4. A simulation study

We evaluated the performance of different model-selecting CRMs, i.e. the CRM based on the PMP (CRM-PMP), the CRM based on the PPL (CRM-PPL), and the CRM based on the DIC (CRM-DIC), through a simulation study with nine scenarios representing the underlying dose–toxicity relationships used in Section 3 of Yin and Yuan [15]. Two additional scenarios were studied, in which either a flat dose–toxicity curve up to the targeted MTD or a steep dose–toxicity curve were considered.

Under each scenario eight dose levels were chosen, and four sets of initial working models shown in the following:

$$\left\{\begin{array}{l}(0.02, 0.06, 0.08, 0.12, 0.20, 0.30, 0.40, 0.50),\ \text{working model 1},\\(0.01, 0.05, 0.09, 0.14, 0.18, 0.22, 0.26, 0.30),\ \text{working model 2},\\(0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80),\ \text{working model 3},\\(0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75),\ \text{working model 4}\end{array}\right\}.$$

We refer to the individual CRMs using each of these toxicity probabilities, respectively, as CRM 1, CRM 2, CRM 3 and CRM 4. We set not only these CRMs but also the CRM based on the Bayesian model averaging (CRM-BMA), proposed by Yin and Yuan [15], as controls for comparisons with our proposed model-selecting CRMs. In the CRM-BMA, the prior working model probability for each of CRM 1–4 was assigned $\frac{1}{4}$, which was the same as that of Yin and Yuan [15]. The number of simulated trials was 1000. In each trial the first patient was allocated to d_1, and the maximum sample size with a stopping rule was 30. For the model-selecting CRMs and the CRM-BMA, the stopping rules were based on equations (5) and (6), respectively. For CRMs 1–4, equation (5) was used for each pre-specified working model. The performance was evaluated through the percentage of correct selection (PCS), i.e. the percentage of the recommended dose levels at the end of the trial, associated with the true MTD. In addition we estimated the mean-squared error (MSE) of the estimated toxicity probability to the true at the estimated MTD as

$$\sum_{N=1}^{1000}\frac{\{(\text{The estimated toxicity probability})_N-(\text{The true toxicity probability})\}^2}{1000}.$$

The target toxicity probability was $\theta^*=0.3$. In Tables I and II, the scenarios, the PCS, the average number of DLTs, and the average sample size are given.

Scenario 1 had the MTD at d_7. The CRM-DIC as well as CRM 3 had a PCS over 70 per cent, and all the other designs had PCSs >53 per cent. In Scenario 2, CRM 3 behaved the best, whereas CRM 2 the least well. Our proposed model-selecting CRMs and the CRM-BMA lie between the two designs. In particular the CRM-DIC and the CRM-BMA showed PCSs about 45.6 per cent, with almost 30 per cent of patients treated at the MTD (shown in the supplementary material).[‡] In Scenario 3, all designs except CRM 1 had high PCSs (>62 per cent). In Scenarios 4 and 5, all designs except CRM 1 had close PCSs. If we had to compare among the model-selecting CRMs and the CRM-BMA, the CRM-PPL behaved the best overall. CRM 2 performed the least well in Scenario 6, associated with a PCS almost 10 per cent lower than the other designs. In Scenario 7, the CRM-BMA had a relatively higher PCS, compared with the model-selecting CRMs. In Scenario 8, the CRM-PMP and the CRM-BMA had a relatively higher PCS, but the CRM-PPL and the CRM-DIC gave lower PCSs compared with other designs. In Scenario 9, in which d_1 was associated with a toxicity probability higher than the target, showed that the used stopping rules could stop the trial earlier with no dose recommendation in over 55 per cent of trials. Using a flat dose–toxicity relationship as in Scenario 10 showed that all designs except CRM-DIC overestimated or underestimated the MTD. In Scenario 11, in which the underlying dose–toxicity curve had a steep slope, the CRM-DIC as well as CRM 4 yielded a PCS >70 per cent. For all the scenarios, the average number of observed toxicities was similar within each scenario.

The more spaced out the dose levels, in terms of coding, the easier for the method to distinguish one level from another. In practice our suggestion would be to roughly divide the interval (0,1) into $I+1$

[‡]*Supporting information may be found in the online version of this article.*

Table I. The percentage of correct selection (PCS), the average number of DLTs and the average sample size under Scenarios 1–6. The scenario and the PCS are given in italic and in bold, respectively.

	Percentage of correct selection										
Design	1	2	3	4	5	6	7	8	None	Ave. # tox.	Ave. # pats
Scenario 1	*0.02*	*0.03*	*0.04*	*0.06*	*0.08*	*0.1*	**0.3**	*0.5*			
CRM 1	0.0	0.0	0.0	0.0	0.4	14.3	**67.6**	17.7	0	8.6	30
CRM 2	0.0	0.0	0.0	0.0	2.0	20.7	**53.5**	23.8	0	8.8	30
CRM 3	0.0	0.0	0.0	0.3	1.6	18.7	**73.4**	6.0	0	6.1	30
CRM 4	0.0	0.0	0.0	0.2	1.6	19.6	**64.3**	14.3	0	6.4	30
CRM-PMP	0.0	0.0	0.0	0.3	1.0	15.9	**65.7**	17.1	0	8.4	30
CRM-PPL	0.0	0.0	0.0	0.2	1.5	18.1	**67.9**	12.3	0	6.2	30
CRM-DIC	0.0	0.0	0.0	0.2	0.6	20.4	**71.1**	7.7	0	7.4	30
CRM-BMA	0.0	0.0	0.0	0.1	1.0	17.0	**66.0**	15.9	0	8.0	30
Scenario 2	*0.02*	*0.06*	*0.08*	*0.12*	*0.2*	***0.3***	*0.4*	*0.5*			
CRM 1	0.0	0.0	0.1	1.4	23.0	**45.2**	25.1	5.2	0	9.3	30
CRM 2	0.0	0.0	0.5	6.1	24.5	**36.8**	23.1	9.0	0	9.2	30
CRM 3	0.0	0.0	0.5	4.1	30.4	**50.6**	14.0	0.4	0	7.3	30
CRM 4	0.0	0.0	0.5	5.0	34.1	**40.7**	17.3	2.4	0	7.0	30
CRM-PMP	0.0	0.1	0.6	4.3	20.5	**42.8**	26.7	5.0	0	9.0	30
CRM-PPL	0.0	0.0	0.5	4.7	31.6	**43.2**	18.2	1.8	0	7.0	30
CRM-DIC	0.0	0.0	0.2	5.2	26.6	**45.6**	19.3	3.1	0	8.4	30
CRM-BMA	0.0	0.0	0.2	3.2	24.8	**45.6**	22.3	3.9	0	8.5	30
Scenario 3	*0.06*	*0.15*	***0.3***	*0.55*	*0.6*	*0.65*	*0.68*	*0.7*			
CRM 1	0.8	26.7	**51.6**	18.8	1.5	0.2	0.0	0.0	0.4	10.8	29.9
CRM 2	0.1	24.1	**62.4**	12.4	0.4	0.1	0.0	0.0	0.5	10.2	29.9
CRM 3	0.3	19.2	**68.2**	11.4	0.4	0.1	0.0	0.0	0.4	9.4	29.9
CRM 4	0.3	19.2	**69.4**	10.2	0.4	0.0	0.0	0.0	0.5	9.0	29.9
CRM-PMP	0.4	20.7	**63.8**	13.5	1.0	0.2	0.0	0.0	0.4	10.2	29.9
CRM-PPL	0.4	19.9	**69.3**	10.0	0.0	0.0	0.0	0.0	0.4	9.0	29.9
CRM-DIC	0.1	21.2	**65.7**	12.1	0.4	0.1	0.0	0.0	0.4	10.1	29.9
CRM-BMA	0.5	22.1	**64.2**	12.2	0.5	0.0	0.1	0.0	0.4	10.0	29.9
Scenario 4	*0.2*	***0.3***	*0.4*	*0.5*	*0.6*	*0.65*	*0.7*	*0.75*			
CRM 1	20.3	**40.9**	21.4	8.6	1.2	0.1	0.0	0.0	7.5	10.4	28.2
CRM 2	13.4	**49.1**	23.5	4.1	0.7	0.0	0.0	0.0	9.2	9.8	28.1
CRM 3	15.5	**45.7**	26.5	5.1	0.3	0.0	0.0	0.0	6.9	9.5	28.3
CRM 4	16.8	**45.7**	23.5	4.5	0.3	0.1	0.0	0.0	9.1	8.9	27.8
CRM-PMP	15.4	**45.1**	25.2	4.8	0.0	0.0	0.0	0.0	9.5	10.0	27.6
CRM-PPL	16.7	**46.8**	23.7	4.6	0.5	0.0	0.0	0.0	7.7	9.4	28.3
CRM-DIC	16.0	**45.1**	25.3	4.3	0.7	0.1	0.0	0.0	8.5	10.1	28.1
CRM-BMA	18.3	**45.3**	23.4	4.3	0.7	0.0	0.0	0.0	8.0	10.1	28.1
Scenario 5	*0.1*	*0.2*	***0.3***	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*			
CRM 1	1.3	27.7	**31.8**	30.5	7.1	0.4	0.1	0.0	1.1	10.6	29.7
CRM 2	1.2	27.1	**41.4**	23.1	4.9	0.8	0.0	0.0	1.5	9.9	29.6
CRM 3	1.0	23.0	**45.9**	25.0	3.6	0.3	0.0	0.0	1.2	9.4	29.7
CRM 4	1.2	22.4	**48.5**	24.0	2.2	0.1	0.0	0.0	1.6	8.9	29.6
CRM-PMP	1.2	23.7	**46.3**	23.0	4.0	0.5	0.0	0.0	1.3	10.0	29.7
CRM-PPL	1.3	24.6	**46.7**	21.8	3.8	0.2	0.0	0.0	1.6	8.9	29.6
CRM-DIC	1.4	22.1	**46.8**	24.6	3.8	0.0	0.0	0.0	1.3	9.8	29.7
CRM-BMA	0.9	23.5	**43.8**	24.3	6.0	0.3	0.0	0.0	1.2	9.8	29.7
Scenario 6	*0.02*	*0.03*	*0.05*	*0.07*	***0.3***	*0.5*	*0.7*	*0.8*			
CRM 1	0.0	0.0	0.0	11.3	**70.6**	17.8	0.3	0.0	0	9.7	30
CRM 2	0.0	0.0	0.1	20.5	**57.1**	20.5	1.7	0.1	0	9.4	30
CRM 3	0.0	0.0	0.0	11.1	**73.2**	15.7	0.0	0.0	0	8.4	30
CRM 4	0.0	0.0	0.0	11.5	**71.1**	16.9	0.5	0.0	0	7.9	30
CRM-PMP	0.0	0.0	0.0	11.8	**69.7**	18.2	0.3	0.0	0	9.4	30
CRM-PPL	0.0	0.0	0.0	11.9	**70.6**	17.0	0.5	0.0	0	7.9	30
CRM-DIC	0.0	0.0	0.0	12.8	**70.6**	16.5	0.0	0.1	0	8.9	30
CRM-BMA	0.0	0.0	0.0	13.7	**68.8**	17.4	0.1	0.0	0	9.0	30

Table II. The percentage of correct selection (PCS), the average number of DLTs and the average sample size under Scenarios 1 to 6. The scenario and the PCS are given in italic and in bold, respectively.

