An Approximate Critical Point for a Test in the Growth Curve Model: A Conditional Approach

Jemila Seid Hamid and Dietrich von Rosen

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2005:5
ISSN 1651-8543

An Approximate Critical Point for a Test in the Growth Curve Model: A Conditional Approach

Jemila Seid Hamid[1] and Dietrich von Rosen

Centre of Biostochastics
Swedish University of Agricultural Sciences
Box 7013, SE-750 07, Uppsala, Sweden

Abstract

A test proposed by Seid Hamid & von Rosen (2005b) is considered. The critical point is calculated using a conditional approach. The conditional distributions under the null and alternative hypotheses are shown to be represented as sums of weighted central and non central chi-square random variables, respectively. Under the null hypothesis, the Satterthwaite approximation is used to a get an approximate critical point. An approximate estimated power is given using a Satterthwaite kind of approximation together with some new ideas. A numerical example is given to illustrate the results.

Keywords: Ancillary statistics, conditional test, growth curve model, residuals, Satterthwaite approximation.

[1]E-mail address to the correspondence author: Jemila.Seid.Hamid@bt.slu.se

1. Introduction

Quite often models that are proposed and fitted to data are "incorrect" to some extent and need to be treated carefully. Moreover, the models proposed are usually based on several assumptions. An important part of modelling is diagnosing the flaws in these models as well as checking if the assumptions are true or rather check if the data violates the assumptions.

In most model fitting problems, whether they are linear or non linear, diagnosing the model or the model assumptions is performed by examining the residuals. Residuals are also used to detect outliers and/or influential observations in the data. Residuals are the part of the data which is left unexplained by the fitted model. They are defined as the difference between the observed and fitted values:

$$e_i = y_i - \hat{y}_i.$$

However, in the Growth Curve Model which was introduced by Potthof & Roy (1964) and is usually referred to us the Potthoff & Roy model, the ordinary residuals which are defined as above consist of two parts. One part gives information about the between individual structure. The second part gives information about the within individual structure which in fact is a part of the residual that tells us if the estimated model fits the data. This part can in fact be shown to be the difference between the observed and estimated means.

Therefore, when dealing with the Growth Curve Model one should be careful when examining the ordinary residuals. For example, the two parts mentioned in the previous paragraph may happen to have opposite signs and could cancel with each other and give an impression that the model fits the data well when it in fact does not. That is why there is a need to define other residuals which take the bilinear structure in the model into consideration. This has been done by von Rosen (1995) and the residuals are presented below. First, let us give the definition of the Growth Curve Model.

Let $X : p \times n$ and $B : q \times k$ be the observation and parameter matrices, respectively, and let $A : p\times q$ and $C : k\times n$ be the within and between individual design matrices, respectively. Suppose that $q \leq p$ and $\rho(C) + p \leq n$, where $\rho(.)$ denotes the rank of a matrix. The Growth Curve model is given by

$$X = ABC + \epsilon, \tag{1.1}$$

where the columns of ϵ are assumed to be independently p-variate normally distributed with mean zero and an unknown positive definite covariance matrix Σ.

Taking the bilinear structure into account three residuals for the above model were defined by von Rosen (1995):

$$R_1 = A(A'S^{-1}A)^- A'S^{-1}X(I - C'(CC')^- C),$$
$$R_2 = (I - A(A'S^{-1}A)^- A'S^{-1})X(I - C'(CC')^- C),$$
$$R_3 = (I - A(A'S^{-1}A)^- A'S^{-1})XC'(CC')^- C,$$

where $S = X(I - C'(CC')^- C)X'$.

In the same paper it was suggested that one can use one of the above residuals or a combination of two to check the between and within individuals assumptions. Specifically one can use R_3, which is the difference between the observed and estimated means, to check if the estimated model fits the data. For detailed information about the interpretation of residuals in the Growth and Extended Growth models we refer to Seid Hamid & von Rosen (2005a).

In this paper we are going to consider a statistic for testing the hypothesis that the model fits the data which can formally be presented as:

$$\begin{aligned} H_o &: B = 0 \\ H_1 &: B \neq 0. \end{aligned} \tag{1.2}$$

The test which is a function of R_3 is proposed by Seid Hamid & von Rosen (2005b) and is given as:

$$\phi(X) = tr\{XC'(CC')^- CX'S^{-1}A(A'S^{-1}A)^- A'S^{-1}\}. \tag{1.3}$$

The hypothesis is rejected when $\phi(X) > c$ where c is obtained from

$$P_{H_o}(\phi(X) > c) = \alpha, \tag{1.4}$$

where α is the desired level of significance.

By looking at the test given in (1.3), it is not easy to see that the test is a function of the residual, R_3. However, if one refers to the paper by Seid Hamid & von Rosen (2005b) and looks at the steps they used in obtaining the test, one could see that the test statistics is equivalent to

$$\frac{\exp\{-\frac{n}{2}tr\{S^{-1}XC'(CC')^- CX'\}\}}{\exp\{-\frac{n}{2}tr\{S^{-1}R_3R_3'\}\}}, \tag{1.5}$$

where the hypothesis is rejected for small values of the ratio, i.e., when the values are close to zero. As we can see from the above expression, the numerator is a function of the observed mean structure, $XC'(CC')^- C$. On the other hand the denominator is a function of the residual, R_3 which is obtained by subtracting the estimated mean structure from the observed mean, i.e.,

$$R_3 = XC'(CC')^- C - A\hat{B}C.$$

That means the test compares the observed mean and the residuals. In other words, the test compares the observed and estimated means and rejects the hypothesis when they are "close" to each other, i.e., when the residuals are very "small".

