Using the SIMEX Method to Estimate Temporal Change for a High-scoring Group

Andrea Piesse[1] and Cindy X. Feng[2]
[1]Westat, 1650 Research Blvd., Rockville, MD 20850
[2]Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, V5A 1S6

Abstract
This paper discusses issues with the estimation of temporal change in the presence of measurement error – a problem commonly referred to as "regression toward the mean". We propose the SIMEX method as a technique for dealing with measurement error in this case and present the results of applying the method in a simulation study.

Keywords: Measurement error, regression toward the mean, SIMEX method

1. Introduction

Most longitudinal studies have the estimation of gross change as one of their objectives, i.e., change over time at the individual level. However, such analyses can be problematic if the measurement of the variable of interest is unreliable. The effect of random additive measurement error in the predictor variable in a simple linear regression model has been well documented (see, for example, Fuller, 1987). The ordinary least squares estimate of the slope coefficient is attenuated toward zero by the ratio of the variance of the true value of the predictor to the variance of the observed value (i.e., the true value plus measurement error). This ratio, r, is known as the "reliability" of the observed variable. However, when multiple predictor variables are measured imprecisely, the error structure is complex; also in non-regression settings, the effects of measurement error are often not easy to determine. Analysts sometimes fail to recognize that the "results" of their study may simply be due to one or more statistical effects, in particular to the statistical regression artifact (Campbell and Kenny, 1999).

The motivation for our work comes from the field of education, where there is often great interest in estimating change in student academic performance over time. In particular, there may be interest in studying students at one end of the distribution of academic performance; those most in need of improvement or those with very high test scores. For a variety of reasons, test scores are an imperfect measure of student ability. As a result, such scores are expected to regress toward the mean when any two time periods are compared.

Suppose, for example, that we want to estimate the percentage of students from the top test quartile at one grade who are still in the top test quartile based on their test score at a later grade. A naïve survey-weighted estimate, based only on the observed proportion of students from the top test quartile who remain there, is likely to under-estimate the true percentage based on actual ability. If we also wish to compare the percentages of persistent high-performance for students in two groups (e.g., low vs. high income, whites vs. other racial/ethnic backgrounds), inference is further complicated to the extent that regression toward the mean affects the groups differently.

One way to mitigate the effects of measurement error in measures of academic performance is to analyze change using the average of two or more test scores for the same student, e.g., math and reading. Since the amount of measurement error in the average of the two test scores is smaller than that in either of the individual scores, this approach is helpful but does not eliminate the problem altogether. In addition, the availability of more than one test score is not guaranteed in any given study. Other suggestions for how to account for imperfect measures include shrinking time 1 scores toward a group mean. However, particularly in analyses involving group comparisons, the choice of the group to which a particular student truly belongs then comes into question.

2. The SIMEX Method

Cook and Stefanski (1995) have proposed the simulation extrapolation (SIMEX) method for dealing with measurement error in regression analyses. The method assumes that the measurement error variance is either known or can be estimated from other available data. It consists of four main steps:

1. In the first step, a known amount of measurement error is added to each observation, X_i, in the data set. This is accomplished by choosing $\lambda > 0$ and letting

   $$W_i(\lambda) = X_i + \sqrt{\lambda}\sigma_e Z_i$$

   where σ_e^2 is the measurement error variance associated with the observed data, and the Z_i have a standard Normal distribution, are mutually independent, and are independent of the observed data. The amount of measurement error associated with the augmented data points, $W_i(\lambda)$, is $(1+\lambda^2)\sigma_e^2$. Keeping λ fixed, this process is then repeated a number of times to create, say, B simulated data sets having the same known amount of measurement error.

2. In the second step of the SIMEX method, an estimate of the parameter of interest, θ, is computed from each of the data sets generated in step one. Then, $\hat{\theta}(\lambda)$ is computed as the average of the estimates across the B simulated data sets having the same amount of measurement error.

3. Step three of the SIMEX method consists of repeating the first two steps for different values of λ, corresponding to different known amounts of additional measurement error. Then a trend is established between the amount of measurement error and the parameter estimate, $\hat{\theta}(\lambda)$.

4. In the final step of the SIMEX method, the trend established in step three is extrapolated back to $\lambda = -1$ to find the estimate of the parameter of interest that would correspond to the situation where the data contain no measurement error. That is, $\hat{\theta}_{SIMEX} = \hat{\theta}(-1)$ is an estimate of the true parameter. Typically, $\hat{\theta}(\lambda)$ is well approximated by a quadratic, however other functional forms may also work. We comment further on this aspect of the method in the next section.

