Consequences of measurement error for inference in cross-lagged panel design – The example of the reciprocal causal relationship between subjective health and SES

Hannes Kröger†, Rasmus Hoffmann and Eduwin Pakpahan

European University Institute, Florence, Italy

Summary. We discuss the problem of random measurement error in two variables when using a cross-lagged panel design. We apply the problem to the question of the causal direction between socioeconomic status (SES) and subjective health, known also as health selection versus social causation. We plot the bias of the ratio between the social causation and the health selection coefficient as a function of the degree of measurement error in subjective health and SES for different scenarios which might occur in practice. Using simulated data we give an example of a Bayesian model for the treatment of measurement error that relies on external information about the degree of measurement error.

Keywords: Measurement error; Cross-lagged panel design; INLA; Bias; Health selection; Social causation

1. Introduction

It has long been accepted that a life course perspective presents an adequate framework for the analysis of the causal relationship between socioeconomic status (SES) and health (Adler et al., 1994; Ben-Shlomo and Kuh, 2002). A life course perspective allows researchers to investigate how SES at one point in life influences future health and vice versa. However, a unidirectional analysis might under- or over-estimate the importance of SES for health.

†*Address for correspondence*: Hannes Kröger, Via dei Roccettini 9, I-50014 San Domenico di Fiesole.
email: hannes.kroger@eui.eu

Under the life course framework, many different concrete research questions have been investigated, for example whether childhood health influences SES in adulthood or if childhood SES influences adult health (Power et al., 1996), the relationship between occupational status and health (Chandola et al., 2003), and how unemployment and mental health influence each other (Heponiemi et al., 2007).

All these studies have in common that they try to jointly analyze the effect of health at time t_1 on SES at t_2, and of SES at t_1 on health at t_2. The measures and time points are different, but the general logic is the same. If SES has an impact on health, it is seen as support for the social causation hypothesis; if health has an impact on SES, it lends support to the health selection hypothesis. This popular research design is known as cross-lagged panel design (CLPD) (Kenny, 2005). Even though studies using CLPD can be quite heterogeneous in content, they can run into a common problem of differences in measurement error in the variables, which is the focus of this paper.

Almost all studies dealing with the dual causal relationship between SES and health over the life course use data gathered in surveys. The problem is that even surveys of the highest quality contain considerable amounts of measurement error (ME) in their responses. In this paper, we will show that the common problem of errors-in-variables might systematically bias studies on SES and health even though we are "only" dealing with random measurement error, which is often viewed as the easy type of ME – in contrast to systematic ME, or so-called reporting heterogeneity. The consequences of random measurement error in a single variable are well known (Bollen, 1989). We will show that this understanding only holds for a unidirectional analysis. As soon as a reversed causal structure is modelled in a CLPD, random ME presents a serious obstacle for causal inference. We further present a method suitable for correcting the bias in cases in which it is not possible to estimate the degree of measurement error from the data. Based on Bayesian statistics this correction model uses expert knowledge or information from literature reviews to adjust the analyses for the impact of measurement error. Our paper adds to the literature dealing with statistical problems in a life-course perspective, especially the role of measurement error in such analyses (Stavola et al., 2006). It expands discussion on the effect of measurement error in cross-lagged panel designs to non-contemporaneous effects and

with explicit reference to differences in the degree of measurement error (Lorenz et al., 1995).

1.1. Structure of the paper

First, we present the cross-lagged panel design that is used in many studies on SES and health over the life course. Subsequently, we describe the consequences of random ME in a unidirectional analysis with one variable that is measured with error, before extending this to a bidirectional case with two variables measured with error. For the bidirectional case, we show that differences in the degree of ME will lead to a bias in the estimator of the relative strength of the coefficients on the health selection and social causation paths. The bias of the relative strength depends on several quantities other than the ME, making assessment of it very complex. We therefore illustrate the impact of the degree of ME graphically for certain constellations which are relevant for applied social science research. For this purpose, we draw on literature that reports degrees of measurement error for measures of SES and subjective health to plot plausible scenarios of bias in the ratio between the social causation and health selection paths. Given certain defined circumstances, the health selection hypothesis will be deemed less important relative to social causation in explaining health inequalities in subjective health than it should be, purely due to differences in random ME in the SES and subjective health variable.
We briefly address the most common strategies to address the problem. These are sensitivity analysis, instrumental variables regression, multiple imputation for measurement error (MIME), and measurement models in a structural equation modelling framework. We point out that all variants of these methods except the first need additional data. For a case in which the necessary additional data is not available we propose a systematic version of sensitivity analysis within a Bayesian estimation framework. It has the advantage that it uses external information on measurement error in a compact way compared to common deterministic approaches to sensitivity analyses. The implementation of the approach uses the Integrated Nested Laplace Approximation (INLA) for Bayesian estimation making it computationally efficient compared to the use of Markov Chain Monte Carlo (MCMC) methods.
In the final section, we draw conclusions on the application of CLPD when measurement error in two variables is to be expected.

2. Health selection versus social causation and cross-lagged panel design

The question of how health inequalities arise has been discussed intensively in recent decades. One core aspect of this debate is the distinction between health selection and social causation as mechanisms that can explain health inequalities. The questions that derive from this debate are: Does social status determine health or does health influence social status? Or are both true? And if both are true, which is more important?

As health can be measured in many different ways we have to restrict ourself to one category, namely subjective health. In the remaining part of the study we always refer to subjective health measures, but often refer simply to health for brevity's sake. Health selection (HS) is defined as the process in which differences in health status lead to differences in social position. Those who are in good health are able to achieve higher positions in society, those in poor health have worse chances and will only achieve low status positions. This is one explanation for the general finding that there is a social gradient in almost all dimensions of health. This form of explanation is more intuitive for researchers from health economics, less so for those from public health or medical sociology. The latter usually favour explanations which can be subsumed under the umbrella term social causation (SC). Here, the idea is that social circumstances in higher status positions are more beneficial to health than in lower status positions. The social gradient is therefore created by differences in resources, support, knowledge, behaviour or other factors which are socially stratified and benefit those who are well off.

The two hypotheses are not, of course, mutually exclusive. However, many studies either explicitly or implicitly ask the question of which of the two causal pathways is more important. This often happens with a cross-lagged panel design (CLPD) or auto-regressive panel model.

Before we continue with a definition of the CLPD we want to remark on the possibility of doing causal inference within this framework. It is clear that most of the studies applying the framework cannot claim to test causality in the sense of potential outcomes (Rubin, 2005). Health and SES are not randomly assigned and the research question usually only arose, because they are entangled in a complex pattern. Although it is actually rarely discussed explicitly in the studies, we think that if causal inference is drawn it should be understood as what Heckman called *econometric causality* (Heckman, 2008). The aim is to model the treatment, so that given

control variables SES and health are quasi-randomized. We do not aim to judge in which cases in the literature such an assumption is plausible and which cases it is not. It suffices to say that a lot of times, what is really estimated should rather be regarded as a conditional association of SES and health one time point later and vice versa. Nevertheless, the argument of our paper also holds for these cases, as the relative strength of these conditional associations will also be influenced in a complex pattern by measurement error in SES and health.

Continuing after this small excursion, we define CLPD as any attempt to estimate the impact of health at t_1 on SES at t_2, and SES at t_1 on health at t_2, within one study using any kind of regression analysis. Formally, this would mean that the following equations are estimated. (Indices for individuals and any further covariates are left out for the sake of simplicity):

$$\mathbf{SES}_{t_2}|\mathbf{H}_{t_1} \sim N(\beta_{0,SES_{t_2}}\mathbf{1} + \beta_{H_{t_1}}\mathbf{H}_{t_1} + \gamma_{SES_{t_1}}\mathbf{SES}_{t_1}, \sigma_{\epsilon_{SES_{t_2}}}\mathbf{I}) \quad (1)$$

$$\mathbf{H}_{t_2}|\mathbf{SES}_{t_1} \sim N(\beta_{0,H_{t_2}}\mathbf{1} + \beta_{SES_{t_1}}\mathbf{SES}_{t_1} + \gamma_{H_{t_1}}\mathbf{H}_{t_1}, \sigma_{\epsilon_{H_{t_2}}}\mathbf{I}) \quad (2)$$

In these equations, the error-terms can be allowed to be correlated (σ stands for the covariance), capturing non-modelled common influences on health and SES at t_2 through observed covariates. We consider only the case in which error terms are uncorrelated ($\sigma_{(\epsilon_{SES_{t_2}},\epsilon_{H_{t_2}})} \neq 0$). $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ an all-ones vector. The β coefficients are the coefficients of interest, in this case representing the health selection ($\beta_{H_{t_1}}$) and social causation ($\beta_{SES_{t_1}}$) paths. The γ coefficients are the predictive power of health and SES on health and SES at t_2, sometimes referred to as autocorrelation or state dependency (Contoyannis et al., 2004). Note that we specify the model in terms of variances: $\sigma_{\epsilon_{SES_{t_2}}}$ is the error variance of the SES equation, $\sigma_{\epsilon_{H_{t_2}}}$ is the error variance of the health equation. The reason is that the consequences on measurement error can be easier understood in terms of error variances. However, in the third part of the paper we will pick up the a notation using precisions (τ) instead of variances, because the method we use in the third part specifies the errors in terms of precisions. The precision is

defined as the inverse of the variance:

$$\tau_{SES_{t_2}} = \frac{1}{\sigma_{\epsilon_{SES_{t_2}}}},$$
$$\tau_{H_{t_2}} = \frac{1}{\sigma_{\epsilon_{H_{t_2}}}}$$

Usually the CLPD will be applied to the study of reciprocal effects in adulthood using repeated measures of the same items, because it can be argued that this is the period in which an on-going reciprocal influence of SES and health can be found. However, there are instances when the interrelation of SES and health in childhood and SES and health in adult are analysed (Haas, 2006; Stansfeld et al., 2011). In these cases the measures of SES and health at t_1 and t_2 are not the same. Regardless of whether one thinks this is a useful approach, the logic of the model remains the same. In any study applying this approach, random measurement error might create severe problems for the assessment of the merits of the social causation versus the health selection hypotheses and similar research requestions. In the next sections we explain why.

2.1. Measurement error – one variable measured with error

In our discussion about the consequences of ME, we first consider the traditional problem of one variable measured with error. For this purpose, we take only social causation into account. For our illustrations, we refer to linear regression. However, the problem appears in a similar fashion in all regression models including limited dependent variable models (e.g. logit, probit, and survival analysis; see Rabe-Hesketh et al. (2004)).

We assume that all problems of causal identification have been addressed with the exception of random ME. This mainly involves correcting for spurious correlation through covariates, or models which account for correlated unobserved factors, but also systematic forms of measurement error (e.g. justification bias in subjective health measures).