Unfortunately, the exact distribution for the test statistic given in (1.3) is difficult to obtain. As a result, in practical situations one needs to come up with other ways to calculate the critical point. Two suggestions were made in Seid Hamid & von Rosen (2005b). The first one is to approximate the density of the statistic using the first two moments. In this paper, we shall discuss the second alternative which is based on conditioning using a natural ancillary statistic. That is, we calculate the critical point for a given S, where $S = X(I - C'(CC')^{-}C)X'$ is an ancillary statistic for the parameter of interest, B. We show that the resulting conditional distribution can be written as a linear combination of independent chi-square random variables which allows us to use existing results for such sums including a well known approximation by Satterthwaite (1946).

Apart from a great simplification provided by conditioning, conditioning like sufficiency and invariance, leads to a reduction of the data (Lehmann, 1986). When the problem involves ancillary statistics conditioning is appropriate since it makes the inference more relevant to the situation at hand.

Ancillary statistics is a statistics whose distribution doesn't depend on the parameter of interest. The term ancillary was first used by Fisher (1956) and those statistics are referred as non-informative since they do not provide any information about the parameter of interest.

In the presence of an ancillary statistics Z, i.e., a statistic with a distribution independent of the parameter, one can think of the observation X (with distribution P) as obtained from a two-stage experiment (Lehmann, 1986):

i) Observe the ancillary statistic Z with distribution F.

ii) Given Z, observe a quantity X with distribution $P(X|Z)$.

The resulting X is distributed according to the original distribution P. Under these circumstances conditioning is appropriate since it makes the inference more relevant to the situation at hand (Lehmann, 1986). It was also suggested there that the above argument is valid even if the distribution of the ancillary statistic depends on parameters other than the parameter of interest. Such statistic is usually called S-ancillary or partial ancillary statistic. However, in this paper we use the term ancillary even if the statistic depends on other parameters. By ancillary we mean it is ancillary for the parameter of interest.

For more details about ancillary statistics we refer, among others, to papers by Fisher (1956) and Basu (1964).

The conditioning variables are not always restricted to ancillary statistics. For brief discussions and further references about conditioning with variables other than ancillary statistics and concepts of relevant subsets we refer to Lehmann (1986).

Now let us come to the Growth Curve Model given in (1.1), $S = X(I - C'(CC')^{-}C)$ has a Wishart distribution with parameters Σ and $n - \rho(C)$, which is independent of the parameter of interest B, although it depends on the covariance matrix Σ. This shows that S is ancillary for B. What we shall do in this section is to find a critical point for the test defined in (1.3) by conditioning on the ancillary statistic S.

It is important to note that our approach is different from other conditional inferences in which conditioning is usually made at an early stage, mainly for eliminating nuisance parameters, see for example Basu (1977). Whereas, in this paper we condition after the test has been constructed using the restricted followed by estimated likelihood approaches. Moreover, in most conditional approaches, the statistic which is used for conditioning is a partial sufficient statistic which gives the advantage that the resulting conditional distribution depends only on the parameter of interest, see Basu (1978) about partial sufficiency. However, in our case, we have already eliminated the nuisance parameter using the restricted maximum likelihood approach. The main reason for conditioning here, unlike most cases, is to make the distribution relatively easier so that we could be able to calculate the critical point.

Moreover, S is ancillary for B and as a result it contains no information about B. Consequently, we do not lose any information about B by conditioning on S. In fact, by conditioning on S, we will reduce the data to make the inference more relevant to the situation at hand without losing any information about B.

Furthermore, in problems of testing Fisher (1956) uses ancillary statistics for the determination of the true level of significance. He recommends that, in the presence of ancillary statistics, the level of significance of a test should be computed by refereing to the conditional sample space determined by the set of all possible samples for which the value of the ancillary statistics is the one presently observed, see Basu (1964).

2. The critical point: a conditional approach

Suppose the Growth Curve model given in (1.1) has been fitted to data and we want to evaluate the model through the hypothesis presented in (1.2). A test statistic was constructed by Seid Hamid & von Rosen (2005b) using restricted likelihood followed by an estimated likelihood approach. The test statistic is given in (1.3). The hypothesis is rejected when $\phi(X) > c$, where c is calculated such that

$$P_{H_o}(\phi(X) > c) = \alpha, \tag{2.1}$$

where α is the desired level for the test. As mentioned in the previous section, obtaining the critical point from (2.1) requires the distribution of $\phi(X)$ which unfortunately is difficult to obtain. In the same paper, two alternating approaches in finding an approximate critical point, which could be applied in practical situations, were suggested. What we present and discuss here is the conditional approach and we will calculate a critical point for a given value of S. That is to approximate c by $c(S)$, where $c(S)$ is obtained such that

$$P_{H_o}(\phi(X|S) > c(S)) = \alpha. \tag{2.2}$$

The conditional distribution $\phi(X|S)$ is relatively easy to deal with. Moreover, in the presence of an ancillary statistic, which is the case here, conditioning is appropriate for the reasons explained in the previous section. The conditional distribution is given in the theorem below as a weighted sum of independent chi-square random variables which enables us to use existing results for such a sum.

