Adjustment of Recall Errors in Duration Data Using SIMEX

Jose Pina-Sánchez[1]

Abstract

It is widely accepted that due to memory failures retrospective survey questions tend to be prone to measurement error. However, the proportion of studies using such data that attempt to adjust for the measurement problem is shockingly low. Arguably, to a great extent this is due to both the complexity of the methods available and the need to access a subsample containing either a gold standard or replicated values. Here I suggest the implementation of a version of SIMEX capable of adjusting for the types of multiplicative measurement errors associated with memory failures in the retrospective report of durations of life-course events. SIMEX is a method relatively simple to implement and it does not require the use of replicated or validation data so long as the error process can be adequately specified. To assess the effectiveness of the method I use simulated data. I create twelve scenarios based on the combinations of three outcome models (linear, logit and Poisson) and four types of multiplicative errors (non-systematic, systematic negative, systematic positive and heteroscedastic) affecting one of the explanatory variables. I show that SIMEX can be satisfactorily implemented in each of these scenarios. Furthermore, the method can also achieve partial adjustments even in scenarios where the actual distribution and prevalence of the measurement error differs substantially from what is assumed in the adjustment, which makes it an interesting sensitivity tool in those cases where all that is known about the error process is reduced to an educated guess.

1   Introduction

Applied quantitative searchers commonly assume that variables included in their models are measured perfectly. This often implicit assumption is, however, difficult to maintain when using survey data as interviewer effects, interviewee fatigue, social desirability bias, lack of cooperation, or plain deceit inevitably introduce measurement error (ME). This is especially true for surveys using a retrospective design, which collect information about past events

---

[1] School of Law, University of Leeds, J.PinaSanchez@leeds.ac.uk

from a single contact with respondents. The advantages of retrospective designs, in comparison with prospective studies[2], are well known: a) immune to problems of attrition; b) cheaper to administer; and c) more capable of detecting transitions occurring in short periods. Retrospective questions are however prone to ME as they require respondents to both interpret the question correctly and recall events that took place in the past.

The consequences of using data affected by ME are both difficult to estimate and potentially disastrous (Nugent, Graycheck & Basham, 2000; and Vardeman et al, 2010)[3]. Unfortunately, the latter is rarely acknowledged, and in certain cases it is directly misunderstood For example, Carroll et al (2006) point at the widespread belief that ME affecting an explanatory variable will only attenuate the regression estimate of that variable[4]. Even amongst researchers that acknowledge the potential consequences of ME very little is done to tackle the problem besides mentioning it as a caveat. There are two reasons for this: the requirement of additional data and the complexity of the adjustment methods available.

Generally, methods for the adjustment of ME need to be informed about the true unobserved values using additional data. For example, multiple imputation (Rubin, 1987, and Cole, Chu & Greenland, 2006) requires access to a validation subsample where the true values are observed. Regression calibration (Carroll & Stefanski, 1990; and Glesjer, 1990) needs at least repeated measurements, while two stage least squares (Theil, 1953) requires instrumental variables. However, researchers' access to this type of data tends to be the exception rather than the norm. In addition, these three methods belong to the family of functional methods – i.e. those that do not make any assumptions about the distribution of the true values. A second group of methods known as structural methods are technically more complex, amongst other things because they require specifying the probability function of the unobserved true values. Examples of structural methods are likelihood based adjustments, either Bayesian or Frequentist. These methods account directly for the ME mechanism in place, which tend to involve ad hoc specifications, in turn increasing the complexity of the adjustment.

---

[2] See Solga (2001) for a comparison of data quality derived from prospective and retrospective questions.
[3] *"Measurement error is, to borrow a metaphor, a gremlin hiding in the details of our research that can contaminate the entire set of estimated regression parameters."* (Nugent, et al. 2000: 60). *"Even the most elementary statistical methods have their practical effectiveness limited by measurement variation."* (Vardeman et al., 2010: 46).
[4] *"Despite admonitions of Fuller (1987) and others to the contrary, it is a common perception that the effect of ME is always to attenuate the line. In fact, attenuation depends critically on the classical additive ME model."* (Carroll et al., 2006: 46).

In this paper I will use simulated data to study the effectiveness of multiplicative Simulation Extrapolation Method (SIMEX) (Carroll et al. 2006; and Biewen, Nolte & Rosemann, 2008). This is an extension of the standard SIMEX method (Cook & Stefanski, 1994) capable of adjusting for the recall errors that are typically observed in the retrospective reports of life-course events. SIMEX implementation is relatively simple in that only requires an estimate of the reliability of the variable affected by ME. This is normally obtained using a subsample of replicated data. Here I will assume that such information is not available to the researcher, as it is often the case. Instead I will use this technique to show its potential to carry out sensitivity analysis when the reliability ratios have to be assumed.

That is, I will demonstrate how the problem of recall errors so ubiquitous in retrospective data can be effectively dealt with by researchers who do not have neither the technical background to carry out complex adjustments, nor access to additional sources of data. In so doing my ultimate goal is to encourage a wider audience of survey researchers both to reflect about the implications of relying variables affected by ME and to consider the possibility of assessing the robustness of their findings.

In the following section I review the theory regarding the types of errors that can be expected from retrospective questions and the models that have been normally used to specify them. In Section 3, I present the simulated data that will be used in the analysis and illustrate the implications of using an explanatory variable affected by multiplicative errors in different outcome models. Section 4 lays out the functioning of the standard SIMEX and the extension considered to accommodate multiplicative ME. In Section 5 the results of the analysis are presented, and in Section 6 I conclude with a discussion of the relevance of the main findings.

2 Modelling Memory Failures in Retrospective Questions on Life-Course Events

Most studies aiming to assess the implications of ME or to adjust for them assume a simple error mechanism known as the classical ME model. This model was first formally defined by Novick (1966) as follows,

$$X_i^* = X_i + V_i \tag{2.1}$$

where X^* is the observed variable, equal to the true variable X, plus the ME term V, which fulfils six important assumptions:

$$Classical\ Model \begin{cases} E(V) = 0\,; & null\ expectancy \\ Var(V_i) = Var(V)\,; & homoscedasticity \\ V \sim N\big(0, Var(V)\big)\,; & normally\ distributed \\ Cov(X,V) = 0\,; & independence\ error\ and\ true\ value \\ Cov\big(V_i, V_j\big) = 0\,; & independence\ between\ errors \\ Cov(\in, V) = 0\,; & non-differentiality \end{cases} \quad (2.2)$$

1. Null expectancy refers to the assumption that the error term is non-systematic, or in other words, the expected value of the error term is zero, $E(V) = 0$.

2. The assumption of homoscedasticity indicates that the variance of the error term is assumed to remain constant across subjects, $Var(V_i) = Var(V) = \sigma_V^2$.

3. In addition to having an expectation of zero and constant variance the error term is normally distributed, $V \sim N(0, \sigma_V^2)$.

4. The correlation between the true value and the error term is assumed to be zero, $Cov(X, V) = 0$.

5. Furthermore, the correlation between different values of the error term is also assumed to be zero, $Cov\big(V_i, V_j\big) = 0$, where V_i and V_j represent any two values of the error term, for subjects i and j.

6. The last assumption, non-differentiality, only becomes relevant when X^* is used in a regression model to specify a response variable, Y. It indicates that, given the true value, the ME is not associated with the residual term from the regression model, $\in$. That is, $E(Y|X, X^*) = E(Y|X)$, or alternatively, $Cov(\in, V) = 0$.

The second and sixth assumptions were not originally established by Novick (1966), but they have been included here because they are often required in the application of the adjustment methods. Assumptions 1 and 4 (null expectancy and independence of the error and true value) can be used to define the expected value and the variance of the true value as follows,

$$E(X^*) = E(X) + E(V) = E(X) \quad (2.3)$$

and,

$$Var(X^*) = Var(X) + Cov(X,V) + Var(V) = Var(X) + Var(V) \qquad (2.4)$$

which can in turn be used to define the reliability of an observed variable affected by classical ME, ρ_{X^*}, as the ratio of the true to observed variance,

$$\rho_{X^*} = \frac{Var(X)}{Var(X^*)} = \frac{Var(X)}{Var(X) + Var(V)} \qquad (2.5)$$

Notice that in order to calculate the reliability ratio either one of the unobserved variances ($Var(X)$ or $Var(V)$) needs to have been previously estimated.

The classical model reflects nicely the type of ME that we can expect to find in measurement processes that are prone to random errors. However, it is not the most appropriate model that we can use to reflect the memory failures associated with retrospective reports of developmental milestones, transitions of events, or ages at onset (Pickles, Pickering & Taylor, 1996), such as the recall of age at menarche, or the date since last employed. Those reports convey duration (or time-to-event) data, which by definition cannot be negative, a possible outcome in the classical model when V_i is negative and bigger than X_i.

Furthermore, although determining the prevalence of ME in these types of questions is not always straightforward, it makes sense to think that the distance from the interview has something to do with the magnitude of the ME. Pickles et al (1998) point at different views on this issue. On the one hand, different authors (Golub, Johnson & Labouvie, 2000; Johnson & Schultz, 2005) have detected telescoping effects in reports of dates of onset of a particular event. The term telescoping was coined by Neter and Waksberg (1964) to refer to the temporal displacement of an event whereby people perceive recent events as being more remote than they are (known as backward telescoping or time expansion) and distant events as being more recent than they are (forward telescoping or time compression).[5] Golub, Johnson & Labouvie, (2000) and Johnson & Schultz (2005) have detected telescoping effects in reports of dates of onset of a particular event.

Conversely, another group of researchers (Huttenlocher, Hedges & Prohaska, 1988; Rubin and Baddeley, 1989; or Bradburn, Huttenlocher & Schwarz, 1994) have argued that rather than distorted time perceptions recall errors take the form of non-systematic ME around the reported date with its size being proportional to the distance between the day of the interview

---

[5] For a review of the cognitive processes resulting in telescoping see Janssen and Chessa (2006).

and the reported date. That is, the further away the date of the event to be reported the harder its recall and therefore the bigger the ME. Both these findings question the appropriateness of applying the additive ME model to retrospective designs.

