Table 1. Parameter Estimates and Their Standard Errors by Four Estimation Methods (OLS: ordinary least squares; EV1: $\sigma_{xi}^2 = \sigma_{zi}^2 = \sigma^2$; EV2: $\sigma_{xi}^2 = \sigma_x^2$ and $\sigma_{zi}^2 = \sigma_z^2$ but $\sigma_x^2 \neq \sigma_z^2$; and NEV: all error variances vary between measurements)

	$\hat{\beta}_0$	s.e. $(\hat{\beta}_0)$	$\hat{\beta}_1$	s.e. $(\hat{\beta}_1)$
OLS	−6.189	.673	.277	.047
EV1	−6.399	1.393	.298	.098
EV2	−6.281	1.633	.287	.114
NEV	−10.619	.380	.864	.150

error variances in the Draize test have no relevance to biases. However, the biases of EV2 are small when error variances in the CAM-TB are large.

Third, we examined how differences in the numbers of chemical substances or the number of eggs influenced biases. The comparisons between the same numbers of chemical substances and different numbers of eggs (the comparisons between top and bottom in Figure 3) show that on all of four methods, biases when the number of eggs = 5 are larger than those for when the number of eggs = 10. On the other hand, the comparisons between the different numbers of chemical substances and same numbers of eggs (the comparisons between left and right in Figure 3) show that the magnitude of bias does not change with changes in the number of chemical substances. These results imply that biases become smaller if the number of eggs used in the CAM-TB increases but not smaller even if the number of chemical substances increases.

In Figure 4, we can see that MSEs also have features similar to biases. However, NEV produces large MSEs in the case in which the number of chemical substances is 36 and the number of eggs is 5.

5. APPLICATION OF THE PROPOSED METHOD TO REAL DATA

We applied the proposed method and the other 3 methods to the data described in Section 2.

Table 1 summarizes the regression parameters and their standard errors estimated by the four methods described in Section 3. Figure 5 shows their fitted curves. $\hat{\beta}_0$ and $\hat{\beta}_1$ by OLS, EV1, and EV2 (solid line, large broken line, and small broken line in Figure 5) provided almost identical results. Those provided by NEV (solid thin line in Figure 5) were considerably differed from the other three methods. As Figure 5 indicates, the four methods do not fit the data well.

Our objective is to formulate a model to predict the score in the Draize test by the measurement in the CAM-TB. Accurate prediction of Draize test scores is much more important when CAM-TB produces moderate measurements. Large CAM-TB measurements indicate high Draize test scores, even though the prediction is not very precise. Though the prediction is loose, it is possible to identify chemical substances that have strong toxicity. A chemical substance having strong toxicity cannot be adopted for use in cosmetics. When the CAM-TB test produces moderate measurements, however, we should be careful to avoid underestimating the toxicity of the chemical substance being tested. This is what happens

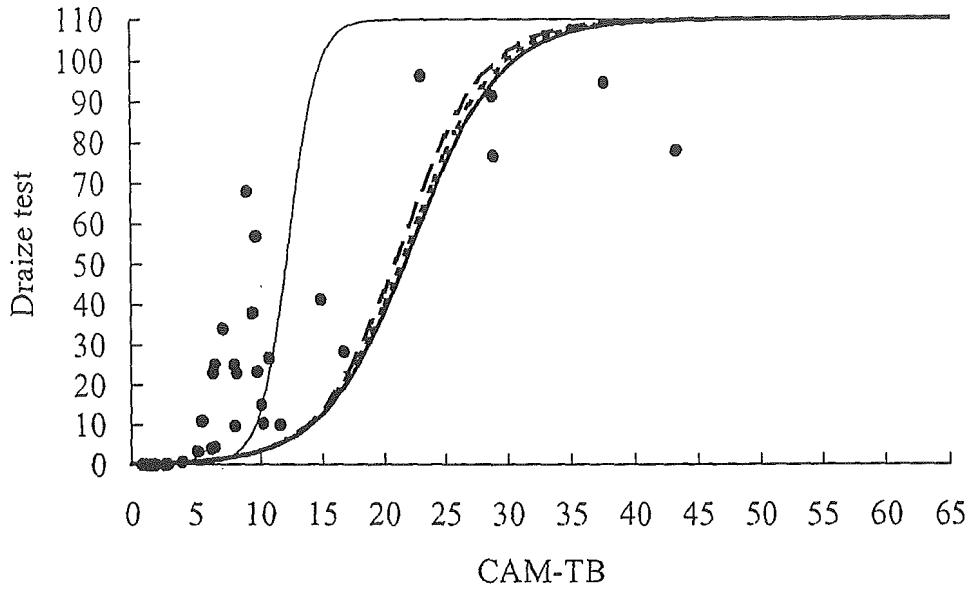

Figure 5. Fitted curves (solid line: OLS, large broken line: EV1, small broken line: EV2, and solid thin line: NEV).

when the score in the Draize test is underestimated.