In the CLPD, the regression coefficient of the social causation path is defined as:

$$\beta_{SES_{t_1}} = \frac{\sigma_{(SES_{t_1},H_{t_2})}}{\sigma_{SES_{t_1}}} - \left(\frac{\sigma_{(H_{t_1},H_{t_2})}}{\sigma_{H_{t_1}}} \times \frac{\sigma_{(SES_{t_1},H_{t_1})}}{\sigma_{SES_{t_1}}}\right). \quad (3)$$

Random ME occurs if the observed variable does not exactly represent the true concept being measured, but has a random error term ("white noise"). Note that if a variable is actually limited in the values it can take, but is treated in the analysis as a quasi-linear variable (e.g. 5-point self-rated health or number of symptoms as a measure for subjective health) there will be floor and ceiling effects. These are an important type of measurement error, but they are not random. Limiting the possible responses to an item at the top and the bottom of the distribution results in the fact that persons in poor health can only over-report their health status and persons in excellent health can only under-report the health status. This leads to a negative correlation of the error term and the true score of the variable (Siegel and Hodge, 1968). Floor and ceiling effects should be taken into account as systematic measurement error. However, as mentioned above we only consider cases in which the error term is independent of the true value and consequently the variance of the observed variable is the sum of the variances of the true variable plus the variance of the measurement error. This is the classical model of measurement error that holds for almost all cases in survey research. The Berkson type of random measurement error model is not considered in this paper. Lower case letters indicate an observed variable, upper case letters refer to the true concept without error:

$$\begin{aligned} \mathbf{ses}_{t_1} &= \mathbf{SES}_{t_1} + \mathbf{u}_{ME_{SES_{t_1}}}, \qquad \mathbf{u}_{ME_{SES_{t_1}}} \sim N(0, \sigma_{ME_{SES_{t_1}}} \mathbf{I}), \\ \sigma_{ses_{t_1}} &= \sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}. \end{aligned} \tag{4}$$

One measure we will often refer to is the degree of measurement error. It reports how much of the variance of the observed variable is due to measurement error. The degree of measurement error is therefore defined as the ratio of the variance of the measurement error term to the variance of the observed variable:

$$DME_{ses_{t_1}} = \frac{\sigma_{ME_{SES_{t_1}}}}{\sigma_{ses_{t_1}}} = \frac{\sigma_{ME_{SES_{t_1}}}}{\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}}. \tag{5}$$

If SES (but not health) is measured with error, but the regression is conducted only on the

variables observed, we get an estimated coefficient defined as:

$$\hat{\beta}_{SES_{t_1}} = \frac{\sigma_{(SES_{t_1},H_{t_2})}}{\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}} - \left(\frac{\sigma_{(H_{t_1},H_{t_2})}}{\sigma_{H_{t_1}}} \times \frac{\sigma_{(SES_{t_1},H_{t_1})}}{\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}} \right). \tag{6}$$

It follows that

$$E(\hat{\beta}_{SES_{t_1}}) \neq \beta_{SES_{t_1}}.$$

This estimated coefficient $\hat{\beta}_{SES_{t_1}}$ is always smaller in magnitude than the real coefficient $\beta_{SES_{t_1}}$. This fact is known as attenuation bias (Bollen, 1989, Ch.5).

The reason why random ME is often considered to be of less consequence than systematic ME is that the direction of the bias of the estimator is known, as already pointed out by an early discussion of consequences of measurement errors in survey research (Asher, 1974). We will always underestimate the true impact of SES on health. Under some circumstances, this could be considered a "conservative" estimate of health inequalities.

2.2. Measurement error – two variables measured with error

Answering the question of how health inequalities arise requires us to consider the reversed causal relationship as well. We want to assess the relative strength of HS versus SC. What is usually done is a comparison of the strength of the coefficients of the HS and SC paths. Formally speaking, we interpret a ratio calculated by dividing the SC coefficient by the HS coefficient. Sometimes a ratio for the standardized coefficients is chosen (indicated by a tilde: $\sim$), because differences in metrics in the health and SES variable can render an unstandardized ratio meaningless:

$$\begin{aligned} \Delta_{(SES,H)} &= \frac{\beta_{SES_{t_1}}}{\beta_{H_{t_1}}}, \\ \tilde{\Delta}_{(SES,H)} &= \frac{\tilde{\beta}_{SES_{t_1}}}{\tilde{\beta}_{H_{t_1}}}. \end{aligned} \tag{7}$$

If the ratio is greater than 1, SC is more important; below 1 HS is more important, provided that both $\beta_{H_{t_1}}$ and $\tilde{\beta}_{H_{t_1}} \neq 0$. If the ratio is, e.g., 0.5, we could say that the impact of health on SES is twice as strong as the impact of SES on health.

Now, what happens if we consider a regression on observed variables measured with error? This is not an easy step, because as soon as a regression equation contains two variables measured with error the direction of the bias is no longer clear. Since the motivation for this article is the proposition that common survey data for SES and health always contain ME, we have to stick with this assumption. The biased SES coefficient can be expressed as:

$$\hat{\beta}_{SES_{t_1}} = \beta_{SES_{t_1}} + \frac{\gamma_{H_{t_1}}\sigma_{ME_{H_{t_1}}}\sigma_{(H_{t_1},SES_{t_1})} - \beta_{SES_{t_1}}\sigma_{ME_{SES_{t_1}}}(\sigma_{H_{t_1}} + \sigma_{ME_{H_{t_1}}})}{(\sigma_{H_{t_1}} + \sigma_{ME_{H_{t_1}}})(\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}) - \sigma^2_{(H_{t_1},SES_{t_1})}}. \tag{8}$$

A derivation can be found in the appendix B.

In order to simplify, we assume that all other possible variables in the equation are measured without error and so we leave them out of the argument to keep it manageable. We also only discuss coefficient estimates. For problems related to standard errors in the case of two variables measured with error, see Brunner and Austin (2009). The biased health coefficient is

$$\hat{\beta}_{H_{t_1}} = \beta_{H_{t_1}} + \frac{\gamma_{SES_{t_1}}\sigma_{ME_{SES_{t_1}}}\sigma_{(H_{t_1},SES_{t_1})} - \beta_{H_{t_1}}\sigma_{ME_{H_{t_1}}}(\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}})}{(\sigma_{H_{t_1}} + \sigma_{ME_{H_{t_1}}})(\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}) - \sigma^2_{(H_{t_1},SES_{t_1})}}. \tag{9}$$

Now, if the measures are standardized (by using the standard deviation of both the independent and the dependent variables) the resulting ratio of the standardized coefficients is

$$\begin{aligned}
\frac{\tilde{\beta}_{SES_{t_1}}}{\tilde{\beta}_{H_{t_1}}} &= \frac{A \times B}{C \times D}, \text{ where} \\
A &= \beta_{SES_{t_1}} + \frac{\gamma_{H_{t_1}}\sigma_{ME_{H_{t_1}}}\sigma_{(H_{t_1},SES_{t_1})} - \beta_{SES_{t_1}}ME_{SES_{t_1}}(\sigma_{H_{t_1}} + \sigma_{ME_{H_{t_1}}})}{(\sigma_{H_t} + \sigma_{ME_{H_t}})(\sigma_{SES_t} + \sigma_{ME_{SES_{t_1}}}) - \sigma^2_{(H_{t_1},SES_{t_1})}}, \\
B &= \sqrt{\sigma_{SES_{t_2}} + \sigma_{ME_{SES_{t_2}}}}\sqrt{\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}}, \\
C &= \beta_{H_{t_1}} + \frac{\gamma_{SES_{t_1}}\sigma_{ME_{SES_{t_1}}}\sigma_{(H_{t_1},SES_{t_1})} - \beta_{H_{t_1}}ME_{H_{t_1}}(\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}})}{(\sigma_{H_{t_1}} + \sigma_{ME_{H_{t_1}}})(\sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}) - \sigma^2_{(H_{t_1},SESt_1)}}, \\
D &= \sqrt{\sigma_{H_{t_2}} + \sigma_{ME_{H_{t_2}}}}\sqrt{\sigma_{H_{t_1}} + \sigma_{ME_{H_{t_1}}}} \qquad .
\end{aligned} \tag{10}$$

It is obvious that any intuitive interpretation of this equation is lost. We can see that the bias of the ratio not only depends on the degree of measurement error in health and SES, but also

on the covariance of health and SES at t_1. Further, it is influenced by the state dependency of SES and health (the coefficients $\gamma_{H_{t_1}}$ and $\gamma_{SES_{t_1}}$) and, of course, by the degree of ME in the two dependent variables at time t_2. Due to this complexity, we can no longer make simple statements based on an equation.

2.3. Plotting the bias

We therefore plot several scenarios in which we keep certain determinants of the bias constant at plausible values while varying the degree of ME in SES. The quantity of interest that is plotted is always the degree of bias in the ratio of the (standardized or unstandardized) coefficients expressed as a percentage overestimation of the SES coefficient (or social causation pathway). The formula underlying the figures is the deviation of (10) from the true ratio. For the unstandardized ratio the standardization terms B and D for (10) are left out. For example, a plotted value of 40 on the y-axis of the graph means that for the given degree of ME in SES (expressed as a percentage on the x-axis) the SES coefficient is overestimated by 40% relative to the health coefficient. If the value is -35 it means that the SES coefficient is underestimated by 35%. At 0, the ratio of the coefficients based estimates from variables containing ME is the same as the true ratio. Note that it is always a bias in the ratio, not a bias in the coefficients themselves, which is plotted.