An attractive feature of the SIMEX method is that it can be easily extended beyond the classic additive model to measurement error problems that are far more general. However, the ease with which estimates can be obtained is somewhat offset by their complexity, which has implications for the computation of standard errors. The bootstrap variance estimator is a natural choice but is computationally intensive. Under suitable assumptions, an estimating equation approach can be used. Alternatively, the variance of $\hat{\theta}_{SIMEX}$ itself can be estimated using the SIMEX method. Using this approach, the variance is decomposed into a component due to sampling variability and a component due to measurement error variability, and the following quantity is modeled and then extrapolated to $\lambda = -1$ to produce $\mathrm{var}_{SIMEX}(\hat{\theta}_{SIMEX})$:

$$\frac{1}{B}\sum_{b=1}^{B}\mathrm{var}(\hat{\theta}_{b\lambda}) - \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_{b\lambda} - \hat{\theta}_{\lambda})^2$$

The reader is referred to Carroll et al. (1995) for a more detailed discussion of variance estimation methods.

To the best of our knowledge, use of the SIMEX method for survey estimates of change over time has not previously been proposed. However, it is an intuitively appealing solution to problems such as our motivating example, since an estimate of the reliability of educational test scores (and therefore their associated measurement error) is generally known. In the next section, we explore via simulation the use of the SIMEX method to estimate the percentage of students from the top test quartile at one grade who remain there at a later grade.

3. Simulation Study

First, a population of size 100,000 was generated. From this, 1,000 simple random samples were selected. Three different sample sizes were investigated: 200, 1,000, and 10,000. True test scores at time 1 and time 2 were generated from a standard Normal distribution, such that the correlation between the true scores over time was ρ. Letting U_{ti} denote the true score for unit i at time t, observed test scores were generated as $X_{ti} = U_{ti} + \sigma_e$, where σ_e^2 represents the measurement error variance.

The population quantity of interest was assumed to be

$$\theta = P\left(U_2 \text{ in top quartile at time 2} \mid U_1 \text{ in top quartile at time 1}\right)$$

Under simple random sampling, the naïve estimate of θ, $\hat{\theta}_{NAIVE}$, is

$$\hat{\theta}_{NAIVE} = \frac{\begin{pmatrix}\text{number in sample for which } X_1 \text{ in top quartile of time 1 values and} \\ X_2 \text{ in top quartile of time 2 values}\end{pmatrix}}{\left(\text{number in sample for which } X_1 \text{ in top quartile of time 1 values}\right)}$$

To compute the SIMEX estimate of θ we applied the method described in Section 2, except that in step one, the measurement error was added to the observed test scores at both time 1 *and* time 2. That is, for each unit in the sample, $\lambda > 0$, and $b = 1,\ldots,100$, we computed

$$W_{1ib}(\lambda) = X_{1i} + \sqrt{\lambda}\sigma_e Z_{1ib}$$

$$W_{2ib}(\lambda) = X_{2i} + \sqrt{\lambda}\sigma_e Z_{2ib}$$

where the Z_{tib} have a standard Normal distribution, are mutually independent, and are independent of the observed data.

We then computed

$$\hat{\theta}_b(\lambda) = \frac{\begin{pmatrix}\text{number in sample for which } W_{1b} \text{ in top quartile of time 1 values and} \\ W_{2b} \text{ in top quartile of time 2 values}\end{pmatrix}}{\left(\text{number in sample for which } W_{1b} \text{ in top quartile of time 1 values}\right)}$$

and

$$\hat{\theta}(\lambda) = \frac{\sum_{b=1}^{100} \hat{\theta}_b(\lambda)}{100}$$

for λ = 0.4, 0.8, 1.2, 1.6, and 2. In the extrapolation step, we tried fitting $\hat{\theta}(\lambda)$ using linear, quadratic, spline, and exponential functions of λ.

The simulation was repeated using three different values for the correlation in the true scores over time: $\rho = 0.5$, 0.7, and 0.9; and three different assumptions for the amount of measurement error in the observed scores: $\sigma_e^2 = 0.09$, 0.16, and 0.25, corresponding to reliability ratios of 0.92, 0.86, and 0.80, respectively. These reliabilities seem sensible based on national psychometric reports (NCES, 1995 and 2002).