Theorem 2.1. *Consider the hypothesis given in (1.2). Under the null hypothesis the conditional test $\phi(X|S)$ can be described as*

$$\phi(X|S) \equiv \sum W_{ii}\Lambda_{ii}, \tag{2.3}$$

where, $W_{ii}'s$ are independently distributed as chi-square random variables with $\rho(C)$ degrees of freedom and $\Lambda_{ii}'s$ are non negative constants which are functions of S. The "$\equiv$" in equation (2.3) represents equivalence in distribution and $\rho(C)$ denotes the rank of the between individual design matrix C.

Proof. Consider the test given in (1.3) and condition on S. The conditional test equals

$$\phi(X|S) = tr\{XC'(CC')^- CX'S^{-1}A(A'S^{-1}A)^- A'S^{-1}\}. \tag{2.4}$$

Observe that the above expression is identical with that of $\phi(X)$ given in (1.3). However, it is important to note that here S is no longer a part of the random data instead it has become a constant.

Now let us assume, without loss of generality, that $\Sigma = I$. This is possible due to the fact that the null distributions of both the conditional and unconditional tests given in (1.3) and (2.4), respectively, are independent of Σ. We refer to Seid Hamid & von Rosen (2005b) for the proof.

Therefore, under the null hypothesis, i.e., when $B = 0$, we have

$$W = XC'(CC')^- CX' \sim W(I, \rho(C)),$$

where $W(I, \rho(C))$ represents a Wishart distribution with parameters I and $\rho(C)$. We may therefore rewrite $\phi(X|S)$ as

$$\phi(X|S) = tr\{WP\},$$

where $P = S^{-1}A(A'S^{-1}A)^{-}A'S^{-1}$. P is a symmetric positive semi-definite matrix, as a result we could decompose it as

$$P = \Gamma\Lambda\Gamma', \tag{2.5}$$

where Γ is an orthogonal matrix, Λ is a diagonal matrix where the diagonal elements Λ_{ii} are the ith eigenvalues of P.

On the other hand W can be written as the sum of $\rho(C)$ independent random matrices as,

$$W = \sum_{i=1}^{\rho(C)} w_i w_i', \tag{2.6}$$

where $w_i \sim N_p(\mathbf{0}, I)$, see Kollo & von Rosen (2005).

Consequently,

$$\begin{aligned}\phi(X|S) &\equiv tr\{W\Gamma\Lambda\Gamma'\}\\ &\equiv tr\{W\Lambda\},\end{aligned}$$

where the last statement was possible since $tr\{AB\} = tr\{BA\}$ for any two matrices, the Wishart distribution is rotation invariant and Γ is an orthogonal matrix which is independent of W. Now using the property of the trace function we get

$$\phi(X|S) \equiv \sum W_{ii}\Lambda_{ii},$$

where the W_{ii}'s are the diagonal elements of W. Moreover, using the representation in (2.6), it is possible to show that they are independently distributed as a chi-square distribution with $\rho(C)$ degrees of freedom. □

A weighted sum of independent chi-square random variables arise very frequently in practical situations, see for example Johnson & Kotz (1968), Mathai (1982) and Moschopoulos (1985). Exact distribution for the sum has been given as an infinite series form in Kotz et al. (1967), Mathai (1982) and Moschopoulos (1985) where in the latter the applications in different areas such as queue type problems and engineering were given. It was mentioned that their representation is computationally convenient since the coefficients are easily computed by simple recursive relations.

However, the exact distribution is too complicated to be applied in practical situations which brings a need for a good and reasonable approximation. Several approximations has been proposed, see for example Moschopoulos (1985), where a series representation of the exact distribution is given and they suggested that one can use a truncated version of the series where the truncation error is readily obtainable.

However we are going to use the celebrated and well known Satterthwaite approximation which will be presented briefly below. For more details about the approximation and extension of the approximation to linear combination of independent Wishart random variables see Statterthwaite (1949) and Tan & Gupta (1983), respectively. In the latter paper, some Monte Carlo results were given to demonstrate the closeness of the approximation and the studies indicate that the approximation in general is quite good.

Let

$$Z = \sum_{i=1}^{p} a_i \sigma_i^2 \chi_{f_i}^2,$$

where the $\chi_{f_i}^2$'s are independent chi-square random variables and the a_i's positive constants. The well known approximation of Z is given by (see Tan and Gupta, 1983):

$$Z \sim a\chi_f^2, \tag{2.7}$$

$$a = \frac{\sum_{i=1}^{p} a_i^2 f_i \sigma_i^4}{\sum_{i=1}^{p} a_i f_i \sigma_i^2}, \tag{2.8}$$

$$f = \frac{(\sum_{i=1}^{p} a_i f_i \sigma_i^2)^2}{\sum_{i=1}^{p} a_i^2 f_i \sigma_i^4}. \tag{2.9}$$

The f and a in the above two equations are obtained by equating the first two moments of both sides of equation (2.7). However, in practical situations the σ_i^2's are unknown. In this case, a and f are estimated by replacing the σ_i^2's by their estimates. This approximation is known as the Satterthwaite approximation (Satterthwaite, 1946).