For the creation of the figures2-6 the following determinants of the bias in the ratio were kept constant at these values unless one of the determinants is plotted on the x-axis of a plot:

(a) Correlation of SES at t_1 and Health at t_1: 0.2

(b) Degree of ME in health at t_2: 45%

(c) Degree of ME in SES at t_2: 15%

(d) State dependency of health: $\gamma_{H_{t_1}} = 0.6$

(e) State dependency of SES: $\gamma_{SES_{t_1}}$ 0.6

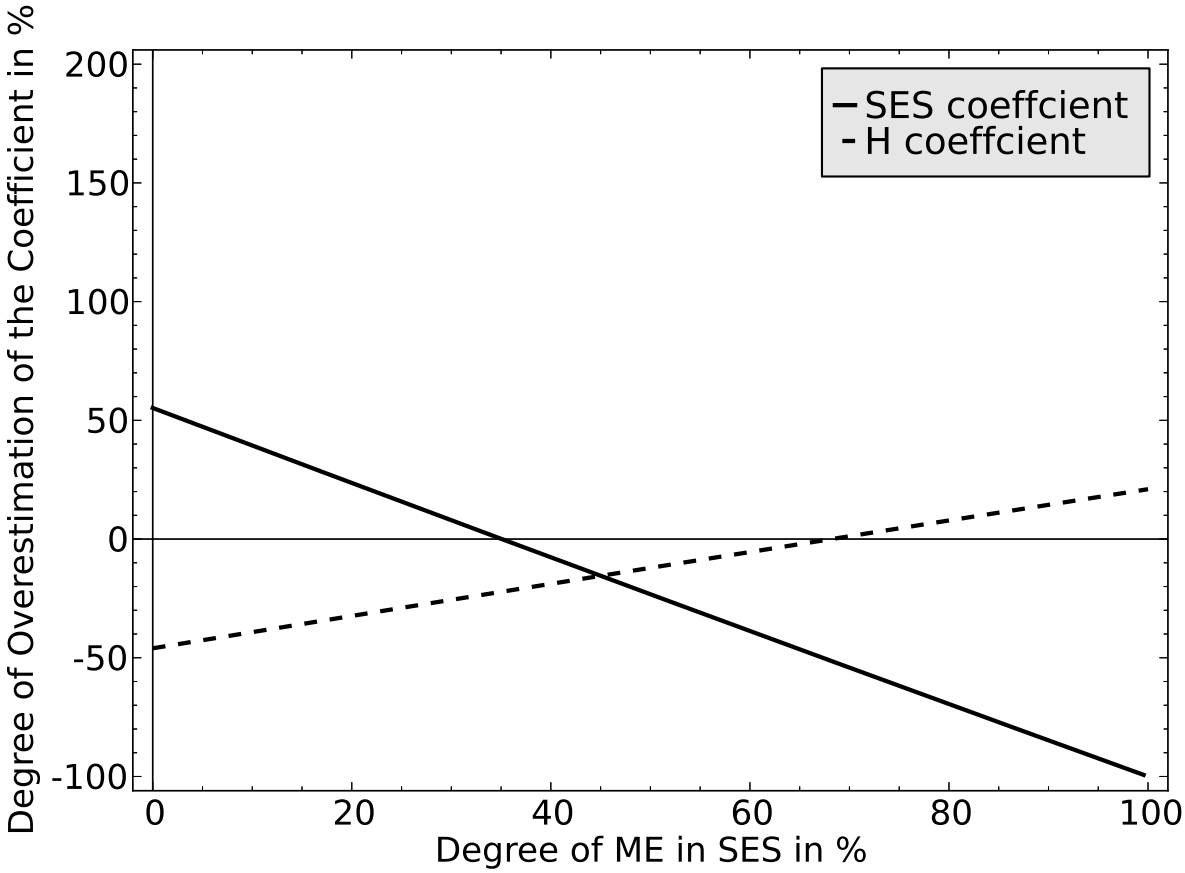
(f) The true coefficients of SES ($\beta_{SES_{t_1}}$) and health ($\beta_{H_{t_1}}$) were set at 0.1 each, resulting in a true ratio of $\Delta_{(SES,H)}$ =1.

In the monte-carlo analyses if the third part of the paper some of these values will be varied to see how they influence conclusions about the ratio of the coefficients. Before we look at the ratio it is worth looking on the effects of measurement error on the individual coefficients. Figure 1 graphically represents the degree of bias in the health coefficient and the SES coefficient dependent on the degree of measurement error in SES. We can see that the size of the SES coefficient (solid line) is overestimated at low values of ME in SES, but decreases steadily in size if the degree of measurement error increases until it is strongly underestimated. The relation of the bias in health coefficient (dashed line) to measurement error in SES is the opposite. At low values of ME in SES the size of the health coefficient is strongly underestimated. The more measurement error the SES variable contains, the larger the estimated of health coefficient gets until it is overestimated at very high levels of ME in SES.

These two opposing relationships lead to a complex relationship of the ratio of the two coefficients to the degree of measurement error. Figure 2 shows the bias of the standardized and the unstandardized ratios depending on the degree of ME in SES fixed at the above-mentioned values. The bias in favour of the SES coefficient decreases for both the standardized (dotted line) and the unstandardized (solid line) ratios. At high degrees of ME in SES, the bias starts to favour the health coefficient relative to the SES coefficient (negative values on the y-axis). Note that the solid line which represents the unstandardized ratio shows much higher values of bias and a steeper curve than that for the standardized values. The process of standardization thus seems to level the bias, at least for the configuration chosen in this example. The unstandardized ratio is unbiased if the degree of ME in SES is the same as the fixed degree of ME in health (45%). The same is not true for the standardized ratio due to the additional influence of the ME in the dependent variables.

In a simulation of a CLPD design we found that the same principle holds true for binary dependent variables using the above mentioned parameters and dichotomised the dependent variables (results not shown).

Fig. 1. Bias in the unstandardized coefficients depending on the degree of ME in SES



2.4. Correlation of SES and health as an intervening factor

Unfortunately, the problem does not end here. As explained above, the bias also depends on other quantities. If we fix the degrees of ME and let the correlation between SES and health at t_1 vary, we get the association described by figure 3. The degree of ME of SES is fixed at 15%. All other quantities take the values described above. The figure shows that the bias in the standardized and unstandardized ratios continuously increases with the correlation of SES and health at time point t. The bias in the estimated ratio is therefore amplified by the strength of the association between SES and health at time t.

Fig. 2. Bias in the standardized and unstandardized ratio depending on the degree of ME in SES

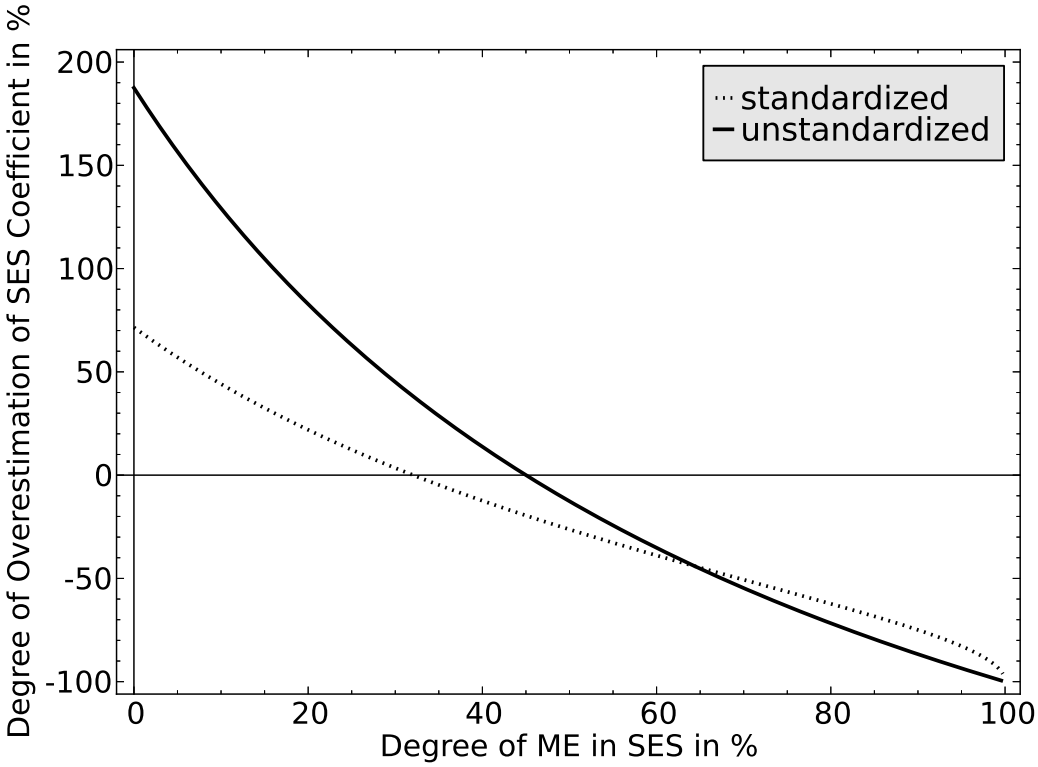
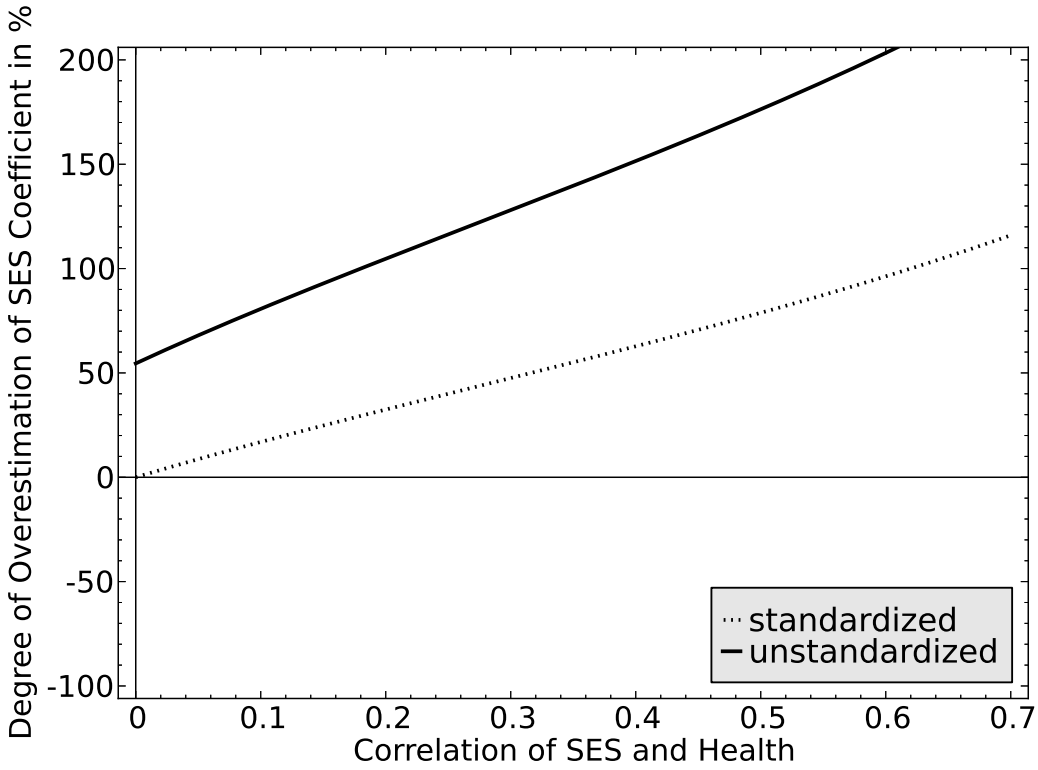


Fig. 3. Bias in the standardized and unstandardized ratio depending on the correlation of SES and health

