Paper 262-26

Approximations to Distributions of Test Statistics in Complex Mixed Linear Models Using SAS® Proc MIXED

G. Bruce Schaalje, Department of Statistics, Brigham Young University, Provo, UT
Justin B. McBride, 3M Electronic Handling and Protection Division, Austin, TX
Gilbert W. Fellingham, Department of Statistics, Brigham Young University, Provo, UT

ABSTRACT

The MIXED procedure of SAS® has made use of the linear mixed model accessible to researchers. However, a sticky problem for the procedure has been the specification of appropriate denominator degrees of freedom for test statistics for fixed effects in both balanced designs with simple covariance structures and complex designs involving complicated covariance structures, unbalanced data and/or small sample sizes. This paper compares the denominator degrees-of-freedom options in Proc MIXED.

INTRODUCTION

It is no exaggeration to claim that statistical practice has been changed by the development of the MIXED procedure ('Proc MIXED') of SAS® (Littell et al. 1996). Because of its generality and relative ease of use, Proc MIXED has made linear mixed model technology accessible to researchers in a wide variety of fields and with a wide range of statistical training. Recent books discuss linear mixed models in general (McCulloch and Searle 2001), as well as the Proc MIXED implementation of mixed model calculations (Littell et al. 1996, Verbeke and Molenberghs 1997, Brown and Prescott 1999).

With a little instruction and experience, researchers can use Proc MIXED to intelligently model and analyze data involving multiple random terms, heterogeneous variances, and correlations resulting from clustered and serially measured units. Although there are sometimes modifications to be made when the data are unbalanced, Proc MIXED in general is equally easy to use for balanced or unbalanced (even seriously so) data.

To use Proc MIXED, the linear model for the means or fixed effects is specified in one statement (the MODEL statement) and the variance-correlation structure is specified in one or more separate statements (RANDOM and REPEATED statements). Proc MIXED uses either maximum likelihood or residual maximum likelihood together with estimated generalized least squares to estimate the variance-covariance and mean parameters. Both procedures are statistically defensible, and have well-known optimality properties. Proc MIXED also provides sensible test statistics for linear hypotheses involving fixed effects.

Proc MIXED and linear mixed models in general have a weakness, however. Null distributions of the test statistics are often unknown, and p-values cannot be computed exactly. This is not invariably true because in many balanced data situations with simple covariance structures, test statistics are in fact known to follow F distributions with specific denominator degrees of freedom (ddf). Nonetheless, in a great many situations the distributions are unknown. For both estimation methods, test statistics asymptotically follow the chi-square distribution; but this fact is of little help because small sample distributions are commonly encountered in applications of linear mixed models and can be very different from chi-square.

Some progress can be made by assuming that test statistics approximately follow an F-distribution with a carefully calculated ddf. This paper reviews methods of calculating the ddf in Proc MIXED of SAS® version 8. Methods appropriate for balanced designs are reviewed, and two recent general methods, the Fai-Cornelius (1996) and the Kenward-Roger (1997) methods are described in detail. Past simulation studies of these methods are reviewed, as are results of a new study involving the effects of covariance structures, imbalance, and sample size on adequacy of the Proc MIXED implementation of these methods.

EXAMPLE

In a recent study of differential olfactory responses of male and female lady beetles, Hamilton et al. (1999) examined the antennae of a small sample of male and female lady beetles using electron microscopy. For a sample of 3 males and 3 females, they counted the number of 'sensilla' (small hair-like attachments) on each of 10 segments of an antenna. Although the design was balanced, it was a small sample repeated measures design with an apparently heterogeneous-variance correlation structure for the repeated measurements. Particular attention was focused on the last segment, and its relationship to the other segments and sex. The Proc MIXED statements used in the analysis were:

```
proc mixed data=beetle covtest;
class sex seg indiv last;
model sensilla = last seg(last) sex last*sex / ddfm=      ;
random indiv(sex);
repeated seg / subject=indiv type=arh(1);
```

Even though the design was balanced, the complex covariance structure dictated that the test statistics follow an unknown null distribution. The output below indicates how the tests of fixed effects varied depending on the ddf options available in Proc MIXED. The default option in this case was *containment*.

