

# Inverse prediction for multivariate mixed models with standard software

Lynn R. LaMotte<sup>1</sup>  · Jeffrey D. Wells<sup>2</sup>

Received: 7 October 2015 / Revised: 25 July 2016 / Published online: 4 August 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Inverse prediction (IP) is reputed to be computationally inconvenient for multivariate responses. This paper describes how IP can be formulated in terms of a general linear mixed model, along with a flexible modeling approach for both mean vectors and variance–covariance matrices. It illustrates that results can be had as standard output from widely-available statistical computing packages.

**Keywords** Heteroscedastic multivariate models · Multivariate calibration · Forensic entomology

**Mathematics Subject Classification** 62H15 · 62J05 · 62H30

## 1 Introduction

In a model relating a response  $y$  to a factor  $t$ , *inverse prediction* refers to a process for inferring  $t$  from the observed response  $y_*$  for a subject with an unknown value of  $t$ . The inference is based on a model fit to *training data* from subjects at known values of  $t$ . The subject in question, with unknown  $t$ , will be termed the *mystery specimen* (MS) here, and its true (but unknown) value of  $t$  will be denoted  $t_*$ . It is presumed that the training data are from a process that is credibly similar to the one that produced the mystery specimen.

---

✉ Lynn R. LaMotte  
llamot@lsuhsc.edu

Jeffrey D. Wells  
jedwell@fiu.edu

<sup>1</sup> LSU Health Sciences Center, New Orleans, LA, USA

<sup>2</sup> Florida International University, Miami, FL, USA

In applications, usually there are other factors that affect the response and the relation between the response and the principal factor  $t$ . Models for the response would include potential effects of those factors. However, to simplify and focus the developments here, we shall not explicitly include these other factors in formulations.

Our investigation into this statistical problem is prompted particularly by the objective, to estimate the time since death when a body is discovered and there is suspicion of criminal activity. Larvae of carrion-feeding flies grow and develop in regular patterns over time, and so their sizes can serve as a biological clock to guess their ages. Their growth is affected by multiple conditions, of which temperature is usually a major factor. As an example to correspond to the setting described above,  $y$  might be length and  $t$  time since egg deposition. The MS found at the scene has length  $y_*$ . Other dimensions of the size of the larva can be measured (weight, for example), and they may provide more information relevant to the age of the MS. Further, such larvae develop through identifiable stages, and so some information is categorical. Training data might already exist from rearing experiments on larvae of the same species. Or they might be produced after the discovery, perhaps even from larvae or adults found at the scene.

In practice, this setting is complex. See [Catts \(1992\)](#). The temperature profile often is not known, and it must be guessed. How, and which, specimens are collected at the scene can affect inferences. Multiple other conditions affect the rate of decomposition and the presence and growth of insect larvae. Although the useful species seldom lay eggs on living bodies, the time elapsed since the body was exposed and when the eggs were laid is not known. Thus, even if the age of a larva were known exactly, one could say only that the body had been exposed at least that long. In forensic terms, the age of a larva provides a minimum postmortem interval.

Despite these complexities, the crux of the statistical problem is the comparison of the response from the MS to the training data at any proposed time  $t_0$ , in answer to the question, "Is it tenable to think that this specimen could be of age  $t_0$ ?" We assume that the response variable is quantitative, as opposed to categorical. We consider it to be multivariate in this paper. [LaMotte and Wells \(2016\)](#) give a corresponding development for univariate responses.

The problem has been approached in three broad ways for univariate responses, and these approaches carry over into multivariate responses, denoted here by  $y$ . The first is to fit a family of functions of  $t$ ,  $f(t; \beta)$ , to responses  $y$  to get a representation like  $\hat{y}(t) = f(t; \hat{\beta})$ , where the estimate  $\hat{\beta}$  of the vector of parameters  $\beta$  results from fitting the model to the training data  $y_1, \dots, y_n$ . Then estimate  $t_*$  by minimizing an appropriate norm of  $f(\hat{t}_*; \hat{\beta}) - y_*$  and construct an interval estimate in the form  $\hat{t}_* \pm q_\alpha \text{SE}(\hat{t}_*)$ , where  $q_\alpha$  is the  $\alpha$  quantile of the standard normal distribution or of a Student's  $t$  distribution, and  $\text{SE}(\hat{t}_*)$  is an approximate standard error of  $\hat{t}_*$ . The second approach is to fit a family of functions of  $y$ ,  $h(y; \gamma)$ , to  $t$ , resulting in  $\hat{t}(y) = h(y; \hat{\gamma})$ , and then estimate  $t_*$  as  $\hat{t}_* = h(y_*; \hat{\gamma})$ , along with a prediction interval as if  $t$  were the random response variable.

