

Comments on: Goodness-of-fit tests in mixed models

Axel Munk · Tatyana Krivobokova

Published online: 14 May 2009

© Sociedad de Estadística e Investigación Operativa 2009

The authors are to be congratulated for an interesting paper which highlights the need for more formal procedures on the assessment of goodness of fit in the area of mixed models. The authors give a comprehensive overview of diagnostic techniques and goodness-of-fit tests for random effects in mixed models and propose a number of new nonparametric tests with omnibus character. Some of them are extensions of goodness-of-fit methods in the area of linear (regression) models and consequently share various properties of those. The paper addresses lack of fit tests for the distribution of the random effects and for the error distribution in a mixed model. The importance of this topic is vast. The application area of mixed models is broad—from longitudinal data to smoothing—and correct distributional assumptions are crucial for inference (see, e.g., Verbeke and Lesaffre 1997). Also, for the generalized linear mixed models, a misspecified distribution of random effects can lead to the substantial loss of efficiency (see, e.g., Agresti et al. 2004).

We would like to address three issues in the following:

1. Smoothing splines and goodness of fit
2. Adjustment for ill-posedness
3. Model validation

Smoothing splines and goodness of fit An interesting application of mixed models, which have become very popular over the last decade is smoothing with penalized splines (Eilers and Marx 1996; Brumback et al. 1999), showed that a penalized

This comment refers to the invited paper available at: doi:[10.1007/s11749-009-0148-8](https://doi.org/10.1007/s11749-009-0148-8).

A. Munk (✉) · T. Krivobokova
Georg-August-Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany
e-mail: munk@math.uni-goettingen.de

T. Krivobokova
e-mail: tkrivob@gwdg.de

spline estimator is equivalent to the Best Linear Unbiased Predictor (BLUP) in a certain mixed model. More details and applications of the mixed model formulation of penalized splines can be found in Ruppert et al. (2003). Thereby the random effects—which are the spline coefficients in this context—are assumed to be normally distributed. Kauermann (2005) investigated the smoothing parameter estimator resulted from the mixed model representation of penalized splines and found that it is asymptotically biased with respect to the mean squared error minimizer and the corresponding penalized spline estimator tends to undersmooth the data. This suggests that the normality assumption for the spline coefficients is not fully appropriate. Indeed, simple diagnostic plots of the estimated spline coefficients confirm that their normality is often violated and invite for more formal tests discussed by Claeskens and Hart. Some recent results of Zhao (2000) for the Bayesian smoothing splines, which are closely related to the mixed model representation of penalized splines, allow one to conjecture that modeling spline coefficients with an appropriate mixture of normal distributions can result in a more efficient penalized spline estimator. Therefore, an interesting direction for further research is the application of the procedures presented in Sect. 5.3.3 in the smoothing context.

The violation of the normal distribution assumption for the spline coefficients has implications not only for the efficiency of a penalized spline estimator but also for the simultaneous inference. Indeed, Krivobokova et al. (2009) found that the simultaneous confidence bands for penalized spline estimators obtained in the mixed model framework tend to be conservative. Thus, a proper modeling of random effects is crucial here as well.

Also the mixed model representation of generalized penalized spline smoothing, which is much less studied, offers a further promising area of application for the Claeskens and Hart toolbox.

Another interesting question is how robust are the tests considered by Claeskens and Hart if the variance of random effects tends towards zero. For example, in the mixed model representation of penalized splines this implies that the spline coefficients are close to zero and the data can be described by a parametric model. One can expect that in this case the identification of the random effects distribution can become unfeasible.

