



IAW-Diskussionspapiere

Discussion Paper

| 39 |

Multiplicative Measurement Error and the Simulation Extrapolation Method

Elena Biewen
Sandra Nolte
Martin Rosemann

January 2008

ISSN: 1617-5654

INSTITUT FÜR
ANGEWANDTE
WIRTSCHAFTSFORSCHUNG

Ob dem Himmelreich 1
72074 Tübingen
T: (0 70 71) 98 96-0
F: (0 70 71) 98 96-99
E-Mail: iaw@iaw.edu
Internet: www.iaw.edu

IAW-Diskussionspapiere

Das Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen ist ein unabhängiges außeruniversitäres Forschungsinstitut, das am 17. Juli 1957 auf Initiative von Professor Dr. Hans Peter gegründet wurde. Es hat die Aufgabe, Forschungsergebnisse aus dem Gebiet der Wirtschafts- und Sozialwissenschaften auf Fragen der Wirtschaft anzuwenden. Die Tätigkeit des Instituts konzentriert sich auf empirische Wirtschaftsforschung und Politikberatung.

Dieses **IAW-Diskussionspapier** können Sie auch von unserer IAW-Homepage als pdf-Datei herunterladen:

<http://www.iaw.edu/Publikationen/IAW-Diskussionspapiere>

ISSN 1617-5654

Weitere Publikationen des IAW:

- IAW-News (erscheinen 4x jährlich)
- IAW-Wohnungsmonitor Baden-Württemberg (erscheint 1x jährlich kostenlos)
- IAW-Forschungsberichte

Möchten Sie regelmäßig eine unserer Publikationen erhalten, dann wenden Sie sich bitte an uns:

IAW Tübingen, Ob dem Himmelreich 1, 72074 Tübingen,
Telefon 07071 / 98 96-0
Fax 07071 / 98 96-99
E-Mail: iaw@iaw.edu

Aktuelle Informationen finden Sie auch im Internet unter: <http://www.iaw.edu>

Der Inhalt der Beiträge in den IAW-Diskussionspapieren liegt in alleiniger Verantwortung der Autorinnen und Autoren und stellt nicht notwendigerweise die Meinung des IAW dar.

Multiplicative Measurement Error and the Simulation Extrapolation Method

Elena Biewen*
IAW

Sandra Nolte[†]
University of Konstanz
CMS

Martin Rosemann[‡]
IAW

January 16, 2008

Abstract

Whereas the literature on additive measurement error has known a considerable treatment, less work has been done for multiplicative noise. In this paper we concentrate on multiplicative measurement error in the covariates, which contrary to additive error not only modifies proportionally the original value, but also conserves the structural zeros.

This paper compares three variants to specify the multiplicative measurement error model in the simulation step of the Simulation-Extrapolation (SIMEX) method originally proposed by Cook and Stefanski (1994): i) as an additive one without using a logarithmic transformation, ii) as the well-known logarithmic transformation of the multiplicative error model, and iii) as an approach using the multiplicative measurement error model as such. The aim of the paper is to analyze how well these three approaches reduce the bias caused by the multiplicative measurement error. We apply three variants to the case of data masking by multiplicative measurement error, in order to obtain parameter estimates of the true data generating process. We produce Monte Carlo evidence on how the reduction of data quality can be minimized.

JEL classification: C13, C21

Keywords: Errors-in-variables in nonlinear models, disclosure limitation methods, multiplicative error

*IAW, Ob dem Himmelreich 1, D-72074 Tuebingen, Germany (elena.biewen@iaw.edu).

[†]Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany (sandra.lechner@uni-konstanz.de).

[‡]IAW, Ob dem Himmelreich 1, D-72074 Tuebingen, Germany (martin.rosemann@iaw.edu). For helpful comments we would like to thank Helmut Küchenhoff, Winfried Pohlmeier and Gerd Ronning. Financial support by the DFG is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

The growing demand of firm-level data from the empirical research side creates a tradeoff for the data collecting institutions between providing a maximum amount of information to the scientific users, and guarantying confidentiality and privacy to the firms. To protect the identity of firms, statistical offices apply different disclosure limitation techniques, where the most popular one consists in the addition of independent noise to the covariates. This leads to the well-known error-in-variables problem, where the effects of additive measurement error on the properties of linear estimators are well understood (see e.g. Fuller (1987)). The monograph of Carroll, Ruppert, and Stefanski (1995) surveys various approaches to errors-in-variables for nonlinear models. Nevertheless, even if this anonymization method can easily be implemented, it does not minimize the probability of re-identifying a single observation. Additive measurement error indeed only slightly modified the original value, especially when the original value is high.

This paper is concerned with multiplicative measurement error, which contrary to additive measurement error not only modifies proportionally the original value, so that a single observation is better protected against disclosure, but also conserves the structural zeros contained in the dataset.

Whereas the literature on additive measurement error has known a considerable treatment, less work has been done for multiplicative noise. Hwang (1986) derives a consistent estimator for the slope parameter in a linear regression model in the presence of multiplicative measurement error in the regressors. This model is extended by Lin (1989) to include also multiplicative error in the dependent variable. Schafer (1990) analyzes a quasi-likelihood approach when no distributional assumptions are made about the true and the mismeasured variables. Lyles and Kupper (1997) compare three methods of adjusting for multiplicative measurement error to obtain consistent estimators of the true parameters: a simple ordinary least squares correction, a regression of the dependent variable on the covariates and on the conditional expectation of true variable given the mismeasured one, and a quasi-likelihood approach. Iturria, Carroll, and Firth (1999) derive a consistent moment estimator for a polynomial regression model in the presence of multiplicative measurement error.

