Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 3164–3175
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Prediction intervals for general balanced linear random models

T.Y. Lin[a,*], C.T. Liao[b]

[a]*Department of Applied Mathematics, Feng Chia University, Taichung 40724, Taiwan*
[b]*Division of Biometry, Institute of Agronomy, National Taiwan University, Taipei 10617, Taiwan*

Received 18 April 2007; received in revised form 31 December 2007; accepted 8 January 2008
Available online 17 January 2008

Abstract

The main interest of prediction intervals lies in the results of a future sample from a previously sampled population. In this article, we develop procedures for the prediction intervals which contain all of a fixed number of future observations for general balanced linear random models. Two methods based on the concept of a generalized pivotal quantity (GPQ) and one based on ANOVA estimators are presented. A simulation study using the balanced one-way random model is conducted to evaluate the proposed methods. It is shown that one of the two GPQ-based and the ANOVA-based methods are computationally more efficient and they also successfully maintain the simulated coverage probabilities close to the nominal confidence level. Hence, they are recommended for practical use. In addition, one example is given to illustrate the applicability of the recommended methods.

Keywords: Generalized confidence interval; Generalized p-value; Variance component

1. Introduction

Prediction is one of the most important statistical inferences. A statistical interval which contains the values of a specified number of future observations with a fixed level of confidence is known as a prediction interval. There are meaningful applications for such intervals in various practical problems arising in quality control, industrial production process, environmental monitoring, and many other areas. For example, a manufacturer may wish to establish an interval that, with a high degree of confidence, will contain the performance values for some future units of a particular type of product, based on the observed performance of similar past units (Hahn and Meeker, 1991). In the development of environmental monitoring procedures, the simultaneous prediction interval approach is often interesting to many investigators and governmental regulation agencies (Davis and McNichols, 1987, and references cited therein).

Despite their practical importance, prediction intervals have received only limited attention in standard textbooks, primarily in the context of regression problems. The reader may refer to the book by Miller (1981). The most basic problem of computing prediction intervals is for the simple univariate normal distribution and this has been extensively studied in the literature. Typically, there are two types of prediction intervals discussed. One is to contain all the specified number of future observations, say m, e.g., see Chew (1968) and Hahn (1969, 1970). The other is to contain at least k out of m future observations, e.g., see Hall and Prairie (1973), Fertig and Mann (1977) and Odeh (1990). Additional work on simultaneous prediction intervals can be found in Chou and Owen (1986) and Davis and McNichols (1987).

* Corresponding author. Tel.: +886 4 24517250; fax: +886 4 24510801.
E-mail address: linty@fcu.edu.tw (T.Y. Lin).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.01.001

An excellent review of construction and application of such prediction intervals with other univariate distributions like lognormal, exponential and Weibull can be found in Hahn and Meeker (1991).

For more complex situations, to the best of our knowledge, very little work has been done on prediction intervals. Only Wang (1992) considers the balanced one-way random model and provides good approximate prediction intervals. In the same article, he specifically imposes a certain structure on the covariance matrix of the future observations to avoid complex computations required by his method. In particular, he takes into account only two special cases, one in which future observations are restricted to be from a single batch, and another involving multiple batches, but with exactly one observation per batch.

In this article, we generalize the problem as follows. We seek prediction intervals which contain all the m future observations for general balanced linear random models. Our procedures are developed mainly based on the concept of a generalized pivotal quantity (GPQ), which has been frequently used to obtain confidence intervals in situations where conventional methods are difficult to apply or fail to provide good solutions. The GPQ and its counterpart generalized test variable are introduced by Weerahandi (1993) and Tsui and Weerahandi (1989), respectively. The resulting generalized confidence interval (GCI) and generalized p-value are now available in the literature for many applications. See, for instance, Weerahandi (1995), Khuri et al. (1998), Hamada and Weerahandi (2000), Liao and Iyer (2004), Iyer et al. (2004), Liao et al. (2005), Mathew and Webb (2005), Lin and Liao (2006), Lin et al. (2007) and Burch (2007). The current study focuses on prediction intervals rather than confidence intervals.

The rest of this article is organized as follows. Section 2 introduces the definitions of the prediction interval of interest and GPQ. Section 3 proposes methods for constructing prediction intervals for general balanced linear random models and contains an exploratory simulation study to compare the performance of the proposed methods. Based on the results of this study one GPQ-based and one ANOVA-based methods are recommended for practical use. Section 4 gives an example to illustrate the applicability of the recommended methods. A more detailed simulation study is conducted to further evaluate their performance in Section 5. Discussion and final remarks are provided in the last section.

2. Preliminaries

In this section, we formally define the prediction interval of interest and review the concept of a GPQ.

A prediction interval can be defined and interpreted as follows. Let $Y_1, Y_2, \ldots, Y_n$ be an observed (past) random sample from a distribution F. Consider two statistics $L(\boldsymbol{Y})=L(Y_1, Y_2, \ldots, Y_n)$ and $U(\boldsymbol{Y})=U(Y_1, Y_2, \ldots, Y_n)$ satisfying $L(\boldsymbol{Y}) < U(\boldsymbol{Y})$ with probability one. Moreover, let $Y_1^\star, Y_2^\star, \ldots, Y_m^\star$ be a future random sample from the same distribution F. It is assumed that the future sample is drawn independently of the past sample. An interval $[L(\boldsymbol{Y}), U(\boldsymbol{Y})]$ is said to be a two-sided $100\gamma\%$ prediction interval containing all the m future observations $Y_1^\star, Y_2^\star, \ldots, Y_m^\star$ if

$$Pr[L(\boldsymbol{Y}) < Y_i^\star < U(\boldsymbol{Y}); i = 1, 2, \ldots, m] = \gamma,$$

where γ is called the *prediction probability*. One can assert that a $100\gamma\%$ prediction interval procedure, in the long run, will be correct $100\gamma\%$ of times in claiming that all the m future observations will be contained within the prediction interval. The corresponding one-sided prediction interval can be defined in a same manner.