Effect	Num DF	Den DF	F Value	Pr > F
ddfm=residual				
last	1	48	9713.06	<.0001
seg(last)	8	48	61.52	<.0001
sex	1	48	7.26	0.0097
sex*last	1	48	6.77	0.0123
ddfm=contain				
last	1	44	9713.06	<.0001
seg(last)	8	44	61.52	<.0001
sex	1	4	7.26	0.0544
sex*last	1	44	6.77	0.0126
ddfm=betwithin				
last	1	4	9713.06	<.0001
seg(last)	8	40	61.52	<.0001
sex	1	4	7.26	0.0544
sex*last	1	4	6.77	0.0599
ddfm=satterth				
last	1	3.76	9713.06	<.0001
seg(last)	8	9.84	61.52	<.0001
sex	1	3.96	7.26	0.0549
sex*last	1	3.91	6.77	0.0613
ddfm=kenwardroger				
last	1	3.76	9709.68	<.0001
seg(last)	8	11.3	44.35	<.0001
sex	1	3.96	4.90	0.0918
sex*last	1	3.91	4.61	0.1000

THE LINEAR MIXED MODEL

The linear mixed model can be represented as $\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}$ where **y** is the vector of n responses, **X** is the design matrix of known constants for fixed effects, $\boldsymbol{\beta}$ is the vector of unknown fixed effects, **Z** is the design matrix of known constants for random effects, $\boldsymbol{\gamma}$ is the vector of unknown random effects, and $\boldsymbol{\varepsilon}$ is another vector of random effects or errors. It is assumed that $\boldsymbol{\gamma} \sim \mathbf{N}(\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}_1))$ and $\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_2))$ where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are vectors of variance parameters. Letting $\boldsymbol{\theta} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2$, it can be shown that the covariance matrix of **y** is $\mathbf{V}(\boldsymbol{\theta}) = \mathbf{ZGZ}' + \mathbf{R}$. In Proc MIXED, the model statement is used to set up **X**, random statements are used to set up **Z** and specify the structure of **G**, and the repeated statement is used to specify the structure of **R**. Clearly, there are often alternative ways to specify a given **V**.

The restricted maximum likelihood (REML) estimate (Patterson and Thompson 1971) of $\boldsymbol{\theta}$ is denoted as $\hat{\boldsymbol{\theta}}$, and an estimate of the variance-covariance matrix of **y** is thus $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}})$. $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$, an estimate of the covariance matrix of $\hat{\boldsymbol{\theta}}$, is the inverse of the Hessian matrix of the restricted log likelihood function. Given $\hat{\mathbf{V}}$, the estimated generalized least squares estimate of $\boldsymbol{\beta}$ can be obtained as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$, and the usual approximate covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$.

A commonly used test statistic for fixed effects hypotheses of the form $H_0 : \mathbf{C\beta} = \mathbf{0}$, where **C** is a matrix of contrasts of rank q, is $F = \frac{1}{q}[(\mathbf{C}\hat{\boldsymbol{\beta}})'(\mathbf{C}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{C})^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})]$.

'EXACT' DEGREES OF FREEDOM METHODS

Unlike classical balanced ANOVA algorithms, the likelihood-based algorithms for fitting a linear mixed model and computing the test statistics for fixed effects do not generate correct ddf values automatically, even for those cases in which the distribution is exactly the F distribution. Proc MIXED has incorporated three methods of computing ddf for situations in which the design is balanced and the variance-covariance structure is simple (ddfm=residual, ddfm=contain and ddfm=betwithin).

In the *residual* method, all tests are performed using the residual degrees of freedom, n-rank(**XZ**). This method provides correct degrees of freedom only for 'iid' designs. In other words, this method ignores the covariance structure of the model. For truly mixed models, this method could only be considered useful in very large sample designs where asymptotic distributions are good approximations. This method was the default in very early releases of Proc MIXED, but is no longer ever the default.

The *containment* method is the default when one or more random statements are used to specify the variance-covariance structure. Even if a repeated statement is also used, this method is the default. Under this method, the denominator degrees of freedom for a test of a fixed effect is calculated by searching the random statements for random terms that syntactically (that is, in terms of the way the random effects are *written*) contain the fixed effect of interest. The ddf is calculated as the minimum contribution to the rank of **XZ** of random effects containing the fixed effect of interest. If no such random terms can be found, the ddf is set equal to the residual degrees of freedom.