In our case, the distribution of $\phi(X|S)$ does not depend on the unknown covariance matrix, consequently, the distribution of $\sum W_{ii}\Lambda_{ii}$ is free of any unknown parameters. The resulting a and f, therefore, are free of any unknown parameters. The distribution of $\sum W_{ii}\Lambda_{ii}$ is approximated by $a\chi_f^2$, where χ_f^2 is a chi-square random variable with f degrees of freedom and a is a positive constant. The unknown parameters a and f are then obtained by equating the first two moments of $\sum W_{ii}\Lambda_{ii}$ and $a\chi_f^2$. We shall present this result in the following theorem.

Theorem 2.2. *Consider the test statistic given in (1.3). Its distribution for a given S can be approximated by that of $a\chi_f^2$ where χ_f^2 is a chi-square random variable with f degrees of freedom and a is positive constant. The constants a and f are given by*

$$a = \frac{\sum_{i=1}^{p} \Lambda_{ii}^2}{\sum_{i=1}^{p} \Lambda_{ii}}, \tag{2.10}$$

$$f = \frac{\rho(C)(\sum_{i=1}^{p} \Lambda_{ii})^2}{\sum_{i=1}^{p} \Lambda_{ii}^2}, \tag{2.11}$$

where Λ_{ii} are the eigen values given in Theorem 2.1, and $\rho(C)$ is the degrees of freedom of the chi-square random variables in Theorem 2.1, which is equal to the rank of the between individual design matrix C.

The critical point $c(S)$ can then be calculated from

$$P_{H_o}(a\chi_f^2 > c(S)) = \alpha.$$

It could also be of interest to calculate the power of the test which may be used as a measure of performance for the test. What we shall do here is to calculate the conditional power for a given S. That is the power is obtained from

$$P_{H_1}(\phi(X|S) > c(S)).$$

The distribution of $\phi(X|S)$, under the alternative hypothesis, is easier to handle than the unconditional one because it can be written as a sum of weighted non-central chi-square random variables. This fact is given in the following theorem.

Theorem 2.3. *Suppose $\phi(X|S)$ as given in (2.4). Let $\mu = \sqrt{n}\ \Sigma^{-\frac{1}{2}} AB\underline{C}$, where $\underline{C}$ is a matrix of full rank constructed from the linearly independent columns of the between individual design matrix C. The distribution of $\phi(X|S)$ under the alternative can be written as*

$$\phi(X|S) \equiv \sum T_{ii}\Lambda_{ii}; \tag{2.12}$$

where Λ_{ii} are positive constants, $W_{ii}'s$ are independently distributed as a non-central chi-square random variable with $\rho(C)$ degrees of freedom and $\lambda_i = \sum_{j=1}^{\rho(C)} \mu_{ij}^2$ is the non-centrality parameter with μ_{ij} as the $(i,j)^{th}$ element of μ.

Proof. Consider the conditional test given in (2.4). It can be rewritten as

$$\phi(X|S) = tr\{\{\Sigma^{-\frac{1}{2}} XC'(CC')^{-} CX'\Sigma^{-\frac{1}{2}}\}\{\Sigma^{\frac{1}{2}} S^{-1} A(A'S^{-1}A)^{-} A'S^{-1}\Sigma^{\frac{1}{2}}\}\}, \tag{2.13}$$

where S is considered as a constant.

Now consider the first part of the above expression. It is possible to show that

$$W = \Sigma^{-\frac{1}{2}} X C'(CC')^{-} C X' \Sigma^{-\frac{1}{2}} \sim W_p(I, \rho(C), \Delta),$$

where $\Delta = \mu\mu'$.

Let

$$P = \Sigma^{\frac{1}{2}} S^{-1} A (A' S^{-1} A)^{-} A' S^{-1} \Sigma^{\frac{1}{2}}.$$

P is a symmetric positive semi-definite matrix. As a result we can decompose it as

$$P = \Gamma \Lambda \Gamma',$$

where Γ is an orthogonal matrix and Λ is a diagonal matrix where its diagonal elements are the eigenvalues of P. Under the alternative hypothesis, $\phi_{H_1}(X|S)$ can therefore be written as

$$\begin{aligned}
\phi(X|S) &= tr\{WP\} \\
&= tr\{W\Lambda\} \\
&= \sum W_{ii}\Lambda_{ii}.
\end{aligned}$$

However, using a representation similar to (2.6) for W, it is possible to show that $W_{ii}'s$ are independently distributed as non-central chi-square with $\rho(C)$ degrees of freedom and non-centrality parameter $\lambda_i = \sum_{j=1}^{\rho(C)} \mu_{ij}^2$. □

As we can see from the above theorem, the distribution of the conditional test under the alternative hypothesis depends on the covariance matrix Σ, which in many practical situations is unknown. Consequently, an estimator is needed to get an estimated power. A brief discussion about two alternative estimators is given in the next section.