4. Results

A nice feature of the SIMEX method is that it provides a visual tool for demonstrating the effects of measurement error. Consider, for example, the following illustration based on one of the simulated sets of conditions.

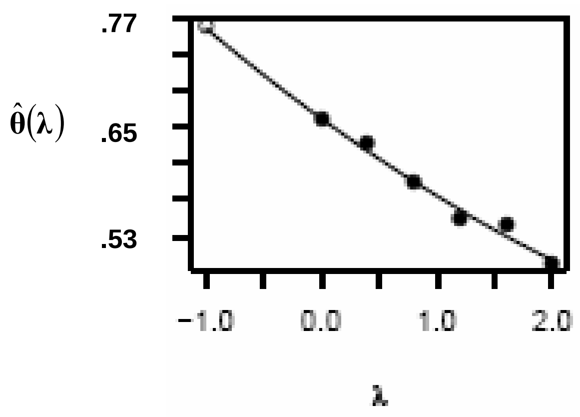

Figure 1: Illustration of the SIMEX method

The solid circle in Figure 1 corresponding to $\lambda = 0$ represents the naïve estimate of θ, whereas the open circle corresponding to $\lambda = -1$ represents the SIMEX estimate. It would appear that the naïve method underestimates the true parameter of interest, an idea which is reinforced by the observation that the estimates get smaller as the data upon which they are based contain increasing amounts of measurement error.

We evaluated the performance of the SIMEX method in terms of the bias and mean square error (MSE) of the point estimate. Table 1 shows the percentage relative bias and MSE of the naïve and SIMEX estimators for the three different sample sizes and levels of reliability, when the correlation in the true scores over time is 0.9. Table 2 shows the corresponding statistics for the different degrees of correlation over time when the reliability of the observed scores is 0.8. The summary statistics reported for the SIMEX estimator are based on quadratic extrapolation.

It is apparent that the bias in the point estimate of θ was substantially smaller with the SIMEX method than the naïve method for all the scenarios considered in the simulation. However, when both the correlation over time and the amount of measurement error were high, some bias remained even with the SIMEX method. The relative bias was largely unaffected by the sample size. As expected, for each simulated scenario, the MSE of both the naïve and SIMEX estimators became smaller as the sample size increased. The SIMEX estimator appeared to benefit much more from increased sample size in this respect than did the naive estimator. We speculate that this is because the decrease in the variance of the naïve estimator with larger n was not sufficient to offset the large relative bias in the point estimate. Using MSE as the overall measure of performance, the SIMEX method performed better than the naïve method for larger sample sizes, and for smaller sample sizes when the correlation in the true scores over time was high and the reliability was not high (i.e., the amount of measurement error in the observed data was not small).

As illustrated in Figure 2, the choice of functional form in the extrapolation step did influence the results, in particular the variance of the SIMEX estimator. The grey vertical dashed line represents the true value of θ. For this particular application, fitting a cubic to $\hat{\theta}(\lambda)$ tended to produce the largest variance but smallest bias, and the exponential function produced the smallest variance but largest bias, among the SIMEX estimators. Overall, the quadratic function seemed the best choice in terms of minimizing MSE.

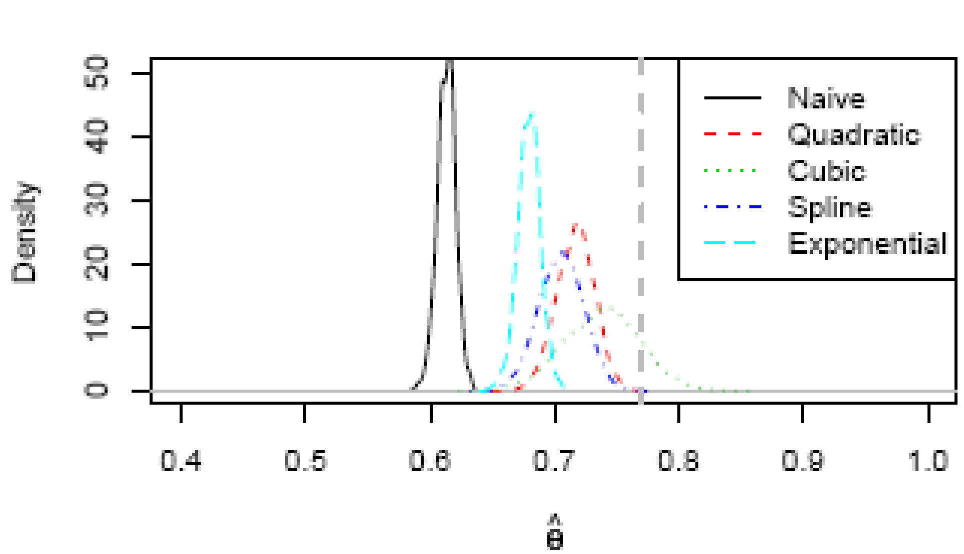

Figure 2: Density of the naïve and SIMEX estimators for $n = 10{,}000$, $\rho = 0.9$, and $r = 0.8$.