	Percentage of correct selection										
Design	1	2	3	4	5	6	7	8	None	Ave. # tox.	Ave. # pats
Scenario 7	*0.03*	*0.07*	*0.1*	*0.15*	*0.2*	***0.3***	*0.5*	*0.7*			
CRM 1	0.0	0.0	0.2	3.7	28.4	**54.9**	12.6	0.2	0	9.2	30
CRM 2	0.0	0.0	1.0	9.9	30.8	**41.5**	15.2	1.6	0	9.1	30
CRM 3	0.0	0.0	1.0	6.7	30.5	**54.0**	7.8	0.0	0	7.3	30
CRM 4	0.0	0.0	0.9	9.6	35.5	**42.6**	11.3	0.1	0	7.1	30
CRM-PMP	0.0	0.0	0.5	6.9	27.8	**52.5**	12.0	0.3	0	8.8	30
CRM-PPL	0.0	0.0	1.1	8.9	32.8	**45.2**	11.8	0.2	0	7.1	30
CRM-DIC	0.0	0.0	0.7	8.3	32.3	**50.6**	8.1	0.0	0	8.1	30
CRM-BMA	0.0	0.0	0.4	6.1	27.9	**53.3**	12.2	0.1	0	8.5	30
Scenario 8	*0.02*	*0.03*	*0.05*	*0.06*	*0.07*	*0.09*	*0.1*	***0.3***			
CRM 1	0.0	0.0	0.0	0.0	0.0	0.8	13.3	**85.9**	0	7.1	30
CRM 2	0.0	0.0	0.0	0.0	0.1	1.3	12.6	**86.0**	0	7.3	30
CRM 3	0.0	0.0	0.0	0.3	1.5	4.9	31.3	**62.0**	0	4.4	30
CRM 4	0.0	0.0	0.0	0.3	1.2	3.2	24.4	**70.9**	0	5.0	30
CRM-PMP	0.0	0.0	0.0	0.1	0.7	1.6	12.7	**84.9**	0	7.0	30
CRM-PPL	0.0	0.0	0.0	0.3	1.1	3.2	25.9	**69.5**	0	4.7	30
CRM-DIC	0.0	0.0	0.0	0.2	0.5	2.7	27.1	**69.5**	0	6.3	30
CRM-BMA	0.0	0.0	0.0	0.1	0.7	1.1	14.8	**83.3**	0	6.7	30
Scenario 9	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*0.95*	*0.99*			
CRM 1	35.7	4.7	0.2	0.1	0.0	0.0	0.0	0.0	59.3	8.6	18.1
CRM 2	19.5	6.5	0.2	0.0	0.0	0.0	0.0	0.0	73.8	8.0	16.5
CRM 3	37.2	7.2	0.4	0.0	0.0	0.0	0.0	0.0	55.2	8.5	18.7
CRM 4	30.4	6.0	0.2	0.0	0.0	0.0	0.0	0.0	63.4	7.6	16.9
CRM-PMP	29.1	8.0	0.4	0.0	0.0	0.0	0.0	0.0	62.5	8.3	17.0
CRM-PPL	31.2	5.2	0.2	0.0	0.0	0.0	0.0	0.0	63.4	8.2	18.2
CRM-DIC	28.4	8.1	0.6	0.0	0.0	0.0	0.0	0.0	62.9	8.4	18.1
CRM-BMA	32.2	6.3	0.0	0.1	0.0	0.0	0.0	0.0	61.4	8.4	18.3
Scenario 10	*0.2*	*0.21*	*0.22*	*0.23*	*0.24*	*0.25*	***0.3***	*0.35*			
CRM 1	1.8	2.8	2.7	5.1	12.4	16.8	**25.5**	27.9	5.0	8.1	28.7
CRM 2	1.9	4.5	5.4	9.0	10.9	11.1	**18.4**	33.5	5.3	8.2	28.7
CRM 3	2.8	6.4	8.1	14.1	16.3	24.7	**18.2**	4.5	4.9	6.9	28.7
CRM 4	2.3	7.9	12.5	19.1	16.3	14.4	**13.9**	6.7	6.9	6.7	28.3
CRM-PMP	1.9	4.7	5.7	8.0	9.2	14.5	**22.1**	28.1	5.8	7.9	28.6
CRM-PPL	2.6	8.5	11.4	19.1	16.3	16.3	**13.4**	6.2	6.2	6.8	28.5
CRM-DIC	1.7	5.2	6.1	9.9	10.7	18.4	**22.1**	20.8	5.1	7.8	28.7
CRM-BMA	5.9	4.3	6.4	9.5	12.2	17.9	**20.0**	22.6	1.2	7.6	28.8
Scenario 11	*0.01*	*0.05*	*0.10*	***0.3***	*0.5*	*0.6*	*0.7*	*0.8*			
CRM 1	0.0	0.7	15.1	**63.6**	20.1	0.4	0.1	0.0	0.0	10.4	30
CRM 2	0.0	0.3	18.2	**59.0**	20.0	2.3	0.2	0.0	0.0	9.7	30
CRM 3	0.0	0.0	13.6	**69.8**	16.3	0.3	0.0	0.0	0.0	9.0	30
CRM 4	0.0	0.0	13.1	**70.9**	15.8	0.2	0.0	0.0	0.0	8.5	30
CRM-PMP	0.0	0.0	14.9	**66.6**	17.3	1.2	0.0	0.0	0.0	9.8	30
CRM-PPL	0.0	0.2	14.1	**67.8**	17.5	0.4	0.0	0.0	0.0	8.5	30
CRM-DIC	0.0	0.0	13.9	**70.2**	15.1	0.8	0.0	0.0	0.0	9.5	30
CRM-BMA	0.0	0.1	14.3	**65.1**	20.1	0.4	0.0	0.0	0.0	9.4	30

equal intervals (for the examples here taken from Yin and Yuan [15] this is $8+1=9$). If we assess the performance in terms of the percentage of recommendations, at the MTD, one level below the MTD and one level above the MTD, then there is very little to choose between the different models. They behave much the same. If we only consider the percentage recommendation at the MTD itself, then the working model CRM 2, in scenarios 1, 2 and 6, while performing well, does perform less well than other model choices. For the model CRM 2, the spacing is much tighter than for the other choices and this is most likely the explanation for the differences in performance. Model choice CRM 1 performs less well than the other choices in scenarios 3, 4 and 5, and, again, in these cases, the MTD was located in a region where the model choice bunched up the dose levels. At other scenarios, when the MTD was located in an area corresponding to more spread determined by the model, the

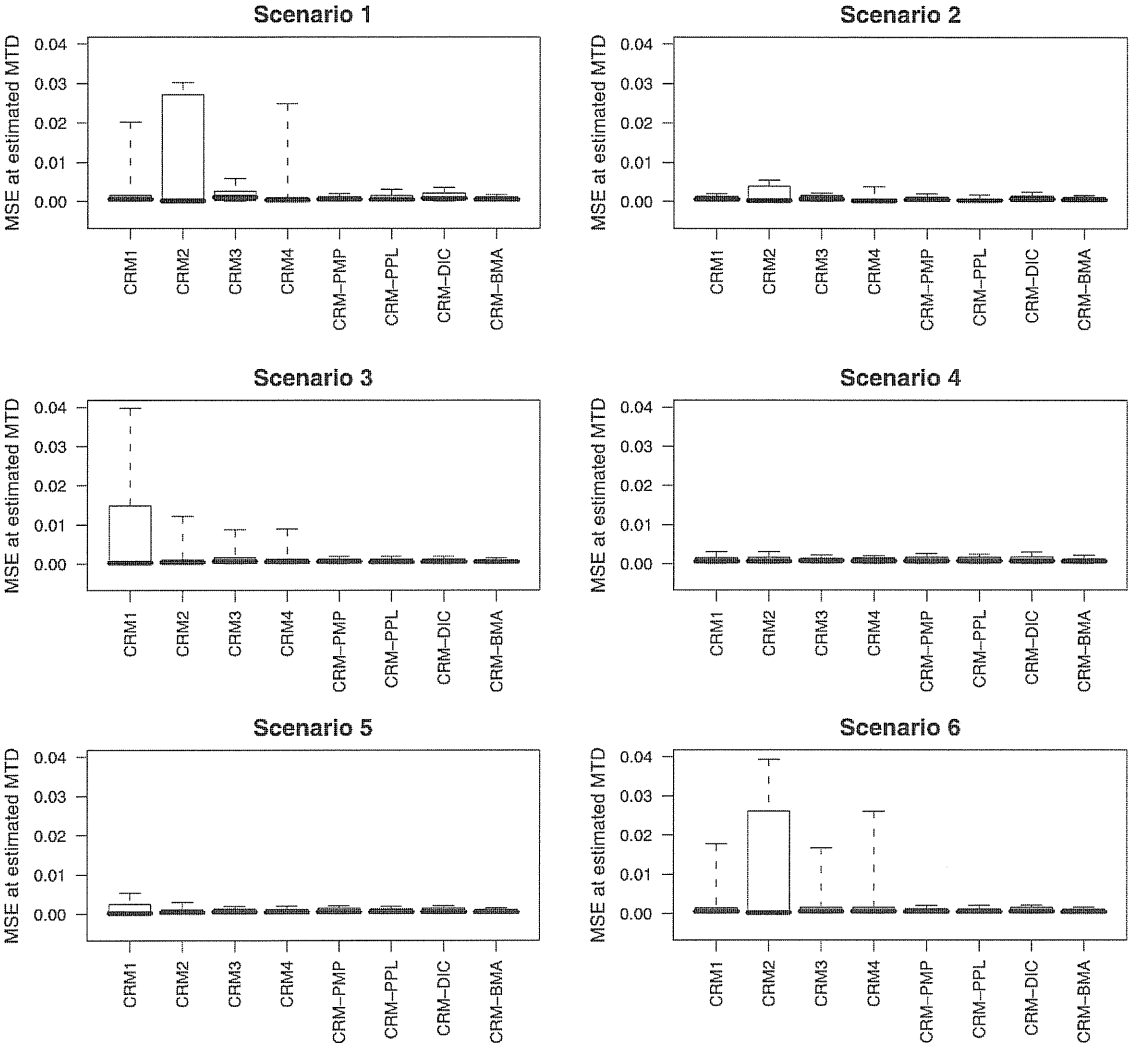

Figure 1. Comparison of the MSEs at the estimated MTD under Scenarios 1–6.

performance was comparable. For model choice CRM 3 and CRM 4, where the levels were spread widely, the performance was in all cases satisfactory. The model-selecting CRMs and the CRM-BMA may help reduce the influence of a less than fortunate model specification. In practice, when we utilize the model-selecting CRMs or the CRM-BMA, one way to identify a less than optimal working model is to plot values of the used criterion, say, PMP, PPL, or DIC, for each pre-specified working model, against accumulating patients or cohorts (see [15]). A great difference in the values exceeding some threshold might give us a reason for being cautious in the use of a particular working model during the course of a trial.