Cook and Stefanski (1994) introduce the Simulation Extrapolation method (SIMEX), which is well suited to estimate and reduce the bias due to additive measurement error. In this paper, we compare three variants to specify the multiplicative measurement error model in the simulation step of the SIMEX method, and analyze their effects on the estimates. First, we consider the approach of Ronning et al. (2005) and Rosemann (2006), who interpret the multiplicative error model as an additive one. Then we consider the most popular method to deal with multiplicative measurement error, which applies a logarithmic transformation to the multiplicative measurement error model in order to get an additive one (Carroll, Ruppert, and Stefanski (1995)). And finally, we consider the approach of Nolte (2007) who extends the SIMEX approach to the multiplicative noise case, without using a logarithmic transformation.

The outline of the paper is as follows. Section 2 gives a short review of the original SIMEX method. Section 3 presents all three ways to specify the multiplicative measurement error in the simulation step of the SIMEX method. Using Monte Carlo design for data which are masked by multiplicative measurement error, we analyse in Section 4 the small sample properties of the three proposed variants of the SIMEX procedure applied to a binary choice example. Finally, Section 5 summarizes the main results and addresses further research questions.

2 Additive Measurement Error: SIMEX Approach

In this section, we briefly sketch the idea of the SIMEX approach developed by Cook and Stefanski (1994), which was suggested in order to estimate and reduce the bias due to additive measurement error. In the simulation step additional measurement error is added to the mismeasured variable, so that the statistician can infer in which way the estimation bias is affected by the increase of variance of the measurement error. In the extrapolation step, the estimated parameters are modelled as a function of the magnitude of the variance of the measurement error and extrapolated to the case of no measurement error.

Without loss of generality, let us consider the simple case, where rather than observing an explanatory variable X_i , we observe a masked explanatory variable X_i^m defined as:

$$X_i^m = X_i + u_i, \quad i = 1, \dots, N, \quad (2.1)$$

where u_i is an independent normally distributed random variable with $E[u_i|X_i] = 0$ and $V[u_i|X_i] = \sigma_u^2$, that is added to the original variable in order to mask it.

In the simulation step, B new covariates $X_{i,b}^m(\lambda_t)$ are generated by the rule:

$$X_{i,b}^m(\lambda_t) = X_i^m + \sqrt{\lambda_t} u_{i,b}, \quad b = 1, \dots, B, \quad t = 0, \dots, T, \quad i = 1, \dots, N, \quad (2.2)$$

where $\lambda_0 < \lambda_1 < \lambda_2 < \dots < \lambda_T^1$ are given parameters controlling for the variance of the measurement error, and $\{u_{i,b}\}_{b=1}^B$ are iid computer simulated normal random numbers with mean zero and variance σ_u^2 .²

An average estimate $\hat{\beta}(\lambda_t) = \frac{1}{B} \sum_{b=1}^B \hat{\beta}_b(\lambda_t)$ is finally computed, where $\hat{\beta}_b(\lambda_t)$ denotes the naive parameters estimates obtained by regression of Y on $\{X_b^m(\lambda_t)\}$ for each λ_t , and for $b = 1, \dots, B$.

In the extrapolation step each component $\hat{\beta}(\lambda_t)$ is modelled as a function of λ_t for $\lambda_t \geq 0$. The SIMEX estimator is defined as the extrapolation of $\hat{\beta}(\lambda_t)$ to $\hat{\beta}(\lambda_t = -1)$,

¹The value $\lambda_T = 2$ is recommended by Carroll, Ruppert, and Stefanski (1995)

²Note that the variance of the explanatory variable $X_{i,b}^m(\lambda_t)$ increases with the control parameter λ_t , and is given by

$$V[X_{i,b}^m(\lambda_t)] = \sigma_x^2 + (1 + \lambda_t)\sigma_u^2, \quad (2.3)$$

where σ_x^2 is the variance of X .

which represents the bias free estimate of β , since at that point the variance of the mismeasured variable is equal to the variance of the original one. Cook and Stefanski (1994) suggest to use a linear, quadratic or a nonlinear extrapolation function. All of them will be used and specified later in the paper (See Section 4).

To estimate the variance of the SIMEX estimator, one can either use the delta method (Carroll, Kuechenhoff, Lombard, and Stefanski (1996)), the jackknife (Stefanski and Cook (1995)) or the bootstrap.³ Carroll, Kuechenhoff, Lombard, and Stefanski (1996) derive the asymptotic distribution of the SIMEX estimator for parametric models.

Note: Here we see that the SIMEX algorithm does not depend on the functional form of the model, so that this method is well suited for linear regression models as well as for nonlinear regression models.

3 Multiplicative Measurement Error

In this section we focus on the multiplicative error. We first define the general structure of a multiplicative measurement error model, and then present the three specifications used in this paper.

3.1 General Model

Without loss of generality, let us consider the following multiplicative measurement error model:

$$X_i^m = X_i \cdot w_i, \tag{3.1}$$

where w_i is an independent random variable with $E[w_i | X_i] = 1$ and $V[w_i | X_i] = \sigma_w^2$, that is multiplied with the original variable. We suppose that $E[w_i | X_i] = 1$, because in the context of disclosure limitation techniques it is noticeable that the mean of the masked data is equal to the mean of the original one.

³Usually the latter is preferred.

3.2 SIMEX Approach in the Multiplicative Case

In the previous section we have described the SIMEX approach for the additive measurement error model. This method can also be extended to the multiplicative case. In the following we present three approaches to handle the simulation extrapolation method in the case of multiplicative measurement error.

3.2.1 Interpretation of the Multiplicative Measurement Error Model as an Additive One (Variant 1)

Originally the SIMEX method was designed for the case of additive measurement errors. Therefore, it seems illuminating to write the multiplicative measurement errors as an additive one (Ronning et al. (2005), Rosemann (2006)). This might be important particularly if real data are used since the structure of the measurement error usually is unknown.