In the lady beetle example, the individuals were coded 1-6 rather than 1-3 within each sex. Hence, the random statement could have been written as

```
random indiv;
```

rather than

```
random indiv(sex);
```

If it had been specified as in the first statement, however, the ddf for the test of the 'sex' effect would have been calculated as 44 rather than 4 even though the F statistic would have been exactly the same, because the random term did not syntactically contain 'sex'. Also note in the lady beetle example that ddf for the other fixed effects were calculated as 44 rather than 48 (as per the residual method). This appears to be because the variance component for individuals was estimated as 0, so that **ZG'Z** would have been **0** even if **Z** were not **0**. In such cases the residual method apparently considers the rank contribution of **Z** to **XZ** to be 0 in such cases while the containment method does not.

Although the containment method cleverly mimics classical degrees of freedom rules for balanced ANOVA situations, it should only be considered to give exact ddf when the design is balanced, there is no structure in the **R** matrix, and care has been taken to ensure that the random statement has been written so that the relationships between fixed and nested random effects is clear. In some other cases, such as those with moderate imbalance, the method may give adequate approximate results.

The *between-within* method is the default when a repeated statement is used to specify the variance-covariance structure, and no random statements are used. Ignoring the covariance structure specified by the 'type=' option (unless type=un), this method divides the residual degrees of freedom into between-subjects and within-subjects values. According to the documentation (SAS Institute 1999), fixed effects that do not change within subjects (as specified in the repeated statement) are assigned the between-subjects value as their ddf. All others are assigned the within-subjects value as their ddf. In the lady beetle example, only antennae segments were assigned the within-subjects value even though levels of the factor 'last' also change within individuals. This apparently is because there are only two levels of 'last' and the levels are replicated within subjects. The between-within method should only be considered to give exact ddf values for balanced repeated measures designs with the compound symmetry covariance structure (type=cs) for which levels of the within-subjects effect are not replicated within any of the subjects. In other cases, the method is at best approximate, and can be unpredictable. This method should be used with care.

APPROXIMATE DEGREES OF FREEDOM METHODS

For complex mixed models involving complicated covariance structures and/or unbalanced data sets, the test statistics for fixed effects follow unknown distributions. Assuming that the test statistics approximately follow an F-distribution with carefully calculated degrees of freedom, Proc MIXED has incorporated two recently proposed methods for computing ddf (or otherwise modifying the test statistics) for cases such as these. Fai and Cornelius (1996) proposed a method for multi-degree-of-freedom tests in unbalanced split plot designs, and Kenward and Roger (1997) proposed a method for tests in mixed linear models based on any covariance structure. In SAS® version 8, Proc MIXED specifies the Fai-Cornelius (FC) and Kenward-Roger (KR) methods as ddfm=satterth and ddfm=kenwardroger, respectively.

The FC method involves the spectral decomposition of $(\mathbf{C}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{C}')^{-1}$ to yield $\mathbf{P}'(\mathbf{C}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{C}')^{-1}\mathbf{P} = \mathrm{diag}(\lambda_m)$ where columns of **P** are normalized eigenvectors and the λ_m are the corresponding eigenvalues of $(\mathbf{C}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{C}')^{-1}$. Using this decomposition, Q = qF can be written as a sum of q approximate squared *t*-variables,

$$Q = \sum_{m=1}^{q} \frac{(p'_m \mathbf{C}\hat{\boldsymbol{\beta}})^2}{\lambda_m} = \sum_{m=1}^{q} t^2_{U_m}$$

where $\mathbf{p}'_m$ is the m^{th} eigenvector and υ_m is the approximate degrees of freedom for the m^{th} independent single degree of freedom t-test. Once the decomposition is performed, the method computes υ_m values by repeatedly applying a method for single degree of freedom contrasts proposed by Giesbrecht and Burns (1985). If **c** is a vector of constants for the single-degree-of-freedom test of $H_0 : \mathbf{c}'\boldsymbol{\beta} = 0$, Giesbrecht and Burns follow Satterthwaite's (1941) premise to recommend

$$\frac{2(\mathbf{c}'\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{c})^2}{[\mathrm{var}(\mathbf{c}'\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{c})]}$$