Theorem 2.3 also shows that the distribution of the conditional test depends on the parameter B. It is important to note that this dependence is through μ and hence through the non-centrality parameters λ_i's.

The representation given in Theorem 2.3 enables us to use existing results for a weighted sum of independent non-central chi-square random variables. Linear combinations of non-central chi-square random variables were considered among others by Press (1966) and the exact distribution was given there. It was also mentioned that this kind of distribution arises in classifying an unknown vector into one of two multivariate normal populations with unequal means and covariance matrices.

However, the distribution is too complicated to be used in practical applications although there exist several algorithms to numerically solve the series, see for example Imhof (1961). Here we are going to use an approximation similar to Satterthwaite's approximation. This kind of approximation, as to our knowledge, has not been done for a weighted sum of non-central chi-square random variables. Moreover, we shall show later that our approach is somehow different and new ideas has been implemented to get the approximation.

Suppose

$$Z = \sum_{i=1}^{p} a_i \chi^2_{f_i,\lambda_i},$$

where $\chi^2_{f_i,\lambda_i}$ is distributed as non-central chi-square with f_i degrees of freedom and non-centrality parameter λ_i.

The idea is to approximate the distribution of Z by

$$Z \sim a\chi^2_{f,\lambda}. \tag{2.14}$$

In order that the approximate distribution is completely specified, we need to specify the values of a, f and λ. If we want to use a similar approach as in Satterthwaite (1946), the parameters will be obtained by equating the first three moments on both sides of (2.14). However, this involves solving three equations in a, f, and λ where one of the equations is a third degree polynomial in all the three parameters.

Our approach is now presented shortly and involves decomposing the non-central chi-square random variable into two independent components. One part which is distributed as a non-central chi-square distribution with one degree of freedom and non-centrality parameter λ. The second component is distributed as a central chi-square distribution with $f-1$ degrees of freedom.

Let $x_i \sim N(\mu_i, 1)$, i=1,2,...,f. Then it is well known that

$$\sum x_i^2 \sim \chi^2_{f,\lambda},$$

where $\lambda^2 = \sum \mu_i^2$.

We can make an orthogonal transformation (Rao, 1973) form $x_1, x_2, ..., x_f$ to $y_1, y_2, ..., y_f$ such that $y_1, y_2, ..., y_f$ are independently normally distributed with a unit variance and $\boldsymbol{E}(y_1) = \lambda$, and $\boldsymbol{E}(y_i) = 0$ for $i = 2, 3, ..., f$.

Therefore,

$$\sum_{i=1}^{f} x_i^2 = \sum_{i=1}^{f} y_i^2 = y_1^2 + \sum_{i=2}^{f} y_i^2. \tag{2.15}$$

Observe that y_1^2 is a non-central chi-square random variable with 1 degree of freedom and non-centrality parameter λ. Whereas, $\sum_{i=2}^{f} y_i^2$ is distributed as central chi-square with $f-1$ degrees of freedom. Moreover, it is important to note that the two terms in (2.15) are independent.

Consequently, we can write $\chi^2_{f,\lambda}$ as

$$\chi^2_{f,\lambda} = \chi^2_{1,\lambda} + \chi^2_{f-1}. \tag{2.16}$$

We can therefore write $a\chi^2_{f,\lambda}$ and $\sum_{i=1}^{p} a_i\chi^2_{f_i,\lambda_i}$ as

$$a\chi^2_{f,\lambda} = a\chi^2_{1,\lambda} + a\chi^2_{f-1}, \tag{2.17}$$

$$\sum_{i=1}^{p} a_i\chi^2_{f_i,\lambda_i} = \sum_{i=1}^{p} a_i\chi^2_{1,\lambda_i} + \sum_{i=1}^{p} a_i\chi^2_{f_i-1}. \tag{2.18}$$

Now, a and f in (2.14) are obtained by equating the first two moments of $a\chi^2_{f-1}$ and $\sum_{i=1}^{p} a_i\chi^2_{f_i-1}$. The non-centrality parameter λ will then be obtained by equating the first moments of $a\chi^2_{1,\lambda}$ and $\sum_{i=1}^{p} a_i\chi^2_{1,\lambda_i}$, where a is replaced by its estimator. The result for $\phi(X|S)$ is presented in the following theorem.

Theorem 2.4. *The distribution of $\phi(X|S)$ under the alternative hypothesis can be approximated by $a\chi^2_{f,\lambda}$, where a, f and λ are given by*

$$\begin{aligned}
a &= \frac{\sum_{i=1}^{p} \Lambda_{ii}^2}{\sum_{i=1}^{p} \Lambda_{ii}},\\
f &= \frac{(\rho(C)-1)[\sum_{i=1}^{p} \Lambda_{ii}]^2}{\sum_{i=1}^{p} \Lambda_{ii}^2} + 1,\\
\lambda &= \frac{\{\sum_{i=1}^{p} \Lambda_{ii}(1+\lambda_i)\}\{\sum_{i=1}^{p} \Lambda_{ii}\}}{\sum_{i=1}^{p} \Lambda_{ii}^2} - 1,
\end{aligned}$$

where the Λ_{ii}'s and λ_i's are as given in Theorem (2.3), and $\rho(C)$ is the degrees of freedom of the non-central chi-square random variables, W_{ii}'s.