Let $\boldsymbol{y}$ be the realized value of the observable random vector $\boldsymbol{Y}$ and $\boldsymbol{\xi}$ be the vector of model parameters. Furthermore, let ϕ be a function of $\boldsymbol{\xi}$ for which a confidence interval is sought. According to Weerahandi (1993), a function $R=r(\boldsymbol{Y}; \boldsymbol{y}, \boldsymbol{\xi})$ of $\boldsymbol{Y}$, $\boldsymbol{y}$ and $\boldsymbol{\xi}$ is called a GPQ for ϕ if it satisfies the following two conditions:

(i) R has a probability distribution that is free of unknown parameters.
(ii) The observed R, namely $r = R(\boldsymbol{y}; \boldsymbol{y}, \boldsymbol{\xi})$, depends on $\boldsymbol{\xi}$ only through ϕ. Namely, r is only a function of $(\boldsymbol{y}, \phi)$.

A two-sided equal tailed generalized $(1-\alpha)$-confidence interval for ϕ is given by $\{\phi : R_{\alpha/2} \leqslant r \leqslant R_{1-\alpha/2}\}$, where R_α is the (100α)th percentile of the distribution of R. The required percentiles of R can be estimated using Monte-Carlo sampling. Specifically, if the observed quantity $r = \phi$, then the GPQ is called the *fiducial GPQ* and the GCI based on such GPQs, under some mild conditions, are proven to have asymptotically correct frequentist coverage probability (Hanning et al., 2006). Since the observed quantity $r = \phi$, a location estimate such as the mean or median of a Monte Carlo sampling distribution of the fiducial GPQ should serve as a reasonable point estimate for ϕ. Liao et al. (2005) use the median for such an application of fiducial GPQs in construction of tolerance intervals. In the current study, the proposed GPQ-based methods for the prediction intervals are also based on this concept.

3. Proposed prediction interval methods

In this study, we restrict ourselves to general balanced linear random effects models and are interested in prediction intervals which contain all the m future observations for a pre-specified value m. Following the textbook of Graybill (1976), a general linear random model with balanced data can be described as follows:

$$\boldsymbol{Y} = \mathbf{1}_n \mu + \sum_{i=1}^{r} \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}, \tag{3.1}$$

where $\boldsymbol{Y}$ is an $n \times 1$ observable random vector; μ is the overall mean; and $\mathbf{1}_n$ is the vector of length n with all entries equal to 1; $\boldsymbol{\beta}_i$ is a random vector and $\boldsymbol{X}_i$ is a Kronecker product of identity matrices and the vectors with all entries equal to 1. Moreover, it is assumed that the random error $\boldsymbol{\epsilon}$ and the random effects $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_r$ are pairwise uncorrelated, $\boldsymbol{\epsilon}$ follows a normal distribution $\mathrm{N}(\mathbf{0}_n, \sigma_e^2 \boldsymbol{I}_n)$ and $\boldsymbol{\beta}_i$ also follows a normal distribution $\mathrm{N}(\mathbf{0}_{q_i}, \sigma_i^2 \boldsymbol{I}_{q_i})$, for $i = 1, 2, \ldots, r$. Here $\mathbf{0}_n$ and $\boldsymbol{I}_n$ denote the zero vector of length n and the identity matrix of order n, respectively, and q_i denotes the number of levels of $\boldsymbol{\beta}_i$. For convenience, let σ_{r+1}^2 denote σ_e^2. Then, the observable total sum of squares for the model can always be partitioned into quadratic forms $\boldsymbol{Y}'\boldsymbol{A}_i\boldsymbol{Y}$ such that

(1) $U_i = S_i^2/\psi_i^2 = \boldsymbol{Y}'\boldsymbol{A}_i\boldsymbol{Y}/\psi_i^2$ is distributed as $\chi_{f_i}^2$, for $i = 1, 2, \ldots, r+1$, where ψ_i^2 are linear combinations of $\sigma_1^2, \sigma_2^2, \ldots, \sigma_{r+1}^2$; matrices $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_{r+1}$ are nonnegative definite; and $\chi_{f_i}^2$ denotes the chi-squared distribution with degrees of freedom f_i;
(2) $U_0 = (\overline{Y} - \mu)^2/\tau^2$ is distributed as χ_1^2, where $\overline{Y}$ denotes the sample mean of the data and τ^2 denotes the variance of $\overline{Y}$; and
(3) $U_0, U_1, \ldots, U_{r+1}$ are jointly independent.

Notice that $\overline{Y}, S_1^2, S_2^2, \ldots, S_{r+1}^2$ are complete sufficient statistics for a general balanced linear random model. Also, there is a one-to-one relationship between $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{r+1}^2)$ and $(\psi_1^2, \psi_2^2, \ldots, \psi_{r+1}^2)$, namely, there exist constants c_{ij} such that $\sigma_i^2 = \sum_{j=1}^{r+1} c_{ij}\psi_j^2$, $i = 1, 2, \ldots, r+1$, for a balanced linear random model.

We seek a two-sided prediction interval with prediction probability γ to contain all the m future observations $Y_1^\star, Y_2^\star, \ldots, Y_m^\star$, based on the n original observations $\boldsymbol{Y}$. In particular, we seek a prediction interval of the form

$$Pr[\overline{Y} - D < Y_i^\star < \overline{Y} + D; i = 1, 2, \ldots, m] = \gamma, \tag{3.2}$$

where $\overline{Y}$ is the sample mean of the original observations and D is the margin of error that needs to be determined. Expression (3.2) can be rewritten as

$$Pr[-D < Y_i^\star - \overline{Y} < D; i = 1, 2, \ldots, m] = \gamma. \tag{3.3}$$

The original n observations $Y_1, Y_2, \ldots, Y_n$ and the additional m future observations $Y_1^\star, Y_2^\star, \ldots, Y_m^\star$ are assumed to be independent. Also, they are all sampled from the same random model. Let $\boldsymbol{Y}^\star = (Y_1^\star, Y_2^\star, \ldots, Y_m^\star)'$. From model (3.1), we have $\boldsymbol{Y}^\star \sim \mathrm{N}(\mu\mathbf{1}_m, \boldsymbol{\Sigma}^\star)$, where $\boldsymbol{\Sigma}^\star$ denotes the covariance matrix of $\boldsymbol{Y}^\star$ that is a function of the variance components. Also let $\overline{Y} \sim \mathrm{N}(\mu, \tau^2)$. It follows that

$$\hat{\boldsymbol{\theta}} = \boldsymbol{Y}^\star - \overline{Y}\mathbf{1}_m \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{3.4}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\star + \tau^2 \boldsymbol{J}_m$. Here $\boldsymbol{J}_m$ is a square matrix of order m with all entries equal to 1.