as the ddf. The denominator of this expression is approximated using the multivariate delta method (Lehman 1998). For the test with q>1, Fai and Cornelius (1996) note that each υ_m can be approximated by the Giesbrecht-Burns single degree of freedom method. Using the relationship $E(F_{FC_{q,\upsilon}}) = \frac{\upsilon}{\upsilon - 2}$ for $\upsilon > 2$, they then find υ (=ddf) such that $q^{-1}Q \sim F_{q,\upsilon}$ approximately. Since the t_{υ_m} can be regarded as having independent *Student's t* distributions with υ_m degrees of freedom,

$$E(Q) = \sum_{m=1}^{q} E(t^2_{\upsilon_m}) = \sum_{m=1}^{q} E(F_{q,\upsilon})$$

$$= \sum_{m=1}^{q} \frac{\upsilon_m}{\upsilon_m - 2}$$

$$= E_Q \text{ (say).}$$

Now, since

$$\frac{1}{q}E_Q = \frac{\upsilon}{\upsilon - 2}$$

it can be shown that

$$\upsilon = \frac{2E_Q}{E_Q - q}.$$

The KR method first implements an adjustment to $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$ to account for small sample bias and incorporate the variability in $\hat{\boldsymbol{\theta}}$. Kenward and Roger (1997) extend Kackar and Harville's (1984) approximation to obtain $\hat{\boldsymbol{\Sigma}}^*_{\hat{\boldsymbol{\beta}}}$. They then calculate a scale factor, δ, and the approximate ddf, υ. The test statistic is

$$F^* = \delta F_{KR} = \frac{\delta}{q}(\mathbf{C}\hat{\boldsymbol{\beta}})'(\mathbf{C}\hat{\boldsymbol{\Sigma}}^*_{\hat{\boldsymbol{\beta}}}\mathbf{C}')^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}}).$$

Kenward and Roger use a second order Taylor series expansion of $(\mathbf{C}'\hat{\boldsymbol{\Sigma}}^*_{\hat{\boldsymbol{\beta}}}\mathbf{C})^{-1}$ about $\boldsymbol{\theta}$ and conditional expectation relationships to yield $E(F_{KR})$ and $V(F_{KR})$ approximately. The approximate moments of F^* are then generated and equated to the moments of an F distribution to solve for δ and υ. This yields

$$\upsilon = 4 + \frac{q+2}{q\gamma - 1} \quad \text{and} \quad \delta = \frac{\upsilon}{\tilde{E}[F_{KR}](\upsilon - 2)}$$

where $\gamma = \frac{\tilde{V}[F_{KR}]}{2\tilde{E}[F_{KR}]^2}$.

SIMULATION STUDIES

Fai and Cornelius (1996) tested their ddf method, along with two other very closely related methods, using simulation studies with 'split-plot designs'. In split plot designs the variance-covariance structure is always compound symmetric (one of the simplest structures) due to the presence of independent whole plots and conditionally independent nested subplots. In their studies, Fai and Cornelius varied the degree of imbalance and the relative magnitudes of the whole-plot and sub-plot error variances from 1:2 to 8:1. They used four basic designs, all of which had 3 levels of the whole-plot factor, and either 3 or 4 levels of the subplot factor. The designs had between 9 and 18 whole plots, and there were between 1 and 4 subplots per whole plot. Only one design had instances of 1 subplot per whole plot. The designs were all highly unbalanced, but usually one level of the whole plot factor was combined several times with every level of the subplot factor. The designs were all of moderate sample size, having between 25 and 48 total observations. One design had 25 observations, and the others had at least 35 observations.

In all of their simulation studies, Fai and Cornelius found that the observed Type I error rates for the whole-plot factor were reasonably close to nominal values, regardless of the relative magnitudes of the error variances or the ddf method. Design 3, with the smallest sample size, tended to produce Type I error rates slightly larger than nominal values (eg. .01 rates varied from .014 to .019). Design 4, with instances of whole plots with only 1 subplot, tended to produce Type I error rates slightly smaller than nominal values (eg. .05 rates varied from .040 to .047). In general, though, the Fai-Cornelius method appeared to be very successful in approximating the null distribution of the test statistics for highly unbalanced split plot designs.

