed effect model that was presented in equation 1. The extent and direction of the bias will be evaluated in a stepwise process using three different model specifications. The results for the three models are shown in figure 4-1 (The respective tables are available on request.). The figure shows the model coefficients and 95% confidence intervals for UB-II-receipt and unemployment on self-rated health for men and women. For the first model specification, models are calculated using survey data and using the complete analysis sample (C). For the second model specification, the models are recalculated for those respondents that could be linked (L), using the survey entries for the UB II indicator. By comparing the first two models, one can assess a possible selectivity of results due to the linkage. In a third step, the information for UB-II-receipt in the linked sample is replaced with the values of the register entries of the individuals. The coefficients of this register models (R) can then be compared with the results of the linked sample. Since these two models are otherwise identical, differences between the linked (L) and the register (R) model can only be caused by the measurement error for UB-II-receipt. Comparing the effect estimates for UB-II-receipt for the complete analysis sample (C) with the linked analysis sample (L), no differences can be seen for the estimates for women in the right graph. For men, a restriction on the linked sample causes a larger negative effect for UB-II-receipt. The estimates for unemployment are not affected by the restriction on the linked sample for men and women. In order to assess the size of the bias, the relative bias for the estimates for UB II is calculated using the coefficients of the register models as the benchmark.

Using register information instead of survey information, measurement error does not bias the results for women considerably. A relative bias of 20% is found when comparing the coefficients of the register model and the linked model. Results for men, however, are seriously biased by measurement error. Using survey information, UB-II-receipt has a negative, significant effect on health. Using register information, UB-II-receipt has a zero effect on health. Thus, the measurement error causes a relative bias of 150%. For this empirical example, measurement error does not cause an attenuation for the effects estimate. Instead it causes a considerable overestimation. Additionally, it causes a slight attenuation of the effect of unemployment for men. This also stresses the sometimes overlooked possibility, that measurement error in one variable can also bias the regression coefficients of other correlated variables. For women on the other hand, the coefficient for unemployment remains unaffected by the measurement error. The coefficients for the remaining control variables are not affected for both subgroups, despite being correlated with the measurement error. It is also interesting to note that the associations between the measurement error and the dependent variables in table 4-1 are similar for men and women, yet the effects of the error differ.

For this empirical example, minor bias is caused by the selective linkage but considerable bias is caused by measurement error for the male subpopulation. The original differences between the effect estimates for men and women seemed to be mainly caused by the measurement error in the survey data. While the survey coefficients for UB-II-receipt differ considerably between the male and the female sample, the coefficients, using register information, are equal to zero in both groups.

## 5. Conclusion

This paper provides new insights regarding measurement error for welfare receipt and its impact on longitudinal panel models. Linking register data with survey data on an individual level, the measurement error for welfare receipt is analyzed for up to five panel waves. Thus it was possible to evaluate the classical assumptions for measurement error as well as ascertain interdependencies between measurement error and model variables.

UB-II-receipt as the main type of welfare benefit in Germany and its measurement error are analyzed. UB-II-receipt receipt is underreported and to a lesser extent overreported. The measurement error was correlated across panel waves. The extent of measurement error decreases significantly over panel waves and thus cross-sectional data quality increases over panel waves. However, this causes a high number of erroneous transitions into UB-II-receipt. Transitions are a necessary prerequisite for all kinds of longitudinal analyses and necessary conditions for fixed-effect models. It is also shown that the measurement error is highly differential. Thus, classical measurement error cannot be assumed when using the survey response regarding UB-II-receipt.

Previous research on the impact of measurement error on panel models found, that fixed-effect models tend to be severely attenuated in the presence of measurement error. For this study, this could not be replicated for the male subgroup when estimating a fixed effect model analyzing the association between UB-II-receipt and subjective health. A significant negative association of UB-II-receipt on subjective health is observed due to measurement error in the survey data. Using validation data, the association between the two variables is close to zero. Thus for men, measurement error causes a relevant shift for the effect estimates for UB-II-receipt away from the zero and an overestimation of the true association. For women, on the other hand, the measurement error in the survey data did not cause any shifts in the effect estimates. Using register data decreases also the differences between the associations for the male and female samples. The impact of the measurement error could only be determined for a selective sample that could be linked to the register data. However, since the restriction on the linked sample causes only minor shifts in effect estimates, it is assumed that the bias due to measurement error can be generalized on the complete sample and is not caused by a selection effect due to the linkage.

This study provides additional evidence, that the properties and effects of measurement errors are context-specific and that classical measurement error is not the usual case (Hyslop & Imbens, 2001). According to these results, it can be argued that measurement error for current UB-II-receipt is not caused by a random process like cognitive decay. One cause of the misreporting might be the influence of social desirability as individuals might misreport their UB-II-status to avoid the stigma attached to welfare (Booth & Scherschel, 2010).