Figures 1 and 2 show the boxplots of the MSEs at the estimated MTD under each scenario (except Scenario 9 with early stopping). Our proposed model-selecting CRMs and the CRM-BMA tended to yield the narrower quartile ranges than the worst-performed CRM.

5. Conclusions and discussions

The results of the above section show that the poor pre-specification of the toxicity probabilities in a single CRM design could have a negative impact on both the MTD estimation and the dose allocation to each patient. The model-selecting CRMs as well as the CRM-BMA may help to reduce the potential for errors following such specification.

We have investigated a model-selecting CRM design, in which we use multiple working models together with model selection, based on the Bayesian model selection criteria, such as the PMP, the

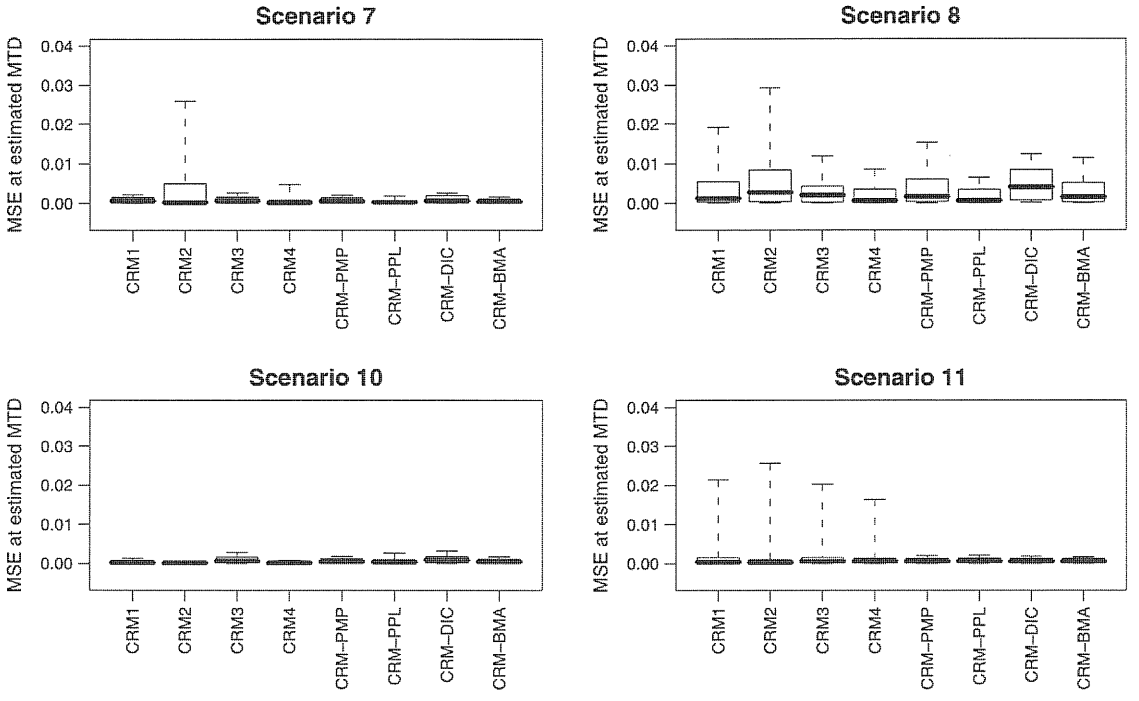

Figure 2. Comparison of the MSEs at the estimated MTD under Scenarios 7–11 (except Scenario 9).

PPL and the DIC. The results of the simulation study show that all of the three proposed designs can provide a compromise solution between the best-performing and worst-performing model. In other words, it may be possible to improve the robustness of the CRM to poor pre-specification of a single working model. The Bayesian model averaging CRM will average several working models into a single one unlike the model-selecting CRMs. From the results of the simulation study, it appears that the former performs better under some scenarios and the latter under others. Roughly, over the class of working models, there are two kinds of simple summaries, the mean and the mode; the Bayesian model averaging CRM being the mean and the model-selecting CRMs being the mode. An immediate question is which one of these summaries is to be preferred. For the cases we have studied here there is no clear cut answer to this and no clear cut winner. Although only speculative, we believe that if it were possible to identify and include all possible dose–toxicity curves, as well as all potential working models, in a comparative study, there would still be no clear cut winner. The comparison is more complex but not fundamentally all that different from a comparison of summary measures for some given unimodal random variable U with density $f(u)$. We could consider $\mathrm{E}(U)$ or $u_m : f(u_m) > f(u), u \in \mathscr{U}$ where $\mathscr{U}$ is the support of U. If the variable is symmetric then the two measures coincide. If positively skewed then $E(U)$ will be more sensitive to larger values of u than u_m. This may be an argument in its favor when we would like the impact of the larger values to be given due weight but an argument against it when these larger values are considered potentially too influential. In that case u_m would be much less sensitive to the larger values. At the time of writing not enough is known about the distribution of the variable of interest—in this case the set of potential working models—to be able to provide firm recommendations. As more becomes known through real studies and accumulated experience, it may be possible in the future to provide sharper indications.

Although it appears too difficult to rank the three proposed model-selecting CRMs, our simulations provide some insight into their different characteristics. For example, when the true dose–toxicity relationship is quite flat up to the targeted MTD (in this situation some existing methods fail in the MTD estimation [20]), the model-selecting CRM based on the DIC works well. The model-selecting CRM based on the PPL appears to provide a conservative dose escalation, resulting in the number of patients allocated to the higher more toxic dose levels being relatively smaller than with the other model-selecting techniques. This may be a result of the PPL being based on the posterior predictive distribution allowing for larger variability than the posterior distribution. The model-selecting CRM based on the PMP could be quite similar in performance to the model averaging CRM because both use the PMP.

An interpretation of the working model in terms of Bayesian prior point estimates of the probabilities of toxicity is quite common in this area. Although such an interpretation is not only needed but may not correspond in practice to anything very concrete (since operating characteristics and final recommendations remain invariant to these prior point estimates being raised to any positive power), many clinicians and experimenters are reassured by the interpretation. When this is so, an additional reassurance can be obtained by using several working models as opposed to a single one, in particular in cases when there are differences in opinion among investigators. Rather than force the choice to any one particular working model it is possible to say that all the suggested potential working models can be employed and then, as the observations are accrued, the method itself will help select the most plausible given this information. In reality, this will still end up being equivalent to having worked with a particular simple model choice. However, for those investigators who may be concerned at the apparent arbitrariness of working model selection, and the worry that this may impact the relative performance of the method, the use of several models enables them to gain confidence in the resulting MTD estimate, based as it is on a number of rival models.

Acknowledgements

This research was supported partially by the Health Labour Sciences Research Grant (Research on Measures for Intractable Diseases) of the Ministry of Health, Labour and Welfare, and by Grant-in-Aid for Researchers, Hyogo College of Medicine. We thank the two referees and the editors for their helpful comments that improved the quality of the paper.

References

1. Storer BE. Design and analysis of phase I clinical trials. *Biometrics* 1989; **45**:925–937.
2. Lin Y, Shih WJ. Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. *Biostatistics* 2001; **2**:203–215.
3. Ivanova A, Montazer-Haghighi A, Mohanty SG, Durham SD. Improved up-and-designs for Phase I trials. *Statistics in Medicine* 2003; **22**:69–82.
4. Ivanova A. Escalation up-and-down and $A+B$ designs for dose-finding trials. *Statistics in Medicine* 2003; **25**: 3668–3678.
5. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; **46**:33–48.
6. Whitehead J, Williamson D. Bayesian decision procedures based on logistic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics* 1998; **8**:445–467.
7. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 1998; **17**:1103–1120.
8. Shen LZ, O'Quigley J. Consistency of continual reassessment method under model misspecification. *Biometrika* 1996; **83**:395–405.
9. O'Quigley J. Another look at two phase I clinical trial designs. *Statistics in Medicine* 1999; **18**:2683–2690.
10. Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in Phase I dose-finding studies. *Clinical Trials* 2008; **5**:465–477.
11. O'Quigley J, Zohar S. Retrospective robustness of the continual reassessment method. *Journal of Biopharmaceutical Statistics* 2010; **20**(5):1013–1025.
12. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684.
13. Cheung YK, Chappell R. A simple technique to evaluate model sensitivity in the continual reassessment method. *Biometrics* 2002; **58**:671–674.
14. Lee SM, Cheung YK. Model calibration in the continual reassessment method. *Clinical Trials* 2009; **6**:227–238.
15. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. *Journal of the American Statistical Association* 2009; **104**:954–968.
16. Gelfand A, Ghosh S. Model choice: a minimum posterior predictive loss approach. *Biometrika* 1998; **85**:1–11.
17. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 2002; **64**:583–639.
18. Kass RE, Raftery AE. Bayes factors and model uncertainty. *Journal of the American Statistical Association* 1995; **90**:773–795.
19. Lunn DJ, Thomas A, Best NG, Spiegelhalter DJ. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**:325–337.
20. Cheung YK. Sequential implementation of stepwise procedures for identifying the maximum tolerated dose. *Journal of the American Statistical Association* 2007; **102**:1448–1461.

Communications in Statistics—Theory and Methods, 38: 193–207, 2009

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920802187414

Taylor & Francis
Taylor & Francis Group

Relative Curvature Measure for Heteroscedastic or Non Normal Nonlinear Regression

TAKASHI DAIMON[1], TOSHIHIRO YOSHIKAWA[1], TOMINORI KOBAYASHI[1], AND MASASHI GOTO[2]

[1]Department of Drug Evaluation and Informatics, School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan
[2]Non Profit Organization Biostatistical Research Association, Osaka, Japan

The Bates–Watts relative curvature measure can assess the validity of the linearized approximation in nonlinear regression models. However, it is developed based on an ordinary nonlinear regression in which the observation is assumed to be homoscedastically and normally distributed. In this article, we extend the original Bates–Watts relative curvature measure to one that can be applicable to nonlinear regression with heteroscedastic or non normal data, based on the transformation-both-sides (TBS) approach. In pharmacokinetic models, a diagnostic use of their measures is illustrated. By means of a simulation experiment, the performance of the relative curvature measure for the TBS approach is evaluated.