Figure 5 shows that OLS, EV1, and EV2 underestimate the score in the Draize test when the CAM-TB test produces moderate measurements. Figure 5 also shows that NEV makes the degree of underestimation smaller in the moderate measurements in comparison with the other three methods. NEV is a worse fit than the other three methods for large measurements. As discussed earlier, however, it is not important that the fit for these measurements is wrong. Consequently, we consider that the proposed method, NEV, results in a better prediction formula than the other three methods.

6. DISCUSSION

In this article, we extended the estimating function method proposed by Amari and Kawanabe (1997) to the case in which error variances vary between measurements. The estimators for the proposed method were incidentally identical to the maximum likelihood estimators.

We also conducted a simulation study to compare the regression parameters estimated by the proposed method with those estimated by three other methods under the assumption of homogeneous error variances. The parameters estimated by the proposed method, NEV, may be imprecise when sample size is small and error variances are large. However, the results of the simulation study suggest that, in comparison to the other three methods, NEV results in smaller biases and MSEs. As shown in Figure 4, NEV produced large MSE in the case in which the number of chemical substances was 36 and the number of eggs was 5. The reason is that there are few cases in which NEV takes a very large value of $\hat{\beta}_1$ when the error variances are large. For example, under $(\mathrm{SD}_x, \mathrm{SD}_y) = (\xi, .005\eta(110 - \eta))$, NEV took 16.93 as the maximum value of $\hat{\beta}_1$.

In the simulation study, normal random numbers were generated on the original scales before the logit transformation, while the scales after the logit transformation were modeled. Our primary objective was to see if the proposed method could have worked in the validation study. In the validation study, measurement errors occurred at the original scale. This was because we considered normal error structures at the original scale in the simulation study. If we knew the true error structures, it would be appropriate to use a nonlinear measurement error model. Note that to develop such a nonlinear measurement error model is troublesome with heterogeneous error variances. As a result, in the simulation study, we used somewhat misspecified models. Nevertheless, the simulation study showed that the proposed method, NEV, gave the smaller biases and MSEs than the other three methods.

The goal of the validation study was to predict Draize test scores from CAM-TB values. When a fitted curve shows greater deviation in the direction of underestimating the score in the Draize test, there is a greater risk of underestimating a chemical substance's toxicity. From the results of the simulation study shown in Section 3, it is reasonable to conclude that the proposed method is less biased than the other three methods in the validation study. The results of the simulation study also indicate that we can make bias and MSE smaller if larger numbers of eggs are used in the CAM-TB. This shows that if alternative methods such as the CAM-TB are carried out with large numbers of experiment materials for each chemical substance, we can estimate the regression parameters more precisely.

In validation studies, both the measurements in animal experiments and in alternative methods are measured with error. We need to adopt measurement error models when we predict measurements in animal experiments by alternative methods. In order to investigate whether or not the method proposed in this paper is useful in evaluating alternative methods, we need to examine other alternative methods in addition to CAM-TB.

APPENDIX

A.1 THE DERIVATION OF ASYMPTOTIC VARIANCE (3.3)

If the estimating equation for $\hat{\beta}_j (j = 0, 1)$, $\sum_{i=1}^{n} g_j(x_i, z_i, \hat{\beta}_j, \beta_k) = 0\,(j \neq k)$, is expanded with a Taylor series on the true parameter β_j, we get

$$0 = \sum_{i=1}^{n} g\left(x_i, z_i, \beta\right) + \sum_{i=1}^{n} \frac{\partial g(x_i, z_i, \beta)}{\partial \beta_j} \left(\hat{\beta}_j - \beta_j\right) + O_p\left(\left|\hat{\beta}_j - \beta_j\right|^2\right). \qquad \text{(A.1)}$$

$\hat{\beta}_j$ takes the almost same value as β_j when n becomes very large. Then, since the higher order terms, $O_p\left(\left|\hat{\beta}_j - \beta_j\right|^2\right)$, are negligible, Equation (A.1) is transformed as

$$\sqrt{n}\left(\hat{\beta}_j - \beta_j\right) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(x_i, z_i, \beta\right) \Bigg/ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g\left(x_i, z_i, \beta\right)}{\partial \beta_j}. \qquad \text{(A.2)}$$

By applying the law of large numbers, the denominator on the right hand side in Equation (A.2) converges to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E_{\beta,\xi_i}\left[\frac{\partial g_j(x_i,z_i,\beta)}{\partial\beta_j}\right].$$

By the central limit theorem, the numerator on the right hand side in Equation (A.2) is distributed according to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_j(x_i,z_i,\beta)\sim N\left(0,E_{\beta,\xi_i}\left[g_j(x_i,z_i,\beta)^2\right]\right).$$