ME induced by memory failures can instead be better represented as a multiplicative ME model (Holt, McDonald & Skinner, 1991; Pickles et al., 1996 and 1998; Skinner & Humphreys, 1999; Augustin, 1999; Glewwe, 2007; and Dumangane, 2007) that builds upon the classical ME model as follows,

$$X_i^* = X_i \cdot V_i \tag{2.6}$$

Here the multiplicative relation between X and V reflects that the effect of V on the observed variable X^* is proportional to the value of X. In addition, for the specific case of non-systematic multiplicative ME, most of the assumptions about the error term described in equation 2.2 apply. I will refer to this type of error as classical multiplicative from here on. The only exceptions are items 1 and 3. Here, V follows a log-normal distribution bounded from 0 to ∞, and with mean equal to 1. This way the ME has a relatively symmetric effect across the true values and maintains the scale used in duration data. Note as well that, this same model can also be used to account for backward and forward telescoping effects by shifting the distribution of V to the right or left, so its mean goes below or above 1.

3   The Impact of Classical Multiplicative Measurement Error in Regression Analyses

Gustafson (2003) traced out analytically the impact of classical multiplicative ME affecting an explanatory variable, X^*, in a linear model where a second explanatory variable, Z, is measured without error,

$$Y = \beta_0 + \beta_1 X^* + \beta_2 Z + \in \tag{3.1}$$

Classical multiplicative ME produces attenuation in the β_1 regression coefficient directly proportional to the variance of the ME term, increasing the more skewed to the right X is, and as the correlation between the true variable X and Z grows stronger. Gustafson (2003) does not evaluate, however, the impact on the β_2 error-free regression coefficient or consider non-linear outcome models.

To address these issues, this section explores the impact of different types of multiplicative ME on three different generalised linear model specifications – linear, logit and Poisson regressions - containing error-prone and perfectly measured explanatory variables, X^* and Z. A simulated dataset of 1000 observations allows sufficient statistical power to detect moderate parameter estimates while keeping a low computational burden given the number of scenarios that will be explored. To reflect the features of duration data, the true variable X is taken to be exponentially distributed with mean 8.75 and range (1.15, 44.45); Z follows a standard normal distribution, and both of them are associated with the response variable of the linear model, Y, which is also shaped as a standard normal distribution. In addition, to generate the response variables for the logit and Poisson models, Y is recoded as a binary variable, Y_{ca},

$$Y_{ca} \begin{cases} 1, & Y \geq 0 \\ 0, & Y < 0 \end{cases}$$

and as a count variable, Y_{co},

$$Y_{co} \begin{cases} 0, & Y < 0 \\ 1, & 0 \leq Y > 1 \\ 2, & 1 \leq Y > 2 \\ 3, & 2 \leq Y > 3 \\ 4, & Y \geq 3 \end{cases}$$

The four simulated ME scenarios are represented by the variables, X^*, X_h^*, X_u^* and X_o^*. In each of these scenarios X is subject to normally distributed classical multiplicative ME. I choose to simulate normal instead of log-normal errors (as explained in equation 2.6) to ensure that they are perfectly symmetric around their mean (the latter are skewed to the right to a certain extent). Figure 1 shows the probability and mass functions for each of the variables simulated, while the specific code used in R is shown in Appendix I.

**Figure 1:** Probability Density and Mass Functions of the Simulated Variables
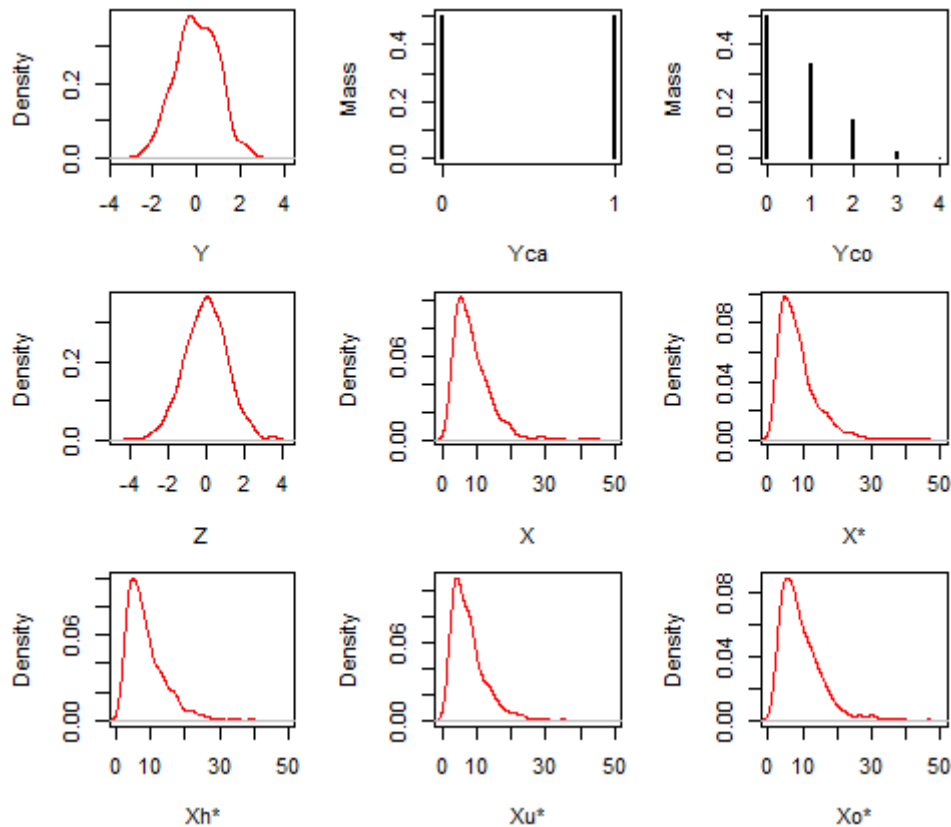
In the first ME scenario I simulate non-systematic errors distributed as a $N(1,.25)$. The multiplicative effect of these errors results in a new variable X^* with a reliability ratio (RR) of .816. In the second scenario I explore the effect of heteroscedastic ME by changing the distribution of the errors from $N(1,.15)$ to $N(1,.35)$ when $Z > 0$. This is a type of ME that could take place when different survey modes are used. For example, Roberts (2007) - after reviewing the literature - concluded that telephone interviews place a higher cognitive demand on the interviewee than face-to-face interviews, which tend to make them more prone to measurement error. In the third scenario I study the effect of systematically underreported durations by simulating errors distributed as $N(.9,.25)$. These are the types of errors that could be expected in the presence of forward telescoping bias (e.g. Golub et al., 2000, and Johnson and Schultz, 2005, found evidence of these types of errors in reports of onset of drug usage and smoking, respectively), but also in the report of durations of socially undesirable events (e.g. Pina-Sánchez, Koskinen & Plewis, 2013 and Pina-Sánchez, Koskinen & Plewis, 2014, found an increased tendency to underreport the longer spells of unemployment). Lastly, I explore the opposite scenario, one where the errors are distributed

as $N(1.1, .25)$ to reflect overreported durations, which could be expected in the presence of backward telescoping or in reports of socially desirable events.

The effects of this four types of simulated ME are shown using scatter-plots in Figure 2. Notice that the top-right plot uses Z instead of X.in the y-axis.

Figure 2: Scatterplots of the effect of the different types of measurement error considered
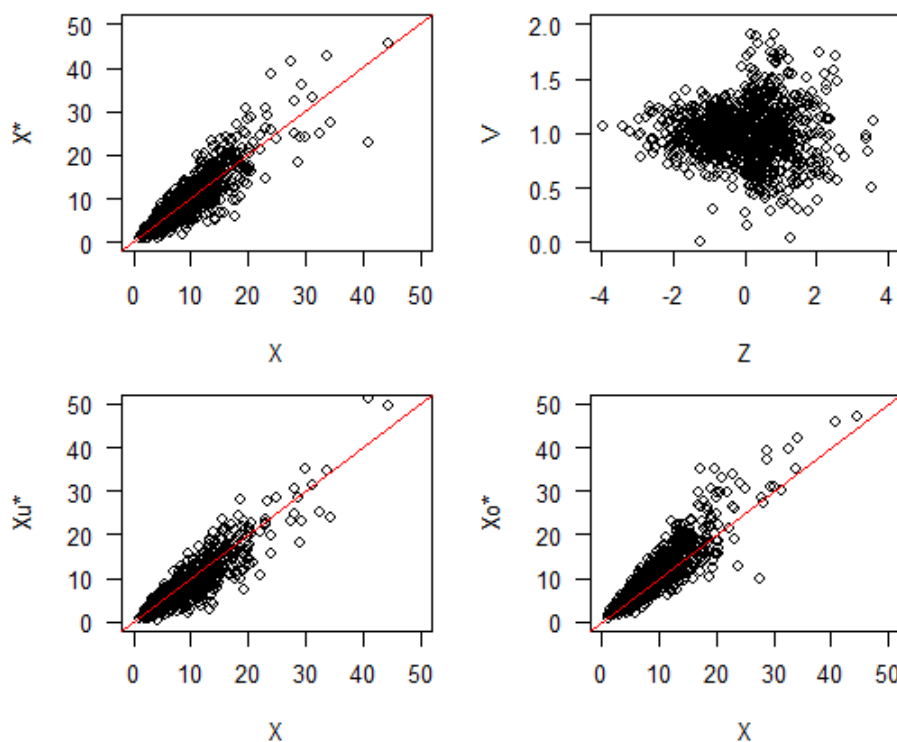
To assess the impact that these types of errors have on the regression coefficient of a linear, a logit, and a Poisson model, I compare the results from each of these models when X^* is used (the naïve model) instead of X (the true model). Specifically, I focus on the bias in the regression coefficients,

$$BIAS = \hat{\beta}_n - \hat{\beta}_t \tag{3.2}$$

where the subscript n stands for the naïve model and t for the true model. In addition, to compare the impact of ME across models and across regression coefficients using different scales, I calculate a relative measure of the bias as follows,

$$R.BIAS = \frac{|(\hat{\beta}_n - \hat{\beta}_t)100|}{|\hat{\beta}_t|} \tag{3.3}$$

Results for the different models studied and the impact generated by the different types of ME are presented in Table 1. In all of the scenarios studied the effect of ME was reflected in a downward bias for β_1 (the coefficient for the variable X^*), and in upward biases for β_0 and β_2 (the coefficients for the constant and Z). In addition to the observed differences in the direction of the biases across coefficients there are also strong differences in their intensity. The size of the bias for β_2 is about twice as large as the bias for β_0 and β_1, reaching levels as alarming as 94.8% for the logit model with heteroscedastic ME, although the average size of the bias across all the scenarios is 39.5%.