There have been attempts to correct for measurement error in panel data. Measurement error models depend on the assumption of a random process causing measurement error, implying a simple structure for measurement error (Küchenhoff et al., 2006). For UB-II-receipt, these assumptions were shown to be violated. The use of correctional methods is also hard to justify, if the coefficient for one half of the population are not affected by the measurement error. Thus, using those methods could cause more harm than good, at least for outcomes like welfare receipt, where simple assumptions about the error-generating process are not met. Instead, when faced with a variable, that is known to be affected with non-standard measurement error, one should conduct a range of subgroup analyses and sensitivity checks in order to assess the robustness of the results. This could be combined with the use of different types of estimators as the impact of measurement error depends on the model specifications (Angrist & Pischke, 2008). Each approach would be associated with some restrictions, but one could assess the consistency of the results across the range of the different approaches. Other possibilities to decrease the bias could target the data entries of the study population. An easy approach to decrease the measurement error bias could have been the discarding of the entries of the first wave of panel participation for each individual. As measurement error decreases over panel waves, errors in transitions are mostly a problem in early panel waves. However, when the information for the first wave of panel participation is dropped, the bias due to measurement error does not decrease (Results not shown).

While UB II is a German welfare benefit, measurement error for welfare receipt can be found in most surveys in western countries (Bound et al., 2001). It would be surprising, if less complex patterns for the error would be the normal scenario. This study provides further evidence that common assumptions about statistical properties for error models are not necessarily met and should only be assumed after some deliberation and that an attenuation of model coefficients due to measurement error is not necessarily the usual case.

## References


The American Association for Public Opinion Research (2009), *Standard Definitions Dispositions of Case Codes and Outcome Rates for Surveys*, 6th edition, AAPOR.

Angrist, J., & Pischke, J.-S. (2008), *Mostly harmless econometrics*. Princeton: Princeton University Press.

Beste, J. (2011), “Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten”, Record Linkage mit Daten des Panels "Arbeitsmarkt und soziale Sicherung" und administrativen Daten der Bundesagentur für Arbeit *FDZ-Methodenreport No.9*.

Booth, M., & Scherschel, K. (2010), *The impact of activating labor market policies on labor market orientations and institutions: Economic Sociology Jena*.

Bound, J., Brown, C., & Mathiowetz, N. (2001), “Measurement Error in Survey data”, in J. J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 5, pp. 3705 - 3843), Amsterdam: Elsevier.

Bound, J., & Krueger, A. B. (1991), “The extent of measurement error in longitudinal earnings data: Do two wrongs make it right?”, *Journal of Labor Economics, 9*, 1-24.

Bruckmeier, K., Müller, G., & Riphahn, R. T. (2014), “Who misreports welfare receipt in surveys?”, *Applied Economics Letters, 21*(12), 812-816.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Craniceanu, C. M. (2006), *Measurement error in nonlinear models* (2 ed.), New York: Chapman & Hall.

Chowdhury, G., & Nickell, S. (1985), "Hourly Earnings in the United States: Another Look at Unionization, Schooling, Sickness, and Unemployment Using PSID Data", *Journal of Labor Economics*, *3*(1), 38-69.

Czajka, J. L. (2013), "Can administrative records be used to reduce nonresponse bias?", *The Annals of the American Academy of Political and Social Science, 645*, 171-184.

Eggs, J. (2013), "*Unemployment benefit II, unemployment, and health IAB",* Discussion Paper 12-2013: Institut für Arbeitsmarkt- und Berufsforschung.

Freeman, R. B. (1984), "Longitudinal Analyses of the Effects of Trade Unions", *Journal of Labor Economics, 2*(1), pp. 1-26.

Groves, R. M. (1991), "Measurement Error across the Disciplines", in P. Biemer, R. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Sudman (Eds.), *Measurement Error ins Surveys*: Wiley.

Hyslop, D., & Imbens, G. W. (2001), "Bias From Classical and Other Forms of Measurement Error", *Journal of Business & Economic Statistics, 19*(4), 475-481.

Jäckle, A., Eggs, J., & Trappmann, M. (2012), "*Will respondents eventually get it right? Changes in measurement error across waves of a panel survey using dependent interviewing",* paper presented at the Annual Meeting of the American Association for Public Opinion Research, Orlando, USA.

Jaro, M. A. (1989), "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida", *Journal of the American Statistical Association, 84*(406), 414-420.

Köhler, M., & Thomsen, U. (2009), "Data Integration and Consolidation of Administrative Data From Various Sources. The Case of Germans' Employment Histories", *Historical Social Research, 34*(3), 215-229.

Kreuter, F., Mueller, G., & Trappmann, M. (2010), "Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data", *Public Opinion Quarterly, 74*(5), 880-906.

Küchenhoff, H., Mwalili, S. M., & Lesaffre, E. (2006), "A General Method for Dealing with Misclassification in Regression: The Misclassification SIMEX", *Biometrics, 62*, 85-96.

Lynn, P., Jäckle, A., Jenkins, S. P., & Sala, E. (2012), "The Impact of Interviewing Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study", *Journal of the Royal Statistical Society, Series A, 175*, 289-308.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000), *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Trappmann, M., Beste, J., Bethmann, A., & Müller, G. (2013), "The PASS panel survey after six waves", *Journal for Labour Market Research, 46*(4), 275-281. doi: 10.1007/s12651-013-0150-1