The third, like the first, fits a family of functions of  $f(t; \beta)$  to  $y_1, \dots, y_n$ . Then, for each value  $t_0$  over a range of values of  $t$ ,  $y_*$  is tested as an outlier against  $f(t_0; \hat{\beta})$ . If such tests can be had at the  $\alpha$  level of significance, then the range of values  $t_0$  not

rejected constitutes a  $100(1 - \alpha)\%$  confidence set on the true  $t_*$  (Lehmann 1959, p. 79). This is the basic approach taken here. It is the same as the approach that Oman and Wax (1984, p. 951) designate (ii) and attribute to Brown (1982), and it corresponds to Eq. (5.14) in Brown (1993, p. 88) and Eq. (2.11) in Sundberg (1999, p. 168). It is the direct extension of the test statistic for univariate  $y_*$  as an outlier at  $t_0$  in a linear regression of  $y$  on  $t$ . There, the confidence set comprises the range of values  $t_0$  such that a horizontal line at height  $y_*$  intersects the  $100(1 - \alpha)\%$  prediction interval on  $y$  at  $t_0$ .

The first two approaches yield interval estimates of  $t_*$ . The third may yield an interval, or a range fixed at only one end, or a collection of intervals, or an empty set. Under assumptions of independence, homogeneous variance, and normality, the coverage probability of confidence sets by the third approach is the nominal  $100(1 - \alpha)\%$ . There is no such finite-sample probabilistic property of the other two approaches, although simulation results indicate that in some settings they perform satisfactorily. See, for example, Krutchkoff (1967) and other works spawned by that paper. LaMotte (2014) compares coverage rates of confidence sets on  $t_*$  for the second ( $t$  vs.  $y$ ) and third ( $y$  vs.  $t$ ) approaches. In simulation results reported there, the performance characteristics of confidence sets on  $t_*$  based on the second approach degraded when the variance–covariance matrix of bivariate  $y$  was not constant in  $t$ , while the performance using the third approach did not.

There is a considerable literature on inverse prediction and calibration (a term also used for the same sort of methodology). See Osborne (1991) and Sundberg (1999) for comprehensive reviews and Brown (1993) for thorough development. For the most part, there is the tacit assumption that the variance or the variance–covariance matrix of the response is constant over  $t$ . The paper by Oman and Wax (1984) was one of the first papers to illustrate the application of multivariate calibration. Although they modeled the variance–covariance matrix as constant, they applied two adjustments to correspond to greater variance and lesser correlation as  $t$  increased. Clarke (1992) dealt with a multivariate response and a nonlinear model for the mean vector, along with constant variance–covariance matrix. Liao (2005) devised formulations of IP confidence sets based on mixed models that included random effects of batches; however, variances did not differ with the target of inference ( $t$  here). See also Brown (1982) for seminal methodological development, also with the variance assumed to be constant.

Sundberg (1999, pp. 168–169) summarizes the discussion provoked by a feature that Brown (1993, pp. 88–90) pointed out. The numerator sum of squares in the test statistic under the third approach can be expressed as the sum of two squared norms. One is of  $t_0 - \hat{t}_*$ , with  $\hat{t}_*$  as described above for the first approach. The other, which Brown (1993) denotes by  $R$ , is the squared norm of  $y_* - \hat{y}(\hat{t}_*)$ , the difference between the MS  $y_*$  and the predicted value of  $y$  given by the model fit to training data, evaluated at  $\hat{t}_*$ . This is the minimum value of the norm mentioned above in the description of the first approach to find  $\hat{t}_*$ . It is 0 if the response is univariate and the model is linear in  $t$  with non-zero slope. Otherwise, it is a measure of “multivariate inconsistency” (a term attached by a reviewer), of the failure of the model to be able to match all components of  $y_*$  simultaneously at a single value of  $t$ . Thus rejection of the proposed  $t_0$  as untenable in light of the data depends both on how close  $t_0$  is to the estimated true value  $\hat{t}_*$  and on the magnitude of this multivariate inconsistency. Brown (1993, p. 89)

says, “This unsatisfactory behavior sullies the procedure’s exact confidence property.” Sundberg (1999, p. 169) says,