Keywords Box–Cox transformation; Compartment models; Nonlinearity.

Mathematics Subject Classification Primary 62K05, 65C60, 65D17; Secondary 46N55, 90C30, 94C30.

1. Introduction

The inference of nonlinear regression models such as chemical reactor model, kinetic model, and growth model is generally based on a linearized approximation with respect to the model parameters (e.g., Bates and Watts, 1988; Davidian and Giltinan, 1995; Seber and Wild, 1989). One of the useful tools for diagnosing the validity of the linearized approximation or assessing the nonlinearity underlying the nonlinear regression model is the relative curvature measure developed by Bates and Watts (1980). This measure is constructed on an ordinary nonlinear regression model, where the error is assumed to be independently, identically, and normally distributed. However, in general, the observed data (or the errors) fitted

Received February 15, 2007; Accepted May 7, 2008
Address correspondence to Takashi Daimon, Department of Drug Evaluation and Informatics, School of Pharmaceutical Sciences, University of Shizuoka, 52-1, Yada, Suruga-ku, Shizuoka 422-8526, Japan; E-mail: daimon@u-shizuoka-ken.ac.jp

to the above-mentioned nonlinear regression models exhibit the heteroscedasticity or non normality. For example, in pharmacokinetics, the blood-drug concentration data observed after administration of a drug generally lead to heteroscedasticity, depending on the method or instrument used for the assay (e.g., Davidian and Giltinan, 1995; Oberg and Davidian, 2000). Therefore, the heteroscedasticity or non normality needs to be taken into account to utilize the Bates–Watts relative curvature measure in practice. In this article, the relative curvature measure and associated measures using the Transformation-Both-Sides (TBS) approach (Carroll and Ruppert, 1988; Oberg and Davidian, 2000) have been proposed to assess the nonlinearity underlying the model and to ensure the validity of the inference, without being influenced by the heteroscedasticity or non normality of the data.

2. Relative Curvature Measure for TBS Approach

2.1. *Relative Curvature Measure for TBS Approach*

The TBS approach applied to the inference of the nonlinear regression model is represented as (see Carroll and Ruppert, 1988):

$$Y^{(\lambda)} = f^{(\lambda)}(x; \beta) + \varepsilon, \tag{1}$$

where $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ is the $n \times 1$ random vector of the observations, n is the number of observations, and x is the vector of the explanatory variables with any dimensions. Y_j is observed with jth vector of the explanatory variables x_j. $Y^{(\lambda)} = (Y_1^{(\lambda)}, \ldots, Y_n^{(\lambda)})^{\mathrm{T}}$ is the $n \times 1$ random vector of the Box and Cox (1964)'s power transformation of Y and is represented as

$$Y^{(\lambda)} = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \log Y, & \lambda = 0, \end{cases} \tag{2}$$

where λ is the power-transforming parameter; $f^{(\lambda)}(x; \beta) = (f^{(\lambda)}(x_1; \beta), \ldots, f^{(\lambda)}(x_n; \beta))^{\mathrm{T}}$ is the $n \times 1$ vector of the power-transformed functions of $f(x; \beta) = (f(x_1; \beta), \ldots, f(x_n; \beta))^{\mathrm{T}}$; and f is the function of the nonlinear regression model that is derived from the theory related to specific fields. β is the $p \times 1$ vector of parameters given by $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$. $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ is the $n \times 1$ vector of errors; ε is assumed to be independently, identically, and normally distributed as $\mathrm{N}(0, \sigma^2)$, where σ is the scale parameter.

The ordinary nonlinear regression approach corresponds to setting $\lambda = 1$ in (1). Therefore, in this article, the TBS approach with $\lambda = 1$ is called the non transformation (NT) approach. The variance function approach assumes that $\lambda = 1$ in (1) and that ε is usually distributed as $\mathrm{N}(0, \sigma^2 g\{f(x; \beta), \theta\})$. In this case, $g\{f(x; \beta), \theta\}$ is the variance function, and θ is the variance parameter with any dimensions. For example, $g\{f(x; \beta), \theta\} = \{f(x; \beta)\}^{\theta_1}$.

The log-likelihood function of β, λ, and σ for (1) is represented as

$$l(\beta, \lambda, \sigma; y \mid x) = -\frac{n}{2} \ln(2\pi\sigma^2) + (\lambda - 1) \sum_{j=1}^{n} \ln y_j - \frac{1}{2\sigma^2} \| y^{(\lambda)} - f^{(\lambda)}(x; \beta) \|^2, \tag{3}$$

where y is the $n \times 1$ vector of observations for Y and is represented as $y = (y_1, \ldots, y_n)^{\mathrm{T}}$. Here, $\| \cdot \|$ is the norm. The parameters β, λ, and σ are estimated by the maximum likelihood method (see Oberg and Davidian, 2000). In the process of parameter estimation, given the estimate of λ, the estimation of β and σ in the TBS approach is carried out based on the linearized approximation about $\hat{\beta}$:

$$f^{(\hat{\lambda})}(x; \beta) \approx f^{(\hat{\lambda})}(x; \hat{\beta}) + \dot{\mathrm{F}}^{(\hat{\lambda})}(\beta - \hat{\beta}), \tag{4}$$

where $\dot{\mathrm{F}}^{(\hat{\lambda})}$ is the $n \times p$ first derivative matrix with respect to β evaluated at $\hat{\beta}$ and is represented as

$$\dot{\mathrm{F}}^{(\hat{\lambda})} = \Delta_{\beta} f^{(\hat{\lambda})}(x; \beta)\big|_{\beta=\hat{\beta}} = \left[\left(\frac{\partial f^{(\hat{\lambda})}(x_j; \beta)}{\partial \beta_k}\right)\Big|_{\beta_k=\hat{\beta}_k}\right], \quad j = 1, \ldots, n; \quad k = 1, \ldots, p.$$

In addition, the hypothesis on the parameter β could be tested in a manner similar to that of an usual linear regression, based on a linearized approximation. For example, given the estimate of λ, the $100(1 - \alpha)\%$ joint confidence region of β is represented as

$$\{(\beta - \hat{\beta})^{\mathrm{T}} \dot{\mathrm{F}}^{(\hat{\lambda})} \dot{\mathrm{F}}^{(\hat{\lambda})\mathrm{T}} (\beta - \hat{\beta})\} < p\tilde{\sigma}^2 \mathrm{F}_{p,n-p,1-\alpha}, \tag{5}$$

where $\tilde{\sigma} = \sqrt{\|Y^{(\hat{\lambda})} - f^{(\hat{\lambda})}(x; \hat{\beta})\|^2/(n-p)}$. $\mathrm{F}_{p,n-p,1-\alpha}$ is the $100(1 - \alpha)$ percentile of the F-distribution with a degree of freedom given by $(p, n - p)$.

2.2. *Relative Curvature Measure of the TBS Approach*

The inference based on the linearized approximation in (4) is applicable to a large sample on assuming that the function of the nonlinear regression model is not misspecified, and the behavior about $\hat{\beta}$ is approximately or locally close to that in the linear model. That is to say, it is important to quantitatively check if this model can be regarded as close to linear or locally linear for a small sample and if the inference based on the linearized approximation is valid. Therefore, it is beneficial to apply the relative curvature measure by Bates–Watts to the diagnosis of the inference of the nonlinear regression model. However, it is noted that this measure is constructed using the NT approach; since the observed data usually have the heteroscedasticity or non normality, we extend this measure in the framework of the TBS approach.

The relative curvature measure for the TBS approach, following the notation by Seber and Wild (1989), is defined as

$$\gamma^{PE}_{\lambda=\hat{\lambda}} = \rho \frac{\left\|h^{\mathrm{T}} \ddot{\mathrm{F}}^{PE(\hat{\lambda})} h\right\|}{\left\|\dot{\mathrm{F}}^{(\hat{\lambda})} h\right\|^2} \quad \text{and} \quad \gamma^{IN}_{\lambda=\hat{\lambda}} = \rho \frac{\left\|h^{\mathrm{T}} \ddot{\mathrm{F}}^{IN(\hat{\lambda})} h\right\|}{\left\|\dot{\mathrm{F}}^{(\hat{\lambda})} h\right\|^2}, \tag{6}$$

where $\gamma^{PE}_{\lambda=\hat{\lambda}}$ is the parameter-effects curvature in the direction h of $\hat{\beta}$ and $\gamma^{IN}_{\lambda=\hat{\lambda}}$ is the intrinsic curvature in the direction h of $\hat{\beta}$. Further, h is the $p \times 1$ vector in any direction of a line through the parameter estimates vector, which is represented as $h = (h_1, h_2, \ldots, h_p)^{\mathrm{T}}$. ρ is used to make the curvature

scale-invariant and $\rho = \tilde{\sigma}\sqrt{p}$, where $\tilde{\sigma} = \sqrt{\|Y^{(\hat{\lambda})} - f^{(\hat{\lambda})}(x; \hat{\beta})\|^2/(n-p)}$. $\ddot{\mathrm{F}}^{PE(\hat{\lambda})} = [\ddot{\mathrm{F}}_j^{PE(\hat{\lambda})}] = [(\ddot{f}_{jkl}^{PE(\hat{\lambda})})]$, $j = 1, \ldots, n$; $k = 1, \ldots, p$; $l = 1, \ldots, p$ is the $n \times p \times p$ array. Further, $\ddot{f}_{kl}^{PE(\hat{\lambda})} = (\ddot{f}_{1kl}^{PE(\hat{\lambda})}, \ldots, \ddot{f}_{nkl}^{PE(\hat{\lambda})})^{\mathrm{T}}$ is the $n \times 1$ vector represented as $\ddot{f}_{kl}^{PE(\hat{\lambda})} = \dot{\mathrm{F}}^{(\hat{\lambda})}(\dot{\mathrm{F}}^{(\hat{\lambda})\mathrm{T}}\dot{\mathrm{F}}^{(\hat{\lambda})})^{-1}\dot{\mathrm{F}}^{(\hat{\lambda})\mathrm{T}}\ddot{f}_{kl}^{(\hat{\lambda})}$. In this case, $\ddot{f}_{kl}^{(\hat{\lambda})}$ is the $n \times 1$ vector represented as $\ddot{f}_{kl}^{(\hat{\lambda})} = (\ddot{f}_{1kl}^{(\hat{\lambda})}, \ldots, \ddot{f}_{nkl}^{(\hat{\lambda})})^{\mathrm{T}}$. $\ddot{\mathrm{F}}^{IN(\hat{\lambda})} = [\ddot{\mathrm{F}}_j^{IN(\hat{\lambda})}] = [(\ddot{f}_{jkl}^{IN(\hat{\lambda})})]$, $j = 1, \ldots, n$; $k = 1, \ldots, p$; $l = 1, \ldots, p$ is the $n \times p \times p$ array. $\ddot{f}_{kl}^{IN(\hat{\lambda})} = (\ddot{f}_{1kl}^{IN(\hat{\lambda})}, \ldots, \ddot{f}_{nkl}^{IN(\hat{\lambda})})^{\mathrm{T}}$ is the $n \times 1$ vector represented as $\ddot{f}_{kl}^{IN(\hat{\lambda})} = \{\mathrm{I}_n - \dot{\mathrm{F}}^{(\hat{\lambda})}(\dot{\mathrm{F}}^{(\hat{\lambda})\mathrm{T}}\dot{\mathrm{F}}^{(\hat{\lambda})})^{-1}\dot{\mathrm{F}}^{(\hat{\lambda})\mathrm{T}}\}\ddot{f}_{kl}^{(\hat{\lambda})}$, where I_n is an $n \times n$ identity matrix. The superscripts *PE* and *IN* represent the tangential and orthogonal components of f, respectively.