First, suppose $\boldsymbol{\Sigma}$ is known. From (3.3) and (3.4), the problem under consideration reduces to finding a margin error D satisfying

$$\gamma = \int_{-D}^{D}\int_{-D}^{D}\cdots\int_{-D}^{D} f(\hat{\boldsymbol{\theta}})\,\mathrm{d}\hat{\boldsymbol{\theta}} = \int_{-D}^{D}\int_{-D}^{D}\cdots\int_{-D}^{D} (2\pi)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\mathrm{e}^{-(1/2)\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}\,\mathrm{d}t_1\,\mathrm{d}t_2\ldots\mathrm{d}t_m,$$

where $f(\hat{\boldsymbol{\theta}})$ is the joint density function of $\hat{\boldsymbol{\theta}}$. Equivalently, D can be represented as $\Phi_{1+\gamma/2}^{-1}(\mathbf{0}_m, \boldsymbol{\Sigma})$, the $(100\frac{1+\gamma}{2})$th "percentile" of the multivariate normal cdf with mean $\mathbf{0}_m$ and covariance matrix $\boldsymbol{\Sigma}$, denoted by $\Phi(\mathbf{0}_m, \boldsymbol{\Sigma})$.

Then the interval

$$\overline{Y} \pm \Phi^{-1}_{\frac{1+\gamma}{2}}(\mathbf{0}_m, \boldsymbol{\Sigma}) \tag{3.5}$$

is the required two-sided prediction interval. However, the two-sided equi-coordinate critical point D depends upon the covariance matrix $\boldsymbol{\Sigma}$ which is a function of unknown variance components $\sigma_1^2, \sigma_2^2, \ldots, \sigma_{r+1}^2$. A natural approach is to replace $\boldsymbol{\Sigma}$ by a suitable estimate of it and then compute D treating $\boldsymbol{\Sigma}$ as known. Still, the numerical difficulty of working with the multivariate normal probability integral to find D needs to be overcome. We shall use the numerical method and the computer program provided by Genz (1999) to resolve this problem.

We now present three different methods for obtaining the critical value D.

3.1. Method I

It may be reasonable to estimate D by the median of the distribution of a GPQ for D, i.e., a GPQ of $\Phi^{-1}_{\frac{1+\gamma}{2}}(\mathbf{0}_m, \boldsymbol{\Sigma})$. For this, we first need to obtain GPQs for the variance components $\sigma_1^2, \sigma_2^2, \ldots, \sigma_{r+1}^2$. This can be done as follows. GPQs for $\psi_1^2, \psi_2^2, \ldots, \psi_{r+1}^2$ are given by

$$R_{\psi_i^2} = \frac{s_i^2}{U_i} = \frac{\psi_i^2 s_i^2}{S_i^2}, \quad i = 1, 2, \ldots, r+1, \tag{3.6}$$

where $s_1^2, s_2^2, \ldots, s_{r+1}^2$ denote the observed values of $S_1^2, S_2^2, \ldots, S_{r+1}^2$. Hence,

$$R_{\sigma_i^2} = \sum_{j=1}^{r+1} c_{ij} R_{\psi_j^2}, \quad i = 1, 2, \ldots, r+1. \tag{3.7}$$

From the first expression of (3.6), $R_{\psi_i^2}$ has distribution that is free of model parameters. When $s_1^2, s_2^2, \ldots, s_{r+1}^2$ are substituted for the observable random variables $S_1^2, S_2^2, \ldots, S_{r+1}^2$ in the second expression of (3.6), $R_{\psi_i^2}$ become ψ_i^2. Hence $R_{\psi_i^2}$ and $R_{\sigma_i^2}$ fulfill the requirements for being GPQs for ψ_i^2 and σ_i^2, respectively. Similarly, we subsequently substitute $R_{\boldsymbol{\Sigma}}$ in $\Phi(\mathbf{0}_m, \boldsymbol{\Sigma})$ so as to obtain the GPQ $R_{\Phi(\mathbf{0}_m, \boldsymbol{\Sigma})}$ given by

$$R_{\Phi(\mathbf{0}_m, \boldsymbol{\Sigma})} = \Phi(\mathbf{0}_m, R_{\boldsymbol{\Sigma}}). \tag{3.8}$$

Consequently, the required median of the distribution of the GPQ of $\Phi^{-1}_{\frac{1+\gamma}{2}}(\mathbf{0}_m, \boldsymbol{\Sigma})$ may be estimated using the following Monte Carlo algorithm.

Step 1: Choose a large simulation sample size, say $K = 10,000$. For k equal to 1 through K, carry out the following steps.

Step 2: Generate mutually independent chi-squared random deviates $U_{1,k}, U_{2,k}, \ldots, U_{r+1,k}$ with $f_1, f_2, \ldots, f_{r+1}$ degrees of freedom, respectively.

Step 3: Calculate $R_{\psi_i^2,k}$ using (3.6) for $i = 1, 2, \ldots, r+1$.

Step 4: Calculate $R_{\sigma_i^2,k}$ using (3.7) for $i = 1, 2, \ldots, r+1$.

Step 5: Calculate $R_{\boldsymbol{\Sigma},k}$ from substituting $R_{\sigma_i^2,k}$ of Step 4 in $\boldsymbol{\Sigma}$.

Step 6: Applying the numerical method of Genz (1999), calculate the percentile $R_{\Phi^{-1}_{\frac{1+\gamma}{2}}(\mathbf{0}_m, \boldsymbol{\Sigma}),k}$ using (3.8).