Kenward and Roger (1997) used simulation to investigate performance of their ddf method under a variety of variance-covariance structures. Study 1 was a 4-treatment 2-period crossover design, with 12 main units and 24 total observations. The relative magnitudes of the main-unit and sub-unit error variances varied from 1:4 to 4:1. Study 2 was a partially balanced incomplete row-column design with 3 replicates of 12 treatments in 3X4 arrays. Thus there were 36 total observations. The row and column variances were equal to each other, but the relative magnitudes of the row-column variance and the residual variance ranged from 1:4 to 4:1. Study 3 was a 2-level straight-line random coefficients model with 24 subjects split into 3 groups of 8, and 3 repeated observations on each subject. The study was highly unbalanced with observations taken at completely different times for each of the 3 groups. There were 72 total observations. The variances of the slopes and intercepts were set equal to each other, and the magnitude of this variance relative to the residual error variance was set at 1:2 and 1:1. Finally, study 4 was a repeated measures design with the ante-dependence covariance structure (a non-linear structure) for the repeated measurements. There were 2 groups of 10 units, each measured at the same 6 times, except that there were several dropouts at various times. There were 90 total observations. One specific set of values was chosen for the parameters of the variance-covariance structure.

In study 1of Kenward and Roger, observed Type 1 error rates were very close to nominal values, but were possibly slightly large (the .05 rates varied from .052 to .055). In study2, the observed Type I error rates for several pairwise contrasts were also close to but slightly larger than nominal values (.05 rates varied from .054 to .059). In study 3, the observed Type I error rates for tests of the intercept and the slope were very close to nominal values, and apparently not biased. In study 4, tests of polynomial coefficients and their interactions had Type I error close to but slightly larger than nominal values (.05 rates varied from .050 to .054).

Even though the Fai-Cornelius (1996) and Kenward-Roger (1997) simulations are informative, SAS documentation (SAS

Institute Inc., 1999) warns that full information about the performance of these approximate ddf methods is not available. For example, the documentation states that properties of the FC method have not been fully examined for the various covariance structures available in the procedure, especially for small samples. Because of this, McBride (2000) carried out additional simulations studies of the FC and KR ddf methods. Unlike the previous studies, McBride used these methods as implemented in Proc MIXED.

All designs in the McBride study were repeated measures designs. Whole units were assigned to one of 3 levels of a fixed factor, and subunits were assigned to one of 3 levels of another fixed factor. Data sets were generated using 20 combinations of 4 sample size options and 5 covariance structures. The 4 sample size options consisted of combinations of two numbers of whole plots per treatment, namely 3 and 5, and two numbers of sub-plots per whole plot, also 3 and 5. The total number of observations in these simulations thus ranged from 27 to 75. Because the subunit factor had 3 levels, when there were 5 subunits per whole unit the design was unbalanced. In such cases, each sub-plot treatment appeared at least once, but two of the three appeared twice within each whole plot.

The 5 covariance structures were Compound Symmetry, Toeplitz, Heterogeneous Compound Symmetry, First-Order Heterogeneous Autoregressive, and First-Order Ante-dependence (SAS Institute Inc., 1999) with specified values of the parameters. The First-Order Ante-dependence parameter values were taken from the Kenward and Roger (1997) simulation study so that results from this simulation could be compared to their results. Diagonal elements for the First-Order Heterogeneous Autoregressive and Heterogeneous Compound Symmetry structures were the same as in the First-Order Ante-dependence structure in order to be able to evaluate the effect of covariance complexity on effectiveness of the methods.

For the FC method, Type I error rates were close to the nominal values for the CS structure regardless of sample size, type of treatment being tested, or balance (Table 1). These results agree with previous simulation results of Fai and Cornelius (1996) and act as a check that the simulations were set up properly. The FC method gave dramatically different results when other covariance structures were used. In fact, there was only one scenario in which Type I error rates were close to nominal values, namely the balanced design with 5 whole plots per treatment and 3 subplots per whole plot based on the TOEP structure. For covariance structures more complex than CS, the FC method apparently produces approximate degrees of freedom that are too large because in all other cases, Type I error rates were inflated. Although the FC method produced valid tests for unbalanced designs using the CS structure, imbalance seems to affect the FC method when the TOEP structure is used because a larger unbalanced design exhibited inflated Type I error rates while a smaller balanced design did not. Even though Type I error target values were not achieved for most simulations using the TOEP, CSH, ARH(1) and ANTE(1) structures, simulated Type I error rates were always closer to target values for balanced designs than for unbalanced designs. Also, when comparing designs of similar balance, Type I error rates were always closer to target values for designs with larger sample sizes.