The power of the test, under the alternative hypothesis, is then calculated as

$$P_{H_1}(a\chi^2_{f,\lambda} > c(S)).$$

Observe that the test depends on B only through the non-central parameter λ, and this non-centrality parameter increases the more B differs from zero. We have shown this in the numerical example given in the next section. It is also interesting to see that the power is a monotone function of λ.

3. Estimating Σ for power calculations

Theorem 2.3 shows that the distribution of the conditional test under the alternative hypothesis depends on the unknown covariance matrix Σ. As a result, the approximate distribution also depends on Σ, see Theorem 2.4. In practical situations, one needs to find a reasonable estimator for Σ to obtain an estimator for the approximate power.

One possible estimator for Σ is S/n. This estimator was obtained by maximizing the part of the likelihood, and was used to get the estimated likelihood when defining the test in Seid Hamid & von Rosen (2005b). Moreover, it is possible to show that S provides sufficient information in absence of knowledge about the parameter B. See a paper by Sprott (1975) on marginal and conditional sufficiency. One can also use the unbiased estimator $S/(n-\rho(C))$. Furthermore, the two estimators mentioned above are functions of the ancillary statistic and are considered as constants.

However, for a completely specified alternative, S does not provide sufficient information about Σ. Another estimator that might be used is

$$\hat{\Sigma} = \frac{1}{n}\{S + (XC'(CC')^{-}C - ABC)(XC'(CC')^{-}C - ABC)'\}. \tag{3.1}$$

This estimator gives more information than S, provided that there is some knowledge about B. This is particularly true when B is known. However, using (3.1) as an estimator for Σ brings complications to the conditional approach which will not be discussed in this paper. Nevertheless, one can use the estimate after the data has been obtained to get an estimate for the power. This will be shown in the next section using a numerical example.

4. Numerical Illustration

In this section we give a numerical example to illustrate the results presented in the previous sections. We consider the Potthoff & Roy (1964) data. This data was considered by von Rosen (1995) to illustrate how one can use his residuals. The data consist of dental measurements on eleven girls and sixteen boys at four ages (8, 10, 12 and 14). Each measurement is the distance, in millimeters, from the center of pituitary to pteryo-maxillary fissure. The data is reproduced at the end of the paper.

Suppose the Potthoff & Roy model has been fitted to the data with the assumption that the mean growth for both the girls and boys is linear. The observation, design and parameter matrices are given by

$$\mathbf{X}_{4\times 27} = \begin{pmatrix}
21 & 21 & 20.5 & 23.5 & 21.5 & 20 & 21.5 & 23 & 20 & 16.5 \\
24.5 & 26 & 21.5 & 23 & 20 & 25.5 & 24.5 & 22 & 24 & 23 \\
27.5 & 23 & 21.5 & 17 & 22.5 & 23 & 22, & & & \\
20 & 21.5 & 24 & 24.5 & 23 & 21 & 22.5 & 23 & 21 & 19 \\
25 & 25 & 22.5 & 22.5 & 23.5 & 27.5 & 25.5 & 22 & 21.5 & 20.5 \\
28 & 23 & 23.5 & 24.5 & 25.5 & 24.5 & 21.5, & & & \\
21.5 & 24 & 24.5 & 25 & 22.5 & 21 & 23 & 23.5 & 22 & 19 \\
28 & 29 & 23 & 24 & 22.5 & 26.5 & 27 & 24.5 & 24.5 & 31 \\
31 & 23.5 & 24 & 26 & 25.5 & 26 & 23.5, & & & \\
23 & 25.5 & 26 & 26.5 & 23.5 & 22.5 & 25 & 24 & 21.5 & 19.5 \\
28 & 31 & 26.5 & 27.5 & 26 & 27 & 28.5 & 26.5 & 25.5 & 26 \\
31.5 & 25 & 28 & 29.5 & 26 & 30 & 25 & & &
\end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} b_{01} & b_{02} \\ b_{11} & b_{12} \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix}, \quad \text{and} \quad \mathbf{C}_{2\times 27} = \begin{pmatrix} \mathbf{1}_{11} & \mathbf{0}_{16} \\ \mathbf{0}_{11} & \mathbf{1}_{16} \end{pmatrix}.$$

Observe how we presented the observation matrix, commas are used to separate the rows.

We are now interested to check the assumed linear curves for the mean growth. That is to check if the growth for the mean can be regard as linear for both the girls and the boys. The hypothesis can be formulated as

$$H_o : B = 0$$
$$H_1 : B \neq 0.$$

The proposed conditional test is given by

$$\phi(X|S) = tr\{XC'(CC')^{-}CX'S^{-1}A(A'S^{-1}A)^{-}A'S^{-1}\}.$$

The hypothesis is rejected when $\phi(X|S) > c(S)$, where $c(S)$ is obtained from

$$P_{H_o}(a\chi_f^2 > c(S)) = \alpha,$$

where a and f are as given in Theorem 2.2.