D is estimated by the median of the $K = 10,000$ realizations of $R_{\Phi^{-1}_{\frac{1+\gamma}{2}}(\mathbf{0}_m, \boldsymbol{\Sigma}),k}$ generated from Step 6, because the distribution of the realizations appears to be skewed.

Remark. The above procedure is computationally intensive. In particular, the need for calculating D, a large number of times using the code of Genz (1999) in Step 6 imposes a severe computational burden. Hence there is a need to develop some other practical alternatives. The following two plug-in methods use estimates of the variance components in $\boldsymbol{\Sigma}$ so as to save much computing time.

3.2. Method II

ANOVA estimates of $\sigma_1^2, \sigma_2^2, \ldots, \sigma_{r+1}^2$ are natural candidates for use in estimating $\boldsymbol{\Sigma}$. Then $\Phi^{-1}_{\frac{1+\gamma}{2}}(\mathbf{0}_m, \boldsymbol{\Sigma})$ can be estimated by the corresponding percentile of the resulting $\Phi(\mathbf{0}_m, \hat{\boldsymbol{\Sigma}})$ using Genz (1999) method, treating $\hat{\boldsymbol{\Sigma}}$ as known. However, the coverage probability is anticipated to be not sufficient based on the multivariate normal distribution for the small sample sizes, because the uncertainty in the ANOVA estimator is not incorporated. Therefore, we consider replacing the multivariate normal distribution with a multivariate t distribution here. The two-sided prediction interval in (3.5) is thus adapted to be

$$\overline{Y} \pm T^{-1}_{\frac{1+\gamma}{2}}(\hat{\boldsymbol{\Sigma}}, \nu), \tag{3.9}$$

where $T^{-1}_{\frac{1+\gamma}{2}}(\hat{\boldsymbol{\Sigma}}, \nu)$ is the $(100\frac{1+\gamma}{2})$th "percentile" of the multivariate t cdf with covariance matrix $\hat{\boldsymbol{\Sigma}}$ and degrees of freedom ν, denoted by $T(\hat{\boldsymbol{\Sigma}}, \nu)$. Note that ν is determined by the Satterthwaite (1946) approximation and the required percentile of the multivariate t cdf is obtained from the algorithm presented in Genz and Bretz (2002). It is interesting to point out that the resulting interval for the specific case of univariate normal distribution using this method is reduced to the exact prediction interval.

3.3. Method III

A GPQ-based method is also considered. We use the same Monte Carlo algorithm of Steps 1–5 of Method I to generate $K = 10,000$ realizations of $R_{\boldsymbol{\Sigma},k}$. Then $\boldsymbol{\Sigma}$ is replaced with $\hat{\boldsymbol{\Sigma}}$ which is taken to be the average $\sum_{k=1}^{K} R_{\boldsymbol{\Sigma},k}/K$, because the median is not well defined for a matrix of random variables. Then $\Phi^{-1}_{\frac{1+\gamma}{2}}(\mathbf{0}_m, \boldsymbol{\Sigma})$ is computed using the resulting $\Phi(\mathbf{0}_m, \hat{\boldsymbol{\Sigma}})$ in Genz (1999) method, treating $\hat{\boldsymbol{\Sigma}}$ as known.

The following simulation study is conducted to explore the performance of the proposed methods. Consider the balanced one-way random model given by

$$Y_{ij} = \mu + \alpha_i + e_{ij} \tag{3.10}$$

for $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, where μ is the constant term, α_i the batch effects and e_{ij} the measurement errors. We suppose α_i and e_{ij} are random effects that are normally distributed with means 0 and variances equal to σ_α^2 and σ_e^2, respectively. Also, let $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$ denote the intraclass correlation coefficient. Using the notation described earlier, let $\psi_1^2 = b\sigma_\alpha^2 + \sigma_e^2$, $\psi_2^2 = \sigma_e^2$, $S_1^2 = \text{SSA}$ (sum of square between batches) with $S_1^2/\psi_1^2 \sim \chi^2_{a-1}$ and $S_2^2 = \text{SSE}$ (error sum of square) with $S_2^2/\psi_2^2 \sim \chi^2_{a(b-1)}$.

First, we choose $\gamma = 0.95$, $\sigma_\alpha^2 = 1.0$ and let ρ take values from 0.1 through 0.9 in increments of 0.2. For each fixed past sample (fixed a, b and $n = ab$) and a given parameter combination, the observed data ($\bar{y}$, ssa, sse) are generated according to $\overline{Y} \sim \text{N}(0, (b\sigma_\alpha^2 + \sigma_e^2)/n)$, $\text{SSA} \sim (b\sigma_\alpha^2 + \sigma_e^2)\chi^2_{a-1}$ and $\text{SSE} \sim \sigma_e^2\chi^2_{n-a}$. The value of μ is set to 0 without any loss in generality. Then, we repeat the procedure 10,000 times for the setting. For prediction, we consider $m = a^\star b^\star$ future observations, where $a^\star$ and $b^\star$ separately denote the number of batches and the number of observations within each single batch, respectively. Hence, the covariance matrix of $\boldsymbol{Y}^\star$ is given by $\boldsymbol{\Sigma}^\star = \sigma_\alpha^2(\boldsymbol{I}_{a^\star} \otimes \boldsymbol{J}_{b^\star}) + \sigma_e^2(\boldsymbol{I}_{a^\star} \otimes \boldsymbol{I}_{b^\star})$, where $\otimes$ denotes the Kronecker product of matrices. Note that $\boldsymbol{\Sigma}^\star = (\boldsymbol{I}_{a^\star} \otimes \boldsymbol{J}_{b^\star}) + ((1-\rho)/\rho)(\boldsymbol{I}_{a^\star} \otimes \boldsymbol{I}_{b^\star})$ in the setting. The future observations are thus generated according to the specified $\text{N}(\mathbf{0}_m, \boldsymbol{\Sigma}^\star)$. For each of the three proposed methods, the percentage of the 10,000 simulation runs where the computed prediction interval contains all m future samples and the average length of the interval are recorded. Table 1 displays the simulation results.