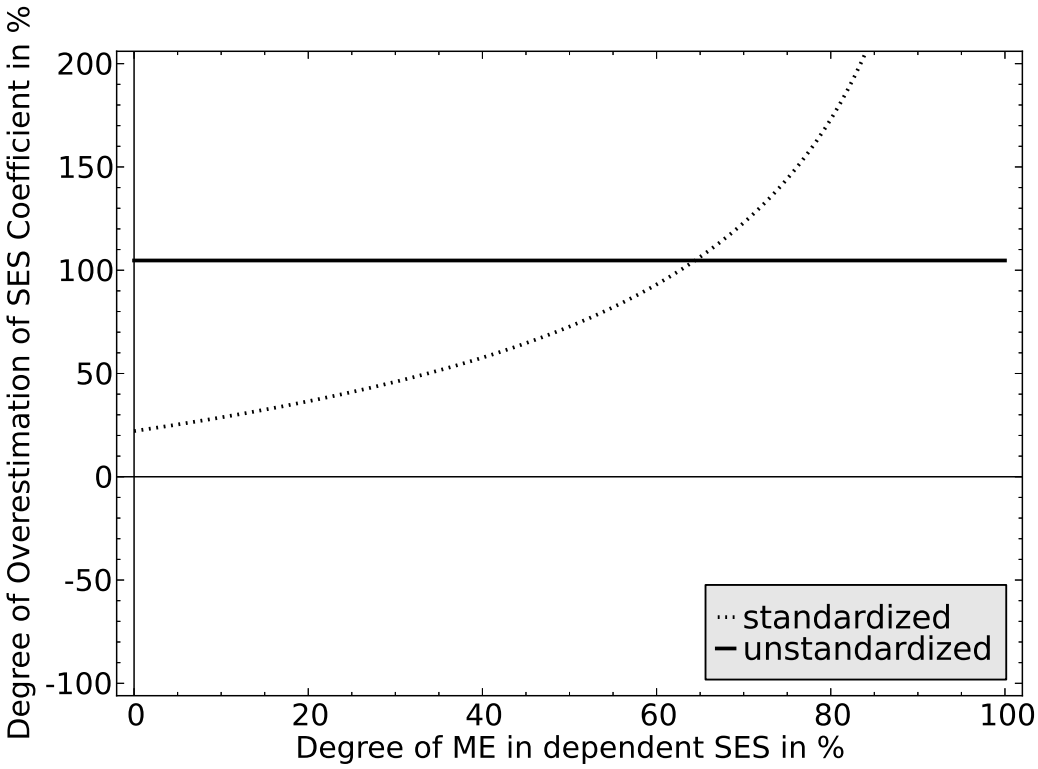


2.5. Measurement error in the dependent variable

Another question that needs to be addressed is the role of the dependent variables, which are also very likely to be measured with error because they represent the same constructs as the independent variables, but at time t_2 instead of time t_1. Figure 4 shows the bias in the standardized and unstandardized ratios depending on the degree of ME in the dependent SES variable. The degree of ME in SES is fixed at 15%. We can see that the solid line is completely flat and the bias does not change with the degree of ME in SES at t_2. The unstandardized ratio is therefore not affected. However, the bias in the standardized ratio (dotted) is influenced by the degree of ME in dependent SES. In contrast to the effect of ME in the independent SES variable, more ME in the dependent SES variable leads to a relative overestimation of the SES coefficient. The fact that the ME in the dependent variable and the ME in the independent

variable have opposite effects on the bias of the ratio is an important result that can lead to counter-intuitive directions of the bias.

Fig. 4. Bias in the standardized and unstandardized ratio depending on the degree of ME in the dependent SES variable



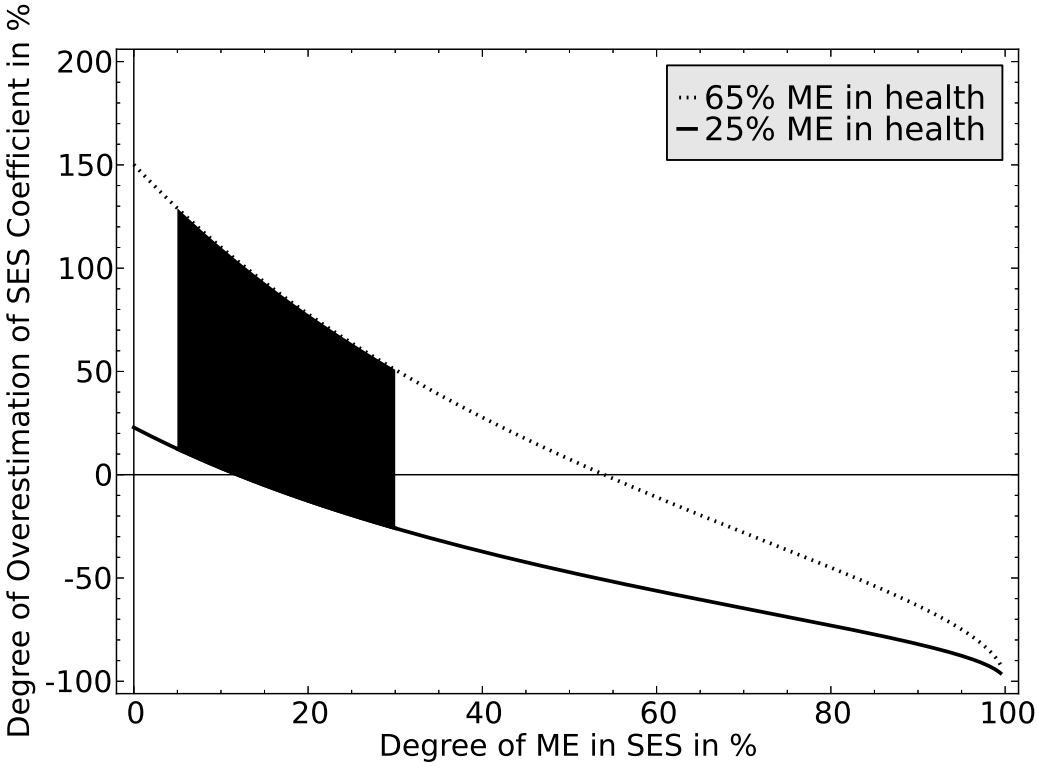
2.6. Plausible scenarios of degree of measurement error

To reduce this very complex multidimensional association of an almost infinite number of possible values of the determinants of the bias, we next plot the bias for specific scenarios which are likely to appear in applied survey research. For this purpose, we first need a plausible range of the degree of ME in health and SES from the literature. A brief overview of the relevant literature can be found in the appendix A. Given the evidence from the literature, we think that in a regular survey it is likely that subjective health data will contain more ME than measures of SES. Estimates for ME in subjective health range between 25% and 65%. For measures of ME in SES, the results range between 5% and 30%.

2.7. Plotting plausible scenarios

Despite the complex association of bias and its determinants, we aim to show a realistic scenario based on the synthesis of the literature discussed in the appendix A. Figure 5 shows two lines which represent the bias depending on the degree of ME in SES. The upper dotted line is true for an assumed degree of ME in health of 65%, and the solid line for a degree of ME in health of 25%. These two figures are the range of plausible values of ME in a subjective health indicator derived from the above-mentioned survey of the literature. In general, the two lines show that if the ME in SES increases, this leads to a relative underestimation of the SES coefficient. The same is true for the health coefficient and the ME in health respectively. The grey area is limited by 5% and 30% ME in SES, the limits of the plausible values for a survey indicator of SES (such as income). The area therefore indicates the bias for what we think are likely scenarios of combinations of degree of ME in SES and health in real world survey data sets. Based on the literature reviewed, it seems rather unlikely that any given survey is not covered by this range of combinations. Following this approach, we can read plausible ranges of bias in a typical survey from this area. With high ME in health (65%) and a very low degree of ME in SES (5%) we get a relative overestimation of the SES coefficient of approximately 130% which is so large that it can fundamentally alter substantive conclusions about the importance of health selection and social causation. With an increasing ME in SES and decreasing ME in health, we see that the grey area approaches and even crosses the zero line. This means that for some configurations the relative bias might be low or close to zero (e.g. degree of ME in health 25% and degree of ME in SES ca. 11%). With the highest degree of ME in SES and the lowest degree of ME in health, we would even underestimate the SES coefficient instead of the health coefficient, by about 20%.

Fig. 5. Range of plausible values for the bias in the standardized ratio



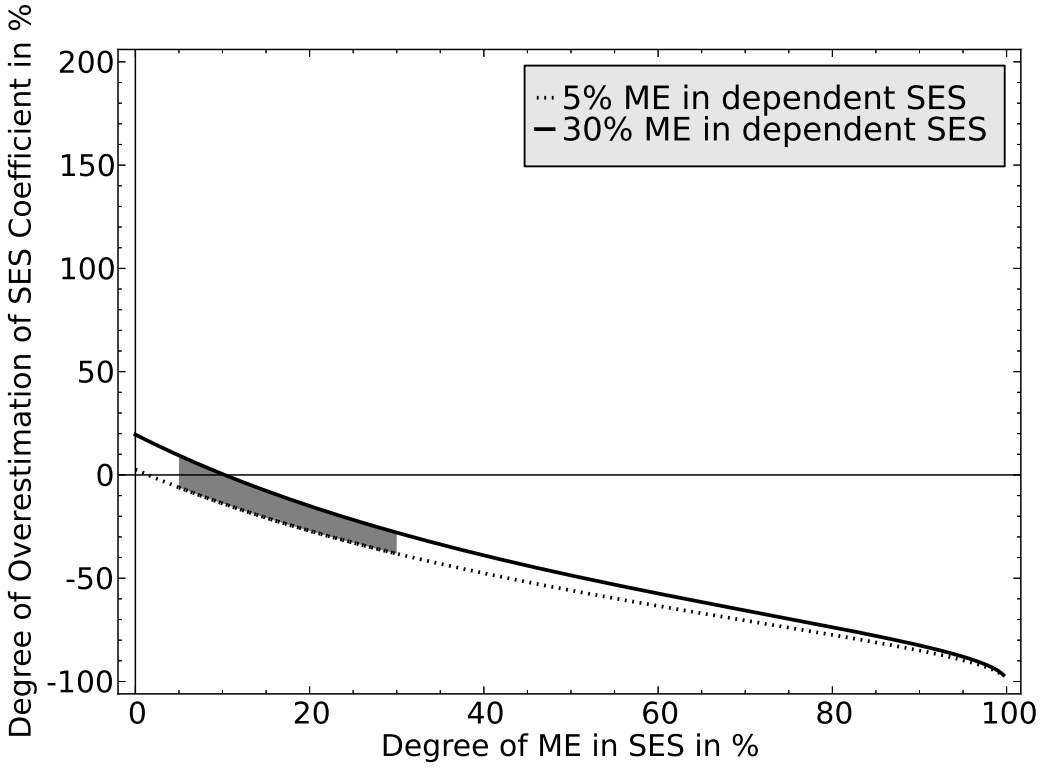
2.8. One variable measured without error

Another scenario to consider is what happens if one construct, like health, is measured without error, but the other construct, like SES, still contains ME. This scenario should not be mistaken for the consequences of ME in bivariate analysis. The ME will influence the estimate of the coefficient of the variable without ME as well. In addition, the ME in the dependent variable will also factor in, when we are dealing with standardized regression coefficients.