We can write for the variable X^m measured with multiplicative error:

$$X_i^m = X_i \cdot w_i = X_i + u_i. \quad (3.2)$$

Thus, for the additive measurement error holds:

$$u_i = X_i \cdot w_i - X_i = X_i(w_i - 1), \quad i = 1, \dots, N, \quad (3.3)$$

whereas the expectation of u_i is given by

$$\mathbb{E}[u_i] = \mathbb{E}[X_i(w_i - 1)] = \mathbb{E}[X_i] \mathbb{E}[w_i - 1] = 0 \quad (3.4)$$

In (3.4) we exploited the fact that the measurement error w_i is independent of X .

The variance of u_i is

$$\begin{aligned} \mathbb{V}[u_i] &= \mathbb{V}[X_i(w_i - 1)] \\ &= \mathbb{E}[(X_i(w_i - 1))(X_i(w_i - 1))] \\ &= \mathbb{E}[X_i^2 w_i^2 - 2w_i X_i^2 + X_i^2] \\ &= (\sigma_x^2 + \mu_x^2)(\sigma_w^2 + 1) - (\sigma_x^2 + \mu_x^2) \\ &= (\sigma_x^2 + \mu_x^2) \sigma_w^2, \end{aligned} \quad (3.5)$$

where μ_x denotes the mean of X and σ_x^2 denotes the variance of X .

In this approach only the variance of the measurement error u_i differs from the additive case and one can use the SIMEX method for additive measurement error presented in Section 2, whereas the variance of the explanatory variable $X_{i,b}^m(\lambda_t)$ is now given by

$$V[X_{i,b}^m(\lambda_t)] = \sigma_x^2 + (1 + \lambda_t) (\sigma_x^2 + \mu_x^2) \sigma_w^2. \quad (3.6)$$

3.2.2 Logarithmic Transformation of the Multiplicative Measurement Error Model (Variant 2)

The easiest way to deal with multiplicative measurement error is to transform the measurement error model into an additive one, through a logarithmic transformation, as pointed out by Hwang (1986). Here we follow the suggestion of Carroll, Ruppert, and Stefanski (1995) and perform the simulation step of the SIMEX procedure on the logarithms of the masked covariates

$$X_{i,b}^m(\lambda_t) = \exp\{\log(X_i^m) + \sqrt{\lambda_t} \log(w_{i,b})\}, \quad b = 1, \dots, B, \quad t = 0, \dots, T, \quad i = 1, \dots, N. \quad (3.7)$$

The rest of the procedure remains unchanged.

The disadvantage of this approach lies in the fact that it can not be used with negative values of the covariates, since we are taking the logarithm of them. In the next section, we will propose a development of this approach, which did not lead to any restrictions concerning the range of variable values.

3.2.3 Multiplicative SIMEX Approach (Variant 3)

Finally we consider the multiplicative SIMEX estimator proposed by Nolte (2007). In the simulation step, the mismeasured variable is multiplied by an additional measurement error, such that B new covariates $X_{i,b}^m(\lambda_t)$ are generated by the rule

$$X_{i,b}^m(\lambda_t) = X_i^m \cdot w_{i,b}^{\lambda_t}, \quad b = 1, \dots, B, \quad t = 0, \dots, T, \quad i = 1, \dots, N, \quad (3.8)$$

where $0 = \lambda_0 < \lambda_1 < \lambda_2 < \dots < \lambda_T$, are positive parameters controlling for the variance of the measurement error, and $\{w_{i,b}\}_{b=1}^B$ are independent and identically distributed computer simulated log-normally distributed random numbers with mean one and variance σ_w^2 .⁴ The vector of average estimates is computed in the same way as in the additive case. The extrapolation step is finally equivalent to the extrapolation step of the original SIMEX approach described in Section 2.

For positive values of the mismeasured variable X_i^m this approach is identical to the approach proposed in Subsection 3.2.2, so that we expect that the results will be the same.

Note: In the additive or multiplicative measurement error case, the data collecting institutions only have to provide the variance-covariance matrix σ_u^2 or σ_w^2 to the data users, which is sufficient to get a consistent estimate of the parameters of interest. It is important to mention, that this additional information does not increase the probability of re-identifying a single observation.

4 Monte Carlo Simulations

4.1 Simulation Design

In this section we investigate the small sample properties of the three proposed variants of the SIMEX procedure, when multiplicative measurement error occurs in the covariates in a binary choice model. Without loss of generality, let us focus on the following binary probit model with one regressor, which has been perturbed by multiplicative noise

$$Y_i = \mathbb{1}[\alpha + \beta X_i + \varepsilon_i \geq 0], \quad (4.1)$$

$$X_i^m = X_i \cdot w_i, \quad i = 1, \dots, N, \quad (4.2)$$

⁴Note that for each λ_t the simulation step creates B additional datasets (replication samples) with the same dependent variable Y_i and the explanatory variable $X_{i,b}^m(\lambda_t)$ with variance equal to

$$V[X_{i,b}^m(\lambda_t)] = E\left[\left(u_{i,b}^{\lambda_t+1}\right)^2\right] E[X_i^2] - E\left[u_{i,b}^{\lambda_t+1}\right]^2 E[X_i]^2. \quad (3.9)$$

See Nolte (2007) for a detailed derivation.

where ε_i follows a standard normal distribution. We determine that the true value of α and β are -2.5 and 0.6 respectively. In order to be able to compare all three variants, the support of the independent variable needs to be positive. That is why we use $X_i \sim \text{Log-N}(4.35, 1.75^2)$. The multiplicative error w_i follows a log-normal distribution with mean 1 and variance $\sigma_w^2 = \{0.01^2, \dots, 0.2^2\}$. The choice of the log-normal distribution is due to the fact that the data collecting institutions want to preserve the same sign for the original variable as for the anonymized one. This avoids erroneous values of some observations, for example that the number of employees becomes negative after the anonymization procedure.