...if  $R$  is high enough the region will be empty. In principle it is OK that a confidence region is empty when data do not fit the model, but here the shrinkage of the region with increasing  $R$  is misleading when we think of the size of the region as reflecting the precision of the estimation procedure. A number of alternatives without this annoying feature have been proposed. ...The time does not yet, if ever, appear ripe for declaring one region superior to the others.

Point estimation of  $t_*$  is not an explicit goal in the third approach. It addresses the question, whether it is unreasonable to assert that  $\mathbf{y}_*$  might have come from  $t_0$ . In our opinion, splitting the squared norm of the residual into these two parts is not an essential part of addressing that question. Further, we opine that “think[ing] of the size of the region as reflecting the precision of the estimation procedure,” in this context of inverse prediction, is itself off-target and slightly misleading. That interpretation is widespread, and it is at times a convenient shortcut. However, it conflates not rejecting a hypothesized value  $t_0$  with asserting that  $t_0$  could be the true value, which then comes out as saying that the true value is in the confidence set, and hence that the smaller is the confidence set, the more precise is the inference on  $t_*$ . These interpretations are based on a single realization, not on the probabilistic properties of the methodology.

The primary statistical performance characteristic of a method to construct confidence sets is its coverage profile, which describes, for each true value  $t_*$  and each possible value  $t$ , the probability that the set contains  $t$ . Ideally, given  $t_*$ , that probability would be the stated level of confidence at  $t = t_*$ , and it would fall off monotonically, the steeper the better, as  $t$  recedes from  $t_*$ . It is unclear how that performance might be related to the possibility that a set might be empty. And how a single realized set depends functionally on  $R$  does not say anything about this performance.

On the other hand, if, in some sense,  $R$  persistently accounts for a considerable part of the numerator sum of squares for the test statistic, then the assumed form of the model does not adequately mimic the path (in the univariate  $t$ ) of the parameters of the distribution of the multivariate response. As both  $R$  and  $\hat{t}_*$  depend on  $\mathbf{y}_*$ , this feature cannot be assessed with the training data alone. One diagnostic approach would be to treat each observation in the training data as a  $\mathbf{y}_*$ , fit the model to the remaining data, and compute  $R$  and the numerator sum of squares to assess the influence of this multivariate inconsistency. This may indicate that changes to the model are needed. It must be stressed, though, that  $\hat{t}_*$  and  $R$  are not produced as part of the output for standard mixed-models programs, and so an effort like this would require special ad-hoc programming.

The methods we present here extend established methods that have known, exact properties under some conditions. We do not attempt to document their performance characteristics here, but because of their grounding, there is no a priori reason to think that their performance would not be satisfactory when compared to others. Furthermore, no comprehensive methodology has otherwise been proffered, as far as we have been able to find, for constructing confidence sets in the setting that has motivated our developments.

In the setting of forensic entomology mentioned above, sizes of fly larvae and their variances and covariances change greatly with age ( $t$ ). Wells and LaMotte (1995) adapted Satterthwaite's approximation to deal with  $t$ -dependent variances, by modeling both means and variances as linear interpolates between sampling points for a univariate response. LaMotte (2014) sketched how  $t$ -dependent variance–covariance matrices could be accommodated in linear-interpolation models for multivariate responses, using the Fai and Cornelius (1996) multivariate extension of Satterthwaite's approximation. The developments in those two papers (1995, 2014) required sample means and sample variance–covariance matrices at each sampled  $t$ , and they did not consider models for the means and variance–covariance matrices beyond linear interpolation between adjacent  $t$  points.

The objective of this paper is to describe how IP can be performed with routine, default computations in widely-available mixed-models programs, for a general setting in which  $\mathbf{y}$  is multivariate and its variance–covariance matrix is modeled as varying with  $t$  and other factors, as needed. Flexible models, using polynomial splines in  $t$ , are described for both means and variances.