The observed value of the test for the above data is $\phi(x|s) = 175.12$. Therefore, we reject the hypothesis if $c(s)$ is greater than 175.12.

The calculated values of a and f are 0.02 and 3, respectively. Suppose the level of significance $\alpha = 0.05$, $c(s)$ is then obtained from

$$P_{H_o}(0.02\chi_3^2 > c(s)) = 0.05,$$

and the value of the $c(s)$ obtained is 0.16, which is much smaller than the observed value for the test. That is, the data gives strong evidence towards rejecting the hypothesis that $B = 0$. Therefore, we conclude that linear growth curves seem appropriate to describe the mean structure for both the girls and the boys.

We would like to note that this conclusion was reached by Potthoff & Roy when they analyzed the data for the first time. We could also see this by looking at the residuals defined by von Rosen (1995). The residuals which are obtained as a difference between the observed and estimated means are very small which leads to the conclusion that the assumed linear curves seem to fit the data well.

Let us now calculate the estimated power for the above test. First, we shall replace the unknown covariance matrix Σ by $\frac{1}{n}S$. That is the matrices μ and

P given in Theorem 2.3 become

$$\mu = \sqrt{n} S^{-\frac{1}{2}} ABC$$
$$P = \frac{1}{n} S^{-\frac{1}{2}} A(A'S^{-1}A)^{-}A'S^{-\frac{1}{2}}$$

Recall that both the exact and approximate powers depend on B. This is also true for the estimated power. That means we need to specify the value of B under the alternative hypothesis. Suppose that we are testing the hypothesis that $B = 0$ against the alternative

$$\mathbf{B} = \begin{pmatrix} 7.43 & 5.84 \\ 0.48 & 0.83 \end{pmatrix}.$$

The calculated values of a and f for the data under the alternative hypothesis are 0.04 and 3, respectively. Note that these values do not depend on the value of B and and hence remain the same for all $B \neq 0$. However, the value of the non-centrality parameter depends on the choice of B. Recall that the power actually is a monotone function of this non-centrality parameter. This is also the case for the estimated power. For the data considered and the above specified B, the value of the non-central parameter obtained is $\lambda = 10.07$. Consequently, the estimated power of the test is calculated as

$$P = P_B(0.04\chi^2_{3,10.07} > c(s)),$$

where $c(s) = 0.16$. The estimated power obtained is 0.94 which is reasonably high.

Let us try another alternative, say

$$\mathbf{B} = \begin{pmatrix} 10.71 & 9.32 \\ 0.91 & 0.87 \end{pmatrix}.$$

As mentioned above the values of a and f remain the same, i.e., $a = 0.04$ and $f = 4$. However, we get $\lambda = 21.11$ which is larger than the value obtained for the previous alternative. The estimated power for this alternative is 0.998 which, as expected, is larger than that of the previous alternative.

Suppose now that

$$\mathbf{B} = \begin{pmatrix} 2.91 & 1.32 \\ 0.091 & 0.087 \end{pmatrix}.$$

For this alternative, we get $\lambda = 1.89$ and the estimated power is 0.5, smaller than the above two values.

We have tried B values very close to zero. We have seen that the power is larger than the level of significance which shows that the test is unbiased for this particular data. It could be interesting to show that this is always the case.

We would like to note here that the above values are estimated values after Σ has been replaced by $\frac{1}{n}S$. However the fact that S gives all information about Σ is no longer true when B has a specified value. Consequently, it is possible to show that the above estimator underestimates the approximated power and perhaps it is also true for the exact power too.

The estimator given (2.19) gives more information about Σ under the alternative hypothesis. Unfortunately, this estimator is no longer a function of only the ancillary statistic which brings complications to the problem. However, the estimate may be used at the last stage. That is, after the conditioning has been done and after the data is observed. We have tried to calculate the estimated power using (2.19) as an estimator for Σ, and found values which are larger than the values found above, i.e., with $\hat{\Sigma} = \frac{1}{n}S$. For example, for the last alternative, i.e.,

$$\mathbf{B} = \begin{pmatrix} 2.91 & 1.32 \\ 0.091 & 0.087 \end{pmatrix},$$

we obtained $a = 5.01, f = 2$ and $\lambda = 1.02$. The estimated power obtained is 0.99 which is much larger than the previous value, which was 0.5. It is also important to note that the estimated power now depends on B not only through λ but also through a and f. As a result, the three parameters are different for different B values specified in the alternative.

Finally let us consider another example. The data consists of four rows. Each row consists of a random sample of 27 observations from a standard normal distribution. Moreover, the rows are independent with each other. Suppose that the first 11 observations belong to one group and the remaining 16 belong to another group. Consider each column as repeated measurements from one individual. The design and parameter matrices for these data are similar to that of the Potthoff & Roy's data.

For this random data the observed value for the test statistic obtained is $\phi(x|s) = 0.024$. The values of a and f obtained are 0.09 and 3, respectively. As a result, $c(s)$ is calculated from

$$P_{H_o}(0.09\chi_3^2 > c(s)) = \alpha.$$

For $\alpha = 0.05, c(s) = 0.70$.