From Table 1, Method I appears to be liberal in all the cases of $a \leqslant 9$. When $a = 3$, Method II is slightly liberal especially when the value of ρ gets large. On the other hand, Method III appears to be conservative for these cases. For $a \geqslant 5$, both Methods II and III are quite successful in maintaining the simulated confidence levels close to the nominal value $\gamma = 0.95$ and provide reasonable expected lengths. For the large sample size of $a = 25$, all the three methods perform equally well. In general, Methods II and III have better performance than Method I. Also Method I is computationally inefficient. Hence we recommend Methods II and III for practical use.

Table 1
Simulated confidence coefficients ($\times 10^4$) and expected lengths (in the parentheses) for 95% two-sided prediction intervals

a	b	$a^\star$	$b^\star$	Method	ρ				
					0.1	0.3	0.5	0.7	0.9
3	2	1	2	I	9223 (18.14)	9178 (10.48)	9106 (8.06)	9019 (6.64)	8795 (5.60)
				II	9470 (22.15)	9456 (13.57)	9415 (11.21)	9421 (10.24)	9335 (9.85)
				III	9915 (41.71)	9893 (25.09)	9898 (20.40)	9886 (17.54)	9825 (15.87)
3	2	2	1	I	9343 (18.76)	9176 (10.78)	9146 (8.37)	8899 (6.93)	8512 (5.94)
				II	9567 (23.28)	9448 (14.31)	9432 (12.11)	9322 (11.27)	9231 (11.31)
				III	9932 (44.07)	9900 (26.29)	9892 (21.80)	9847 (19.02)	9772 (17.54)
5	2	1	2	I	9377 (16.30)	9334 (9.42)	9302 (7.25)	9220 (5.98)	9088 (5.04)
				II	9510 (17.47)	9491 (10.30)	9493 (8.19)	9458 (7.07)	9485 (6.38)
				III	9693 (19.46)	9673 (11.35)	9658 (8.84)	9600 (7.41)	9542 (6.39)
5	2	2	1	I	9421 (16.58)	9359 (9.62)	9241 (7.40)	9132 (6.17)	8922 (5.38)
				II	9544 (17.78)	9493 (10.59)	9443 (8.46)	9422 (7.47)	9394 (7.10)
				III	9704 (19.72)	9689 (11.58)	9624 (9.04)	9564 (7.69)	9440 (6.86)
7	2	1	2	I	9422 (15.71)	9375 (9.02)	9353 (6.88)	9298 (5.74)	9296 (4.86)
				II	9508 (16.34)	9479 (9.50)	9495 (7.40)	9489 (6.36)	9568 (5.62)
				III	9625 (17.31)	9593 (9.99)	9589 (7.67)	9565 (6.47)	9572 (5.55)
7	2	2	1	I	9421 (15.82)	9333 (9.12)	9354 (7.07)	9236 (5.92)	9100 (5.13)
				II	9510 (16.45)	9452 (9.64)	9503 (7.67)	9483 (6.66)	9464 (6.10)
				III	9631 (17.38)	9578 (10.08)	9582 (7.88)	9538 (6.68)	9437 (5.88)
9	2	1	2	I	9428 (15.31)	9415 (8.81)	9377 (6.75)	9354 (5.60)	9238 (4.75)
				II	9474 (15.72)	9489 (9.14)	9482 (7.12)	9488 (6.04)	9467 (5.28)
				III	9582 (16.35)	9580 (9.45)	9540 (7.28)	9529 (6.08)	9464 (5.21)
9	2	2	1	I	9417 (15.44)	9418 (8.94)	9374 (6.88)	9310 (5.78)	9160 (5.05)
				II	9473 (15.84)	9500 (9.28)	9500 (7.29)	9484 (6.29)	9474 (5.74)
				III	9574 (16.45)	9593 (9.57)	9556 (7.41)	9505 (6.28)	9431 (5.56)
25	2	1	2	I	9466 (14.52)	9484 (8.39)	9442 (6.43)	9428 (5.35)	9429 (4.56)
				II	9489 (14.62)	9507 (8.47)	9478 (6.53)	9495 (5.48)	9500 (4.72)
				III	9524 (14.79)	9533 (8.55)	9501 (6.57)	9495 (5.48)	9488 (4.69)
25	2	2	1	I	9506 (14.58)	9465 (8.43)	9407 (6.52)	9449 (5.50)	9438 (4.84)
				II	9518 (14.66)	9492 (8.51)	9453 (6.63)	9505 (5.65)	9556 (5.04)
				III	9547 (14.82)	9515 (8.59)	9465 (6.66)	9505 (5.63)	9525 (4.98)

4. An illustrative example

In this section, we present an example to illustrate the applicability of the recommended methods (Methods II and III).

4.1. Example 1. A gauge repeatability and reproducibility experiment

A common industrial application is to use a designed experiment to study the components of variability in a measurement system. These studies are often called gauge repeatability and reproducibility (R & R) studies because these are the components of variability that are of interest. An example of a gauge R & R study is taken from the paper by Houf and Berman (1988). The data are measurements on thermal impedance (in °C per Watt × 100) on a power module for an induction motor starter. There are ten parts, three operators, and three measurements per part. The data can be fitted by the following random model:

$$Y_{ijk} = \mu + P_i + O_j + (PO)_{ij} + \varepsilon_{ijk}$$

for $i = 1, \ldots, a$; $j = 1, \ldots, b$ and $k = 1, \ldots, c$, where P_i, O_j, $(PO)_{ij}$ and ε_{ijk} are all independent random variables separately representing the effect of part, the effect of operator, the interaction effect between part and operator and the random error. Typically, it is assumed that P_i, O_j, $(PO)_{ij}$, and ε_{ijk} are normally distributed with means 0 and variances σ_{P}^2, σ_{O}^2, σ_{PO}^2 and σ_e^2, respectively.