Therefore, $\hat{\beta}_j$ is also asymptotically normally distributed with the asymptotic variance

$$\text{a.v.}\left[\sqrt{n}\left(\hat{\beta}_j-\beta_j\right)\right]=\frac{\lim_{n\to\infty}(1/n)\sum_{i=1}^{n}E_{\beta,\xi_i}\left[g_j(x_i,z_i,\beta)^2\right]}{\left\{\lim_{n\to\infty}(1/n)\sum_{i=1}^{n}E_{\beta,\xi_i}\left[\partial g_j(x_i,z_i,\beta)/\partial\beta_j\right]\right\}^2}.$$

A.2 THE DERIVATIONS OF EQUATIONS (3.4) AND (3.7) THROUGH THE MAXIMUM LIKELIHOOD METHOD

A pair of measurements, (x_i,z_i), $i=1,2,\ldots,n$, have a functional relationship

$$\begin{aligned}x_i &= \xi_i+\delta_i\\ z_i &= \beta_0+\beta_1\xi_i+\varepsilon_i,\end{aligned}$$

where δ_i and ε_i are distributed according to

$$\begin{pmatrix}\delta_i\\ \varepsilon_i\end{pmatrix}\sim N\left(\begin{pmatrix}0\\ 0\end{pmatrix},\begin{pmatrix}\sigma_{xi}^2 & 0\\ 0 & \sigma_{zi}^2\end{pmatrix}\right).$$

The log-likelihood function, $\log L$, is given as

$$\log L=C-\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(x_i-\xi_i)^2}{\sigma_{xi}^2}+\frac{(z_i-\beta_0-\beta_1\xi_i)^2}{\sigma_{zi}^2}\right\},$$

where C denotes a set of constant terms that does not depend on the parameters β_0,β_1, and ξ_i. $\partial\log L/\partial\xi_i=0$ gives the maximum likelihood estimator

$$\hat{\xi}_i=\frac{\sigma_{zi}^2x_i+\hat{\beta}_1\sigma_{xi}^2(z_i-\hat{\beta}_0)}{\sigma_{zi}^2+\hat{\beta}_1^2\sigma_{xi}^2}.$$

By substituting this $\hat{\xi}_i$ into $\partial\log L/\partial\beta_0=0$ and $\partial\log L/\partial\beta_1=0$, Equations (3.4) and (3.7) are derived.

ACKNOWLEDGMENTS

The authors thank Takashi Omori and the referees for their helpful comments.

[Received February 2004. Revised May 2004.]

REFERENCES

Amari, S., and Kawanabe, M. (1997), "Information Geometry of Estimating Functions in Semi-Parametric Statistical Models," *Bernoulli*, 3, 29–54.

Draize, J., Woodard, G., and Galvery, H. (1944), "Methods for the Study of Irritation and Toxicity of Substances Applied Topically to the Skin and Mucous Membranes," *The Journal of Pharmacology and Experimental Therapeutics*, 82, 377–390.

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

Hagino, S., Itagaki, H., Kato, S., and Kobayashi, T. (1993), "Further Evaluation of the Quantitative Chorioallantoic Membrane Test Using Trypan Blue Stain to Predict the Eye Irritancy of Chemicals," *Toxicology in Vitro*, 7, 35–39.

Hagino, S., Itagaki, H., Kato, S., Kobayashi, T., and Tanaka, M. (1991), "Quantitative Evaluation to Predict the Eye Irritancy of Chemicals Modification of Chorioallantoic Membrane Test by Using Trypan Blue," *Toxicology in Vitro*, 5, 301–304.

Hagino, S., Kinoshita, S., Tani, N., Nakamura, T., Ono, N., Konishi, K., Iimura, H., Kojima, H., and Ohno, Y. (1999), "Interlaboratory Validation of In Vitro Eye Irritation Tests for Cosmetic Ingredients (2) Chorioallantoic Membrane (CAM) Test," *Toxicology in Vitro*, 13, 99–113.

Japan Cosmetic Industry Association (1994), *Report of the First Validation Study on Alternative Methods to the Draize Eye Irritation Test*, Japan Cosmetic Industry Association (in Japanese).

Neyman, J., and Scott, E. L. (1948), "Consistent Estimates Based on Partly Consistent Observation," *Econometrica*, 32, 1–32.

OECD Test Guidelines Programme (1996), "Options for a Testing Strategy for the Testing of Skin and Eye Irritancy," Final report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, September 18–19, 1996, the Chateau de la Muette, Paris.

Walter, S. D. (1997), "Variation in Baseline Risk as an Explanation of Heterogeneity in Meta-analysis," *Statistics in Medicine*, 16, 2883–2900.

Wilhelmus, K. R. (2001), "The Draize Eye Test," *Survey of Ophthalmology*, 45, 493–515.