Table 1: Impact of Measurement Error in the Regression Estimates

		Linear				Logit				Poisson			
		Coef	SE	Bias	R.Bias	Coef	SE	Bias	R.Bias	Coef	SE	Bias	R.Bias
True model	β_0	-1.297	.035			-5.997	.388			-1.362	.069		
	β_1	.150	.003			.768	.050			.092	.004		
	β_2	.111	.016			.210	.099			.082	.038		
Naïve: multi.	β_0	-1.013	.038	.284	21.9%	-3.810	.258	2.187	36.5%	-1.198	.065	.284	20.8%
	β_1	.118	.004	-.032	21.2%	.494	.034	-.275	37.5%	.075	.004	-.032	34.6%
	β_2	.156	.019	.045	40.9%	.275	.084	.065	30.9%	.157	.037	.045	55.1%
Naïve: hetero.	β_0	-.998	.039	.372	28.7%	-3.649	.248	2.372	39.5%	-1.178	.065	.372	27.4%
	β_1	.116	.004	-.043	28.8%	.471	.032	-.299	38.9%	.074	.004	-.043	47.0%
	β_2	.169	.020	.067	60.7%	.377	.085	.199	94.8%	.141	.037	.067	81.7%
Naïve: under.	β_0	-.937	.038	.325	25.1%	-3.441	.233	1.962	32.7%	-1.054	.059	.325	23.9%
	β_1	.121	.004	-.024	15.9%	.490	.033	-.183	23.8%	.069	.004	-.024	26.1%
	β_2	.166	.020	.052	46.9%	.321	.084	.124	58.9%	.149	.037	.052	63.2%
Naïve: over.	β_0	-1.028	.038	.245	18.9%	-4.029	.266	1.842	30.7%	-1.110	.061	.245	18.0%
	β_1	.108	.003	-.039	26.1%	.470	.031	-.284	37.0%	.061	.003	-.039	42.7%
	β_2	.150	.019	.053	48.0%	.284	.087	.152	72.4%	.134	.037	.053	64.6%

While the different ME scenarios clearly show attenuated coefficients, none of the coefficients actually became statistically non-significant or changed their sign in comparison to the naïve models. This is partly due to the small effect that ME had on the standard errors, which were underestimated by a third of their size in the true logit model, and only slightly underestimated and overestimated when using a Poisson and a linear model, respectively.

These results are consistent with Biewen et al. (2008) who, in a simulated probit model with one predictor, find an upward bias in the constant and a downward bias in the slope

induced by classical multiplicative ME. These results obtained here serve to reinforce these findings. In the presence of a type of ME different than classical additive or for a model different than simple linear regression the direction of the bias is not always towards the null. The difficulty to anticipate the direction and size of these biases – even in scenarios with moderate prevalence of ME – makes the implementation of adjustment methods an indispensable part analysing survey data prone to these types of ME.

4   Standard SIMEX and Extensions to Account for Classical Multiplicative Measurement Error

The study of the adjustment of multiplicative errors dates back to the decade of the 80s. Fuller (1984) and Hwang (1986) developed a method-of-moments correction for multiplicative ME in the explanatory variables of a linear model. This method assumes that the value of ME variance is known – or that it can be estimated - and is limited to applications where the ME mechanism is affecting one of the explanatory variables only in the context of a linear model. Lyles and Kupper (1997) compared the effectiveness of this method with others such as regression calibration, and a quasi-likelihood approach, which could be applied to other non-linear outcome models.

These methods, as mentioned above, are however of limited use to applied researchers in that they either require additional data in the form of replicated measures or validation subsamples, or are complex to implement. Regression calibration requires additional data in the form of replicated measures or a validation subsample. Quasi-likelihood approaches only need an estimate of the variance of the ME, and much like those relying on Bayesian statistics can be applied when a full likelihood approach is not feasible due to computational intractability. However, their implementation is relatively complex, starting from the need to use specialised software (such as WinBUGS when considering Bayesian adjustments), which discourages many analysts from attempting the implementation of the necessary adjustment.

Due to only requiring an estimate of the variance of the ME, the simplicity of its application, and its generalizability to any other outcome model regardless of its complexity[6], SIMEX represents a very convenient alternative. SIMEX was first presented by Cook and Stefanski (1994) and refined in the following years by Stefanski and Cook (1995) and

---

[6] See for example He et al. (2007) who applied SIMEX to an Accelerated Failure Time models with one of the explanatory variable affected by classical ME, or Battauz et al. (2008) who adjusted for a similar type of ME problem but for an ordinal probit model as the outcome model.

Carroll, Küchenhoff, Lombard, and Stefanski (1996). *"The key idea underlying SIMEX is the fact that the effect of measurement error on an estimator can be determined experimentally via simulation"* (Carroll et al., 2006: 98).

In particular, SIMEX exploits the relationship between the size of the ME affecting a variable and the size of the bias in the regression estimates in the outcome model. Following Fuller (1987) we know that the unadjusted estimator of the slope, $\hat{\beta}_1$, does not converge asymptotically to the parameter β_1 but to:

$$\hat{\beta}_1^* = \beta_1\big(\sigma_X^2/(\sigma_X^2 + \sigma_V^2)\big) \tag{4.1}$$

where σ_X^2 and σ_V^2 represent the variance of the true explanatory variable and the error term. In other words, the estimator of the slope is biased downwards in absolute terms by a factor equal to the reliability ratio, ρ_{X^*} (defined in equation 2.5), of the observed variable, X^*. In this situation, and if ρ_{X^*} or σ_V^2 is known, it would not be practical nor efficient to use SIMEX, since the adjustment would simply be achieved by substituting the variance terms in equation 4.1. However, I will use this simple setting for illustrative purposes.

To facilitate the understanding of the method, the steps involved in its implementation are outlined below using a simple example of bias in the slope of a simple linear regression, where explanatory variable, X^*, is prone to classical additive ME (equations 2.1 and 2.2). The implementation of SIMEX is divided into six phases:

1) The first step involves simulating additional explanatory variables with increasing levels of ME. These new variables are generated in a way that emulates the classical ME model, but with successively larger values of σ_V^2 affecting X. Specifically, K new explanatory variables $X_k^*(\lambda_k)$ are generated by the rule:

$$X_k^*(\lambda_k) = X^* + \sqrt{(\lambda_k)}V \tag{4.2}$$

with $k = 0,1,\dots,K$, the simulated error normally distributed, $V \sim N(0, \sigma_V^2)$, and, $\lambda_0 < \lambda_1 <\dots< \lambda_K$, a set of parameters used to amplify the ME variance (often these are (.5, 1, 1.5, 2)).

2) Once the different variables with added ME have been generated, the outcome model is re-estimated using this new data, and the values of the estimator of interest (i.e. β_1) for the different levels of ME (λ_k) are saved. In particular, for the case of a simple linear model with

the explanatory variable affected by classical ME, and using the data-generating rule described in equation 4.2, the estimator of the slope will now converge to:

$$\hat{\beta}_{1k}^* = \beta_1 \sigma_X^2 / (\sigma_X^2 + (1 + \lambda_k)\sigma_V^2) \tag{4.3}$$

where the bias increases monotonically as λ_k increases.

3) In order to reduce the Monte Carlo error associated with the simulation procedure steps 1 and 2 are repeated B times so a mean estimate of $\hat{\beta}_{1k}^*$ for $b = 1, \dots, B$ can be computed, where the rule of thumb[7] is to use $B = 100$ iterations.

4) At this stage the $\bar{\hat{\beta}}_{1k}^*$ and λ_k values can be paired considering the former as a function of the latter, $G\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k\right)$, known as the extrapolation function, which should be plotted in order to obtain a first insight of its shape.

5) The extrapolation function is estimated using a regression model, with data $\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k\right)$. Carroll et al. (2006) recommend the use of one of three types of simple functional forms.

    a) linear, $\qquad\qquad\qquad\qquad G\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k\right) = \zeta_1 + \zeta_2\,\lambda_k$

    b) quadratic, $\qquad\qquad\qquad G\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k\right) = \zeta_1 + \zeta_2\,\lambda_k + \zeta_3\lambda_k^2$

    c) non-linear or ratio-linear, $\quad G\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k\right) = \zeta_1 + \zeta_2/(\zeta_3 + \lambda_k)$

For the example presented here, and if the extrapolation function is well approximated by the chosen functional form, we would find the following function,

$$E\left(\bar{\hat{\beta}}_{1k}^* | \lambda_k\right) = G\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k\right) = \beta_1 \sigma_X^2 / (\sigma_X^2 + (1 + \lambda_k)\sigma_V^2) \tag{4.4}$$

6) From here, the SIMEX estimate, $\hat{\beta}_{SIMEX}$, can be calculated by extrapolating $G\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k\right)$ to $G\left(\bar{\hat{\beta}}_{1k}^*, \lambda_k = -1\right)$. Note that from equation 4.4 when $\lambda_k = -1$ the bias is cancelled out.