The extrapolation function fits the regression $\hat{\theta}(\lambda) = f(\lambda, \gamma)$, in which $\hat{\theta}(\lambda)$ is point estimate from the naive Maximum-Likelihood estimation of the probit model. The extrapolation function builds up the relationship between $\hat{\theta}(\lambda)$ and the parameter controlling for the variance of the measurement error, λ . In modelling the SIMEX correction we follow Cook and Stefanski (1994) who suggest three different specifications of the extrapolation function.⁵

1. Linear extrapolation function:

$$\hat{\theta}(\lambda) = \gamma_1 + \gamma_2 \lambda \quad (4.3)$$

2. Quadratic extrapolation function:

$$\hat{\theta}(\lambda) = \gamma_1 + \gamma_2 \lambda + \gamma_3 \lambda^2 \quad (4.4)$$

3. Nonlinear extrapolation function:

$$\hat{\theta}(\lambda) = \gamma_1 + \frac{\gamma_2}{\gamma_3 + \lambda} \quad (4.5)$$

The number of observations is 1,000. In our simulation experiments we use 500 Monte Carlo replications for each variant and each extrapolation function. In the SIMEX correction one hundred replications are used for each λ . λ -values for the variants 1 and 2 are $\lambda \in \{0, 0.5, 1.0, 1.5, 2.0\}$, and we take the root of λ for the third variant to be able to compare those results with the other ones.

⁵The examples of different extrapolations functions in a simple linear model are schematically described in Ronning et al. (2005), pp.241-243.

4.2 Results

Let us first consider the results of the naive probit estimation (Table 1). The standard errors of the multiplicative noise are given in the first column. In addition to the average estimator and the standard deviation we display also the average bias and root mean squared error (RMSE). The higher the value of the standard error of the multiplicative measurement error, the greater the bias of the naive estimates becomes, as expected. The RMSE rises strongly with the perturbation degree.

The simulation results of the SIMEX estimation are given in Tables 2, 3 and 4. We see that in all experiments, the SIMEX procedure is able to reduce the bias due to the presence of multiplicative measurement error. Figure 1 shows the RMSE of the point estimate β depending on the standard error of the multiplicative noise σ_w^2 .⁶ Compared with the results of the naive estimation, the RMSE obtained with the SIMEX estimation procedure is considerably reduced regardless which extrapolation function or which variant was used.

In the following we analyze only the results of the SIMEX estimation. Comparing the results in Figure 1 and Tables 2, 3 and 4, we can detect only slight differences between three variants. However, the choice of a certain extrapolation function has a strong impact on the results.

Now, let us consider which extrapolation function yields the best results. With small errors ($\sigma_w = 0.02$), there are no significant differences.⁷ The quadratic extrapolation function is superior to the other functions. Before σ_w reaches 0.12 (at this point the measurement error is already fairly high), the corrected estimators are very close to the true value. The linear function is second best. It delivers good correction results if the error is middling (σ_w about 0.06). The nonlinear function fares worst. However, when the error is very large ($\sigma_w = 0.2$) the nonlinear function comes closer to the linear one.

Thus, the quality of the estimation depends on the choice of the extrapolation function. In addition to that, the margin of error also plays a role. With minimal noise ($\sigma_w = 0.02$) the differences between the variants and extrapolation functions are also

⁶The development for α is similar and therefore not reported here.

⁷However, the nonlinear extrapolation function in variants 2 and 3 has an outlier at the point $\sigma_w^2 = 0.04$ which is probably caused by numerical problems.

minimal (except for an outlier in variants 2 and 3 with the nonlinear extrapolation function). If we increase the range to middling ($\sigma_w = 0.06$), only quadratic and linear extrapolation functions work well. If the multiplicative noise is high ($\sigma_w = 0.08$ and larger) only the variants with quadratic extrapolation result in better estimates.

Table 1: Naive Probit Estimation

Multiplicative Standard Error	Naive Probit Estimation			
	avg. estimator	std.dev	avg. bias	RMSE
0.02	-2.499	0.160	0.001	0.160
	0.600	0.039	0.000	0.039
0.04	-2.469	0.158	0.031	0.161
	0.592	0.038	-0.008	0.039
0.06	-2.420	0.156	0.080	0.175
	0.581	0.038	-0.019	0.043
0.08	-2.355	0.153	0.145	0.211
	0.565	0.037	-0.035	0.051
0.10	-2.277	0.150	0.223	0.268
	0.546	0.037	-0.054	0.065
0.12	-2.219	0.146	0.311	0.344
	0.525	0.036	-0.075	0.083
0.14	-2.093	0.143	0.407	0.431
	0.502	0.035	-0.098	0.104
0.18	-1.890	0.135	0.610	0.625
	0.453	0.033	-0.147	0.150
0.20	-1.788	0.131	0.712	0.724
	0.429	0.032	-0.171	0.174

Figure 1: RMSE of Probit Estimation (β)