The relative curvature measure using the TBS approach has the same interpretation as the original measure by Bates–Watts and is represented by the parameter-effects curvature and intrinsic curvature as a local behavior of the surface (called an expectation surface) of β in $f(x; \beta)$ (see Bates and Watts, 1988; Seber and Wild, 1989). The parameter-effects curvature assesses the nonlinearity depending on the model parameters. For a small curvature, straight parallel equispaced lines on the parameter space map into similar lines on the expectation surface, as in the case of the tangent plane. The parameter-effects curvature can be decreased by suitable reparameterizations. On the other hand, the intrinsic curvature measures the degree of nonlinearity inherent in the model. If this curvature is small, the expectation surface is locally replaced by the tangent plane. However, this curvature cannot be decreased by reparameterizations.

For an omnibus diagnosis of the underlying nonlinearity in the nonlinear regression model, we utilize the maximized parameter-effects curvature measure $\gamma^{PE}_{\max,\,\lambda=\hat{\lambda}}$ and maximized intrinsic curvature measure $\gamma^{IN}_{\max,\,\lambda=\hat{\lambda}}$, which maximize relative curvature measure (6) of $\gamma^{PE}_{\lambda=\hat{\lambda}}$ and $\gamma^{IN}_{\lambda=\hat{\lambda}}$, respectively, with respect to h. In addition, we use rooted mean squared (RMS) parameter-effects curvature measure $\gamma^{PE}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$ and RMS intrinsic curvature measure $\gamma^{IN}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$, which are calculated as rooted mean squares of $\gamma^{PE}_{\lambda=\hat{\lambda}}$ and $\gamma^{IN}_{\lambda=\hat{\lambda}}$, respectively, in all directions with respect to h. The algorithm for calculation of these measures in accordance with Bates and Watts (1980) and Seber and Wild (1989) is given in the Appendix.

One question that might be asked is whether one should treat λ as fixed or known. In practice, after we also apply the simplified transformation, for example, $\lambda = 0$, $\lambda = 0.25$, or $\lambda = 0.5$, etc. for both the observation and the function of the nonlinear regression model, they could deal with the estimation of β and σ. Therefore, we utilize the relative curvature measure, given the power-transforming parameter estimate.

2.3. *Measures Associated with the Relative Curvature Measure of the TBS Approach*

Generally, the nonlinear regression models differ from linear regression models with regard to the fact that the parameter estimators are biased and non-normally distributed. Box (1971) presented a measure for estimating the bias in the parameter estimator employed in the NT approach and Hougaard (1985) measured the asymmetry or skewness in the distribution of the parameter estimator. The above-mentioned diagnostic measures are closely associated with the relative curvature measure by Bates and Watts (1980) and can play roles on the directional diagnosis

of the properties of the parameter estimator. However, these two measures are also constructed on the basis of the NT approach. It is necessary to extend these measures within the framework of the TBS approach.

2.3.1. *Box's Bias Measure of the Parameter Estimator Used in the TBS Approach.* The bias in the parameter estimator $\hat{\beta}_k$, $k = 1, \ldots, p$ of the TBS approach can be approximated as follows (Box, 1971; Seber and Wild, 1989):

$$\text{Bias}[\hat{\beta}_k] = \text{E}[\hat{\beta}_k - \beta_k^*] \approx -\frac{1}{2}\sigma^2 \sum_{l=1}^{p} r_{11,kl}^{(-1)} \left(\sum_{m=1}^{p} \ddot{f}_{**lmm}^{PE(\hat{\lambda})} \right),$$

where β_k^* is the true value of β_k. $r_{11,kl}^{(-1)}$ is the element of the inverse matrix R_{11}^{-1} of R_{11} (see Appendix), $\ddot{f}_{**,lmm}^{PE(\hat{\lambda})}$ is the diagonal element of the $p \times p$ matrix in the face of the lth face of the parameter-effects curvature array $\ddot{\text{F}}_{**}^{PE(\hat{\lambda})}$. This bias can be estimated by replacing σ^2 by the unbiased estimate $\tilde{\sigma}^2 = \|Y^{(\hat{\lambda})} - f^{(\hat{\lambda})}(x; \hat{\beta})\|^2/(n-p)$. The Box's bias is represented as:

$$\widehat{\text{Bias}}[\hat{\beta}_k]/\hat{\beta}_k \times 100.$$

The Box's bias depends on the parameter-effects curvature and can be reduced with suitable reparameterizations. The thumb rule is that the nonlinear behavior of the model parameter cannot be ignored if the absolute value of the Box's bias is greater than unity.

2.3.2. *Hougaard's Skewness Measure of the Parameter Estimator Used in the TBS Approach.* The third-order moment of the parameter estimator $\hat{\beta}_k$, $k = 1, \ldots, p$ can be approximated as follows (Hougaard, 1985; Seber and Wild, 1989):

$$\begin{aligned}\text{Skew}[\hat{\beta}_k] &= \text{E}[\hat{\beta}_k - \text{E}[\hat{\beta}_k]]^3 \\ &\approx -(\sigma^2)^2 \sum_{k'=1}^{p} \sum_{l'=1}^{p} \sum_{m'=1}^{p} \left\{ \left(\sum_{l=1}^{p} r_{11,kl}^{(-1)} r_{11,k'l}^{(-1)} \right) \left(\sum_{l=1}^{p} r_{11,kl}^{(-1)} r_{11,l'l}^{(-1)} \right) \left(\sum_{l=1}^{p} r_{11,kl}^{(-1)} r_{11,m'l}^{(-1)} \right) \right. \\ &\quad \left. \times \left(\left(\sum_{j=1}^{n} \dot{f}_{jk'} \ddot{f}_{jl'm'} \right) + \left(\sum_{j=1}^{n} \dot{f}_{jl'} \ddot{f}_{jk'm'} \right) + \left(\sum_{j=1}^{n} \dot{f}_{m'j} \ddot{f}_{jl'k'} \right) \right) \right\}\end{aligned}$$

where $r_{11,kl}^{(-1)}$, $r_{11,k'l}^{(-1)}$, $r_{11,l'l}^{(-1)}$, and $r_{11,m'l}^{(-1)}$ are the elements of R_{11}^{-1}; $\dot{f}_{jk'}$, $\dot{f}_{jl'}$, and $\dot{f}_{jm'}$ are the elements of $\dot{\text{F}}^{(\hat{\lambda})}$; and $\ddot{f}_{jl'm'}$, $\ddot{f}_{jk'm'}$, and $\ddot{f}_{jl'k'}$ are the elements of $\ddot{\text{F}}^{(\hat{\lambda})}$ (see Appendix). Moreover, the third-order moment can be estimated by replacing σ^2 by its unbiased estimate $\tilde{\sigma}^2$ in a manner similar to that of the Box's bias. Furthermore, the estimate of the third-order moment is standardized by the kth diagonal element of the asymptotic variance–covariance matrix represented as $\tilde{\sigma}^2 \sum_{l=1}^{p} \left(r_{11,kl}^{(-1)}\right)^2$. The Hougaard's skewness is given by

$$\frac{\widehat{\text{Skew}}[\hat{\beta}_k]}{\left(\tilde{\sigma}^2 \sum_{l=1}^{p} \left(r_{11,kl}^{(-1)}\right)^2\right)^{3/2}}.$$

The Hougaard's skewness evaluates the skewness in the distribution of the parameter estimator. Empirically, if the absolute value of the measure is greater than 0.25, then the estimator has a skewed distribution.

3. Examples

In pharmacokinetics, the time course of drug disposition in the human body is identified primarily by sampling concentration-time profile in blood circulation and is often described using a function of a nonlinear regression model. The function of the nonlinear regression model is called as the compartment model (Davidian and Giltinan, 1995; Oberg and Davidian, 2000). The compartment model is derived by solving the ordinary differential equations representing drug disposition, and thus is a function having pharmacokinetic indices, such as volume of distribution, clearance, and rate constants, in a nonlinear form. Through two examples, the application of the relative curvature measure and associated measures for the TBS approach is illustrated, as compared with those for the NT approach. Furthermore, it is noteworthy that assumptions on the error in each approach are satisfactory. Therefore, the Spearman's rank correlation between predicted blood-drug concentrations and absolute values of the residuals is also calculated in order to check the heteroscedasticity of the observed blood-drug concentration data or the error. Skewness and kurtosis are calculated to check the normality of the residuals.

Example 1: Argatroban Data

A pharmacokinetic study of the anticoagulant drug, argatroban was conducted (Davidian and Giltinan, 1995). Each of the 37 subjects received a 4-h intravenous infusion of argatroban with groups of 4 subjects at 9 dose levels between 1 and $5\,\mu g \cdot kg^{-1} \cdot min^{-1}$ in increments of $0.5\,\mu g \cdot kg^{-1} \cdot min^{-1}$ and one subject received the 10th dose level $-4.37\,\mu g \cdot kg^{-1} \cdot min^{-1}$. Serial blood samples were taken at from 11–14 times from these subjects, and the argatroban concentration (Y_j) was measured in each subject using the HPLC. The concentration at time t_j $(=x_j)$; $j = 1, \ldots, n$ is represented as a function of compartment model as follows:

$$f(x_j; \beta) = \frac{\mathrm{D}}{V\kappa_e}\{\exp(-\kappa_e t_j^*) - \exp(-\kappa_e t_j)\},$$
$$t_j^* = 0, \quad t_j \le t_{\mathrm{inf}}; \quad t_j^* = t_j - t_{\mathrm{inf}}, \quad t_j > t_{\mathrm{inf}}, \tag{7}$$

where D is the drug dose and $\beta = (V, \kappa_e)^{\mathrm{T}}$, V is the volume of distribution, κ_e is the eliminated rate constant, t is the time following the infusion, and t_{inf} is the duration of the infusion; in this case, $t_{\mathrm{inf}} = 240\,\mathrm{min}$.