Figure 6 shows a graph similar to Figure 5. The difference with respect to figure 5 is that the ME in health is set at zero. The dotted line is the association between the degree of ME in SES and the bias of the standardized ratio, given a degree of ME in the dependent SES variable (t_2) of 5%. The solid line represents the same association for a degree of ME in the dependent SES (t_2) of 30%. The grey area represents the plausible range of values for the ME in SES at times t_1 and t_2. We can see that a zero measurement error in health would only

shift the problem of the bias in the ratio. Now, for most of the plausible range we have an underestimation instead of an overestimation of the social causation coefficient. Therefore, even if we improve our measurement in one variable we can be left with no improvement with regard to the bias; it just changes its sign.

Fig. 6. Range of plausible values for the bias in the standardized ratio with health measured without error



2.9. Solutions

What can be done about the problem? We discuss four solutions based on frequentist statistics. We do not claim to be exhaustive and refer to the literature for more detailed discussions (Keogh and White, 2014). Two of these solutions might only be feasible for either health or SES (see table 1).

First, using measurement models in a SEM framework can account for ME. Here, the regression is not conducted on the observables, but on the latent variables estimated from a

measurement model that is similar to confirmatory factor analysis. Similar in principal is the regression calibration approach. At least two indicators or measurements of the same construct are needed for both approaches. For health, this has often been done, and given typical survey data can be implemented relatively easily. For SES, the problem becomes more difficult. In the literature there are several examples of studies which use measurement models for SES similar to measurement models for health (Bollen et al., 2007; Garbarski, 2010; Warren, 2009). However, three restrictions must be accepted when choosing such an approach:

(a) The interest lies in general socioeconomic status, not in specific aspects of the different measurements, like income or education.

(b) The theoretical and methodological approach cannot assume that all indicators, e.g., income, education, and occupation are the causes of SES. Instead the opposite has to be assumed for at at least a subset of the variables to identify the measurement model (MIMIC). This implies that an underlying SES causes the observed measures, e.g. income, education, and occupation.

(c) Following from (b), the interest is only in the common variance of the indicators of SES. Additive effects of individual SES indicators are not considered.

Second, it is possible to assume a plausible range of ME for SES and health based on past research and conduct sensitivity analyses within this range. The question that should be addressed in the sensitivity analysis is whether the conclusions drawn from the results presented in the study change substantially within the range of plausible values of ME. If this is the case, statements about the relative strength of health selection or social causation should be made with caution. If the conclusions do not change, this speaks of the robustness of the findings.

A third approach, popular in econometrics, is instrumental variable regression (IV). Here, an external variation is used to exclude the possibility of attenuation bias. While there are many examples of possible instruments for measures of SES, the number of studies which use convincing instruments for subjective health is very low. One way would be to estimate a complex model where health is measured as a latent variable and the measure of SES is instrumented by an exogenous variable. If the data set satisfies these requirements, this is fine. More often than not,

however, researchers are faced with a situation in which they cannot estimate such a complex model.

A fourth strategy is to use a multiple imputation for measurement error (MIME) approach. The idea behind this approach is to treat the true values as missing and to impute the values based on information from all the covariates and the outcome. This can also be applied to the case of two or more variables measured with error through chained equations. Nevertheless, it is a more complicated approach and still relies on a second measurement of the construct. Alternatively, the true values have to be available for a subset of the sample.

3. A Bayesian approach

The methods outlined in the previous section are fairly well known and established in the literature on the treatment of measurement error. All methods described above require additional information in the data set. This can be a repeated measurement of a construct for the measurement model, an instrumental variable for IV-regression, or a validation measurement without ME for MIME. We will briefly outline a newer approach rooted in Bayesian statistics that can be applied if no such additional information is available. It fulfils a similar purpose to that of a sensitivity analysis. However, it is more formalized and follows a different type of statistical thinking (*Bayesian* vs. *frequentist*). Before we outline the approach and give an example, we would like to stress the fact that the method can be combined with the analysis of repeated measurements. Further, using informative priors which take into account ME, as presented here, can be seen as a first part of a multiple bias model for observed data (Greenland, 2005, 2009).

The general modelling approach we use is called Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009; Martins et al., 2013), implemented as a package in R (www.r-inla.org). It is an approach for latent Gaussian models which allows much flexibility in the specification of the model. It is possible to use INLA for such a wide range of models like generalized mixed models, survival analysis, generalized dynamic models, spatial and spatiotemporal models (Martins et al., 2013, p. 68).

The term Gaussian refers to the vector of latent Gaussian variables ($\mathbf{x}$) that is used in all models. The vectors of hyperparameters (Φ) for $\mathbf{x}$ and the observed dependent variables ($\mathbf{y}$) can be

Table 1. Frequentist solutions for the problem of ME with strengths and weaknesses

Approach	Implementation	Advantage	Problems
Measurement Model/ Regression Calibration	several indicators of the same construct; confirmatory factor analysis (MM); reliability parameter as correction (RC)	• easy to implement for many constructs both in surveys and analysis • well-known methods	• for some constructs not possible or unusual
Sensitivity Analysis	checking robustness when different degrees of ME are specified in the model	• easy to implement	• knowing what ranges are plausible, under which external conditions • can complicate the analysis beyond the limit of comprehension
Instrumental Variables	using external variation unrelated to the dependent variable	• solves problems of ME and confounding • established method	• two instruments are hard to find • generalizability of results • standardized results still affected by ME in dependent variable
Multiple Imputation	multiple imputation of the values of variables with error from models based on the true value	• consistent way of getting unbiased results without additional time points or indicators	• needs a measure of the true value at least for a subgroup, or a second measurement • not widespread • complicated

non-Gaussian. Hyperparameters are the parameters of the prior distribution, in contrast to the parameters of the Gaussian model that is estimated. As the name indicates INLA approximates the posterior distribution, first of Φ, then the density of $\mathbf{x}$, given Φ for certain values of Φ. In the last step, the two approximations are combined through numerical integration or an algorithm that does not need numerical integration. The latter comes at the cost of precision of the posterior marginal distribution of the hyperparameters, but is computationally much faster and the standard in the R-INLA implementation (Martins et al., 2013, p. 75). For latent Gaussian models the error of the approximation of the posterior distribution is smaller than in comparable and still more common Markov Chain Monte Carlo (MCMC) methods. Additionally, it is much faster to compute. Despite the fact that INLA rests on a statistically complex and demanding approximation of the posterior distribution, the implementation in R is relatively user friendly (Rue et al., 2009).

We will refer to the specific application of INLA to the problem of measurement error as ME.INLA. Using ME.INLA for a model with errors in one variable is exemplified in more detail in Muff et al. (2015). However, we conduct a straightforward expansion of their work by considering the problem of not one but two variables measured with error. The use of INLA for models with errors in variables is relatively new and has to the best of our knowledge not been used for the study of the reciprocal relationship of health and SES, a CLPD study, or research using survey items measured with error in general. We therefore present the first attempt to apply ME.INLA to problems encountered in a CLPD design, exemplified with the problem of reciprocal causality of subjective health and SES.

The unique characteristic of ME.INLA compared to methods described in the previous section is that external or expert knowledge about the degree of measurement error in health and SES can be used in a Bayesian approach to specify informative prior distributions of the variance of the measurement error term and the unobserved true variable. What is known to the researcher is the observed variance from the data (e.g. σ_{ses}, but not σ_{SES}). As defined above we know that this observed variance can be decomposed in one part true variance and one part random measurement error. What we need to know to specify the prior correctly is how much of the observed variance is due to measurement error. We cannot know the exact figure for any given

data we have. If we had e.g. repeated measures of subjective health or labor market income (for a subset of the data) we could get an empirical estimate of the degree of ME for one or both variables. For our example, we assume that such repeated measures are not available. Therefore we have to rely on external knowledge about the distribution of the degree of measurement error. Again, we draw on our review of the literature on measurement error in SES and subjective health. As above, the plausible range of degree of ME for health is assumed to be 25%-65% and for SES 5%-30%. Based on these ranges the 2.5% lower bound of the prior distribution of the variance of the measurement error in SES ($\sigma_{ME_{SES}}$) is 0.05 times the observed variance of SES (σ_{ses}). The 97.5% upper bound of the prior distribution of the variance of measurement error in SES is 0.3 times the observed variance in SES. The respective values for the prior distribution for the variance of the true SES (σ_{SES}) are 0.95 times observed variance and 0.7 times observed variance. Formally applied to the CLPD the classical measurement error model of ME.INLA used by (Muff et al., 2015, p. 9) is defined in the following way:

$$
\begin{aligned}
\mathbf{SES}_{t_2}|\mathbf{H}_{t_1} &\sim N(\beta_{0,SES_{t_2}}\mathbf{1} + \beta_{H_{t_1}}\mathbf{H}_{t_1} + \gamma_{SES_{t_1}}\mathbf{SES}_{t_1}, \tau_{\epsilon_{SES_{t_2}}}\mathbf{I}),\\
\mathbf{H}_{t_2}|\mathbf{SES}_{t_1} &\sim N(\beta_{0,H_{t_2}}\mathbf{1} + \beta_{SES_{t_1}}\mathbf{SES}_{t_1} + \gamma_{H_{t_1}}\mathbf{H}_{t_1}, \tau_{\epsilon_{H_{t_2}}}\mathbf{I}),\\
\mathbf{H}_{t_1} &= \alpha_{0,H}\mathbf{1} + \alpha_{SES_{t_1}}\mathbf{SES}_{t_1} + \boldsymbol{\epsilon}_{H_{t_1}}, \qquad \boldsymbol{\epsilon}_{H_{t_1}} \sim N(0, \tau_{H_{t_1}}\mathbf{I}),\\
\mathbf{h}_{t_1} &= \mathbf{H}_{t_1} + \mathbf{u}_{ME_{SES_{t_1}}}, \qquad \mathbf{u}_{ME_{SES_{t_1}}} \sim N(0, \tau_{ME_{H_{t_1}}}\mathbf{I}),\\
\mathbf{SES}_{t_1} &= \alpha_{0,SES}\mathbf{1} + \alpha_{H_{t_1}}\mathbf{H}_{t_1} + \boldsymbol{\epsilon}_{SES_{t_1}}, \qquad \boldsymbol{\epsilon}_{SES_{t_1}} \sim N(0, \tau_{SES_{t_1}}\mathbf{I}), \text{ and}\\
\mathbf{ses}_{t_1} &= \mathbf{SES}_{t_1} + \mathbf{u}_{ME_{SES_{t_1}}}, \qquad \mathbf{u}_{ME_{SES_{t_1}}} \sim N(0, \tau_{ME_{SES_{t_1}}}\mathbf{I}).
\end{aligned} \tag{11}
$$