Figure 3 represents the SIMEX process graphically. The solid line denotes the part of the extrapolation function that can be approximately observed through the regression estimates resulting after the outcome model is specified using simulated predictors with increasing

---

[7] This is the number of variations used by default in the SIMEX packages in STATA and R.

levels of ME, and the dashed line represents the extrapolation to the case of no ME, which gives the adjusted estimate.
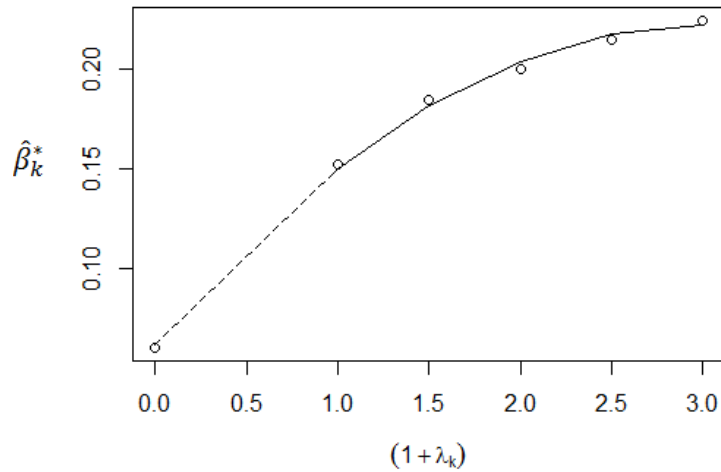
Figure 3: Extrapolation function

Figure 3 also shows some of the limitations of SIMEX. The entire extrapolation function cannot be observed, hence, it is hard to assess the quality of the adjustment. In addition, the extrapolation function needs to be approximated using a simple functional form. Therefore adjustments are only approximated, and their effectiveness depends on how well the extrapolation function is estimated, for which the choice of the right functional form is crucial. In the case depicted by Figure 3 it makes sense to think of the quadratic function as the better approximation, but it might not always be so clear.

Another cause of concern stems from the accuracy of the estimate of σ_V^2 that is used in the simulations. For example, considering the case depicted in Figure 3, if σ_V^2 is underestimated, the extrapolation function will have a flatter slope and the adjustment would only be partial. That is, for an underestimated σ_V^2, lower values of $\hat{\beta}_k^*$ would have been generated for $(1 + \lambda_k) = (1.5, 2, 2.5, 3)$, which would have made the estimated extrapolation function shallower, and produced a bigger - and still biased - adjusted estimate when extrapolating to $\lambda_k = -1$. Such suboptimal adjustment is illustrated in Figure 4 where I compare the extrapolation function shown in Figure 3 with a similar one that would be obtained if σ_V^2 had been underestimated.

Figure 4: Comparison of Extrapolation Functions
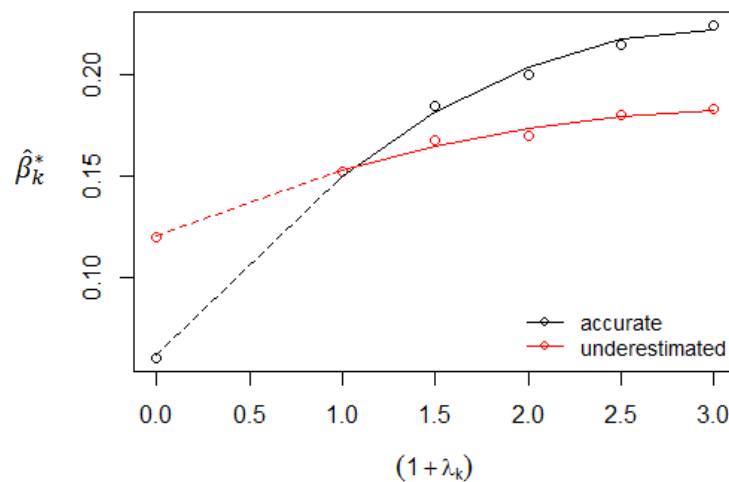
Interestingly the application of SIMEX to any other regression model affected by a problem of error-in-variables would follow the same logic than the example of a simple linear regression model presented here, regardless of the complexity of the outcome model under study. Even more interestingly – at least for this paper's topic of study – is the fact that SIMEX can be applied to ME problems different than the standard classical additive model so long as the ME-generating process can be simulated via Monte Carlo methods (Carroll, 2006). Two remarkable extensions that have proven to be robust in the literature are: 1) MC-SIMEX (Küchenhoff et al. 2006), the application of the SIMEX methodology to problems of misclassification of either the response or an explanatory variable in the outcome model; and 2) SIMEX for classical multiplicative ME (Carroll et al., 2006, and Biewen et al. 2008).

In order to accommodate the standard SIMEX method to account for the classical multiplicative ME setting Carroll et al. (2006) propose a change in the way the simulated variables are increasingly affected by ME, X_k^*, in step 1. In particular, equation 4.3 is substituted by

$$X^*(\lambda_k) = exp\{\log(X^*) + \sqrt{\lambda_k}\log(V)\} \tag{4.5}$$

to represent the multiplicative relationship between the observed durations and the simulated noise, while the rest of the six-step algorithm is implemented as before. However, the expression in equation 4.5 cannot be used to generate negative errors, which in a setting like the one assumed here, where errors are normally (instead of lognormally) distributed, could create complications. To avoid this problem I will use the following error-generating rule suggested by Biewen et al. (2008),

$$X^*(\lambda_k) = X^* \cdot V^{\lambda_k} \qquad (4.6)$$

Lastly, to estimate the standard errors of $\hat{\beta}_{SIMEX}$ I use the bootstrapping pairs algorithm[8], where entire cases covering the response and explanatory variables are resampled with replacement, and for each new sample the SIMEX process is rerun. Bootstrap is only one of the different options available to estimate the variance of the SIMEX estimator. Carroll et al. (1996) suggest using a method based on the sandwich estimator and on the theory of M-estimators to obtain an asymptotic covariance estimator. Specifically, this method is based on the asymptotic equivalence of $\hat{\beta}(\lambda_k)$ and an M-estimator, producing a closed form equation from which the standard errors of $\hat{\beta}_{SIMEX}$ can be directly derived, which avoids the computationally intensive process of replicating the SIMEX procedure for a number of new samples. However, this method has been developed for the specific case of classical additive ME. For extensions of SIMEX to different types of ME processes non-parametric methods such as bootstrap or jackknife become a natural alternative.

5 Effectiveness of the Adjustments for Different Types of Errors and Estimates of their Variance

SIMEX requires an estimate of the variance of the ME, $\hat{\sigma}_V^2$. Ideally the actual parameter is known[9], although in the presence of replicated measures $\hat{\sigma}_V^2$ can also be estimated. However, access to such type of data tends to be the exception rather than the norm. Thus, here I study the effectiveness of SIMEX as a sensitivity tool. That is, I assume that the researcher suspects the presence of ME in one of the variables being used but can only provide an educated guess[10] of the distribution and prevalence of that ME.

For each of the different ME scenarios presented in Section 3, I explore the effectiveness of SIMEX in reducing the bias found in the naïve models when different values of σ_V^2 are tried. In particular I review the extent of the adjustments assuming that $E(V) = 0$, and σ_V^2 is equal to .25, .176, .336, and .412   which are equivalent to assuming reliability ratios of X^*

---

[8] See Keele (2008) for a description of the differences between bootstrapping algorithms.

[9] For example, Biewen et al. (2008) suggested using multiplicative ME as an strategy to anonymised data, while offering the value of σ_V^2 to data users so they can adjust for the implications of the artificially created ME in their analyses.

[10] Ideally based on validation studies looking at similar types of questions available in the literature.

equal to .816, .9, .7, and .6, respectively. I refer to each of these scenarios as assuming a "correct", "overestimated", "underestimated", and "highly underestimated" reliability ratio. In addition, to assess the effectiveness of the adjustment in the presence of forward or backward telescoping effects when the ME process is correctly estimated I take the "correct" adjustments for the systematic negative and positive ME scenarios to use the right ME process, that is $V \sim N(.9, .25)$ and $V \sim N(1.1, .25)$, respectively.

For consistency's sake I use the linear extrapolation function across all of the adjustments[11]. This extrapolation function was chosen instead of the more commonly used quadratic extrapolation to generate more conservative adjustments in scenarios using $\hat{\sigma}_V^2 > \sigma_V^2$, which in my analysis would be the cases of "underestimated" and "very underestimated" reliability ratios. For the estimation of the standard errors of $\hat{\beta}_{SIMEX}$ I run the bootstrapping pairs algorithm for 100 iterations (just like in Biewen et al., 2008)[12], and within each of these iterations I run the six steps of the SIMEX process another 100 times[13]. Lastly, to calculate the effectiveness of the adjustments I use measures of absolute and relative bias as in Section 3. The only differences are in terms of notation: I now substitute $\hat{\beta}_S$ by $\hat{\beta}_{SIMEX}$ in equation 3.2 and 3.3, and for the R.BIAS I also substitute the denominator by the BIAS in the survey. Results are shown in Table 2.

---

[11] The adequacy of the linear extrapolation can be assessed in Figure A2 (Appendix III), where I show the extrapolation functions for the adjustment of non-systematic ME using the correct σ_V^2.

[12] This is less than what would be recommendable to obtain precise estimates of the standard errors but it is sufficiently good considering that the SIMEX process is also computationally intensive and that a compromise needs to be reached.

[13] Figure A1 (Appendix III) includes scatterplots that reflect the simulation of increasing levels of the ME generation process (step 1 of the SIMEX algorithm) when the correct σ_V^2 is used.