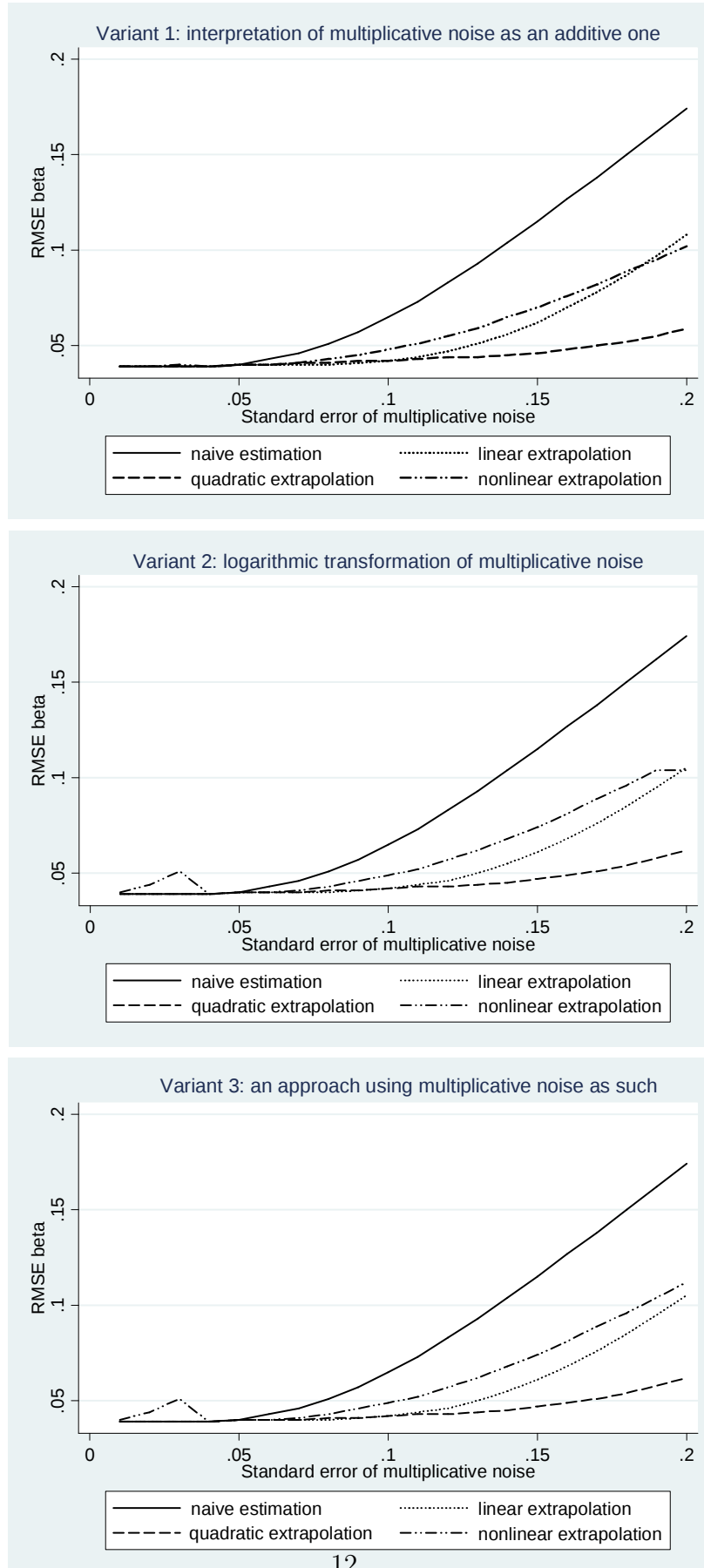


Table 2: Simulation Results: SIMEX with Linear Extrapolation

Multiplicative Standard Error	Probit Estimation with SIMEX Correction Linear Extrapolation Function											
	First Variant				Second Variant				Third Variant			
	avg. estimator	std.dev	avg. bias	RMSE	avg. estimator	std.dev	avg. bias	RMSE	avg. estimator	std.dev	avg. bias	RMSE
0.02	-2.509 0.602	0.161 0.039	-0.009 0.002	0.161 0.039	-2.509 0.602	0.161 0.039	-0.009 0.002	0.161 0.039	-2.509 0.602	0.161 0.039	-0.009 0.002	0.161 0.039
0.04	-2.507 0.602	0.162 0.039	-0.007 0.002	0.162 0.039	-2.507 0.602	0.162 0.039	-0.007 0.002	0.162 0.039	-2.507 0.602	0.162 0.039	-0.007 0.002	0.162 0.039
0.06	-2.499 0.600	0.163 0.040	0.001 -0.000	0.163 0.040	-2.499 0.600	0.163 0.040	0.001 0.000	0.163 0.040	-2.500 0.600	0.163 0.040	0.001 0.000	0.163 0.040
0.08	-2.481 0.595	0.165 0.040	0.019 -0.005	0.165 0.040	-2.480 0.596	0.164 0.040	0.020 -0.004	0.165 0.040	-2.480 0.596	0.164 0.040	0.020 -0.004	0.165 0.040
0.10	-2.449 0.588	0.166 0.040	0.051 -0.012	0.173 0.042	-2.448 0.589	0.165 0.040	0.052 -0.011	0.173 0.042	-2.448 0.589	0.165 0.040	0.052 -0.011	0.173 0.042
0.12	-2.402 0.576	0.165 0.041	0.098 -0.024	0.192 0.047	-2.400 0.577	0.164 0.040	0.100 -0.023	0.192 0.046	-2.400 0.577	0.164 0.040	0.100 -0.023	0.192 0.046
0.14	-2.339 0.562	0.164 0.040	0.161 -0.039	0.230 0.056	-2.337 0.563	0.163 0.040	0.163 -0.037	0.231 0.055	-2.337 0.563	0.163 0.040	0.163 -0.037	0.231 0.055
0.18	-2.176 0.522	0.159 0.039	0.324 -0.078	0.361 0.087	-2.175 0.524	0.160 0.039	0.325 -0.076	0.362 0.085	-2.175 0.524	0.158 0.039	0.325 -0.076	0.362 0.085
0.20	-2.081 0.499	0.156 0.039	0.419 -0.101	0.447 0.108	-2.082 0.502	0.155 0.038	0.418 -0.098	0.446 0.105	-2.082 0.502	0.155 0.038	0.418 -0.098	0.446 0.105