As we can see from the results above $\phi(x|s) = 0.024 < 0.70 = c(s)$. This implies that there is no evidence to reject the hypothesis that $B = 0$. Therefore, the assumption of linear growths for the means is not appropriate. That is actually what we expected since each row is independently taken from a standard normal distribution.

5. Summary

Like in the univariate and ordinary multivariate models, ordinary residuals have been used in checking the adequacy of models in repeated measures and longitudinal analysis and specifically in the Growth Curve Model. However, due to the bilinear structure in the model we believe one should decompose the residuals and examine the different components separately. As to our knowledge, the test considered in this paper is the first in using the right residuals for checking if assumed growth curves fit the mean structures.

It would also be interesting to see various properties of the conditional test presented in this paper. For the Potthoff & Roy model we have seen that the estimated power is larger that the level of significance which was 0.05. We have also tried different levels and found the same results. This indicates that the test might be unbiased. In the future it would be interesting to check if the test really is unbiased. We have also seen that the estimated power obtained by using $\hat{\Sigma} = \frac{1}{n}S$ underestimates the approximate power. It could be interesting to show that this is true for the exact power of the test.

Finally, we would like to mention that the approach utilized in this paper could be extended to the Extended Growth Curve Model. Moreover, the hypothesis could also be formulated in more general form so that it includes a wide range of possibilities.

References

[1] Basu, D. (1964). Recovery of ancillary information. *Sankhya Ser. A* **26**, 3-16.

[2] Basu, D. (1975). Statistical information and likelihood. *sankhya Ser. A* **160**, 268-282.

[3] Basu, D. (1977). On the elimination of nuisance parameters. *J. Amer. statist. Assoc.* **72**, 355-366.

[4] Basu, D. (1978). On partial sufficiency: A review. *J.Stat.Plann. and inf.* **2**, 1-13.

[5] Fisher (1956). *Statistical Methods and Scientific Inference*, Oliver & Boyed, London, UK.

[6] Imhof, J.P. (1961), Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419-426.

[7] Johnson, N.L. & Kotz, S. (1968). Tables of distributions of positive definite quadratic functions in normal vectors. *Ann. Math. Statist.* **37**, 468-479.

[8] Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Mathrices.* Springer, New York.

[9] Kotz, S., Johnson, N.L. & Boyd, D.W. (1967). Series representations of distributions of quadratic forms in normal variables. I. central case, *Ann. Math. Statist.* **38**, 823-837.

[10] Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, Springer-Verlag, New York.

[11] Mathai, A. M. (1982). Storage capacity of a dam with gamma type inputs. *Ann. inst. Statist. Math.* **34**, 3, A, 591-597.

[12] Moschopoulos, P.G. (1985). The didtribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.* **37**, A, 541-544.

[13] Potthoff, R.F. & Roy, S.N. (1964). A generalized multivariate analysis of variance model useful specially for growth curves. *Biometrika* **51**, 313-326.

[14] Press, S.J. (1966) Linear combination of non-central chi-square random variates, *Ann. Math. Statist.* **37**, 480-487.

[15] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, John Wiley & Sons, New York.

[16] von Rosen, D. (1995). Residuals in the growth curve model, *Ann. Inst. Statist. Math.*, **47**, 129-136.

[17] Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics* **2**, 110-114.

[18] Seid Hamid, J. & von Rosen (2005a). Residuals in the extended growth curve model. (Accepted by Scand. J. Statist.)

[19] Seid Hamid, J. & von Rosen, D. (2005b) Hypothesis testing via the residuals in two GMANOVA models. *Research Report 2005:4, Centre of Biostochastics, SLU.*

[20] Sprott, D.A. (1975). Marginal and conditional sufficiency. *Biometrika* **62**, 3, 599-605.

[21] Tan, W.Y. & Gupta, R. P. (1983). On approximating a linear combination of central Wishart matrices with possitive coefficients. *Commun. Statist.-theor. meth.* **12**, 22, 2589-2600.

Table 1. *The Potthoff & Roy (1964) data*

	Age					Age			
Girl	8	10	12	14	Boy	8	10	12	14
1	21.0	20.0	21.5	13.0	1	26.0	25.0	29.0	31.0
2	21.0	21.5	24.0	25.5	2	21.5	22.5	23.0	26.5
3	20.5	24.0	24.5	26.0	3	23.0	22.5	24.0	27.5
4	23.5	24.5	25.0	26.5	4	25.5	27.5	26.5	27.0
5	21.5	23.0	22.5	23.5	5	20.0	23.5	22.5	26.0
6	20.0	21.0	21.0	22.5	6	24.5	25.5	27.0	28.5
7	21.5	22.5	23.0	25.0	7	22.0	22.0	24.5	26.5
8	23.0	23.0	23.5	24.0	8	24.0	21.5	24.5	25.5
9	20.5	21.0	22.0	21.5	9	23.0	20.5	31.0	26.0
10	16.5	19.0	19.0	19.5	10	27.5	28.0	31.0	31.5
11	24.5	25.0	28.0	28.0	11	23.0	23.0	23.5	25.0
					12	21.5	23.5	24.0	28.0
					13	17.0	24.5	26.0	29.5
					14	22.5	25.5	25.5	26.0
					15	23.0	24.5	26.0	30.0
					16	22.0	21.5	23.5	25.0