Adjustment for ill-posedness In Sect. 5.3 the authors elaborate on testing the distribution of γ , the vector of length d of random effects in the mixed model $Y = X\beta + Z\gamma + \epsilon$. To this end it is assumed that the error ϵ is i.i.d. normal. To simplify further let us assume that $\beta = 0$, i.e., no fixed effect is present. The direct analogue to a goodness-of-fit problem in a deconvolution problem becomes obvious. The most simple (and prominent) case occurs if we assume that the random effects are i.i.d. itself. If the dimension d becomes large (as it is the case in the microarray example, Sect. 8) the deconvolution character of the problem and its ill-posedness becomes apparent, and any goodness-of-fit method has to be adjusted for this. For the case of simple deconvolution, this effect has been demonstrated in Bissantz et al. (2009) for AIC and BIC for the order selection test of Eubank and Hart (1992). Roughly speaking, the estimated Fourier coefficients (or the coefficients of any other basis system) have to be reweighted by their singular values and recentered by the sum of

the reciprocal singular values associated to the convolution matrix (resp. spectrum of the L^2 -convolution). It is to be expected that this will affect the choice of the threshold or smoothing parameter M in Sect. 5.3, and it would be of great interest to see this more detailed in future work.

Model validation It is well known and an issue of debate for a long time whether goodness-of-fit testing is appropriate at all when we aim for data-based model building. Box (1976) rephrased this as

“All models are wrong, but some are useful”.

Even more provocative:

“How often do I have to sample in order to get significant deviation from the null model (e.g., normality)?”

The other way around: If the model is not rejected, how can we make sure that this ensures a sufficiently good approximation of the data (in the sense of G.E.P. Box or in any other reasonable sense) and is not due to a lack of power of the test at hand? There is a vast amount of literature on this issue (for a selective overview in the context of regression, see Munk 2002 and Dette and Munk 2003), and we do not want to elaborate on this in depth. However, we would like to mention that treating the “goodness of fit problem” by formal testing methods certainly has to be done with great care. In the context of mixed linear models consider the following simple thought experiment:

Assume that we would like to test whether ϵ is normally distributed as discussed in Sect. 7.1, and we conclude that such a model is appropriate whenever our favorite test does not reject formally. Then, the power of such a test will be very much dependent on the number d of random effect components and their variability, say. As d increases, the power will be inflated, and the probability of misspecification of the model (by formal acceptance of the null) increases. Hence, a wrong decision in favor of a normal model, say, could simply result from too many random effects taken into account into the model. For high-dimensional models, as they, e.g., occur in the analysis of microarray experiments, this matters. It is a difficult and challenging task how to “calibrate” testing procedures properly, so that these become “comparable” for different d in a certain sense.

References

- Agresti A, Caffo B, Ohman P (2004) Examples in which misspecification of a random effects distribution reduces efficiency. *Comput Stat Data Anal* 47(3):639–653
- Bissantz N, Claeskens G, Holzmann H, Munk A (2009) Testing for lack of fit in inverse regression—with applications to biophotonic imaging. *J R Stat Soc, Ser B* 71:25–48
- Box GE (1976) Science and statistics. *J Am Stat Assoc* 71:791–799
- Brumback BA, Ruppert D, Wand MP (1999) Comment on Shively, Kohn and Wood. *J Am Stat Assoc* 94:794–797
- Dette H, Munk A (2003) Some methodological aspects of validation of models in nonparametric regression. *Stat Neerl* 57(2):207–244
- Eilers PHC, Marx BD (1996) Flexible smoothing with B -splines and penalties. *Stat Sci* 11(2):89–121. With comments and a rejoinder by the authors
- Eubank R, Hart JD (1992) Testing goodness-of-fit in regression via order selection criteria. *Ann Stat* 20(3):1412–1425

- Kauermann G (2005) A note on smoothing parameter selection for penalised spline smoothing. *J Stat Plan Inference* 127:53–69
- Krivobokova T, Kneib T, Claeskens G (2009) Simultaneous confidence bands for penalized spline estimators. Technical report, University of Göttingen
- Munk A (2002) Testing the goodness of fit of parametric regression models with random Toeplitz forms. *Scand J Stat* 29(3):501–533
- Ruppert D, Wand M, Carroll R (2003) *Semiparametric regression*. Cambridge University Press, Cambridge
- Verbeke G, Lesaffre E (1997) The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Comput Stat Data Anal* 23(4):541–556
- Zhao LH (2000) Bayesian aspects of some nonparametric problems. *Ann Stat* 28(2):532–552