## 2 IP in mixed models

The formulation of the general mixed linear model shown here follows the notational conventions used in the SAS PROC MIXED documentation (SAS 2012). Suppose the response is  $q$ -variate and that there are  $n$  subjects in the training data with response vectors  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ , with observed values  $\mathbf{y}_1, \dots, \mathbf{y}_n$ . The model for the  $nq$ -vector response  $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_n)'$  is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$$

The  $nq \times qp_m$  matrix  $\mathbf{X}$  and the  $nq \times qp_v$  matrix  $\mathbf{Z}$  are fixed and known. The vector  $\boldsymbol{\beta}$  comprises unknown *fixed-effects* parameters. The *random effects* are the entries in  $\boldsymbol{\gamma}$ , which is assumed to have mean vector  $\mathbf{0}$  and variance–covariance matrix  $\text{Var}(\boldsymbol{\gamma}) = \mathbf{G}$ . The *error term*  $\boldsymbol{\varepsilon}$  is assumed to be independent of  $\boldsymbol{\gamma}$  and to have mean  $\mathbf{0}$  and  $\text{Var}(\boldsymbol{\varepsilon}) = \mathbf{R}$ . Denote the vector of realized values of  $\mathbf{Y}$  by  $\mathbf{y}$ . The matrices  $\mathbf{G}$  and  $\mathbf{R}$  generally are modeled as specified functions of a set  $\boldsymbol{\theta}$  of parameters, sometimes termed variance or covariance components, such that

$$\text{Var}(\mathbf{Y}; \boldsymbol{\theta}) = \mathbf{Z}\mathbf{G}(\boldsymbol{\theta})\mathbf{Z}' + \mathbf{R}(\boldsymbol{\theta})$$

is positive definite. Finally,  $\boldsymbol{\gamma}$  and  $\boldsymbol{\varepsilon}$  are assumed to follow jointly a multivariate normal distribution.

In the setting of inverse prediction (IP) in this framework,  $t$  is a real-valued variable, taking values  $t_1, \dots, t_n$  in the training data, and entries in corresponding rows of  $\mathbf{X}$  and  $\mathbf{Z}$  are functions of  $t$  (and perhaps of other factors as well). Schematically, the training data are  $t_i, \mathbf{y}_i$ ,  $i = 1, \dots, n$ . Each  $t_i$  defines the  $p_m$  columns of the row vector  $\mathbf{x}'_i$  and the  $p_v$  columns of the row vector  $\mathbf{z}'_i$ . The values  $t_1, \dots, t_n$  may include repetitions for multiple observations at the same value of  $t$ . As illustrated here, columns of  $\mathbf{x}'_i$

and  $z'_i$  are polynomial B-splines evaluated at  $t_i$ . From these, the matrices  $X$  and  $Z$  are as shown in the following table. Transpose of  $x$  is denoted  $x'$ , and  $\otimes$  denotes the Kronecker product.

$t$	$y$	$x$	$Z$
$t_1$	$y_1$	$I_q \otimes x'_1$	$I_q \otimes z'_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_n$	$y_n$	$I_q \otimes x'_n$	$I_q \otimes z'_n$

The question that IP addresses is: given the observed response  $y_*$  from a MS, from what value of  $t$  did it come? Or, for each potential value  $t_0$ , is it tenable to think that  $y_*$  came from the population at  $t_0$ ? This can be addressed with a p value for the null hypothesis  $H_0 : E(Y_*) = (I_q \otimes x(t_0)')\beta$  in the model above. This is the same as testing the observed value  $y_*$  as an outlier at  $t_0$ . This is a linear hypothesis, and its test statistic and p value can be had with appropriate statements and options within a general mixed linear models program. However, it is well-known that the same information can be had, usually more efficiently, from results computed by default in practically all statistical packages, as described next.

Formulate the model as above, but with  $y_*$  as an additional observation at  $t = t_0$ ; this appends rows  $y_*, I_q \otimes x'_0,$  and  $I_q \otimes z'_0$  to  $y, x,$  and  $Z,$  respectively, where  $x_0 = x(t_0)$  and  $z_0 = z(t_0)$ . Create  $q$  dummy variables  $d_0$  that are 0 in all rows except the  $q$   $t_0$  rows, where they are  $I_q,$  and include them as  $q$  fixed-effect predictor variables in the model. In the resulting computations and output, the p value for testing  $H_0$  appears as the p value for the  $q$  regression coefficients  $\delta_0$  of the dummy variables: that is, for  $H_0 : \delta_0 = \mathbf{0}$ .