Table 2: Results of the Adjustments

RR		Param.	Linear Coef.	SE	Bias	R.Bias	Logit Coef.	SE	Bias	R.Bias	Poisson Coef.	SE	Bias	R.Bias
Classical	Correct	β_0	-1.283	.017	.014	4.8%	-5.030	.100	.967	44.2%	-1.400	.026	-.038	23.1%
		β_1	.152	.002	.002	7.5%	.662	.014	-.106	38.6%	.097	.003	.005	31.6%
		β_2	.105	.006	-.006	12.9%	.184	.018	-.026	39.8%	.104	.007	.022	29.9%
	Over-estimated	β_0	-1.198	.013	.098	34.7%	-4.714	.087	1.282	58.6%	-1.343	.021	.019	11.7%
		β_1	.141	.002	-.009	26.8%	.617	.012	-.151	54.9%	.090	.002	-.001	7.3%
		β_2	.121	.004	.010	23.2%	.207	.016	-.003	4.4%	.119	.005	.037	48.9%
	Under-estimated	β_0	-1.339	.018	-.043	15.1%	-5.148	.095	.849	38.8%	-1.436	.027	-.074	44.9%
		β_1	.160	.002	.010	32.1%	.681	.014	-.087	31.6%	.101	.003	.010	58.1%
		β_2	.095	.006	-.016	35.5%	.174	.020	-.035	54.6%	.096	.009	.014	18.2%
	Very under-estimated	β_0	-1.361	.018	-.064	22.7%	-5.134	.076	.863	39.4%	-1.451	.030	-.089	54.4%
		β_1	.163	.002	.014	42.7%	.681	.011	-.087	31.6%	.103	.003	.012	71.0%
		β_2	.090	.007	-.020	44.7%	.175	.020	-.034	53.1%	.092	.009	.010	13.2%
Heteroscedastic	Correct	β_0	-1.265	.019	.032	10.6%	-4.790	.111	1.207	51.4%	-1.378	.034	-.016	8.6%
		β_1	.149	.002	<.001	1.1%	.629	.016	-.140	46.9%	.096	.003	.004	22.8%
		β_2	.122	.005	.011	18.8%	.324	.020	.114	68.1%	.082	.009	<.001	0.8%
	Over-estimated	β_0	-1.182	.014	.114	38.3%	-4.481	.093	1.515	64.5%	-1.317	.027	.045	24.3%
		β_1	.139	.002	-.011	32.3%	.585	.013	-.183	61.6%	.089	.003	-.003	15.4%
		β_2	.137	.004	.026	44.4%	.340	.014	.130	77.4%	.098	.006	.016	26.3%
	Under-estimated	β_0	-1.320	.019	-.023	7.8%	-4.925	.090	1.071	45.6%	-1.414	.034	-.052	28.5%
		β_1	.157	.002	.007	21.2%	.650	.013	-.118	39.8%	.100	.003	.008	47.7%
		β_2	.112	.007	.002	3.0%	.315	.020	.105	62.8%	.073	.010	-.009	15.9%
	Very under-estimated	β_0	-1.342	.018	-.045	15.1%	-4.923	.083	1.073	45.7%	-1.427	.035	-.065	35.5%
		β_1	.160	.002	.011	31.1%	.651	.011	-.117	39.3%	.102	.003	.010	58.3%
		β_2	.108	.008	-.002	3.5%	.314	.019	.104	62.1%	.069	.009	-.013	22.6%
Systematic Negative	Correct	β_0	-1.201	.019	.096	26.6%	-4.565	.085	1.432	56.0%	-1.214	.031	.148	47.9%
		β_1	.153	.003	.003	11.4%	.649	.013	-.119	42.9%	.086	.003	-.006	25.2%
		β_2	.115	.005	.004	7.7%	.243	.017	.033	29.9%	.093	.009	.011	16.5%
	Over-estimated	β_0	-1.106	.013	.191	53.0%	-4.207	.072	1.790	70.1%	-1.157	.020	.205	66.6%
		β_1	.144	.002	-.005	18.6%	.606	.011	-.163	58.5%	.081	.002	-.010	44.5%
		β_2	.134	.004	.023	41.5%	.265	.012	.055	49.5%	.111	.007	.029	43.3%
	Under-estimated	β_0	-1.236	.022	.060	16.7%	-4.640	.094	1.357	53.1%	-1.235	.031	.127	41.3%
		β_1	.164	.003	.014	48.1%	.677	.015	-.091	32.9%	.092	.003	<.001	0.9%
		β_2	.109	.006	-.002	3.6%	.239	.020	.029	26.1%	.087	.010	.005	7.2%
	Very under-estimated	β_0	-1.257	.018	.040	11.1%	-4.644	.091	1.352	52.9%	-1.247	.031	.115	37.3%
		β_1	.167	.003	.017	60.5%	.680	.015	-.088	31.8%	.094	.004	.002	9.6%
		β_2	.105	.007	-.006	10.7%	.239	.017	.029	26.1%	.083	.010	.001	1.0%
Systematic Positive	Correct	β_0	-1.278	.019	.018	6.9%	-5.252	.097	.744	37.8%	-1.263	.033	.099	39.2%
		β_1	.141	.002	-.009	21.9%	.632	.012	-.136	45.7%	.079	.003	-.013	41.4%
		β_2	.102	.005	-.008	20.3%	.199	.018	-.011	14.4%	.080	.009	-.002	4.0%
	Over-estimated	β_0	-1.209	.014	.087	32.4%	-4.959	.087	1.038	52.7%	-1.221	.017	.141	56.1%
		β_1	.129	.002	-.021	50.4%	.584	.011	-.185	61.9%	.072	.002	-.019	63.4%
		β_2	.116	.004	.005	13.7%	.217	.014	.007	8.9%	.093	.007	.011	21.8%
	Under-estimated	β_0	-1.355	.023	-.058	21.5%	-5.450	.110	.547	27.8%	-1.307	.038	.055	21.7%
		β_1	.146	.003	-.004	8.7%	.649	.014	-.119	40.0%	.082	.003	-.010	32.7%
		β_2	.089	.006	-.022	55.2%	.188	.022	-.022	29.2%	.066	.010	-.016	30.3%
	Very under-estimated	β_0	-1.378	.020	-.082	30.4%	-5.444	.102	.553	28.1%	-1.322	.034	.040	16.1%
		β_1	.149	.002	.000	0.7%	.650	.013	-.118	39.6%	.084	.003	-.008	26.6%
		β_2	.084	.008	-.027	66.5%	.189	.019	-.021	27.6%	.062	.011	-.021	39.7%

A first point to notice is that compared to results from the true models presented in Table 1, the standard errors are underestimated by a half. This might be due to the small size of the true standard errors (expressed in the second or third decimal point), but it also illustrates that the variance of $\hat{\beta}_{SIMEX}$ using bootstrap can only be approximated.

Regarding the adjustment in terms of the reduction of the biases found in the naïve analyses we can observe varying levels of success. The effectiveness of the adjustment ranged from being able to reduce it to .8% of its size (for β_2 in the Poisson model affected by heteroscedastic ME and using the correct estimate of σ_V^2) to a less impressive figure of 77.4% (for β_2 in the logit model affected by heteroscedastic ME and using an over-estimated reliability ratio). In spite of this variability, the adjustments explored could be considered quite successful since on average they managed to reduce the biases found in the naïve models to 32.8% of their original size.

Figure 5: Adjustments in Terms of R.BIAS[*,**]
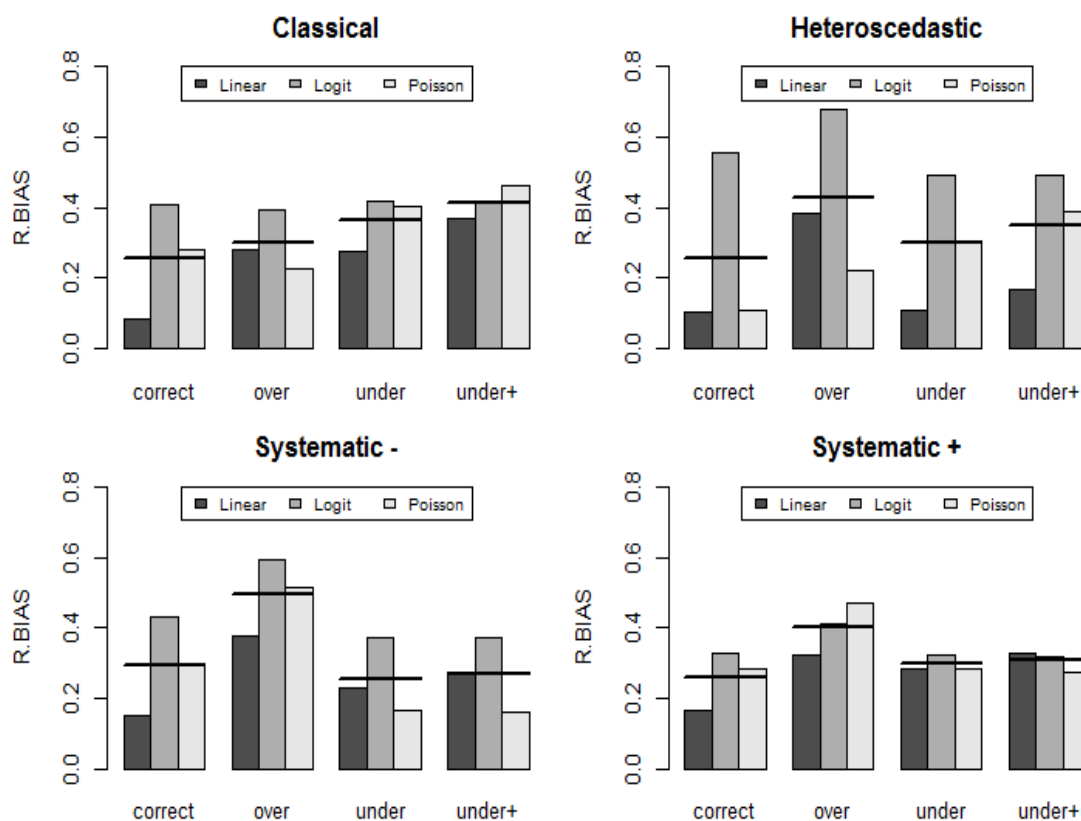
<p style="text-align:center">*The category of barplots "under +" represents a very underestimated reliability ratio of .6.
**The flat lines indicate the average R.BIAS for the three types of models.</p>

To elucidate some trends about the relative effectiveness of SIMEX for the different scenarios explored I have grouped some of the results in Figure 5. Each bar represents the average R.BIAS for the three regression coefficients comprised in each of the models and for

each of the scenarios studied, while the black horizontal lines represent the bars averages over the three outcome models studied.