Table 3: Simulation Results: SIMEX with Quadratic Extrapolation

Multiplicative Standard Error	Probit Estimation with SIMEX Correction Quadratic Extrapolation Function											
	First Variant				Second Variant				Third Variant			
	avg. estimator	std.dev	avg. bias	RMSE	avg. estimator	std.dev	avg. bias	RMSE	avg. estimator	std.dev	avg. bias	RMSE
0.02	-2.509 0.602	0.161 0.039	-0.009 0.002	0.161 0.039	-2.509 0.602	0.161 0.039	-0.009 0.002	0.161 0.039	-2.509 0.602	0.161 0.039	-0.009 0.002	0.161 0.039
0.04	-2.509 0.602	0.162 0.039	-0.009 0.002	0.162 0.039	-2.509 0.602	0.162 0.039	-0.009 0.002	0.162 0.039	-2.509 0.602	0.162 0.039	-0.009 0.002	0.162 0.039
0.06	-2.509 0.602	0.164 0.040	-0.009 0.002	0.164 0.040	-2.508 0.601	0.164 0.040	-0.008 0.001	0.164 0.040	-2.508 0.601	0.164 0.040	-0.008 0.001	0.164 0.040
0.08	-2.507 0.602	0.168 0.041	-0.007 0.002	0.168 0.041	-2.506 0.601	0.168 0.041	-0.006 0.001	0.168 0.041	-2.506 0.601	0.168 0.041	-0.006 0.001	0.168 0.041
0.10	-2.503 0.601	0.172 0.042	-0.003 0.001	0.172 0.042	-2.500 0.599	0.172 0.042	0.000 -0.001	0.171 0.042	-2.500 0.599	0.172 0.042	0.000 0.000	0.171 0.042
0.12	-2.492 0.598	0.177 0.044	0.008 -0.002	0.177 0.044	-2.487 0.595	0.176 0.043	0.013 -0.005	0.176 0.043	-2.487 0.595	0.176 0.043	0.013 -0.005	0.176 0.043
0.14	-2.473 0.594	0.181 0.045	0.027 -0.006	0.183 0.045	-2.466 0.600	0.180 0.044	0.034 -0.010	0.182 0.045	-2.466 0.590	0.179 0.044	0.034 -0.010	0.182 0.045
0.18	-2.405 0.577	0.188 0.047	0.095 -0.023	0.210 0.052	-2.391 0.571	0.184 0.046	0.109 -0.029	0.214 0.054	-2.391 0.571	0.184 0.046	0.109 -0.029	0.214 0.054
0.20	-2.355 0.565	0.190 0.048	0.145 -0.035	0.239 0.059	-2.338 0.558	0.186 0.046	0.162 -0.042	0.247 0.062	-2.338 0.558	0.186 0.046	0.162 -0.042	0.247 0.062

Table 4: Simulation Results: SIMEX with Nonlinear Extrapolation

Multiplicative Standard Error	Probit Estimation with SIMEX Correction Nonlinear Extrapolation Function											
	First Variant				Second Variant				Third Variant			
	avg. estimator	std.dev	avg. bias	RMSE	avg. estimator	std.dev	avg. bias	RMSE	avg. estimator	std.dev	avg. bias	RMSE
0.02	-2.508 0.601	0.168 0.039	-0.008 0.001	0.168 0.039	-2.508 0.601	0.167 0.044	-0.008 0.001	0.167 0.044	-2.508 0.601	0.167 0.044	-0.008 0.001	0.167 0.044
0.04	-2.492 0.598	0.162 0.039	0.008 -0.002	0.162 0.039	-2.492 0.600	0.162 0.039	0.008 -0.002	0.162 0.039	-2.492 0.600	0.162 0.039	0.008 -0.002	0.162 0.039
0.06	-2.467 0.592	0.162 0.039	0.033 -0.008	0.165 0.040	-2.467 0.591	0.162 0.039	0.033 -0.009	0.165 0.040	-2.467 0.591	0.162 0.039	0.033 -0.009	0.165 0.040
0.08	-2.433 0.584	0.162 0.040	0.067 -0.016	0.175 0.043	-2.433 0.583	0.162 0.039	0.067 -0.017	0.176 0.043	-2.433 0.583	0.162 0.039	0.067 -0.017	0.176 0.043
0.10	-2.392 0.574	0.163 0.040	0.108 -0.026	0.196 0.048	-2.391 0.572	0.163 0.040	0.109 -0.028	0.196 0.049	-2.391 0.572	0.163 0.040	0.109 -0.028	0.196 0.049
0.12	-2.345 0.563	0.165 0.040	0.155 -0.037	0.226 0.055	-2.342 0.559	0.164 0.040	0.158 -0.041	0.228 0.057	-2.342 0.559	0.164 0.040	0.158 -0.041	0.228 0.057
0.14	-2.292 0.550	0.166 0.041	0.208 -0.050	0.266 0.065	-2.287 0.545	0.165 0.040	0.213 -0.055	0.270 0.068	-2.287 0.545	0.165 0.040	0.213 -0.055	0.270 0.068
0.18	-2.174 0.522	0.170 0.042	0.326 -0.078	0.367 0.089	-2.162 0.513	0.167 0.041	0.338 -0.087	0.377 0.096	-2.162 0.513	0.167 0.041	0.338 -0.087	0.377 0.096
0.20	-2.112 0.507	0.172 0.043	0.388 -0.093	0.424 0.102	-2.095 0.495	0.168 0.041	0.405 -0.105	0.439 0.112	-2.095 0.495	0.168 0.041	0.405 -0.105	0.439 0.112