Repeating this over a grid of values of  $t_0$  produces a table of values of  $t_0$  and the corresponding p values. Further refinement can be had by interpolating between these points. The set of values of  $t_0$  for which the p value is not less than .05 (say) constitutes—approximately—a 95% confidence set on  $t_*$ .

The structure of the components of the model, augmented by the grid of artificial cases and dummy variables, is shown in the following table, where  $t_{01}, \dots, t_{0g}$  constitute the grid of values of  $t_0$ .

$t$	$y$	$x$	$Z$	$d_{01}$	$\dots$	$d_{0g}$
$t_1$	$y_1$	$I_q \otimes x'_1$	$I_q \otimes z'_1$	$0_{q \times q}$	$\dots$	$0_{q \times q}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_n$	$y_n$	$I_q \otimes x'_n$	$I_q \otimes z'_n$	$0_{q \times q}$	$\dots$	$0_{q \times q}$
$t_{01}$	$y_*$	$I_q \otimes x'_{01}$	$I_q \otimes z'_{01}$	$I_q$	$\dots$	$0_{q \times q}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t_{0g}$	$y_*$	$I_q \otimes x'_{0g}$	$I_q \otimes z'_{0g}$	$0_{q \times q}$	$\dots$	$I_q$

Because of the heteroscedasticity inherent in mixed models, the probability distributions of test statistics under the null hypothesis are approximated by known distributions (usually  $F$ -distributions). The accuracy of these approximations is unknown, and it differs with the setting. Several papers have examined this topic, and they indicate that the [Kenward and Roger \(1997\)](#) approximation provides reasonable accuracy of  $p$  values. It is clear, though, that asymptotics are not much comfort here, because  $\mathbf{y}_*$  is a single observation.

An advantage of this approach, in terms of general mixed linear models, is that there is no need to have training data repeated at each sampled value  $t_i$  of  $t$ , as was required in [Wells and LaMotte \(1995\)](#) and [LaMotte \(2014\)](#). Further, partial multivariate observations, in which some components of  $\mathbf{y}$  are unobserved, are incorporated naturally within the maximum likelihood algorithm, with no need for special remedies, other than the caveat that the missing-ness may be related to the factors under study and may therefore affect inferences.

### 3 Flexible models

The use of polynomial splines to model the mean of a univariate response as a function of  $t$  is well known and widely used. The same can be done for multivariate mean vectors. Choosing the degree of the polynomial (e.g., 1 for linear interpolation, 3 for cubic interpolation) and a set of values of  $t$  to serve as knots causes predictor variables  $x_1, \dots, x_{p_m}$  to be defined so that the mean vector  $\boldsymbol{\mu}(t)$  is modeled as

$$\boldsymbol{\mu}(t) = x_1(t)\boldsymbol{\eta}_1 + \cdots + x_{p_m}(t)\boldsymbol{\eta}_{p_m}.$$

For degree  $d$  and  $k$  knots, the number of spline functions is  $p_m = d + k + 1$ . There are multiple ways to formulate these predictor variables to produce polynomial interpolation. The most widely used functions are B-splines; they are the functions used here. They have the property that the values  $x_j(t)$  are non-negative and they sum to 1. This models the mean vector  $\boldsymbol{\mu}(t)$  as a weighted average of the parameter vectors  $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{p_m}$ .

Within the context of a general mixed linear models package, the same approach can be taken to model the variance–covariance matrix. For the variance–covariance matrix  $\Sigma(t)$  of a  $q$ -variate observation on a single subject at time  $t$ , the model takes the form

$$\Sigma(t) = w_1(t)\Gamma_1 + \cdots + w_{p_v}(t)\Gamma_{p_v},$$

where  $w_1, \dots, w_{p_v}$  are defined by B-splines in  $t$  specified by degree and knots. The variance–covariance matrix of the vector of all observations is block diagonal with  $\Sigma(t_i), i = 1, \dots, n$ , on the diagonal. This is  $ZG(\boldsymbol{\theta})Z'$  in the general form, and  $R(\boldsymbol{\theta})$  is not needed.