On average the adjustments are most effective on the linear model, reducing the R.BIAS to 24.3% of its original size, whereas for the logit models the average adjustment is 43.7%, and for the Poisson is 30.3%. This better performance for the linear outcome model might be related to the linear extrapolation function used across all of the adjustments, which would be the most appropriate function when the changes in bias are proportional to the increased levels of ME. Directly from Figure 5 we can also see that considering the three coefficients of each model, the most successful adjustment was for the linear model when the correct ME process is used, which reduced the bias to just 8.4% of its original size. But even in the worst scenario where heteroscedastic ME is simulated and the reliability ratio of the measure is overestimated the bias can still be reduced to 38.3% of its size. On the other hand, the least promising results were obtained for the adjustments of the heteroscedastic ME when a logit outcome model is used. Here, regardless of the assumed reliability ratio no adjustment could reduce the bias by more than 50.6%.

6 Conclusion

The presence of ME in retrospective questions is widely acknowledged, however, little is done to tackle this problem. A majority of studies using this type of data deal with the ME problem by adding caveats to their findings, whereas those attempting to implement the necessary adjustments are very uncommon. The implications of this problem are truly daunting. Here I have simulated moderate levels of different types of multiplicative errors, which could be expected to arise as a result of memory failures in the report of dates of onset or end of spells, to explore the impact that such ME could have on the regression coefficients of different models. Across the scenarios studied I found an average bias of about 40% the size of the true coefficients, reaching up to 95% in certain cases.

I pointed at two fundamental barriers limiting the implementation of adjustment methods. First, most of these methods require access to additional sources of data in the form of replicated measures, validation subsamples, or instrumental variables, which are rarely available to a majority of researchers. Second, most methods are relatively complex to implement, discouraging researchers from using them. A very illustrative case is that of

adjustments relying on Bayesian statistics; their reliance on MCMC and prior probabilities offers remarkable flexibility to deal with complex ME processes, even in the absence of replicated or validation data. However, it is this complexity together with other practical matters such as the need to use specialised statistical packages that tend to dissuade researchers from using them.

To deal with recall errors in the report of dates of onset when no data to inform about the ME process is available I have suggested the implementation of the more practical SIMEX method. SIMEX is relatively simple to implement, it can be easily replicated to different outcome models regardless of their complexity, and it only requires - in its standard form - knowledge about the variance of the error term. Furthermore, although SIMEX was initially created to adjust for classical additive ME, it can also be extended to account for different types of ME processes so long as these can be simulated using Monte Carlo methods. I have used this feature to assess the effectiveness of SIMEX in the presence of multiplicative errors. In particular, I have explored the application of SIMEX to classical, heteroscedastic, systematic positive and systematic negative multiplicative errors. The types of errors that could be expected from general memory failures, but also those seen when different survey modes are used, or in the presence of backward and forward telescoping effects, respectively.

In the presence of these types of errors, SIMEX adjustments where the distribution of the ME is known have shown satisfactory results, managing to reduce the size of the biases found in the estimates of the naïve models to less than one third of their size on average. But perhaps more interesting is the fact that SIMEX also achieved reasonably good results even when the ME process is assumed to be non-systematic and its variance is only approximated. The quality of the adjustments varied substantially – as it could not be any different given the several scenarios explored – but in each of the 144 estimates studied SIMEX managed to produce positive adjustments, with the worst of them all achieving a reduction of the bias found in the naïve model of 22.6%.

This capacity to obtain partial adjustments even when the type of multiplicative error is only approximated, together with the relative simplicity with which it is implemented, makes SIMEX an ideal method to be used as a sensitivity tool. Researchers concerned of using duration data affected by recall errors could obtain an estimate of the magnitude of the impact, which would allow them to provide more informative caveats regarding the degree to which the validity of their findings is affected. To do that they can use the multiplicative SIMEX process presented in Appendix II. The only two alterations needed would be the re-

specification of the outcome model and the choice of the size of the variance of the error term.

When the latter is not known the method could be run using an educated guess for the reliability ratio of the variable prone to ME. Alternatively, for those questions where previous studies of the validity and reliability of responses are available, the researcher could use an average of the estimates obtained in the literature. The opportunity to use such sensitivity analyses even when no replicated measures or a validation subsample is available also illustrates the importance of studies aiming to assess the prevalence of ME in different types of survey questions. The more we know about the ME processes affecting survey responses the better adjustments could be achieved and the higher the validity of studies using survey data will be.

Acknowledgements

References

[1]   Augustin, T. (1999): Correcting for Measurement Error in Parametric Duration Models by Quasi-likelihood. *Munchen Institut fur Statistik*, from: http://epub.ub.uni-muenchen.de/1546/1/paper_157.pdf.

[2]   Battauz, M., Bellio, R. & Gori, E. (2008): Reducing Measurement Error in Student Achievement Estimation. *Psychometrika*, **73**, 289-302.

[3]   Biewen, E., Nolte, S. & Rosemann, M. (2008): Perturbation by Multiplicative Noise and The Simulation Extrapolation Method. *Advances in Statistical Analysis*, **92**, 375-389.

[4]   Bradburn, N. M., Huttenlocher, J. & Hedges, L. (1994): Telescoping and Temporal Memory. In N. Schwarz et al. (Ed): *Autobiographical Memory and The Validity of Retrospective Reports*, 203-215. New York: Springer.

[5]   Carroll, R. J. & Stefanski, L. A. (1990): Approximate Quasilikelihood Estimation in Models with Surrogate Predictors. *Journal of the American Statistical Association*, **85**, 652-663.

[6]   Carroll, R., Küchenhoff, H., Lombard, F. & Stefanski, L. (1996): Asymptotics for the SIMEX Estimator in Nonlinear Measurement Error Models. *Journal of the American Statistical Association*, **91**, 242-250.

[7]   Carroll, R., Ruppert, D., Stefanski, L. & Crainiceanu, C. (2006): *Measurement Error in Nonlinear Models; a Modern Perspective*, Boca Raton: Chapman and Hall.

[8]   Cook, J. & Stefanski, L. (1994): A Simulation Extrapolation Method for Parametric Measurement Error Models. *Journal of the American Statistical Association*, **89**, 1314–1328.

[9]   Cole, S., Chu, H. & Greenland, S. (2006): Multiple-Imputation for Measurement-Error Correction. *International Journal of Epidemiology*, **35**, 1074-1081.

[10] Da Silva, D. & Skinner, Ch. (2014): The Use of Accuracy Indicators to Correct for Survey Measurement Error. *Journal of the Royal Statistical Society: Series C*, **62**, 303-319.

[11] Dumangane, M. (2007): Measurement Error Bias Reduction in Unemployment Durations. *Centre for Microdata Methods and Practice*, 3, from: http://www.cemmap.ac.uk/wps/cwp0603.pdf.

[12] Efron, B. & Tibshirani, R. J. (1993): *An Introduction to the Bootstrap*. Boca Raton: CRC press.

[13] Fuller, W. (1987): *Measurement Error Models*. New York: John Wiley and Sons.

[14] Glesjer, L. (1990): Improvements of the Naive Approach to Estimation in Nonlinear Errors-in-Variables Regression Models. In P. Brown & W. Fuller (Ed): *Statistical Analysis of Error Measurement Models and Application*, 99-114. Providence: American Mathematics Society.

[15] Glewwe, P. (2007): Measurement Error Bias in Estimates of Income and Income Growth among the Poor: Analytical Results and a Correction Formula. *Economic Development and Cultural Change*, **56**, 163-189.

[16] Golub, A., Johnson, B. D. & Labouvie, E. (2000): On Correcting Biases in Self-Reports of Age at First Substance Use with Repeated Cross-Section Analysis. *Journal of Quantitative Criminology*, **16**, 45-68.

[17] Gustafson, P. (2003): *Measurement Error and Misclassification in Statistics and Epidemiology*. Boca Raton: Chapman and Hall.

[18] He, W., Yi, G. & Xiong, J. (2007): Accelerated Failure Time Models with Covariates Subject to Measurement Error. *Statisitcs in Medicine*, **26**, 4817-4832.

[19] Holt, D., McDonald, J.W. & Skinner, C.J. (1991): The Effect of Measurement Error on Event History Analysis. In P. Biemer (Ed): *Measurement Error in Surveys*, 665-685. New York: John Wiley.

[20] Huber, P. J. (1964): Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, **35**, 73-101.

[21] Huttenlocher, J., Hedges, L. & Prohaska, V. (1988): Hierarchical Organization in Ordered Domains: Estimating the Dates of Events. *Psychological Review*, **95**, 471-484.

[22] Hwang, J. T. (1986): Multiplicative Errors-in-Variables Models with Applications to Recent Data Released by the US Department of Energy. *Journal of the American Statistical Association*, **81**, 680-688.

[23] Janssen, S. M., Chessa, A. G. & Murre, J. M. (2006): Memory for Time: How People Date Events. *Memory & Cognition*, **34**, 138-147.

[24] Johnson, E. O. & Schultz, L. (2005): Forward Telescoping Bias in Reported Age of Onset: An Example from Cigarette Smoking. *International Journal of Methods in Psychiatric Research*, **14**, 119-129.

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

[26] Lyles, R. H. & Kupper, L. L. (1997): A Detailed Evaluation of Adjustment Methods for Multiplicative Measurement Error in Linear Regression with Applications in Occupational Epidemiology. *Biometrics*, 1008-1025.

[27] Neter, J. & Waksberg, J. (1964): A Study of Response Errors in Expenditures Data from Household Interviews. *Journal of the American Statistical Association*, **59**, 18-55.

[28] Novick, M.R. (1966): The Axioms and Principal Results of Classical Test Theory. *Journal of Mathematical Psychology*, **3**, 1-18.

[29] Nugent, W., Graycheck, L. & Basham, R. (2000): A Devil Hidden in the Details: The Effects of Measurement Error in Regression Analysis. *Journal of Social Service Research*, **27**, 53-75.

[30] Pickles, A., Pickering, K. & Taylor, C. (1996): Reconciling Recalled Dates of Developmental Milestones, Events and Transitions: A Mixed Generalized Linear Model with Random Mean and Variance Functions. *Journal of the Royal Statistical Society. Series A*, 225-234.

[31] Pickles, A., Pickering, K., Simonoff, E., Silberg, J., Meyer, J. & Maes, H. (1998): Genetic "Clocks" and "Soft" Events: A Twin Model for Pubertal Development and Other Recalled Sequences of Developmental Milestones, Transitions, or Ages at Onset. *Behavior Genetics*, **28**, 243-253.