5 Conclusion

In this paper we compare three variants to specify the multiplicative measurement error model in the simulation step of the SIMEX method. The choice of the SIMEX approach allows us to consider linear or nonlinear regression models with multiplicative measurement error, because this approach does not depend on the functional form of the model. The only restriction we have in our approach is that the variance of the multiplicative measurement error has to be known. The Monte Carlo experiments show that all three variants work equally well and that the quality of the estimation depends only on the choice of the extrapolation function as well as on the margin of the measurement error. In the light of our Monte Carlo results, we see also that it is always better to use several extrapolation functions and not to restrict oneself to a specific one from the beginning on, as for example the quadratic one which is principally used in applied works, since each of them seems to perform better for a certain values of the variance of the measurement error.

Since our approach is purely descriptive and leaves a lot of questions open a final conclusion concerning the relative ability of the different variants to estimate the parameters of the true data generating process, if multiplicative measurement error occurs, is somewhat premature. More Monte Carlo evidence on various nonlinear and linear regression models is needed. First, in this paper we considered the case of positive noise. From a methodic point of view it might be of interest to analyze how the variants 1 and 3 perform in the the case of uniformly or normally distributed measurement error which adopts also negative values. Second, other specifications of the extrapolation function (for instance, an exponential one) might be used in order to mimic better the relationship between the naive estimates and the control parameter for the variance of the measurement error. Finally, in order to take the outliers or in other words the extreme values of the covariates generated during the simulation step of the algorithm better into account, it will be useful to consider the median of the estimates instead of their mean; this is in progress.

References

- CARROLL, R., H. KUECHENHOFF, F. LOMBARD, AND L. STEFANSKI (1996): “Asymptotics for the Simex Estimator in Structural Measurement Error Models,” *Journal of the American Statistical Association*, 91, 242–250.
- CARROLL, R., D. RUPPERT, AND L. STEFANSKI (1995): *Measurement Error in Nonlinear Models*. Chapman and Hall.
- COOK, J., AND L. STEFANSKI (1994): “A Simulation Extrapolation Method for Parametric Measurement Error Models,” *Journal of the American Statistical Association*, 89, 1314–1328.
- FULLER, W. (1987): *Measurement Error Models*. Wiley.
- HWANG, J. T. (1986): “Multiplicative Errors-in-Variables Models with Applications to Recent Data Released by the U.S. Department of Energy,” *Journal of the American Statistical Association*, 81, 680–688.
- ITURRIA, S., R. CARROLL, AND D. FIRTH (1999): “Polynomial Regression and Estimating Functions in the Presence of Multiplicative Measurement Error,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61, 547 – 561.
- LIN, A. (1989): “Estimation of Multiplicative Measurement-Error Models and Some Simulation Results,” *Economics Letters*, 317, 13–20.
- LYLES, R., AND L. KUPPER (1997): “A Detailed Evaluation of Adjustment Methods for Multiplicative Measurement Error in Linear Regression with Applications in Occupational Epidemiology,” *Biometrics*, 53, 1008–1025.
- NOLTE, S. (2007): “The Multiplicative Simulation Extrapolation Approach,” Working Paper, University of Konstanz.
- RONNING, G., R. STURM, J. HOEHNE, R. LENZ, M. ROSEMAN, M. SCHEFFLER, AND D. VORGRIMLER (2005): *Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten*. Statistisches Bundesamt: Statistik und Wissenschaft, Band 4.
- ROSEMAN, M. (2006): *Auswirkungen datenverändernder Anonymisierungsverfahren auf die Analyse von Mikrodaten*. Institut fuer Angewandte Wirtschaftsforschung: IAW-Report, Nr. 66.
- SCHAFER, D. (1990): “Measurement Error Model Estimation using iteratively weighted least squares,” in Brown P.J. and Fuller W.A. : *Contemporary Mathematics*, Providence Rhode Island, American Mathematical Society.
- STEFANSKI, L., AND J. COOK (1995): “Simulation-Extrapolation: The Measurement Error Jackknife,” *Journal of the American Statistical Association*, 90(432), 1247–1256.

Bisher erschienen:

- Nr. 1 (September 2001)
Das Einstiegsgeld – eine zielgruppenorientierte negative Einkommensteuer: Konzeption, Umsetzung und eine erste Zwischenbilanz nach 15 Monaten in Baden-Württemberg
Sabine Dann / Andrea Kirchmann / Alexander Spermann / Jürgen Volkert
- Nr. 2 (Dezember 2001)
Die Einkommensteuerreform 1990 als natürliches Experiment. Methodische und konzeptionelle Aspekte zur Schätzung der Elastizität des zu versteuernden Einkommens
Peter Gottfried / Hannes Schellhorn
- Nr. 3 (Januar 2001)
Gut betreut in den Arbeitsmarkt? Eine mikroökonomische Evaluation der Mannheimer Arbeitsvermittlungsagentur
Jürgen Jerger / Christian Pohnke / Alexander Spermann
- Nr. 4 (Dezember 2001)
Das IAW-Einkommenspanel und das Mikrosimulationsmodell SIMST
Peter Gottfried / Hannes Schellhorn
- Nr. 5 (April 2002)
A Microeconomic Characterisation of Household Consumption Using Quantile Regression
Niels Schulze / Gerd Ronning
- Nr. 6 (April 2002)
Determinanten des Überlebens von Neugründungen in der baden-württembergischen Industrie – eine empirische Survivalanalyse mit amtlichen Betriebsdaten
Harald Strotmann
- Nr. 7 (November 2002)
Die Baulandausweisungsumlage als ökonomisches Steuerungsinstrument einer nachhaltigkeitsorientierten Flächenpolitik
Raimund Krumm
- Nr. 8 (März 2003)
Making Work Pay: U.S. American Models for a German Context?
Laura Chadwick, Jürgen Volkert
- Nr. 9 (Juni 2003)
Erste Ergebnisse von vergleichenden Untersuchungen mit anonymisierten und nicht anonymisierten Einzeldaten am Beispiel der Kostenstrukturerhebung und der Umsatzsteuerstatistik
Martin Rosemann
- Nr. 10 (August 2003)
Randomized Response and the Binary Probit Model
Gerd Ronning