Here we would compute the prediction interval rather than the variation of the variance components. Using the setting described in the previous section, let $\psi_1^2 = bc\sigma_{\mathrm{P}}^2 + c\sigma_{\mathrm{PO}}^2 + \sigma_e^2$, $\psi_2^2 = ac\sigma_{\mathrm{O}}^2 + c\sigma_{\mathrm{PO}}^2 + \sigma_e^2$, $\psi_3^2 = c\sigma_{\mathrm{PO}}^2 + \sigma_e^2$ and $\psi_4^2 = \sigma_e^2$. Moreover, $S_1^2 = \mathrm{SSP}$ (sum of squares between parts) with $S_1^2/\psi_1^2 \sim \chi_{a-1}^2$, $S_2^2 = \mathrm{SSO}$ (sum of squares between operators) with $S_2^2/\psi_2^2 \sim \chi_{b-1}^2$, $S_3^2 = \mathrm{SSPO}$ (sum of squares between parts and operators) with $S_3^2/\psi_3^2 \sim \chi_{(a-1)(b-1)}^2$ and $S_4^2 = \mathrm{SSE}$ (error sum of squares) with $S_4^2/\psi_4^2 \sim \chi_{ab(c-1)}^2$. From the original data, the ten parts, three operators and three measurements per part ($a = 10$, $b = 3$, $c = 3$, $n = 90$) yield $\bar{y} = 35.8$, $s_1^2 = 3935.96$, $s_2^2 = 39.27$, $s_3^2 = 48.51$ and $s_4^2 = 30.67$. Now suppose we wish to compute a 95% prediction interval that will contain $m = 8$ future units from two parts ($a^\star = 2$), two operators ($b^\star = 2$), and two measurements per part ($c^\star = 2$). For this situation, we have $\boldsymbol{\Sigma}^\star = \sigma_{\mathrm{P}}^2(\boldsymbol{I}_{a^\star} \otimes \boldsymbol{J}_{b^\star} \otimes \boldsymbol{J}_{c^\star}) + \sigma_{\mathrm{O}}^2(\boldsymbol{J}_{a^\star} \otimes \boldsymbol{I}_{b^\star} \otimes \boldsymbol{J}_{c^\star}) + \sigma_{\mathrm{PO}}^2(\boldsymbol{I}_{a^\star} \otimes \boldsymbol{I}_{b^\star} \otimes \boldsymbol{J}_{c^\star}) + \sigma_e^2(\boldsymbol{I}_{a^\star} \otimes \boldsymbol{I}_{b^\star} \otimes \boldsymbol{I}_{c^\star})$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\star + \tau^2 \boldsymbol{J}_m$, where the variance of $\overline{Y}$, $\tau^2 = (bc\sigma_{\mathrm{P}}^2 + ac\sigma_{\mathrm{O}}^2 + c\sigma_{\mathrm{PO}}^2 + \sigma_e^2)/n$.Also we have approximately 9.6 degrees of freedom. Hence the proposed Method II gives a two-sided 95% simultaneous prediction interval as [15.025, 56.575] and the obtained interval of Method III is [13.963, 57.636]. We note that the results based on both methods are in good agreement.

5. A simulation study

To further evaluate the performance of the recommended methods (Methods II and III), we conduct a more detailed simulation study based on the balanced one-way random model. The simulation procedures are the same as described in Section 3 and the results are displayed in Tables 2–4.

From Tables 2–4, the recommended methods appear to be quite successful in maintaining the confidence level close to the nominal value $\gamma = 0.95$ in almost all the simulation scenarios for $a = 5$, 7 and 9. But they become liberal as both the number of future observations under prediction and the value of ρ get large. Fortunately, a small number of future observations is usually of interest in practice (Hahn and Meeker, 1991). On the other hand, when a is small ($a = 3$), the resulting prediction intervals can be conservative, particularly when ρ is small, but the performance may still be acceptable. In general, a more accurate prediction interval can be expected by increasing the number of batches (a) rather than increasing the number of samples within each batch (b). This interesting result is also pointed out in Wang (1992).

Table 2
Simulated confidence coefficients ($\times 10^4$) for 95% two-sided prediction intervals when $a = 5$ of Methods II and III

a	b	$a^\star$	$b^\star$	ρ				
				0.1	0.3	0.5	0.7	0.9
5	2	1	2	9490[a]	9503	9441	9442	9480
				9708[b]	9687	9633	9586	9535
			6	9499	9533	9544	9552	9597
				9643	9633	9621	9610	9580
			10	9539	9581	9544	9612	9652
				9628	9646	9605	9639	9622
		2	1	9527	9493	9482	9411	9418
				9726	9673	9690	9583	9465
			2	9532	9444	9483	9420	9506
				9692	9628	9621	9502	9500
			10	9592	9586	9579	9607	9594
				9582	9548	9541	9540	9442
		3	1	9526	9496	9430	9420	9428
				9697	9651	9608	9542	9415
			2	9518	9473	9464	9470	9497
				9669	9604	9568	9526	9424

Table 2 *(Continued)*

a	b	$a^\star$	$b^\star$	ρ				
				0.1	0.3	0.5	0.7	0.9
	6	1	2	9503	9484	9442	9439	9459
				9665	9657	9638	9606	9545
			6	9507	9501	9467	9474	9537
				9617	9664	9628	9609	9550
			10	9554	9493	9488	9515	9555
				9667	9626	9624	9600	9545
		2	1	9503	9480	9400	9353	9364
				9656	9673	9633	9573	9458
			2	9501	9485	9410	9404	9447
				9653	9661	9600	9561	9485
			6	9488	9463	9419	9365	9513
				9596	9596	9560	9484	9469
			10	9542	9462	9441	9434	9496
				9631	9582	9563	9491	9425
		3	1	9510	9465	9380	9325	9332
				9669	9643	9586	9549	9408
			2	9500	9473	9366	9342	9405
				9623	9655	9557	9504	9394
			6	9489	9446	9355	9343	9507
				9623	9614	9517	9452	9397
	10	1	2	9507	9437	9451	9460	9418
				9633	9625	9646	9638	9529
			6	9502	9435	9445	9490	9513
				9643	9631	9636	9624	9539
			10	9501	9463	9447	9495	9583
				9618	9641	9616	9615	9572
		2	1	9535	9433	9385	9375	9307
				9658	9632	9616	9566	9432
			2	9501	9504	9382	9321	9379
				9619	9674	9605	9516	9435
			6	9490	9487	9377	9375	9472
				9594	9656	9555	9517	9447
			10	9505	9458	9397	9431	9525
				9615	9605	9560	9504	9449
		3	1	9501	9435	9363	9317	9284
				9641	9626	9606	9532	9378
			2	9498	9416	9355	9304	9395
				9653	9622	9556	9495	9383
			6	9516	9407	9369	9286	9487
				9641	9623	9566	9428	9415

[a]Method II.
[b]Method III.