Note that here we use precisions instead of variances, because the R-INLA package works with precisions. For the sake of simplicity we do not introduce covariates, although the model can accommodate covariates that are assumed to be measured without error. The τ parameters indicate the precision, which is the inverse of the variance. The unknowns for the social causation pathway are the latent field $v_{SC} = (\mathbf{SES}_{t_1}^T, \mathbf{H}_{t_1}^T)^T$ and the so-called hyper-parameters $\Psi_{SC} = (\beta_{SES_{t_1}}, \gamma_{H_{t_1}}, \tau_{ME_{H_{t_1}}}, \tau_{ME_{SES_{t_1}}}, \tau_{H_{t_1}}, \tau_{SES_{t_1}}, \tau_{\epsilon_{H_{t_2}}})^T$. The unknowns for the health selection pathway are the same latent field $v_{SC} = v_{HS} = (\mathbf{SES}_{t_1}^T, \mathbf{H}_{t_1}^T)^T$ and the hyper-parameters $\Psi_{HS} = (\beta_{H_{t_1}}, \gamma_{SES_{t_1}}, \tau_{ME_{H_{t_1}}}, \tau_{ME_{SES_{t_1}}}, \tau_{H_{t_1}}, \tau_{SES_{t_1}}, \tau_{\epsilon_{SES_{t_2}}})^T$. For the vectors of parameters

$\theta_{SC} = (\beta_{SES_{t_1}}, \gamma_{H_{t_1}})^T$ and $\theta_{HS} = (\beta_{H_{t_1}}, \gamma_{SES_{t_1}})^T$, no prior knowledge is assumed, corresponding to a prior of $N(0, 10^{-4})$. Given the variance of the observed variable in the sample, we can specify an upper limit and a lower limit of the precision parameters for the measurement error term and the true variables by using the plausible range of degree of ME from the literature. This approach shows that the prior specification and consequently the inference drawn from the posterior distribution are dependent on the external information collected by the researchers. The correction of the ME can only be as good as the information about the size of the degree of ME. This problem will become apparent in our examples and will be discussed again in the conclusion.

We use simulated data to illustrate the utility and some limitations of the proposed method. Our example should not be seen as a systematic investigation of all possible strengths and weaknesses of ME.INLA, because this would be beyond the scope of the paper. Instead we show three scenarios that might arise in a study based on survey data. The data generating model on which the simulations are based is the CLPD as defined in (1) and (2). We set the determinants of the bias in the ratio to the values already used for the graphical evaluation in section 3:

(a) Correlation of SES at t_1 and Health at t_1: 0.2

(b) State dependency of health: $\gamma_{H_{t_1}} = 0.6$

(c) State dependency of SES: $\gamma_{SES_{t_1}}$ 0.6

(d) The true coefficients of SES ($\beta_{SES_{t_1}}$) and health ($\beta_{H_{t_1}}$) were set at 0.1 each, resulting in a true ratio of $\Delta_{(SES,H)} = 1$.

(e) Observed variance in SES and health: $\sigma_{ses} = \sigma_h$ =1

(f) Number of observations per draw: 5000.

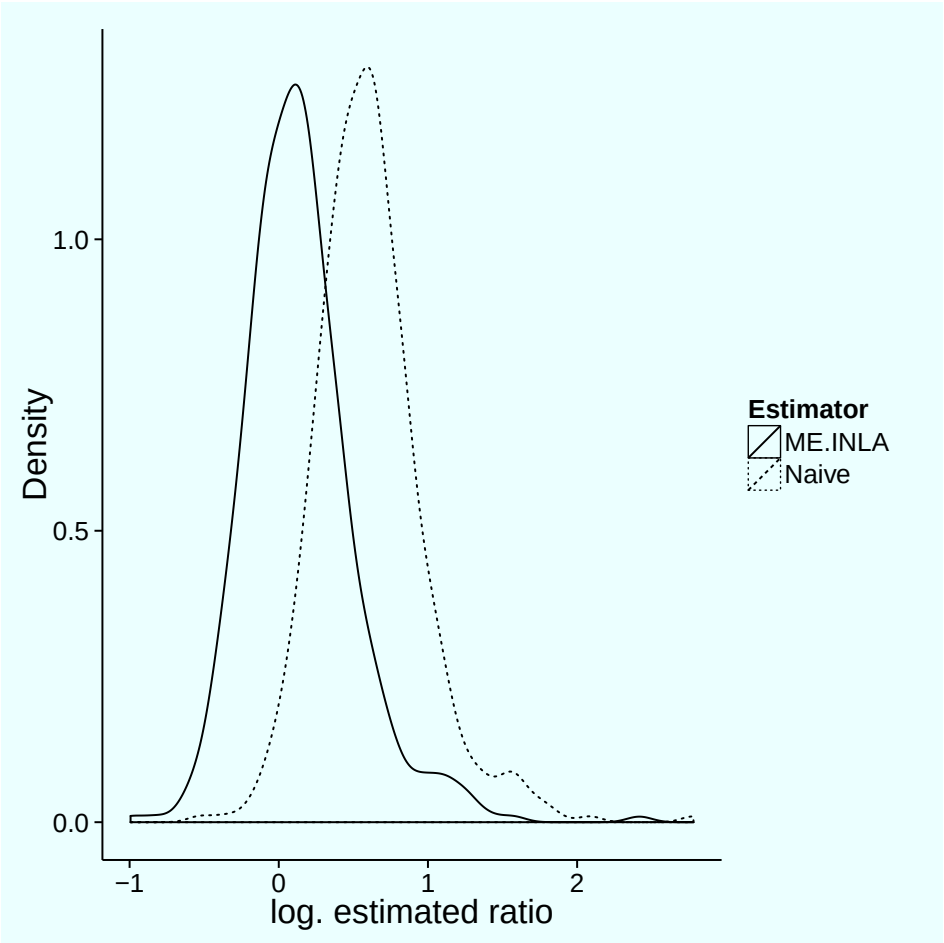
In the first step we assume that the degree of measurement error in health and SES are randomly drawn from a Gamma distribution with the 2.5 % percentile at the lower bound of the plausible range of values for the precision for SES and health ($\frac{1}{0.95}$ and $\frac{1}{0.65}$) and the 97.5% percentile at the upper bound ($\frac{1}{0.3}$ and $\frac{1}{0.25}$). Each simulation was iterated 500 times.
The parameters of the Gamma distribution which correspond to the above-mentioned 2.5% and

97.5% probability limit are found by a numerical optimization process (Muff et al., 2015). For the health variable a Gamma distribution is found that has the two precision values ($\tau = \frac{1}{0.65} = 1.54$ and $\tau = \frac{1}{0.25} = 4$) at the 2.5% and 97.5% probability limits: G(17.3, 6.6). The Gamma distribution for true precision of the health variable with ($\tau = \frac{1}{0.75} = 1.33$ and $\tau = \frac{1}{0.35} = 2.86$) at the 2.5% and 97.5% probability limits is G(26.9,13.3). The respective Gamma distributions for the simulated SES variable are: G(5.34,0.54) and G(164,133).

Table 2 reports the results of the simulations assuming that the degrees of measurement error in SES and health are randomly drawn from the above-mentioned distributions. We compare the deviation from the true ratio of the naïve estimation without correction to the deviation of the ratio estimated using ME.INLA informed by the distributions defined above. The table shows that without correction the ratio is widely overestimated by 80%. Using ME.INLA reduces the bias considerably. However, the ratio is still overestimated by approximately 12%.

From these results we can conclude that given our CLPD setup the INLA method will reduce the bias considerably if we have repeated draws of the data from a correctly specified distribution of degree of measurement error. However, in real world applications we will usually have just one data set and we will not draw repeated samples. In such a case it is also informative to look how the method performs under certain fixed values of measurement error, because they will be fixed on an unknown quantity in a given survey. Figure 7 shows a density plot of the log. ratios of the 500 simulations from the first example for naïve and correction method. It illustrates that in most iterations ME.INLA is much closer to the true value than the naïve estimator. There are, however, several instances in which the estimates of ME.INLA and the naïve estimate overlap indicating that certain constellations of degree of ME drawn from the Gamma distributions might lead to a poor performance of ME.INLA given the prior specification we chose.

Fig. 7. Density plot of the logarithmized ratios using ME.INLA and the naïve estimator



To investigate this further we repeated the simulations for three combinations of degree of measurement error in the SES and health variable. They represent the average from the plausible ranges, and the extreme combinations of high ME in SES and low ME in health, and vice versa:

(a) Scenario 1: average degree of ME in SES = 17.5%, average degree of ME in health =45%

(b) Scenario 2: low degree of ME in SES = 5%, high degree of ME in health =65%

(c) Scenario 3: high degree of ME in SES = 30%, low degree of ME in health =25%

For the combination of average ME we see that the correction method performs well. It reduces the median bias in the ratio from 71 % to 7%. For the high ME in health, low ME in SES

Table 2. Comparison of naïve and Bayesian ME.INLA estimates using simulated data

Degree of ME in SES	Degree of ME in Health	True Ratio	Naïve Ratio	ME.INLA Ratio
G(5.34,0.54)	G(17.3, 6.6)	1	1.80	1.12
17.5%	45%	1	1.71	1.07
5%	65%	1	3.67	2.36
30%	25%	1	0.91	0.58
Observations				5000
Iterations				500

it performs better than the naïve estimator but is still way off the mark (136% overestimation instead of 267%). Last, if ME in health is actually set to be a bit lower than in SES we see that the correction actually performs worse than the naïve estimator (42% underestimation instead of only 9% underestimation for the naïve estimate). This is due to the fact noted above that in case of almost equality of degree of measurement error in both variables the ratio will be unbiased. We can see that the correct specification of the plausible ranges of measurement error is essential to the performance of the method. On the one hand, we get a substantial improvement in the bias reduction if the assumptions on the distribution are correct. If we are off with our prediction we might actually have a poorer estimate than with the naïve method. As indicated above our correction is only as good as our information for the prior specification. Consequently, we cannot simply apply ME.INLA and assume we have eradicated all influences of ME. Rather, we should use the method as a sensitivity check to naïve estimations. We can check whether our conclusions about the relative strength of health selection and social causation still hold after we take into account a plausible range of measurement error?