Davidian and Giltinan (1995, Sec. 9.5) conducted reparameterization in (7), and the mixed effects model with the variance function was applied to the entire subject data. We adopt the TBS approach (1) for the data obtained from a single subject (No. 26; $\mathrm{D} = 4$, $n = 13$); these blood samples were taken at $t_j = (30, 60, 90, 115, 160, 200, 240, 245, 250, 260, 275, 295, 320)$. The results for the measures using the TBS and NT approaches are shown in Table 1.

As shown in Table 1, the power-transforming parameter and its 95% confidence interval based on the profile log-likelihood are estimated as 0.600 and (0.485, 0.678), respectively. The above-mentioned result supports the square root transformation

Table 1
Results for the argatroban data

	TBS approach			NT approach ($\lambda = 1$)		
	Parameter estimate	Box's bias	Hougaard's skewness	Parameter estimate	Box's bias	Hougaard's skewness
V	0.124	1.562	0.539	0.111	1.877	0.529
κ_e	0.050	1.698	0.926	0.054	3.489	1.609
σ^2	79.734			977.070		
$\lambda^{(1)}$	0.600	(0.485, 0.678)				
$\gamma^{PE}_{\text{RMS},\lambda=\hat{\lambda}}$		0.826			1.570	
$\gamma^{PE}_{\max,\lambda=\hat{\lambda}}$		1.311			2.525	
$\gamma^{IN}_{\text{RMS},\lambda=\hat{\lambda}}$		0.086‡			0.136‡	
$\gamma^{IN}_{\max,\lambda=\hat{\lambda}}$		0.141‡			0.222‡	
Log-likelihood		−76.795			−78.163	
Rank correlation[c]		−0.049 [0.873]			0.440 [0.146]	
Skewness[d]		−0.110 [0.194]			−0.222 [0.143]	
Kurtosis[e]		1.311 [<0.001]			1.862 [0.612]	

(a) †: $<1/\sqrt{F_{2,13-2,1-0.05}} = 0.501$, ‡: $<1/2\sqrt{F_{2,13-2,1-0.05}} = 0.251$;
(b) 95% confidence interval (in parentheses);
(c) Spearman's rank correlation between predicted drug concentrations and absolute residuals and the p-value (in parentheses) for the hypothesis testing of H_0: no correlation between predicted drug concentrations and absolute residuals;
(d) Skewness of the residuals and the p-value (in parentheses) for the hypothesis testing of H_0: skewness = 0;
(e) Kurtosis of the residuals and the p-value (in parentheses) for the hypothesis testing of H_0: kurtosis = 3.

for both blood-drug concentration data and function of the compartment model; further, this result supports the rationale reported by Davidian and Giltinan (1995, Sec. 9.5) that the errors in the HPLC measurements exhibit a Poisson-like variance. It is evident from the results that the TBS approach can eliminate the heteroscedasticity in the error and the skewness in the error distribution as compared with the NT approach. In other words, using the NT approach, the heteroscedasticity or non normality of the blood-drug concentrations (or the errors) can have influences on not only the inference and measures but also the interpretation of their results.

The RMS and maximized values of the parameter-effects curvature measure and the intrinsic curvature measure are smaller in the TBS approach than in the NT approach. However, a criterion is needed to judge a satisfactory linearized approximation. The RMS curvature measure calculated within 95% joint confidence region (5) of the parameter β is used as the criterion (e.g., Bates and Watts, 1980, 1988; Seber and Wild, 1989). This value is given by the inverse F-value, $1/\sqrt{F_{p,n-p,1-\alpha}}$, for the significant level α of the F-distribution with a degree of freedom given by $(p, n-p)$. The rule of thumb is that the linearized approximation applied to the expectation surface is satisfactory if each of the values of the curvature measures is less than $1/\sqrt{F_{p,n-p,1-\alpha}}$ or $1/2\sqrt{F_{p,n-p,1-\alpha}}$. All the values of the

intrinsic curvature measures involved in the TBS approach and NT approach are less than $1/2\sqrt{F_{2,13-2,1-0.05}} = 0.251$ and the intrinsic nonlinearity can be ignored.

In both the approaches, the absolute values of the Box's bias are greater than unity. Particularly, the Box's bias in $\hat{\kappa}_e$ of the NT approach is reasonably large. The bias might increase the parameter-effects nonlinearity of the NT approach. Furthermore, it is suggested that the distributions of $\widehat{V}$ and $\hat{\kappa}_e$ for both the approaches are skewed, since the Hougaard's skewness are greater than 0.25. The inference and its interpretation of these parameters, based on the linearized approximation, have to be particularly considered.

Example 2: Indomethacin Data

As a second example, we consider a well-known data set from a study on indomethacin in six healthy volunteers that was reported by Kwan et al. (1976) and analyzed subsequently by Davidian and Giltinan (1995, Sec. 2.5). Each individual received an intravenous dose of $D = 25\,mg \cdot kg^{-1}$. Further, indomethacin was labeled by a radioisotope, ^{14}C, and the concentration (Y_j) was observed $n = 1, \ldots, 11$ times by using a radiation counting device. Using the function of the compartment model, the concentration at time t_j $(=x_j)$; $j = 1, \ldots, 11$ is represented as

$$f(x_j; \beta) = \frac{D(\kappa_1^* - \kappa_{21})}{V(\kappa_1^* - \kappa_2^*)} \exp(-\kappa_1^* t_j) + \frac{D(\kappa_{21} - \kappa_2^*)}{V(\kappa_1^* - \kappa_2^*)} \exp(-\kappa_2^* t_j), \tag{8}$$

where $\beta = (V, \kappa_{12}, \kappa_{21}, \kappa_e)^T$, V is the volume of distribution, κ_{12} and κ_{21} are the transfer rate constants, and κ_e is the eliminated rate constant. In this case, $\kappa_1^* + \kappa_2^* = \kappa_e + \kappa_{12} + \kappa_{21}$, $\kappa_1^* \kappa_2^* = \kappa_e \kappa_{21}$, and $\kappa_1^* > \kappa_2^*$. Davidian and Giltinan (1995, Sec. 2.5) applied the variance function approach on a single subject (No. 5)'s concentration-time profile. Instead of making *a priori* assumptions on the variance function, we adopt TBS approach (1) to the single subject's concentration-time profile. The results of the measures for the TBS and NT approaches are shown in Table 2.

As shown in Table 2, the power-transforming parameter is estimated as 0.170 and its confidence interval based on the profile log-likelihood is calculated as $(-0.120, 0.651)$. In practice, applying log transformation might be suitable. In other words, as shown by Davidian and Giltinan (1995, Sec. 2.5), the observed drug concentrations or errors might have log-normal distributions, or their variances might be in proportion with the square of the values of the predicted blood-drug concentrations. This property of the error distribution might depend on the type of the radiation counting device. The Spearman's rank correlation of the TBS approach is close to zero as compared with that of the NT approach; further, the heteroscedasticity of the TBS approach can be improved. The p-values of the null hypotheses for skewness as well as kurtosis in the TBS approach are greater than the corresponding values in the NT approach; further, normality can be realized in the TBS approach.

As elucidated by the results in Table 1, for the TBS as well as NT approaches, each value of the intrinsic curvature measures is less than that of the parameter-effects curvature measures. In the inference of the compartment model, it might be noteworthy to consider the parameter nonlinearity rather than the inherent

Table 2
Results for the indomethacin data

	TBS approach			NT approach ($\lambda = 1$)		
	Parameter estimate	Box's bias	Hougaard's skewness	Parameter estimate	Box's bias	Hougaard's skewness
V	6.785	0.475*	0.460	6.481	−0.988*	0.043*
κ_e	1.384	0.086*	0.271	1.498	−3.694	−0.281
κ_{12}	1.101	2.899	0.396	1.136	10.867	0.377
κ_{21}	0.332	−0.004*	0.306	0.419	7.807	0.390
σ^2	0.011			0.003		
$\lambda^{(1)}$	0.170	(−0.120, 0.651)				
$\gamma^{PE}_{\text{RMS},\lambda=\hat{\lambda}}$		0.752			1.114	
$\gamma^{PE}_{\max,\lambda=\hat{\lambda}}$		1.884			3.081	
$\gamma^{IN}_{\text{RMS},\lambda=\hat{\lambda}}$		0.109‡			0.252†	
$\gamma^{IN}_{\max,\lambda=\hat{\lambda}}$		0.306†			0.669	
Log-likelihood		21.686			16.459	
Rank correlation[b]		0.036 [0.915]			0.618 [0.042]	
Skewness[c]		0.467 [0.076]			0.977 [0.025]	
Kurtosis[d]		1.853 [0.545]			4.333 [0.068]	

(a) †: $<1/\sqrt{F_{4,11-4,1-0.05}} = 0.493$, ‡: $<1/2\sqrt{F_{4,11-4,1-0.05}} = 0.246$, ⋆: $<1 * <0.25$;
(b) 95% confidence interval (in parentheses);
(c) Spearman's rank correlation between predicted drug concentrations and absolute residuals and the p-value (in parentheses) for the hypothesis testing of H_0: no correlation between predicted drug concentrations and absolute residuals;
(d) Skewness of the residuals and the p-value (in parentheses) for the hypothesis testing of H_0: skewness = 0;
(e) Kurtosis of the residuals and the p-value (in parentheses) for the hypothesis testing of H_0: kurtosis = 3.

nonlinearity of the model; therefore, reparameterization might prove to be useful. Each value of the curvature measures in the TBS approach is less than that in the NT approach. All the values of the intrinsic curvature measures in the TBS approach are less than $1/2\sqrt{F_{4,11-4,1-0.05}} = 0.246$. On the other hand, in the NT approach, only the value of the RMS intrinsic curvature is less than $1\sqrt{F_{4,11-4,1-0.05}} = 0.493$, and the linearized approximation in the NT approach may not be valid.