6. Discussion and final remarks

The concept of GPQs is originally used for interval estimation. In this study, we apply it to point estimation. The Monte Carlo sampling of the distribution of GPQs is similar to the concept of resampling technique used in the parametric Bootstrap (Efron and Tibshirani, 1993). However, they are somewhat different, because the distributions in the parametric Bootstrap are exactly known, but those of GPQs are usually unknown even they are functions of random variables with known distributions. Hence, the point estimation using GPQs may be considered as a generalization of the parametric Bootstrap. Consequently, this can be one reason why the concept of GPQs have been successfully applied in many situations including our current study.

Table 3
Simulated confidence coefficients ($\times 10^4$) for 95% two-sided prediction intervals when $a = 7$ of Methods II and III

a	b	$a^\star$	$b^\star$	ρ				
				0.1	0.3	0.5	0.7	0.9
7	2	1	2	9539[a]	9446	9530	9481	9533
				9650[b]	9557	9620	9540	9534
			6	9511	9494	9512	9585	9554
				9586	9536	9529	9587	9516
			10	9546	9536	9553	9588	9612
				9558	9549	9543	9552	9553
		2	1	9520	9497	9480	9481	9452
				9630	9613	9579	9539	9423
			2	9526	9505	9475	9521	9486
				9620	9577	9522	9531	9412
			6	9536	9507	9544	9561	9574
				9534	9489	9514	9496	9444
		3	1	9489	9502	9464	9483	9490
				9587	9590	9526	9496	9422
			2	9481	9517	9475	9501	9502
				9540	9564	9505	9461	9373
	6	1	2	9529	9448	9485	9476	9456
				9610	9550	9598	9562	9466
			6	9504	9516	9501	9546	9604
				9582	9592	9562	9569	9552
			10	9488	9525	9545	9580	9626
				9542	9599	9592	9587	9565
		2	1	9505	9507	9441	9430	9460
				9595	9621	9543	9503	9449
			2	9440	9521	9440	9451	9487
				9524	9615	9546	9501	9433
			6	9514	9514	9466	9490	9568
				9567	9574	9528	9494	9449
			10	9563	9520	9414	9532	9626
				9583	9567	9449	9473	9457
		3	1	9508	9502	9419	9383	9391
				9579	9602	9536	9464	9345
			2	9542	9496	9452	9403	9475
				9611	9583	9531	9443	9371
			6	9497	9445	9370	9450	9508
				9553	9519	9434	9421	9344
	10	1	2	9463	9519	9462	9475	9493
				9541	9624	9568	9556	9497
			6	9486	9515	9492	9531	9593
				9562	9587	9578	9567	9538
			10	9499	9525	9527	9532	9632
				9564	9614	9590	9537	9564
		2	1	9512	9555	9470	9448	9423
				9599	9629	9578	9548	9422
			2	9537	9484	9477	9429	9452
				9614	9584	9567	9494	9411
			6	9515	9482	9466	9514	9570
				9576	9587	9548	9540	9452
			10	9484	9453	9463	9487	9592
				9530	9527	9521	9450	9439
		3	1	9473	9495	9475	9406	9420
				9562	9611	9614	9487	9377
			2	9506	9489	9448	9398	9470
				9582	9584	9544	9468	9377

Table 3 *(Continued)*

a	b	$a^\star$	$b^\star$	ρ				
				0.1	0.3	0.5	0.7	0.9
			6	9534	9433	9409	9420	9547
				9610	9534	9493	9430	9391

[a]Method II.
[b]Method III.

Table 4
Simulated confidence coefficients ($\times 10^4$) for 95% two-sided prediction intervals when $a = 9$ of Methods II and III

a	b	$a^\star$	$b^\star$	ρ				
				0.1	0.3	0.5	0.7	0.9
9	2	1	2	9525[a]	9492	9503	9472	9520
				9610[b]	9588	9569	9502	9494
			6	9478	9527	9552	9560	9590
				9513	9549	9553	9542	9543
			10	9492	9520	9572	9617	9605
				9494	9501	9535	9584	9547
		2	1	9493	9489	9466	9437	9516
				9587	9570	9532	9471	9471
			2	9541	9520	9485	9486	9523
				9592	9575	9523	9478	9442
			6	9526	9540	9529	9561	9562
				9506	9514	9463	9477	9450
		3	1	9515	9528	9502	9442	9438
				9588	9564	9541	9438	9343
			2	9542	9487	9469	9505	9515
				9576	9508	9464	9441	9369
			6	9500	9533	9471	9533	9580
				9453	9475	9407	9428	9415
	6	1	2	9465	9514	9473	9460	9477
				9541	9594	9532	9511	9460
			6	9453	9497	9532	9568	9603
				9520	9557	9573	9578	9539
			10	9509	9513	9555	9569	9657
				9540	9559	9568	9541	9580
		2	1	9523	9499	9471	9417	9475
				9576	9576	9548	9474	9439
			2	9495	9494	9463	9451	9488
				9556	9567	9532	9467	9417
			6	9534	9510	9471	9520	9599
				9559	9534	9497	9500	9491
			10	9518	9506	9501	9519	9618
				9522	9520	9490	9434	9468
		3	1	9471	9486	9487	9412	9433
				9531	9557	9543	9442	9378
			2	9523	9482	9481	9465	9481
				9559	9553	9530	9465	9386
			6	9549	9485	9475	9491	9555
				9566	9520	9491	9444	9380
	10	1	2	9498	9491	9446	9478	9480
				9549	9566	9533	9523	9464
			6	9500	9511	9524	9550	9597
				9546	9573	9564	9542	9538

Table 4 *(Continued)*

a	b	$a^\star$	$b^\star$	ρ				
				0.1	0.3	0.5	0.7	0.9
			10	9455	9540	9543	9608	9642
				9505	9584	9576	9583	9586
		2	1	9482	9515	9469	9442	9425
				9556	9567	9550	9491	9406
			2	9497	9447	9470	9447	9522
				9547	9521	9529	9479	9457
			6	9511	9503	9472	9515	9571
				9549	9545	9514	9499	9455
			10	9511	9503	9488	9546	9604
				9546	9542	9491	9484	9458
		3	1	9457	9474	9471	9442	9463
				9508	9558	9527	9485	9395
			2	9527	9492	9465	9460	9501
				9562	9557	9520	9454	9401
			6	9519	9463	9452	9504	9534
				9583	9561	9493	9431	9356

[a]Method II.
[b]Method III.