As with the means, the  $w_j(t)$ s are non-negative and they sum to 1. This property is particularly useful for modeling variance–covariance matrices, because nonnegative-definite matrices are closed under such convex combinations.

In the model for the mean vector, the coefficients  $\eta_1, \dots, \eta_{p_m}$  are unknown parameters, and they are components of  $\beta$  in the general formulation. Similarly, the symmetric matrices  $\Gamma_1, \dots, \Gamma_{p_v}$  are unknown parameters; they are components of  $\theta$ .

In concept, the steps in this process leading to confidence sets on  $t_*$  are clear. For the MS observation  $y_*$ , create multiple artificial cases with response  $y_*$  at each value in a grid  $t_{01}, \dots, t_{0g}$ , and create the  $gq$  corresponding dummy variables. Choose polynomial degrees and knots for the models of the mean vector and variance–covariance matrix. Compute  $x_1(t), \dots, x_{p_m}(t)$  and  $w_1(t), \dots, w_{p_v}(t)$  for each value of  $t$  in the training-data set and the grid. Compose  $X$  and  $Z$  from these. The next step is the computation of maximum likelihood estimates and approximate p values for the coefficients  $\delta_{01}, \dots, \delta_{0g}$  of the dummy variables corresponding to the grid on  $t_0$ .

By far the best way to accomplish the computations is to use a tested, stable, debugged program like PROC MIXED in SAS, which implements the Kenward–Roger approximation to variance and degrees of freedom, or corresponding programs in other packages. To use such programs requires specifying the particular model with the syntax of the program.

Commonly, programs communicate in terms of fixed effects, random effects, repeated measures, and factor and interaction effects. While the mathematical models, as described above, are clear, still they must be translated carefully in the program's syntax in order to specify exactly the desired model. In most applications, some data handling is necessary to compose  $X$  and  $Z$ , including the polynomial splines and extra rows and columns for the dummy variables. In order to get the desired model for the variance–covariance matrix, columns of  $Z$  are composed as square roots of the B-spline variables,  $z_j(t) = \sqrt{w_j(t)}$ , and declared to represent random-coefficient effects. For a  $q$ -component response vector, the model must provide for different sets of coefficients in  $\beta$  and different random coefficients for each component of the response, and the forms of variance–covariance matrices for the response and for the random coefficients must be specified. The details will differ from package to package.

## 4 Illustration

Following is an illustration. The response  $y$  is bivariate. Data were simulated from a 'true' model in which the two components of the mean vector and the three components of the variance–covariance matrix changed smoothly with  $t$  going from 0 to 10.

The mathematical model from which the observations were generated was configured to resemble characteristics of size measurements (length and weight, for example) found in fly larvae. They start small, with small variances, go through a period of rapid growth in both dimensions during which variances increase and correlations shift; and as they approach pupation, they cease feeding and their growth slows, and their size may even decrease. The time scale has been shifted and scaled in this simulation to range between 0 and 10; in growth experiments, the first measurements would typically be taken soon after egg-hatch, but not at  $t = 0$ .

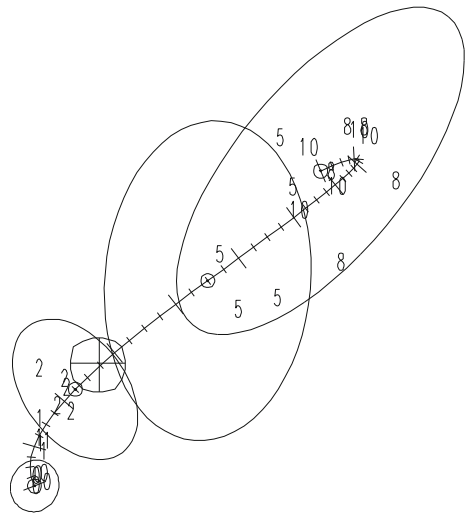
The model for the mean vectors is in terms of cubic B-splines in  $t$  with a single interior knot at  $t = 5$ . This generates  $qp_m = 2 \times 5$  columns in  $x$ . The matrix  $Z$  has  $q \times p_v = 2 \times 5$  columns, in five pairs, each with a pair of random coefficients in