Each absolute value of the Box's bias of the parameters in the TBS approach decreased remarkably as compared with that in the NT approach. Except for $\hat{\kappa}_{12}$, all the absolute values of the Box's biases in the TBS approach are less than unity. It might be noteworthy to consider the inference of κ_{12}. Furthermore, the absolute values of the Box's biases in the TBS approach increase in the order $\hat{\kappa}_{12}$, $\widehat{V}_1$, $\hat{\kappa}_e$, and $\hat{\kappa}_{21}$; however, in the NT approach, the increase is in the order $\hat{\kappa}_{12}$, $\hat{\kappa}_{21}$, $\hat{\kappa}_e$, and $\widehat{V}_1$. The Hougaard's skewness of $\widehat{V}$ in the TBS approach becomes greater than that in the NT approach; hence, the distribution of $\widehat{V}$ in the TBS approach might be right skewed.

4. Simulation Experiment

According to the above pharmacokinetic scenario, we evaluated the influence of heteroscedasticity of the blood-drug concentration data on the accuracy and precision of the relative curvature measure of the TBS approach and compared the results with those of the NT approach. The concentration data were generated from the following model:

$$Y_j = \{f^{\lambda}(x_j; \beta) + \lambda\varepsilon_j\}^{1/\lambda}; \quad j = 1, \ldots, n, \tag{9}$$

where Y denotes the random variable of the blood-drug concentration, and f is the function of the compartment model. In this case, the function of the compartment model is replaced by (7), D = 4, and $t_j = (30, 60, 90, 115, 160, 200, 240, 245, 250, 260, 275, 295, 320)$ with $t_{\text{inf}} = 240$ min, according to the argatroban data. n is the number of observed blood-drug concentrations $-n = 13$. ε is assumed to be independently, identically, and normally distributed as $N(0, \sigma^2)$ and σ is the scale parameter. In (9), the true parameter values of V, κ_e, and λ are 0.125, 0.05, and 0.5, respectively. The true value of σ is substituted as 1, 3, 5, 7, and 9. Therefore, for example, when $\sigma = 1$, the true values $\gamma^{PE*}_{\text{RMS},\,\lambda=0.5}$, $\gamma^{PE*}_{\text{max},\,\lambda=0.5}$, $\gamma^{IN*}_{\text{RMS},\,\lambda=0.5}$, and $\gamma^{IN*}_{\text{max},\,\lambda=0.5}$ are 0.136, 0.214, 0.014, and 0.023, respectively, for the TBS approach, and the true values for the NT approach are 0.015, 0.025, 0.006, and 0.009, respectively. For the other values of σ, the true values of the curvature measures are shown in Table 3.

For the evaluation, the relative bias (RB) and relative average error (RAE) of each curvature measure are estimated. For $N = 1{,}000$ replications of the experiment, for example, the estimated RB and RAE values of the RMS parameter-effects curvature measure of the TBS approach are represented, respectively, as:

$$\widehat{RB}\left[\gamma^{PE}_{\text{RMS},\,\lambda=\hat{\lambda}}\right] = \frac{\gamma^{PE*}_{\text{RMS},\,\lambda=\hat{\lambda}} - \dfrac{\sum_{l=1}^{N}\gamma^{PE}_{\text{RMS},\,\lambda=\hat{\lambda},l}}{N}}{\gamma^{PE*}_{\text{RMS},\,\lambda=0.5}} \quad \text{and}$$

$$\widehat{RAE}\left[\gamma^{PE}_{\text{RMS},\,\lambda=\hat{\lambda}}\right] = \frac{\sqrt{\dfrac{\sum_{l=1}^{N}\left(\gamma^{PE}_{\text{RMS},\,\lambda=\hat{\lambda},l} - \dfrac{\sum_{l=1}^{N}\gamma^{PE}_{\text{RMS},\,\lambda=\hat{\lambda},l}}{N}\right)^2}{N-1}}}{\gamma^{PE*}_{\text{RMS},\,\lambda=0.5}},$$

where $\gamma^{PE}_{\text{RMS},\,\lambda=\hat{\lambda},l}$ represents the estimate of the RMS parameter-effects curvature measure of the TBS approach in the l $(l = 1, \ldots, N)$th sample. Furthermore, the heteroscedasticity in the error is assessed by the means of the Spearman's rank correlations and the p-values of the 1,000 replications. The non normality is assessed based on the skewness and kurtosis and p-values of these replications. The results are summarized in Table 3.

For all the levels of σ, the TBS approach improves the heteroscedasticity and non normality as compared with the NT approach. All the curvature measures of the TBS approach can be less influenced by the heteroscedasticity and non normality than those of the NT approach. Actually, for all the levels of σ, $\widehat{RB}$ as well as $\widehat{RAE}$ of each of the curvature measures in the TBS approach are less than the corresponding values in the NT approach. For both approaches, each value of the $\widehat{RB}$ and $\widehat{RAE}$ for the RMS parameter-effects curvature measure behaves almost similar to that of the maximized parameter-effects curvature measure. Further, the behavior of each value of the $\widehat{RB}$ and $\widehat{RAE}$ for the intrinsic curvature measures is

Table 3
Results for the simulation experiment in the argatroban scenario

		TBS approach			NT approach ($\lambda = 1$)		
		True value	$\widehat{RB}$	$\widehat{RAE}$	True value	$\widehat{RB}$	$\widehat{RAE}$
$\sigma = 1$	$\gamma^{PE}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$	0.136	−0.229	0.350	0.015	17.606	4.428
	$\gamma^{PE}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	0.214	−0.227	0.354	0.025	17.550	4.425
	$\gamma^{IN}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$	0.014	−0.302	0.408	0.006	17.220	4.433
	$\gamma^{IN}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	0.023	−0.302	0.408	0.009	17.222	4.433
	Mean of rank correlation[a]		−0.055 [0.627]			0.387 [0.273]	
	Mean of skewness[b]		0.015 [0.123]			0.270 [0.095]	
	Mean of kurtosis[c]		2.309 [0.394]			3.077 [0.259]	
$\sigma = 3$	$\gamma^{PE}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$	0.407	−0.248	0.326	0.044	16.974	4.488
	$\gamma^{PE}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	0.643	−0.248	0.328	0.075	16.827	4.531
	$\gamma^{IN}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$	0.042	−0.327	0.382	0.017	15.940	4.982
	$\gamma^{IN}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	0.068	−0.327	0.382	0.027	15.946	4.977
	Mean of rank correlation[a]		−0.061 [0.617]			0.389 [0.270]	
	Mean of skewness[b]		−0.002 [0.123]			0.328 [0.092]	
	Mean of kurtosis[c]		2.301 [0.397]			3.103 [0.256]	
$\sigma = 5$	$\gamma^{PE}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$	0.678	−0.341	0.283	0.073	16.362	4.698
	$\gamma^{PE}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	1.071	−0.341	0.284	0.126	16.156	4.839
	$\gamma^{IN}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$	0.069	−0.435	0.340	0.028	14.856	6.004
	$\gamma^{IN}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	0.113	−0.435	0.340	0.046	14.863	5.994
	Mean of rank correlation[a]		−0.063 [0.601]			0.399 [0.257]	
	Mean of skewness[b]		−0.045 [0.124]			0.391 [0.089]	
	Mean of kurtosis[c]		2.314 [0.392]			3.147 [0.252]	
$\sigma = 7$	$\gamma^{PE}_{\mathrm{RMS},\,\lambda=\hat{\lambda}}$	0.950	−0.402	0.266	0.102	15.812	5.105
	$\gamma^{PE}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	1.499	−0.403	0.265	0.176	15.580	5.405
	$\gamma^{IN}_{\mathrm{RMS},\lambda=\hat{\lambda}}$	0.097	−0.507	0.322	0.039	14.034	7.504
	$\gamma^{IN}_{\mathrm{max},\,\lambda=\hat{\lambda}}$	0.158	−0.507	0.322	0.064	14.040	7.486
	Mean of rank correlation[a]		−0.054 [0.593]			0.409 [0.248]	
	Mean of skewness[b]		−0.095 [0.124]			0.454 [0.086]	
	Mean of kurtosis[c]		2.338 [0.387]			3.198 [0.245]	

(continued)

Table 3
Continued

		TBS approach			NT approach ($\lambda = 1$)		
		True value	$\widehat{RB}$	$\widehat{RAE}$	True value	$\widehat{RB}$	$\widehat{RAE}$
$\sigma = 9$	$\gamma^{PE}_{\mathrm{RMS},\lambda=\hat{\lambda}}$	1.221	−0.470	0.242	0.132	15.344	5.713
	$\gamma^{PE}_{\mathrm{max},\lambda=\hat{\lambda}}$	1.928	−0.471	0.241	0.226	15.115	6.249
	$\gamma^{IN}_{\mathrm{RMS},\lambda=\hat{\lambda}}$	0.125	−0.585	0.295	0.050	13.480	9.509
	$\gamma^{IN}_{\mathrm{max},\lambda=\hat{\lambda}}$	0.204	−0.585	0.295	0.082	13.484	9.479
	Mean of rank correlation[a]	−0.031 [0.585]			0.423 [0.235]		
	Mean of skewness[b]	−0.156 [0.121]			0.519 [0.082]		
	Mean of kurtosis[c]	2.400 [0.376]			3.259 [0.239]		

(a) Mean of the Spearman's rank correlations between predicted drug concentrations and absolute residuals and of the p-values (in parentheses) for the hypothesis testing of H_0: no correlation between predicted drug concentrations and absolute residuals;

(b) Mean of skewness of residuals and of the p-values (in parentheses) for the hypothesis testing of H_0: skewness = 0;

(c) Mean of kurtosis of residuals and of the p-values (in parentheses) for the hypothesis testing of H_0: kurtosis = 3.

also similar. The absolute value of the $\widehat{RB}$ in the TBS approach tends to increase with an increase in σ, whereas it tends to decrease in the NT approach. On the other hand, the $\widehat{RAE}$ in the TBS approach tends to decrease with an increase in σ, whereas it tends to increase in the NT approach.

5. Discussion

In this article, we proposed the relative curvature measure and its associated measures extended in the framework of the TBS approach and demonstrated a diagnostics for the inference of several nonlinear regression models, based on their measures. We could develop the measures in the framework of the variance function approach, but the approach takes account of only the heteroscedasticity and does not the non normality explicitly, as discussed by Carroll and Ruppert (1988), Oberg and Davidian (2000), and other authors. Furthermore, the wrong choice of the variance function can have the bad influence on the parameter estimation, as pointed out by Houwelingen (1988). Therefore, we focused on the TBS approach that could improve both of the heteroscedasticity and non normality, since the observed data sometimes have the non normality as well as the heteroscedasticity.