We also experimented with the range of additional measurement error, but discovered that MSE statistics were larger for $\lambda \in (0,1]$ than for $\lambda \in (0,2]$. Our experience based on the simulation study is that using a small number of points in the extrapolation step may be adequate, particularly when sample sizes are large. We used five values of λ to reduce computing time. The number of simulations used in step one of the SIMEX method was varied between $B = 100$ and 200, but the choice had little effect on the results.

5. Discussion and Further Research

We are encouraged by the results of the simulation study into the use of the SIMEX method for estimating change over time and see several directions for future research. Here we assumed that the survey data came from a simple random sample, whereas a more realistic approach would be to study the performance of the SIMEX method assuming a complex sample design. Thinking back to our motivating example, we are also interested in using the method to estimate group effects, for example by comparing the percentage of students who maintain top academic quartile performance over time for different income groups. Finally, we hope to explore the use of the SIMEX method for estimating temporal change in variables with non-Normal distributions.

Acknowledgements

The authors acknowledge the late Randy Sitter for his interest and suggestions during the early phases of the research. We also thank Graham Kalton of Westat for valuable discussions, direction, and comments.

References

Campbell, D.T. and Kenny, D.A. (1999). *A Primer on Regression Artifacts*. The Guilford Press, New York, NY.

Carroll, R.J., Ruppert D., and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall, London, UK.

Cook, J. and Stefanski, L.A. (1995). A simulation extrapolation method for parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314-1328.

Fuller, W.A. (1987). *Measurement Error Models*. Wiley, New York, NY.

U.S. Department of Education, National Center for Education Statistics (2002). *Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for Kindergarten Through First Grade*, NCES 2002–05, by Donald A. Rock and Judith M. Pollack, Educational Testing Service, Elvira Germino Hausken, project officer. Washington, DC.

U.S. Department of Education, National Center for Education Statistics (1995). *Psychometric Report for the NELS:88 Base Year Through Second Follow-up*, NCES 2002–05, by Donald A. Rock and Judith M. Pollack, Educational Testing Service, Peggy Quinn, project officer. Washington, DC.

Table 1: Simulation results for $\rho = 0.9$ using quadratic extrapolation

Estimator	$r = 0.8$		$r = 0.86$		$r = 0.92$	
	RB %	*MSE (x 10^{-2})*	*RB %*	*MSE (x 10^{-2})*	*RB %*	*MSE (x 10^{-2})*
Naïve, $n = 200$	-20.2	2.72	-14.8	1.57	-9.4	0.76
Naïve, $n = 1000$	-20.2	2.49	-14.9	1.36	-9.4	0.58
Naïve, $n = 10000$	-20.3	2.45	-14.9	1.32	-9.5	0.55
SIMEX, $n = 200$	-6.9	1.57	-3.4	1.17	-1.0	0.83
SIMEX, $n = 1000$	-6.6	0.52	-3.5	0.27	-1.1	0.18
SIMEX, $n = 10000$	-6.8	0.29	-3.4	0.09	-1.3	0.03

Table 2: Simulation results for $r = 0.8$ using quadratic extrapolation

Estimator	$\rho = 0.5$		$\rho = 0.7$		$\rho = 0.9$	
	RB %	*MSE (x 10^{-2})*	*RB %*	*MSE (x 10^{-2})*	*RB %*	*MSE (x 10^{-2})*
Naïve, $n = 200$	-10.9	0.58	-14.8	1.08	-20.2	2.72
Naïve, $n = 1000$	-10.6	0.32	-14.5	0.81	-20.2	2.49
Naïve, $n = 10000$	-10.6	0.26	-14.3	0.73	-20.3	2.45
SIMEX, $n = 200$	-2.3	1.22	-3.8	1.28	-6.9	1.57
SIMEX, $n = 1000$	-2.0	0.24	-3.5	0.23	-6.6	0.52
SIMEX, $n = 10000$	-2.1	0.03	-3.0	0.06	-6.8	0.29