**Table 1** P values ( $Pr > F$ ) for tests of  $y_*$  as an outlier at each  $t_0 = 0, \dots, 10$ . Results from PROC MIXED using Kenward-Roger approximation and degrees of freedom

Effect	Type 3 tests of fixed effects		F Value	Pr > F
	Num	Den		
	DF	DF		
d00*comp	2	3.02	193.21	0.0007
d01*comp	2	3.26	181.83	0.0005
d02*comp	2	4.73	6.10	0.0490
d03*comp	2	6.02	0.33	0.7341
d04*comp	2	8.11	3.48	0.0810
d05*comp	2	9.1	7.29	0.0129
d06*comp	2	8.65	11.91	0.0033
d07*comp	2	6.45	17.51	0.0025
d08*comp	2	4.13	22.42	0.0060
d09*comp	2	3.96	26.04	0.0053
d10*comp	2	3.01	17.34	0.0224

**Fig. 1** Training data comprise 5 observations on bivariate  $y$  at each of  $t = 0, 1, 2, 5, 8, 10$ , indicated by plotted numerals. The curved line is the locus of the estimated means. Hash marks indicate  $t$  between 0 and 10 in increments of 0.25. Ellipses depict 95% prediction regions at  $t = 0, 2.25, 4.5$ , and 10, centered at the small circles. The MS is indicated by the large  $\oplus$ . The 95% confidence set on its  $t_*$  extends from about  $t = 2$  to about  $t = 4.25$



the model. Entries in  $Z$  are square roots of cubic B-splines with a single interior knot at  $t = 5$ . The training data set comprised 5 observations each at values 0, 1, 2, 5, 8, and 10 of  $t$ . Dummy variables  $d_{01}I_2, \dots, d_{10}I_2$  and corresponding artificial cases were created for  $t_0$  in increments of 0.25 between 0 and 10. In Table 1, the two components of  $y$  are indexed by `comp`, which takes values 1 and 2 for components 1 and 2. The response  $y_*$  for the MS was generated from the population at  $t_* = 3$ .

Table 1 shows only the tests of the coefficients of the dummy variables from SAS PROC MIXED. Based on the p values in the right-most column, at the 5% level of significance, only  $t_0 = 3$  and 4 are tenable values of  $t_*$ .

Figure 1 shows the data used in this example, and it illustrates estimates of means and variance–covariance matrices (corresponding to the ellipses). It illustrates the p values by showing nominal 95% prediction ellipsoids at selected values of  $t$ .

**Acknowledgments** This project was supported by Award No. 2013-DN-BX-K042, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

## References

- Brown PJ (1982) Multivariate calibration. *J R Stat Soc Ser B* 44:287–321
- Brown PJ (1993) Measurement, regression, and calibration. Clarendon Press, Oxford
- Catts EP (1992) Problems in estimating the postmortem interval in death investigations. *J Agric Entomol* 9(4):245–255
- Clarke GPY (1992) Inverse estimates from a multiresponse model. *Biometrics* 48(4):1081–1094
- Fai AHT, Cornelius PL (1996) Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *J Stat Comput Simul* 54:363–378
- Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53:983–997
- Krutchkoff RG (1967) Classical and inverse regression methods of calibration. *Technometrics* 9(3):425–439
- LaMotte LR (2014) On inverse prediction in mixed linear models. *Commun Stat Simul Comput* 43(9):2106–2116
- LaMotte, LR, and Wells JD (2016). Inverse prediction for heteroscedastic response using mixed models software. *Commun Stat Simul Comput*. (Published online 21 Nov 2015)
- Lehmann EL (1959) Testing hypotheses. Wiley, New York
- Liao JJZ (2005) A linear mixed-effects calibration in qualifying experiments. *J Biopharm Stat* 15:3–15
- Oman SD, Wax Y (1984) Estimating fetal age by ultrasound measurements: an example of multivariate calibration. *Biometrics* 40:947–960
- Osborne C (1991) Statistical calibration: a review. *Int Stat Rev* 59(3):309–336
- Sundberg R (1999) Multivariate calibration—direct and indirect regression methodology. *Scand J Stat* 26:161–207
- SAS Institute Inc. (2012) SAS 9.4. Cary NC, USA
- Wells JD, LaMotte LR (1995) Estimating maggot age from weight using inverse prediction. *J Forensic Sci JFSCA* 40(4):585–590