[32] Pina-Sánchez, J., Koskinen, J. & Plewis, I. (2013): Implications of Retrospective Measurement Error in Event History Analysis. *Metodología de Encuestas*, **15**, 5-25.

[33] Pina-Sánchez, J., Koskinen, J. & Plewis, I. (2014): Measurement Error in Retrospective Work Histories. *Survey Research Methods*, **8**, 43-55.

[34] Poterba, J. M. & Summers, L. H. (1984): Response variation in the CPS: Caveats for the unemployment analyst. *Monthly Labor Review*, **107**, 37-43.

[35] Prentice, R. (1982): Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. *Biometrika*, **69**, 331-342.

[36] Rappaport, S. M. (1991): Assessment of Long-Term Exposures to Toxic Substances in Air. *Annals of Occupational Hygiene*, **35**, 61-122.

[37] Rappaport, S. M., Kromhouta, H. & Symanski, E. (1993): Variation of Exposure Between Workers in Homogeneous Exposure Groups. *The American Industrial Hygiene Association Journal*, **54**, 654-662.

[38] Roberts, C. (2007): Mixing modes of data collection in surveys: A methodological review. *Economic and Social Research Council - National Centre for Research Methods*, from: http://eprints.ncrm.ac.uk/418/.

[39] Rubin, D. C. (1987): *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

[40] Rubin, D. C. & Baddeley, A. D. (1989): Telescoping Is Not Time Compression: A Model. *Memory & Cognition*, **17**, 653-661.

[41] Skinner, C. & Humphreys, K. (1999): Weibull Regression for Lifetimes Measured with Error. *Lifetime Data Analysis*, **5**, 23-37.

[42] Skinner, C. (2000): Dealing with Measurement Error in Panel Analysis. In D. Rose (Ed): *Researching Social and Economic Change*, 113-125. New York: Routledge.

[43] Solga, H. (2001): Longitudinal Survey and the Study of Occupational Mobility: Panel and Retrospective Design in Comparison. *Quality and Quantity,* **35**, 291-309.

[44] Stefanski, L. & Cook, J. (1995): Simulation-Extrapolation: The Measurement Error Jackknife. *Journal of the American Statistical Association*, **90**, 1247-1256.

[45] Theil, H. (1953): *Repeated Least Squares Applied to Complete Equation Systems*. The Hague: Central Planning Bureau.

[46] Valaste, M., Lehtonen, R. & Vehkalahti, K. (2010): Multiple Imputation for Measurement Error Correction in Survey Data. *Q2010 European Conference on Quality in Official Statistics*, 89.

[47] Vardeman, S. B., Wendelberger, J. R., Burr, T., Hamada, M. S., Moore, L. M., Jobe, J. M., Morris, M. D. & Wu, H. (2010): Elementary statistical methods and measurement error. *The American Statistician*, **64**, 46-51.

Appendix I. R Script

Data Simulations

```
set.seed(10)
```

```
#The simulated variables
```

```
Y = rnorm(1000,0,1)
Yca = ifelse(Y>=0,1,0)
Yco = ifelse(Y<0,0,ifelse(Y>=0&Y<1,1,ifelse(Y>=1&Y<2,2,ifelse(Y>=2&Y<3,3,4))))
X1 = exp((Y+4)*.5 + rnorm(1000,0,.25))
X2 = Y*.4 + rnorm(1000,0,1)
```

```
#The classical multiplicative measurement error
```

```
X1star = X1*rnorm(1000,1,.25)
```

```
#The heteroscedastic measurement error
```

```
U = seq(1:1000)
for(i in 1:1000){
        U[i] = ifelse(X2[i]*rnorm(1,1,.5)<0, rnorm(1,1,.15), rnorm(1,1,.30))
        }
U = ifelse(U<=0, .001, U)
plot(X2,U)
X1star = X1*U
```

```
#The systematic negative measurement error
```

```
X1star = X1*rnorm(1000,.9,.25)
```

```
#The systematic positive measurement error
```

```
X1star = X1*rnorm(1000,.1.1,.25)
```

The SIMEX Process

```
##Example assuming classical multiplicative measurement error, a linear outcome model, and known variance
of the error term###########################################################################
```

```
#The outcome models.
```

```
lm.true = lm(Y ~ X1 + X2)
summary(lm.true)
lm.naive = lm(Y ~ X1star + X2)
summary(lm.naive)
```

```
#Estimates of the impact of measurement error
```

```
sum.n = summary(lm.naive)
sum.n = summary(lm.naive)
biases = coef(lm.naive) - coef(lm.true)
```

```
#Matrices to save results from the SIMEX process.
```

```
results = matrix(c(0),nrow=12,ncol=12,byrow=TRUE)
colnames(results) = c("lin.coef", "lin.bias", "lin.rbias")
rownames(results) = c("right.cons", "right.X1", "right.X2", "over.cons", "over.X1", "over.X2", "under.cons",
"under.X1", "under.X2", "very.cons", "very.X1", "very.X2")
SE = matrix(c(0),nrow=100,ncol=3,byrow=TRUE)
```

```
##The SIMEX process############################################################

#noise.5

noise.5 = matrix(c(0),nrow=1000,ncol=3,byrow=TRUE)
for(i in 1:1000){
        ME.5 = rnorm(1000,1,.25)^.5
        X1star.5 = X1star * ME.5
        lm.noise.5 = lm(Y ~ X1star.5 + X2)
        noise.5[i,1] = coef(lm.noise.5)[1]
        noise.5[i,2] = coef(lm.noise.5)[2]
        noise.5[i,3] = coef(lm.noise.5)[3]
        }
avg.5_cons = mean(noise.5[,1])
avg.5_X1 = mean(noise.5[,2])
avg.5_X2 = mean(noise.5[,3])

#noise1

noise1 = matrix(c(0),nrow=1000,ncol=3,byrow=TRUE)
for(i in 1:1000){
        ME1 = rnorm(1000,1,.25)^1
        X1star1 = X1star * ME1
        lm.noise1 = lm(Y ~ X1star1 + X2)
        noise1[i,1] = coef(lm.noise1)[1]
        noise1[i,2] = coef(lm.noise1)[2]
        noise1[i,3] = coef(lm.noise1)[3]
        }
avg1_cons = mean(noise1[,1])
avg1_X1 = mean(noise1[,2])
avg1_X2 = mean(noise1[,3])

#noise1.5

noise1.5 = matrix(c(0),nrow=1000,ncol=3,byrow=TRUE)
for(i in 1:1000){
        ME1.5 = rnorm(1000,1,.25)^1.5
        X1star1.5 = X1star * ME1.5
        lm.noise1.5 = lm(Y ~ X1star1.5 + X2)
        noise1.5[i,1] = coef(lm.noise1.5)[1]
        noise1.5[i,2] = coef(lm.noise1.5)[2]
        noise1.5[i,3] = coef(lm.noise1.5)[3]
        }
avg1.5_cons = mean(noise1.5[,1])
avg1.5_X1 = mean(noise1.5[,2])
avg1.5_X2 = mean(noise1.5[,3])

#noise2

noise2 = matrix(c(0),nrow=1000,ncol=3,byrow=TRUE)
for(i in 1:1000){
        ME2 = rnorm(1000,1,.25)^2
        X1star2 = X1star * ME2
        lm.noise2 = lm(Y ~ X1star2 + X2)
        noise2[i,1] = coef(lm.noise2)[1]
        noise2[i,2] = coef(lm.noise2)[2]
        noise2[i,3] = coef(lm.noise2)[3]
        }
avg2_cons = mean(noise2[,1])
avg2_X1 = mean(noise2[,2])
avg2_X2 = mean(noise2[,3])

#I put the mean regression estimates from each level of simulated measurement error in a dataset
```

```
avg_noiseADJ_cons = NA
avg_noiseADJ_X1 = NA
avg_noiseADJ_X2 = NA
lambda = c(-1, 0, .5, 1, 1.5, 2)
addi1 = c(avg_noiseADJ_cons, coef(lm.naive)[1], avg.5_cons, avg1_cons, avg1.5_cons, avg2_cons)
addi2 = c(avg_noiseADJ_X1, coef(lm.naive)[2], avg.5_X1, avg1_X1, avg1.5_X1, avg2_X1)
addi3 = c(avg_noiseADJ_X2, coef(lm.naive)[3], avg.5_X2, avg1_X2, avg1.5_X2, avg2_X2)
SIMEX = data.frame(lambda, addi1, addi2, addi3)
names(SIMEX) = c("lambda","cons","X1","X2")
```

#I obtain the adjusted SIMEX estimates using a linear extrapolation function

```
SIMEXna = SIMEX[-1,]
SIMEX_cons = lm(SIMEXna$cons ~ SIMEXna$lambda)
SIMEX[1,2] = coef(SIMEX_cons)[1] + (-1)*coef(SIMEX_cons)[2]
SIMEX_X1 = lm(SIMEXna$X1 ~ SIMEXna$lambda)
SIMEX[1,3] = coef(SIMEX_X1)[1] + (-1)*coef(SIMEX_X1)[2]
SIMEX_X2 = lm(SIMEXna$X2 ~ SIMEXna$lambda)
SIMEX[1,4] = coef(SIMEX_X2)[1] + (-1)*coef(SIMEX_X2)[2]
```

#I save the adjusted estimates and the remaining bias

```
results[1,1] = SIMEX[1,2]
results[2,1] = SIMEX[1,3]
results[3,1] = SIMEX[1,4]
bias1 = SIMEX[1,2]-coef(lm.true)[1]
bias2 = SIMEX[1,3]-coef(lm.true)[2]
bias3 = SIMEX[1,4]-coef(lm.true)[3]
results[1,2] = bias1
results[2,2] = bias2
results[3,2] = bias3
```