So far, our approach has only included unstandardized coefficients. In practice, standardized estimates are also often used (Aittomaki et al., 2012; Eaton et al., 2001; Huurre et al., 2005). At the moment, we cannot give a convincing example of an implementation of standardized results using INLA. However, we are certain that this can be integrated in the future. The same is true for the simultaneous estimation of the social causation and health selection equations that would, for example, allow the error terms to be correlated. In future research this can be integrated in a Bayesian analysis of a CLPD with two variables measured with error.

4. Discussion

We have shown in this article that differences in random measurement error can seriously affect the interpretation of causal relationships in a bidirectional analysis. The variables measured with a higher degree of ME will be relatively advantaged with regard to coefficient strength, given certain plausible conditions.

A review of the literature gave a range of measurement error in subjective health of 25-65% and for the three classical measures of SES of 5-30%. This suggests that when using normal survey data without measurement models, the social causation hypothesis will in most cases be systematically overstated relative to the health selection hypothesis when operating within a cross-lagged panel design. If a measurement model is used for health but not for SES, the opposite is true.

As feasible solutions we have suggested measurement models, sensitivity analysis, IV, MIME, or a combination of a measurement model for subjective health plus instrumenting SES measures in a SEM framework if the data allow these methods. As a more recently developed approach, we have given an example of a Bayesian model developed to take measurement error into account and applied it to data simulating realistic degrees of measurement error for survey research. We have shown that the results of analysis of survey data can be substantively different if ME is taken into account. If measurement error cannot be corrected through traditional means, a ME.INLA analysis with informative priors about the degree of measurement error can help to obtain more robust results from a cross-lagged panel design. Compared to deterministic approaches to sensitivity analysis, ME.INLA is more formalized and reduces information through a probabilistic specification of the plausible range of measurement error.

It should be noted that the prior information used in this study is of an exploratory nature. For similar endeavours, researchers should make sure that they consult the literature intensively and look for studies that assess ME in indicators which are as similar as possible, and look at surveys which have used similar survey design. This ensures that the priors are well specified. The results from the simulations underline the importance of well specified priors for the degree of measurement error. If the distribution does not match the realised degree of measurement error in a given study, the ME.INLA method will yield results that are further away from the

true ratio than the naïve estimator. The Bayesian model presented in the example should also be tested for its performance when including control variables for confounding factors. Whether and how standardized results can be obtained from such a Bayesian approach should also be considered. Finally, in the future researchers should strive for an integration of the Bayesian approach into a simultaneous estimation of the two equations of the CLPD.

CLPD is widely used because of its analytic strength. However, complex bias structures can appear if the variables of interest are measured with error. For the question about the relative strength of social causation versus health selection and similar research questions, a careful modelling of measurement error should always be undertaken; otherwise, inference might be seriously biased in an unknown direction.

5. Acknowledgements

The research was funded by the European Research Council (ERC): [ERC: 313532]. The authors would like to thank Kristian Karlson, two anonymous reviewers, the associate editor, and the joint editor for valuable comments on earlier versions of the paper.

References

Abowd, J. M., H. Stinson, and Martha (2011). Estimating measurement error in SIPP annual job earnings: A comparison of census bureau survey and SSA administrative data. SSRN Scholarly Paper ID 1894690, Social Science Research Network, Rochester, NY.

Adler, N. E., T. Boyce, M. A. Chesney, S. Cohen, S. Folkman, R. L. Kahn, and S. L. Syme (1994). Socioeconomic status and health. the challenge of the gradient. *The American psychologist 49*(1), 15–24.

Aittomaki, A., P. Martikainen, M. Laaksonen, E. Lahelma, and O. Rahkonen (2012). Household economic resources, labour-market advantage and health problems - a study on causal relationships using prospective register data. *Social science and medicine 75*(7), 1303–1310.

Asher, H. B. (1974). Some consequences of measurement error in survey data. *American Journal of Political Science 18*(2), 469–485.

Bassi, F., A. Padoan, and U. Trivellato (2008). Inconsistencies in reported employment characteristics among employed stayers. *IZA Discussion Paper* (3908).

Ben-Shlomo, Y. and D. Kuh (2002). A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology 31*(2), 285–293.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons.

Bollen, K. A., J. L. Glanville, and G. Stecklov (2007). Socio-economic status, permanent income, and fertility: A latent-variable approach. *Population Studies 61*(1), 15–34.

Brunner, J. and P. C. Austin (2009). Inflation of type I error rate in multiple regression when independent variables are measured with error. *Canadian Journal of Statistics 37*(1), 33–46.

Chandola, T., M. Bartley, A. Sacker, C. Jenkinson, and M. Marmot (2003). Health selection in the whitehall II study, UK. *Social Science & Medicine 56*(10), 2059–2072.

Chandola, T., P. Clarke, J. N. Morris, and D. Blane (2006). Pathways between education and health: a causal modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 169*(2), 337–359.

Contoyannis, P., A. M. Jones, and N. Rice (2004). The dynamics of health in the british household panel survey. *Journal of Applied Econometrics 19*(4), 473–503.

Crossley, T. F. and S. Kennedy (2002). The reliability of self-assessed health status. *Journal of Health Economics 21*(4), 643–658.

Dowd, J. B. and A. Zajacova (2010). Does self-rated health mean the same thing across socioeconomic groups? evidence from biomarker data. *Annals of Epidemiology 20*(10), 743–749.

Duncan, G. J. and D. H. Hill (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. *Journal of Labor Economics 3*(4), 508–532.

Eaton, W. W., C. Muntaner, G. Bovasso, and C. Smith (2001). Socioeconomic status and depressive syndrome: The role of inter- and intra-generational mobility, government assistance, and work environment. *Journal of Health and Social Behavior 42*(3), 277–294.

Garbarski, D. (2010). Perceived social position and health: Is there a reciprocal relationship? *Social Science and Medicine 70*(5), 692–699.

Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 168*(2), 267–306.

Greenland, S. (2009). Bayesian perspectives for epidemiologic research: III. bias analysis via missing-data methods. *International Journal of Epidemiology 38*(6), 1662–1673.

Heckman, J. J. (2008). Econometric causality. Working Paper 13934, National Bureau of Economic Research.

Heponiemi, T., M. Elovainio, K. Manderbacka, A.-M. Aalto, M. Kivimäki, and I. Keskimäki (2007). Relationship between unemployment and health among health care professionals: Health selection or health effect? *Journal of Psychosomatic Research 63*(4), 425–431.

Huurre, T., O. Rahkonen, E. Komulainen, and H. Aro (2005). Socioeconomic status as a cause and consequence of psychosomatic symptoms from adolescence to adulthood. *Social psychiatry and psychiatric epidemiology 40*(7), 580–587.

Kane, T. J., C. E. Rouse, and D. Staiger (1999). Estimating returns to schooling when schooling is misreported. *National Bureau of Economic Research Working Paper Series* (7235).

Kenny, D. A. (2005). Cross-lagged panel design. In *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd.

Keogh, R. H. and I. R. White (2014). A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Statistics in Medicine 33*(12), 2137–2155.

Kreiner, C. T., D. D. Lassen, and S. Leth-Petersen (2011). Examples of combining administrative records and survey data in validation studies. In *Improving the Measurement of Consumer Expenditures*. University of Chicago Press.

Lorenz, F. O., R. D. Conger, R. L. Simons, and L. B. Whitbeck (1995). The effects of unequal covariances and reliabilities on contemporaneous inference: The case of hostility and marital happiness. *Journal of Marriage and the Family 57*(4), 1049.

Marquis, K. H., M. S. Marquis, and J. M. Polich (1986). Response bias and reliability in sensitive topic surveys. *Journal of the American Statistical Association 81*(394), 381–389.

Martins, T. G., D. Simpson, F. Lindgren, and H. Rue (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis 67*, 68–83.

Mathiowetz, N. A. (1992). Errors in reports of occupation. *The Public Opinion Quarterly 56*(3), 352–355.

Moore, J. C., L. L. Stinson, and E. J. Welniak (2000). Income measurement error in surveys: A review. *Journal Of Official Statistics-stockholm 16*(4), 331–362.

Muff, S., A. Riebler, L. Held, H. Rue, and P. Saner (2015). Bayesian analysis of measurement error models using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, forthcoming.

Newcomb, M. D. and P. Bentler (1987). Self-report methods of assessing health status and health service utilization: A hierarchical confirmatory analysis. *Multivariate Behavioral Research 22*(4), 415–436.

Power, C., S. Matthews, and O. Manor (1996). Inequalities in self rated health in the 1958 birth cohort: lifetime social circumstances or social mobility? *BMJ : British Medical Journal 313*(7055), 449–453.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2004). Generalized multilevel structural equation modeling. *Psychometrika 69*(2), 167–190.

Rodgers, W. L., C. Brown, and G. J. Duncan (1993). Errors in survey reports of earnings, hours worked, and hourly wages. *Journal of the American Statistical Association 88*(424), 1208–1218.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association 100*(469), 322–331.

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 319–392.

Shmueli, A. (2003). Socio-economic and demographic variation in health and in its measures: the issue of reporting heterogeneity. *Social Science & Medicine 57*(1), 125–134.

Siegel, P. M. and Hodge (1968). A causal approach to the study of measurement error. In H. M. Blalock and A. B. Blalock (Eds.), *Methodology in social research*, pp. 28–59. New York: McGraw-Hill.

Stavola, B. L. D., D. Nitsch, I. d. S. Silva, V. McCormack, R. Hardy, V. Mann, T. J. Cole, S. Morton, and D. A. Leon (2006). Statistical issues in life course epidemiology. *American Journal of Epidemiology 163*(1), 84–96.

Warren, J. R. (2009). Socioeconomic status and health across the life course: A test of the social causation and health selection hypotheses. *Social Forces 87*(4), 2125–2153.

Wolff, J. K., A. Brose, M. Lövdén, C. Tesch-Römer, U. Lindenberger, and F. Schmiedek (2012). Health is health is health? age differences in intraindividual variability and in within-person versus between-person factor structures of self-reported health complaints. *Psychology and Aging 27*(4), 881–891.

A. Appendix - Measurement error in subjective health and SES

For this paper, we have defined random ME as the deviation of the observed value from the true value. Therefore the variance of the observed value is the variance of the true value plus the variance of the ME. The degree of ME is therefore the variance of the ME divided by the variance of the observed variable. But how can we know how much ME we have? What is the true value if the responses in a survey are assumed to contain ME?