IAW-Diskussionspapiere

- Nr. 11 (August 2003)
Creating Firms for a New Century: Determinants of Firm Creation around 1900
Joerg Baten
- Nr. 12 (September 2003)
Das fiskalische BLAU-Konzept zur Begrenzung des Siedlungsflächenwachstums
Raimund Krumm
- Nr. 13 (Dezember 2003)
Generelle Nichtdiskontierung als Bedingung für eine nachhaltige Entwicklung?
Stefan Bayer
- Nr. 14 (Februar 2003)
Die Elastizität des zu versteuernden Einkommens. Messung und erste Ergebnisse zur empirischen Evidenz für die Bundesrepublik Deutschland.
Peter Gottfried / Hannes Schellhorn
- Nr. 15 (Februar 2004)
Empirical Evidence on the Effects of Marginal Tax Rates on Income – The German Case
Peter Gottfried / Hannes Schellhorn
- Nr. 16 (Juli 2004)
Shadow Economies around the World: What do we really know?
Friedrich Schneider
- Nr. 17 (August 2004)
Firm Foundations in the Knowledge Intensive Business Service Sector. Results from a Comparative Empirical Study in Three German Regions
Andreas Koch / Thomas Stahlecker
- Nr. 18 (Januar 2005)
The impact of functional integration and spatial proximity on the post-entry performance of knowledge intensive business service firms
Andreas Koch / Harald Strotmann
- Nr. 19 (März 2005)
Legislative Malapportionment and the Politicization of Germany's Intergovernmental Transfer System
Hans Pitlik / Friedrich Schneider / Harald Strotmann
- Nr. 20 (April 2005)
Implementation ökonomischer Steuerungsansätze in die Raumplanung
Raimund Krumm
- Nr. 21 (Juli 2005)
Determinants of Innovative Activity in Newly Founded Knowledge Intensive Business Service Firms
Andreas Koch / Harald Strotmann
- Nr. 22 (Dezember 2005)
Impact of Opening Clauses on Bargained Wages
Wolf Dieter Heinbach

Nr. 23 (Januar 2006)

Hat die Einführung von Gewinnbeteiligungsmodellen kurzfristige positive Produktivitätswirkungen? – Ergebnisse eines Propensity-Score-Matching-Ansatzes

Harald Strotmann

Nr. 24 (März 2006)

Who Goes East? The Impact of Enlargement on the Pattern of German FDI

Claudia M. Buch / Jörn Kleinert

Nr. 25 (Mai 2006)

Estimation of the Probit Model from Anonymized Micro Data

Gerd Ronning / Martin Rosemann

Nr. 26 (Oktober 2006)

Bargained Wages in Decentralized Wage-Setting Regimes

Wolf Dieter Heinbach

Nr. 27 (Januar 2007)

A Capability Approach for Official German Poverty and Wealth Reports: Conceptual Background and First Empirical Results

Christian Arndt / Jürgen Volkert

Nr. 28 (Februar 2007)

Typisierung der Tarifvertragslandschaft – Eine Clusteranalyse der tarifvertraglichen Öffnungsklauseln

Wolf Dieter Heinbach / Stefanie Schröpfer

Nr. 29 (März 2007)

International Bank Portfolios: Short- and Long-Run Responses to the Business Cycles

Sven Blank / Claudia M. Buch

Nr. 30 (April 2007)

Stochastische Überlagerungen mit Hilfe der Mischungsverteilung

Gerd Ronning

Nr. 31 (Mai 2007)

Openness and Growth: The Long Shadow of the Berlin Wall

Claudia M. Buch / Farid Toubal

Nr. 32 (Mai 2007)

International Banking and the Allocation of Risk

Claudia M. Buch / Gayle DeLong / Katja Neugebauer

Nr. 33 (Juli 2007)

Multinational Firms and New Protectionisms

Claudia M. Buch / Jörn Kleinert

IAW-Diskussionspapiere

- Nr. 34 (November 2007)
Within-Schätzung bei anonymisierten Paneldaten
Elena Biewen
- Nr. 35 (Dezember 2007)
What a Difference Trade Makes – Export Activity and the Flexibility of Collective Bargaining Agreements
Wolf Dieter Heinbach / Stefanie Schröpfer
- Nr. 36 (Dezember 2007)
To Bind or Not to Bind Collectively? Decomposition of Bargained Wage Differences Using Counterfactual Distributions
Wolf Dieter Heinbach / Markus Spindler
- Nr. 37 (Dezember 2007)
Neue Ansätze zur flächenschutzpolitischen Reform des Kommunalen Finanzausgleichs
Raimund Krumm
- Nr. 38 (Januar 2008)
Banking Globalization: International Consolidation and Mergers in Banking
Claudia M. Buch / Gayle L. DeLong
- Nr. 39 (Januar 2008)
Multiplicative Measurement Error and the Simulation Extrapolation Method
Elena Biewen / Sandra Nolte / Martin Rosemann