#I calculate the R.BIAS

```
R.BIAS.1 = (abs(coef(lm.naive)[1]-coef(lm.true)[1])*100)/abs(coef(lm.true)[1])
R.BIAS.adj.1 = (abs(SIMEX[1,2]-coef(lm.true)[1])*100)/abs(coef(lm.true)[1])
results[1,3] = R.BIAS.adj.1 / R.BIAS.1
R.BIAS.2 = (abs(coef(lm.naive)[2]-coef(lm.true)[2])*100)/abs(coef(lm.true)[2])
R.BIAS.adj.2 = (abs(SIMEX[1,3]-coef(lm.true)[2])*100)/abs(coef(lm.true)[2])
results[2,3] = R.BIAS.adj.2 / R.BIAS.2
R.BIAS.3 = (abs(coef(lm.naive)[3]-coef(lm.true)[3])*100)/abs(coef(lm.true)[3])
R.BIAS.adj.3 = (abs(SIMEX[1,4]-coef(lm.true)[3])*100)/abs(coef(lm.true)[3])
results[3,3] = R.BIAS.adj.3 / R.BIAS.3
```

##The bootstrap process to calculate the standard errors obtained from the SIMEX adjustment##############

#The double-loop

```
for(l in 1:100){
          boot = data[sample(1:nrow(data), 1000, replace=TRUE),]

          #noise.5

          noise.5 = matrix(c(0),nrow=100,ncol=3,byrow=TRUE)
          for(i in 1:100){
                  boot$ME.5 = rnorm(1000,1,.25)^.5
                  boot$X1star.5 = boot$X1star * boot$ME.5
                  lm.noise.5 = lm(Y ~ X1star.5 + X2, data=boot)
                  noise.5[i,1] = coef(lm.noise.5)[1]
                  noise.5[i,2] = coef(lm.noise.5)[2]
                  noise.5[i,3] = coef(lm.noise.5)[3]
                  }
          avg.5_cons = mean(noise.5[,1])
          avg.5_X1 = mean(noise.5[,2])
          avg.5_X2 = mean(noise.5[,3])
```

```
#noise1

noise1 = matrix(c(0),nrow=100,ncol=3,byrow=TRUE)
        for(i in 1:100){
        boot$ME1 = rnorm(1000,1,.25)^1
        boot$X1star1 = boot$X1star * boot$ME1
        lm.noise1 = lm(Y ~ X1star1 + X2, data=boot)
        noise1[i,1] = coef(lm.noise1)[1]
        noise1[i,2] = coef(lm.noise1)[2]
        noise1[i,3] = coef(lm.noise1)[3]
           }
avg1_cons = mean(noise1[,1])
avg1_X1 = mean(noise1[,2])
avg1_X2 = mean(noise1[,3])

#noise1.5

noise1.5 = matrix(c(0),nrow=100,ncol=3,byrow=TRUE)
for(i in 1:100){
        boot$ME1.5 = rnorm(1000,1,.25)^1.5
        boot$X1star1.5 = boot$X1star * boot$ME1.5
        lm.noise1.5 = lm(Y ~ X1star1.5 + X2, data=boot)
        noise1.5[i,1] = coef(lm.noise1.5)[1]
        noise1.5[i,2] = coef(lm.noise1.5)[2]
        noise1.5[i,3] = coef(lm.noise1.5)[3]
           }
avg1.5_cons = mean(noise1.5[,1])
avg1.5_X1 = mean(noise1.5[,2])
avg1.5_X2 = mean(noise1.5[,3])

#noise2

noise2 = matrix(c(0),nrow=100,ncol=3,byrow=TRUE)
        for(i in 1:100){
        boot$ME2 = rnorm(1000,1,.25)^2
        boot$X1star2 = boot$X1star * boot$ME2
        lm.noise2 = lm(Y ~ X1star2 + X2, data=boot)
        noise2[i,1] = coef(lm.noise2)[1]
        noise2[i,2] = coef(lm.noise2)[2]
        noise2[i,3] = coef(lm.noise2)[3]
           }
avg2_cons = mean(noise2[,1])
avg2_X1 = mean(noise2[,2])
avg2_X2 = mean(noise2[,3])

#I save the adjusted estimates and the remaining bias

avg_noiseADJ_cons = NA
avg_noiseADJ_X1 = NA
avg_noiseADJ_X2 = NA
lambda = c(-1, 0, .5, 1, 1.5, 2)
addi1 = c(avg_noiseADJ_cons, coef(lm.naive)[1], avg.5_cons, avg1_cons, avg1.5_cons, avg2_cons)
addi2 = c(avg_noiseADJ_X1, coef(lm.naive)[2], avg.5_X1, avg1_X1, avg1.5_X1, avg2_X1)
addi3 = c(avg_noiseADJ_X2, coef(lm.naive)[3], avg.5_X2, avg1_X2, avg1.5_X2, avg2_X2)
SIMEX = data.frame(lambda, addi1, addi2, addi3)
names(SIMEX) = c("lambda","cons","X1","X2")

#I obtain the adjusted estimate using a linear extrapolation function

SIMEXna = SIMEX[-1,]
SIMEX_cons = lm(SIMEXna$cons ~ SIMEXna$lambda)
SIMEX[1,2] = coef(SIMEX_cons)[1] + (-1)*coef(SIMEX_cons)[2]
SIMEX_X1 = lm(SIMEXna$X1 ~ SIMEXna$lambda)
SIMEX[1,3] = coef(SIMEX_X1)[1] + (-1)*coef(SIMEX_X1)[2]
```

```
SIMEX_X2 = lm(SIMEXna$X2 ~ SIMEXna$lambda)
SIMEX[1,4] = coef(SIMEX_X2)[1] + (-1)*coef(SIMEX_X2)[2]
#I save the SIMEX adjustment for each of the 100 bootstrap iterations
SE[l,1] = SIMEX[1,2]
SE[l,2] = SIMEX[1,3]
SE[l,3] = SIMEX[1,4]
}
```

#I obtain the standard errors

```
SE1 = sd(SE[,1])
SE2 = sd(SE[,2])
SE3 = sd(SE[,3])
```

Appendix II. Illustrations of the SIMEX Process

Figure A1 shows the effect of the increased levels of simulated measurement error on X using $\sigma_V^2 \sim (0, .25)$ and $\lambda_k = (0.5, 1, 1.5, 2)$.

Figure A1: Scatterplots of X_1 and increasing levels of measurement error
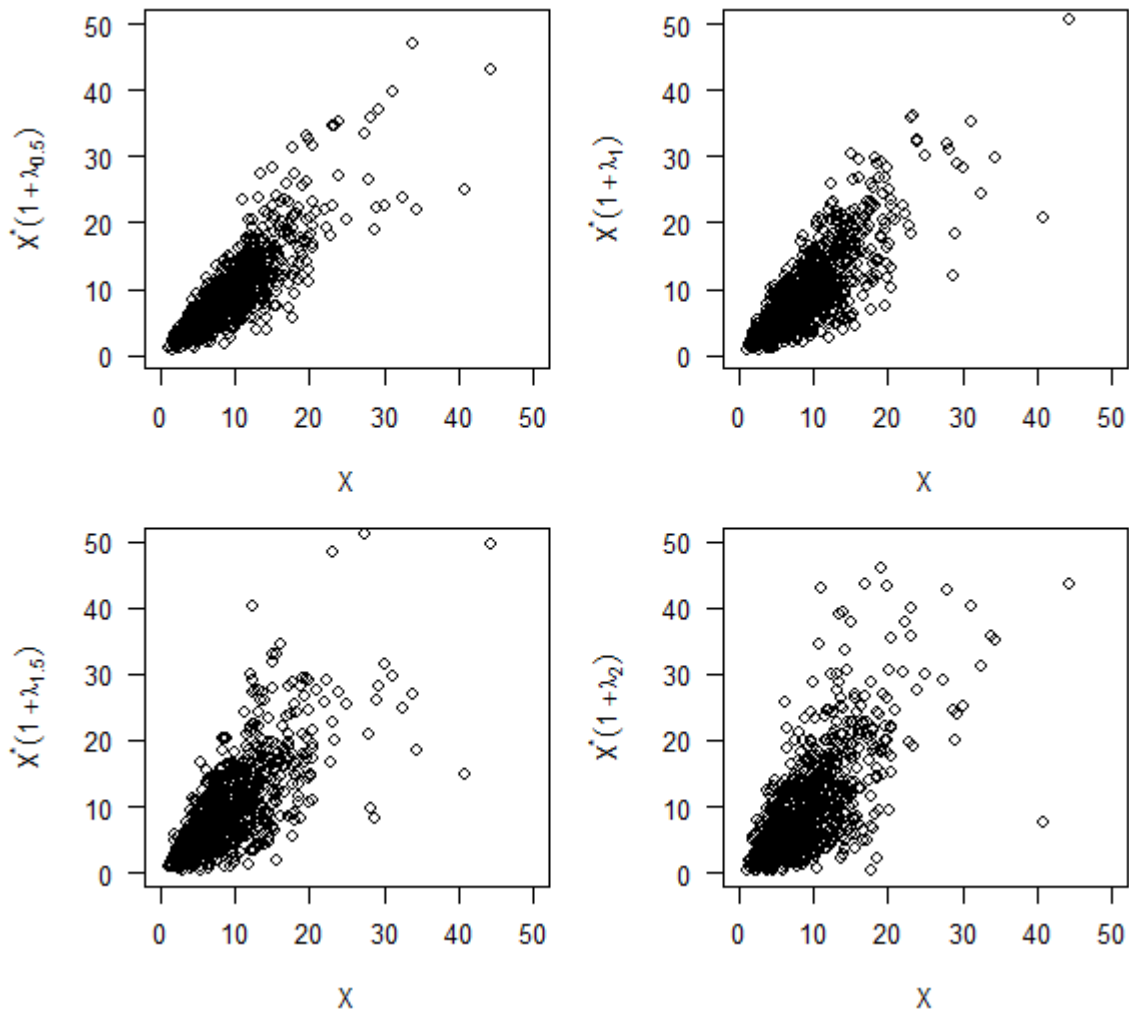

Figure A2 shows the extrapolation functions for β_1 when the outcome model is linear and X_1 is affected by classical multiplicative measurement. Each of the plots represents one of the four scenarios where different reliability ratios were assumed.

Figure A2: Extrapolation functions for the linear model