According to the simulation results of this study, Method II (ANOVA-based method) and Method III (GPQ-based method) are applicable to the balanced one-way random models. Theoretically, these methods can be extended to other more complex balanced linear random models. However, the Satterthwaite approximation can get less accurate when it involves more variance components. On the other hand, the GPQ-based method usually remains fairly satisfactory performance for complex models in interval estimation, e.g., see Liao and Iyer (2004). Moreover, even though we only consider the covariance matrix $\mathbf{\Sigma}^\star$ of future observations derived from the balanced data structure, $\mathbf{\Sigma}^\star$ can also be determined in an obvious manner for unbalanced data structures. So the similar recommended method may be applied to this situation, too. Finally, the proposed GPQ-based methods may also be extended to the unbalanced one-way random model using the GPQs for the variance components in the model presented in Liao et al. (2005), or to the unbalanced one-way random model with heterogeneous variances discussed in Iyer et al. (2004). Nonetheless, it is still needed to evaluate the performance of these possible extensions before they are recommended for practical use.

References

Burch, B.D., 2007. Generalized confidence intervals for proportions of total variance in mixed linear models. J. Statist. Plann. Inference 137, 2394–2404.

Chew, V., 1968. Simultaneous prediction intervals. Technometrics 10, 323–330.

Chou, Y.M., Owen, D.B., 1986. One-sided simultaneous lower prediction intervals for l future samples from a normal distribution. Technometrics 28, 247–251.

Davis, C.B., McNichols, R.J., 1987. One-sided intervals for at least p of m observations from a normal population on each of r future occasions. Technometrics 29, 359–370.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Fertig, K.W., Mann, N.R., 1977. One-sided prediction intervals for at least p out of m future observations for a normal population. Technometrics 9, 167–177.

Genz, A., 1999. Finding critical values using numerical integration. Comput. Sci. Statist. 31, 263–266.

Genz, A., Bretz, F., 2002. Comparison of methods for the computation of multivariate t probabilities. J. Comput. Graph. Statist. 11, 950–971.

Graybill, F.A., 1976. Theory and Application of the Linear Model. Wadsworth, Belmont, CA.

Hahn, G.J., 1969. Factor for calculating two-sided prediction intervals for samples from a normal distribution. J. Amer. Statist. Assoc. 64, 878–888.

Hahn, G.J., 1970. Additional factors for calculating prediction intervals for samples from a normal distribution. J. Amer. Statist. Assoc. 65, 1668–1676.

Hahn, G.J., Meeker, W.Q., 1991. Statistical Intervals. Wiley, New York.

Hall, I.J., Prairie, R.R., 1973. One-sided prediction intervals to contain at least m out of k future observations. Technometrics 15, 897–914.

Hamada, M., Weerahandi, S., 2000. Measurement system assessment via generalized inference. J. Qual. Technol. 32, 241–253.

Hanning, J., Iyer, H.K., Patterson, P., 2006. Fiducial generalized confidence intervals. J. Amer. Statist. Assoc. 101, 254–269.
Houf, R.F., Berman, D.B., 1988. Statistical analysis of power module thermal test equipment performance. IEEE Trans. Components Hybrids Manufacturing Technol. 11 (4), 516–520.
Iyer, H.K., Wang, C.M., Mathew, T., 2004. Models and confidence intervals for true values in interlaboratory trials. J. Amer. Statist. Assoc. 99, 1060–1071.
Khuri, A.I., Mathew, T., Sinha, B.K., 1998. Statistical Tests for Mixed Linear Models. Wiley, New York.
Liao, C.T., Iyer, H.K., 2004. A tolerance interval for the normal distribution with several variance components. Statist. Sin. 14, 217–229.
Liao, C.T., Lin, T.Y., Iyer, H.K., 2005. One- and two-sided tolerance intervals for general balanced mixed models and unbalanced one-way random models. Technometrics 47, 323–335.
Lin, S.H., Lee, J.C., Wang, R.S., 2007. Generalized inferences on the common mean vector of several multivariate normal populations. J. Statist. Plann. Inference 137, 2240–2249.
Lin, T.Y., Liao, C.T., 2006. A β-expectation tolerance interval for general balanced mixed linear models. Comput. Statist. Data Anal. 50, 911–925.
Mathew, T., Webb, D.W., 2005. Generalized p-values and confidence intervals for variance components: applications to army test and evaluation. Technometrics 47, 312–322.
Miller, Jr., R.G., 1981. Simultaneous Statistical Inference, second ed. Springer, New York.
Odeh, R.E., 1990. Two-sided prediction intervals to contain at least k out of m future observations from a normal distribution. Technometrics 32, 203–216.
Satterthwaite, F.E., 1946. An approximate distribution of estimates of variance components. Biom. Bull. 2, 110–114.
Tsui, K.W., Weerahandi, S., 1989. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. J. Amer. Statist. Assoc. 84, 602–607.
Wang, C.M., 1992. Prediction intervals for a balanced one-way random-effects model. Commun. Stat. Simul. Comput. 21, 671–687.
Weerahandi, S., 1993. Generalized confidence intervals. J. Amer. Statist. Assoc. 88, 899–905.
Weerahandi, S., 1995. Exact Statistical Methods for Data Analysis. Springer, New York.