The proposed measures here would be able to assess the nonlinearity underlying the model and diagnose the validity of the inference, without being influenced by the heteroscedasticity, in addition, non normality of the data. Especially, it is very useful to assess the degree of the heteroscedasticity or non normality quantitatively, from the power-transforming parameter estimate.

Appendix A. Calculation of the Relative Curvature Measure of the TBS Approach

The relative curvature measure of the TBS approach is calculated by using the following algorithm:

Step 1 Given the power-transforming parameter estimate $\hat{\lambda}$, calculate the $n \times p$ first-derivative matrix of the function $f^{(\hat{\lambda})}(x; \beta)$, namely,

$$\dot{\mathrm{F}}^{(\hat{\lambda})} = \Delta_\beta f^{(\hat{\lambda})}(x; \beta)\Big|_{\beta=\hat{\beta}} = \left[\left(\frac{\partial f^{(\hat{\lambda})}(x_j; \beta)}{\partial \beta_k}\right)\Big|_{\beta_k=\hat{\beta}_k}\right]$$
$$= [(\dot{f}_{jk}^{(\hat{\lambda})})|_{\beta_k=\hat{\beta}_k}], \quad j = 1, \ldots, n; \quad k = 1, \ldots, p,$$

evaluated at the parameter estimate $\hat{\beta}$. Furthermore, calculate the $n \times p \times p$ second derivative array, namely,

$$\ddot{\mathrm{F}}^{(\hat{\lambda})} = \nabla_\beta f^{(\hat{\lambda})}(x; \beta)|_{\beta_k=\hat{\beta}_k, \beta_l=\hat{\beta}_l} = \left[\left(\frac{\partial f^{(\hat{\lambda})}(x_j; \beta)}{\partial \beta_k \partial \beta_l}\right)\Big|_{\beta_k=\hat{\beta}_k, \beta_l=\hat{\beta}_l}\right]$$
$$= [(\ddot{f}_{jkl}^{(\hat{\lambda})})|_{\beta_k=\hat{\beta}_k, \beta_l=\hat{\beta}_l}], \quad j = 1, \ldots, n; \quad k = 1, \ldots, p; \quad l = 1, \ldots, p,$$

evaluated at the parameter estimate $\hat{\beta}$.

Step 2 Calculate the QR decomposition of $\dot{\mathrm{F}}^{(\hat{\lambda})}$, that is,

$$\dot{\mathrm{F}}^{(\hat{\lambda})} = \mathrm{QR}_1 = [\mathrm{Q}_p, \mathrm{Q}_{n-p}]\begin{bmatrix} \mathrm{R}_{11} \\ \mathrm{O} \end{bmatrix} = \mathrm{Q}_p \mathrm{R}_{11},$$

where Q is the $n \times n$ orthogonal matrix and plays a role of rotating the axes of the sample space. However, $\mathrm{Q}^{\mathrm{T}}\mathrm{Q} = \mathrm{I}_n$, and I_n is the $n \times n$ identity matrix. R_1 is the $n \times p$ matrix. O is the $(n-p) \times p$ zero matrix. Q_p is the $n \times p$ matrix, whose columns form a basis for the tangent plane parallel to the expectation surface. Q_{n-p} is the $n \times (n-p)$ matrix, whose columns form a basis for vectors perpendicular to the tangent plane. R_{11} is the non singular $p \times p$ upper triangular matrix. Since $\mathrm{R}_{11}^{\mathrm{T}}\mathrm{R}_{11}$ is the Cholesky decomposition of $\dot{\mathrm{F}}^{\mathrm{T}}\dot{\mathrm{F}}$, R_{11} is unique if its diagonal elements are all positive or all negative. If R_{11} is unique, Q_p is also unique. In fact, $\mathrm{Q}_p = \dot{\mathrm{F}}^{(\hat{\lambda})}\mathrm{R}_{11}^{-1}$ (however, Q_{n-p} is not unique). Then, $\dot{\mathrm{F}}^{(\hat{\lambda})}h = \mathrm{Q}_p\mathrm{R}_{11}h = \mathrm{Q}_p d$, where $d = \mathrm{R}_{11}h$, $p \times 1$ vector. That is, Q_p for the new coordinate plays the same role of the first derivative matrix $\dot{\mathrm{F}}^{(\hat{\lambda})}$ for the original coordinate.

Step 3 Scale the elements of the matrix $\dot{\mathrm{F}}^{(\hat{\lambda})}$ and array $\ddot{\mathrm{F}}^{(\hat{\lambda})}$ by $\rho = s\sqrt{p}$, where $s = \sqrt{\|Y^{(\hat{\lambda})} - f^{(\hat{\lambda})}(x; \hat{\beta})\|/(n-p)}$.

Step 4 Reparameterize $\gamma^{PE}_{\lambda=\hat{\lambda}}$ and $\gamma^{IN}_{\lambda=\hat{\lambda}}$ in (6) by $\phi = \mathrm{R}_{11}\beta$ and multiply their denominators and numerators by Q^{T}. We can obtain $\|d\| = 1$, since $\gamma^{PE}_{\lambda=\hat{\lambda}}$ and $\gamma^{IN}_{\lambda=\hat{\lambda}}$ depend on the direction of d and h, but do not depend on their sizes. That is, since

$$\left\|Q^T\dot{F}_{\phi}^{(\hat{\lambda})}R_{11}^{-1}d\right\| = 1,$$

$$\gamma_{\lambda=\hat{\lambda}}^{PE} = \frac{\left\|Q^T h^T \ddot{F}^{PE(\hat{\lambda})} h\right\|}{\left\|Q^T\dot{F}^{(\hat{\lambda})}h\right\|^2} = \frac{\left\|Q^T d^T (R_{11}^{-1})^T \ddot{F}^{PE(\hat{\lambda})} R_{11}^{-1} d\right\|}{\left\|Q^T\dot{F}^{(\hat{\lambda})}R_{11}^{-1}d\right\|^2} = \left\|d^T\ddot{F}_{**}^{PE(\hat{\lambda})}d\right\| \quad \text{and}$$

$$\gamma_{\lambda=\hat{\lambda}}^{IN} = \frac{\left\|Q^T h^T \ddot{F}^{IN(\hat{\lambda})} h\right\|}{\left\|Q^T\dot{F}^{(\hat{\lambda})}h\right\|^2} = \frac{\left\|Q^T d^T (R_{11}^{-1})^T \ddot{F}^{IN(\hat{\lambda})} R_{11}^{-1} d\right\|}{\left\|Q^T\dot{F}^{(\hat{\lambda})}R_{11}^{-1}d\right\|^2} = \left\|d^T\ddot{F}_{**}^{IN(\hat{\lambda})}d\right\|, \qquad \text{(A.10)}$$

where $\ddot{F}_{**}^{PE(\hat{\lambda})}$ is the $p \times p \times p$ parameter-effects curvature array, given by $\ddot{F}_{**}^{PE(\hat{\lambda})} = (R_{11}^{-1})^T Q^T \ddot{F}^{PE(\hat{\lambda})} R_{11}^{-1}$ and $\ddot{F}_{**}^{IN(\hat{\lambda})}$ is the $p' \times p \times p$ intrinsic curvature array (p' is at most $p(p+1)/2$), given by $\ddot{F}_{**}^{IN(\hat{\lambda})} = (R_{11}^{-1})^T Q^T \ddot{F}^{IN(\hat{\lambda})} R_{11}^{-1}$. However, it is noted that the following relation in the expansion of (A.10) is satisfied:

$$Q^T\ddot{F}^{(\hat{\lambda})} = Q^T\begin{bmatrix}\ddot{F}^{(\hat{\lambda})PE}\\ \ddot{F}^{(\hat{\lambda})IN}\end{bmatrix} = \begin{bmatrix}Q_p^T\ddot{F}^{PE(\hat{\lambda})}\\ Q_{n-p}^T\ddot{F}^{IN(\hat{\lambda})}\end{bmatrix} = \begin{bmatrix}\ddot{F}_{**}^{PE(\hat{\lambda})}\\ \ddot{F}_{**}^{IN(\hat{\lambda})}\end{bmatrix} = \ddot{F}_{**}^{(\hat{\lambda})},$$

where $\ddot{F}_{**}^{(\hat{\lambda})}$ is the $(p+p') \times p \times p$ curvature array.

Therefore, since there is the one-to-one correspondence as $h = R_{11}^{-1}d$ and R_{11}^{-1} is non singular, maximization of (6) with respect to h is equivalent to that of (A.10) with respect to d. Therefore, the maximized parameter-effects curvature $\gamma_{\max,\,\lambda=\hat{\lambda}}^{PE}$ and the maximized intrinsic curvature $\gamma_{\max,\,\lambda=\hat{\lambda}}^{IN}$ is computed by, respectively,

$$\gamma_{\max,\,\lambda=\hat{\lambda}}^{PE} = \max_h \gamma_{\lambda=\hat{\lambda}}^{PE} = \max_{\|d\|=1} \gamma_{\lambda=\hat{\lambda}}^{PE} = \max_{\|d\|=1} \rho\left\|d^T\ddot{F}_{**}^{PE(\hat{\lambda})}d\right\| \quad \text{and}$$

$$\gamma_{\max,\,\lambda=\hat{\lambda}}^{IN} = \max_h \gamma_{\lambda=\hat{\lambda}}^{IN(\hat{\lambda})} = \max_{\|d\|=1} \gamma_{\lambda=\hat{\lambda}}^{IN} = \max_{\|d\|=1} \rho\left\|d^T\ddot{F}_{**}^{IN(\hat{\lambda})}d\right\|.$$

The Lagrange's multiplier method is used to maximize the above functions (A.10). The rooted mean squared (RMS) parameter-effects curvature $\gamma_{\mathrm{RMS},\,\lambda=\hat{\lambda}}^{PE}$ and the RMS intrinsic curvature $\gamma_{\mathrm{RMS},\,\lambda=\hat{\lambda}}^{IN}$ is computed by the alogorithm of Seber and Wild (1989).

References

Bates, D. M., Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *J. Roy. Statist. Soc.* B42:1–25.

Bates, D. M., Watts, D. G. (1988). *Nonlinear Regression: Analysis and Its Applications.* New York: Wiley.

Box, M. J. (1971). Bias in nonlinear estimation (with discussion). *J. Roy. Statist. Soc.* B33:171–201.

Box, G. E. P., Cox, D. R. (1964). The analysis of transformations (with discussion). *J. Roy. Statist. Soc.* B26:211–252.

Carroll, R. J., Ruppert, D. (1988). *Transformation and Weighting in Regression.* New York: Chapman and Hall.

Davidian, M., Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data.* New York: Chapman and Hall.

Hougaard, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *J. Roy. Statist. Soc.* B47:103–114.