For measures of SES and health, there are different approaches to answering this question. We will give a short overview of findings from the literature on ME in income, and occupation. Afterwards, we will present evidence on measurement error in subjective health.

A.1. Measurement error in income

For income, the approach which is used most often is a link between survey data with official registers. These can be data from tax records, employers' files, or social security systems. Studies using the record linkage method then treat the register data in different ways. One way is to treat them as the true value. The other approach is to assume that even register data are subject to ME, but that the survey and register ME are statistically independent (Kreiner et al., 2011). Regardless of the approach, there are several studies which estimate the amount of ME in survey responses. Moore et al. (2000) provide an extensive review on the subject. The estimates range from 7% (Duncan and Hill, 1985) to a little more than 20% (Rodgers et al., 1993). There are very few studies which report a degree of ME which is above or below this range. The review by Moore et al. is in line with the findings of (Marquis et al., 1986), who performed a review at an earlier point in time. This earlier review is especially valuable because the authors assess different sensitive topics with the same methodology with regard to reliability. They find that estimates for income range between 70% and well over 90% in reliability, meaning that they find between somewhat less than 10% and 30% of measurement error in income. Two more recent studies come to an estimate of about 20% of ME in income (Abowd et al., 2011; Kreiner et al., 2011).

A.2. ME in occupational status

Regarding occupation, there is considerably less evidence. This is probably due to the fact that income is the key variable in most labour economics studies, and that comparison with register data is easier. In an early study, misreporting of occupational status was found to be around 56% on the detailed three digit level, and about 20% on the broad one digit level recall category (Mathiowetz, 1992). Bassi et al. (2008) find that around 10% of the respondents misreport their industry (12 categories) and around 13% misreport their professional status (11 categories). In a comparison of self-reports and transcripts, Kane et al. (1999) find only about 2.5% of persons disagreeing with their transcript record.

If the analysis is not built on very detailed categorizations of occupational groups, the degree of ME in occupation does not exceed that of income. To sum up, for a normal survey it seems feasible to assume that the observed measures of SES contains between 5% and 30% of ME.

A.3. Measurement error in subjective health

For subjective health, the problem of random ME is harder to address. For chronic illnesses, the aforementioned review by Marquis et al. (1986) suggests that ME for such conditions in interviews ranges between 45% and over 80%. However, there are no registers which can record true subjective health. Therefore, the approach which seems most fruitful is to model a latent variable for subjective health and see how it is related to the observed measures of subjective health. One popular approach to this kind of model is the application of confirmatory factor analysis (CFA). While this approach is very popular in many (social-)psychological studies, it is less used in studies from public health, medical sociology or health economics when it comes to subjective health. Therefore, there are no systematic reviews on the issue. We provide evidence from several studies without claiming any exhaustiveness. Note that the studies mentioned often do not address the issue of degree of ME explicitly. Nevertheless, they do provide the necessary statistics.

The study by Shmueli (2003) uses a measurement model of subjective health and estimates 30-60% of ME using data on older Israeli Jews. Chandola et al. (2006) use a measurement model for subjective health and report degrees of measurement error in the indicators between

10% and over 80%. (Wolff et al., 2012) use panel data to estimate within-person reliability and between-person reliability of subjective health complaints. They report estimates of ME of 60%-90% for within-person (meaning change in health complaints) ME and of 30%-75% of ME for between-person variation. For young US adults, an early study by Newcomb and Bentler (1987) estimates the ME of subjective health to be between 40% and 80%.

Using test-retest analysis, one study from the US finds that about 40% of respondents change their subjective health rating within one month (Dowd and Zajacova, 2010). Using a very similar design, Crossley and Kennedy (2002) show that 28% of Australian respondents change their subjective health status within a month. These studies do not allow for a direct assessment of the degree of ME.

To sum up the evidence from the literature, we can say that subjective health indicators will usually have between 25 and 65 percent of ME in a normal survey. There might be exceptions where ME is extremely low or high, but these seem to be rare and might be the result of the particular measurement model specification.

B. Appendix - Derivation of the bias equation

We start with the definition that the observed values (lower case) are composed of the true value and a random measurement error component. Indices for observations are left out for convenience.

$$\begin{aligned} \mathbf{ses}_{t_1} &= \mathbf{SES}_{t_1} + \mathbf{u}_{ME_{SES_{t_1}}}, \qquad \mathbf{u}_{ME_{SES_{t_1}}} \sim N(0, \sigma_{ME_{SES_{t_1}}}\mathbf{I}), \\ \sigma_{ses_{t_1}} &= \sigma_{SES_{t_1}} + \sigma_{ME_{SES_{t_1}}}. \end{aligned} \tag{12}$$

$$\begin{aligned} \mathbf{h}_{t_1} &= \mathbf{H}_{t_1} + \mathbf{u}_{ME_{H_{t_1}}}, \qquad \mathbf{u}_{ME_{H_{t_1}}} \sim N(0, \sigma_{ME_{H_{t_1}}}\mathbf{I}), \\ \sigma_{h_{t_1}} &= \sigma_{H_{t_1}} + \sigma_{ME_{H_{t_1}}}. \end{aligned} \tag{13}$$

The dependent variable is noted as Y_H for the moment to clarify the notation.From this step onward the time index t_1 is also left out to clarify notation. The covariance matrix that is used

to calculate the regression coefficients is defined as:

$$\mathbf{\Sigma} = V \begin{pmatrix} \mathbf{ses}_{t_1} \\ \mathbf{h}_{t_1} \\ Y_H \end{pmatrix} = \begin{pmatrix} \sigma_{ses} & \sigma_{ses,h} & \sigma_{ses,Y_H} \\ \sigma_{ses,h} & \sigma_h & \sigma_{h,Y_H} \\ \sigma_{ses,Y_H} & \sigma_{h,Y_H} & \sigma_{Y_H} \end{pmatrix} \tag{14}$$

$$\boldsymbol{W} = V \begin{pmatrix} \mathbf{ses}_{t_1} \\ \mathbf{h}_{t_1} \end{pmatrix} = \begin{pmatrix} \sigma_{ses} & \sigma_{ses,h} \\ \sigma_{ses,h} & \sigma_h \end{pmatrix} \tag{15}$$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y_H} = \begin{pmatrix} \widehat{\beta}_{SES} \\ \widehat{\gamma}_H \end{pmatrix} = \begin{pmatrix} \frac{\widehat{\sigma}_h\widehat{\sigma}_{ses,Y_H} - \widehat{\sigma}_{ses,h}\widehat{\sigma}_{h,Y_H}}{\widehat{\sigma}_{ses}\widehat{\sigma}_h - \widehat{\sigma}^2_{ses,h}} \\ \\ \frac{\widehat{\sigma}_{ses}\widehat{\sigma}_{h,Y_H} - \widehat{\sigma}_{ses,h}\widehat{\sigma}_{ses,Y_H}}{\widehat{\sigma}_{ses}\widehat{\sigma}_h - \widehat{\sigma}^2_{ses,h}} \end{pmatrix} \tag{16}$$

If we now insert the observed variables based on the true model in the covariance matrix we get:

$$\mathbf{\Sigma} = V \begin{pmatrix} \mathbf{ses}_{t_1} \\ \mathbf{h}_{t_1} \\ Y_H \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{ses} & \sigma_{ses,h} & \sigma_{ses,Y_H} \\ \sigma_{ses,h} & \sigma_h & \sigma_{h,Y_H} \\ \sigma_{ses,Y_H} & \sigma_{h,Y_H} & \sigma_{Y_H} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{SES} + \sigma_{ME_{SES}} & \sigma_{SES,H} & \beta_{SES}\sigma_{SES} + \gamma_H\sigma_{SES,H} \\ \sigma_{SES,H} & \sigma_H + \sigma_{ME_H} & \beta_{SES}\sigma_{SES,H} + \gamma_H\sigma_H \\ \beta_{SES}\sigma_{SES} + \gamma_H\sigma_H & \beta_{SES}\sigma_{SES,H} + \gamma_H\sigma_H & \beta_{SES}\sigma_{SES} + \gamma_H\sigma_H + \sigma_{\epsilon_{Y_H}} \end{pmatrix}$$

where

$$\widehat{\beta}_{SES} = \frac{\widehat{\sigma}_h\widehat{\sigma}_{ses,Y_H} - \widehat{\sigma}_{ses,h}\widehat{\sigma}_{h,Y_H}}{\widehat{\sigma}_{ses}\widehat{\sigma}_h - \widehat{\sigma}^2_{ses,h}} \tag{17}$$

$$\widehat{\beta}_{SES} \rightarrow_{\text{asymptotically}} \frac{\sigma_h\sigma_{ses,Y_H} - \sigma_{ses,h}\sigma_{h,Y_H}}{\sigma_{ses}\sigma_h - \sigma^2_{ses,h}} \tag{18}$$

Finally, we can get to the bias of the coefficient by transforming the estimated coefficient in partly thr true coefficient plus a bias term.

$$\widehat{\beta}_{SES} = \frac{\sigma_h \sigma_{(SES,Y_H)} - \sigma_{(ses,h)} \sigma_{(h,Y_H)}}{\sigma_{ses} \sigma_h - \sigma^2_{(ses,h)}} \tag{19}$$

$$= \frac{(\sigma_H + \sigma_{ME_H})(\beta_{SES}\sigma_{SES} + \gamma_H \sigma_{SES,H}) - \sigma_{ses,h}(\beta_{SES}\sigma_{(SES,H)} + \gamma_H \sigma_H)}{(\sigma_{SES} + \sigma_{ME_{SES}})(\sigma_H + \sigma_{ME_H}) - \sigma^2_{ses,h}} \tag{20}$$

$$= \frac{\beta_{SES}\sigma_{SES}\sigma_H + \gamma_H \sigma_{(SES,H)}\sigma_H + \beta_{SES}\sigma_{SES}\sigma_{ME_H} + \gamma_H \sigma_{(SES,H)}\sigma_{ME_H} - \beta_{SES}\sigma^2_{(SES,H)} - \gamma_H \sigma_{(SES,H)}\sigma_H}{(\sigma_{SES} + \sigma_{ME_{SES}})(\sigma_H + \sigma_{ME_H}) - \sigma^2_{(ses,h)}} \tag{21}$$

$$= \beta_{ses} + \frac{\gamma_H \sigma_{(SES,H)}\sigma_{ME_H} - \beta_{SES}\sigma_{SES}(\sigma_H + \sigma_{ME_H})}{(\sigma_{SES} + \sigma_{ME_{SES}})(\sigma_H + \sigma_{ME_H}) - \sigma^2_{(ses,h)}} \tag{22}$$