The KR method produced Type I error rates very close to nominal values for all simulations involving the CS structure, for most simulations involving the TOEP structure, and for the larger sample size simulations involving the CSH and ARH(1) structures (Table 1). If one were willing to accept error rates as high as 0.07 as reasonable practical approximations to the nominal value 0.05, many of the smaller sample size simulations involving the CSH and ARH(1) structures could also be considered to produce acceptable Type I error rates. Even the larger sample size simulations involving the ANTE(1) structure produced simulated Type I error rates that would be acceptable in this sense. Imbalance apparently also affects the KR method because simulated Type I error rates were usually closer to target values for balanced designs than for unbalanced designs based on the TOEP, CSH, ARH(1), and ANTE(1) structures.

Table 1. Type I error rates (nominal value .05) in McBride (2000)

				Proportion of p-values less than .05 (standard error = .002)				
Meth	*Test*	*Whole*	*Sub*	*CS*	*TOEP*	*CSH*	*ARH*	*ANTE*
FC	*Wh*	5	5	.052	.063	.090	.083	.115
			3	.045	.053	.089	.084	.093
		3	5	.049	.087	.142	.128	.224
			3	.047	.070	.130	.123	.164
	Sub	5	5	.052	.101	.091	.089	.142
			3	.052	.079	.082	.090	.121
		3	5	.049	.151	.141	.142	.261
			3	.049	.102	.126	.133	.201
KR	*Wh*	5	5	.052	.050	.058	.056	.065
			3	.045	.046	.058	.055	.060
		3	5	.049	.054	.082	.071	.117
			3	.047	.051	.071	.070	.083
	Sub	5	5	.052	.061	.056	.054	.066
			3	.052	.055	.051	.055	.065
		3	5	.048	.070	.070	.065	.109
			3	.049	.064	.063	.069	.093

RECOMMENDATIONS

When using Proc MIXED for linear mixed model calculations, the choice of a ddf method for tests of fixed effects should not be taken lightly, even for balanced designs. The default 'exact' method should certainly not be mindlessly accepted. Proc Mixed users should even think about ddf methods when formulating the model and its syntax.

The approximate ddf methods come closer to serving as automatic methods than the exact methods. However, complexity of the covariance structure, sample size, and imbalance affect the performance of both approximate ddf methods. These factors affect the FC method much more than the KR method. The FC method can only be recommended for use with the CS covariance structure or with simple structures such as TOEP when sample sizes are moderately large. The KR method works reasonably well with more complicated covariance structures when sample sizes are moderate to small and the design is reasonably balanced. Even the KR method had problems, however, with structures as complex as ANTE(1) when sample sizes were small.

Even though it worked well in connection with the CS structure, there seems little reason to use the FC method now that the KR method is available. From the simulations, it appears that the KR method works as well as or better than the FC method in all situations. When using the KR method, however, it must be kept in mind that it does not perform perfectly in all situations. When the covariance structure is complex and the sample size is small, the KR method produces inflated Type I error rates. Hence, the p-value produced by the KR method in such situations should be treated with caution, perhaps as a lower bound. Research should continue on methods of approximating the distributions of test statistics in mixed models and on the implementation of these methods in Proc MIXED.

REFERENCES

Brown, H. and Prescott, R. (1999), *Applied Mixed Models in Medicine*, New York: John Wiley and Sons.

Fai, A.H.T. and Cornelius, P.L. (1996), "Approximate F-tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments," *Journal of Statistical Computing and Simulation* 54: 363-378.

Giesbrecht, F.G. and Burns, J.C. (1985), "Two-Stage Analysis Based on a Mixed Model: Large-sample Asymptotic Theory and Small-Sample Simulation Results," *Biometrics* 41: 853-862.

Hamilton, R.M., Dogan, E.B., Schaalje, G.B., and Booth, G.M. (1999), "Olfactory Response Of The Lady Beetle Hippodamia Convergens (Coleoptera: Coccinellidae) To Prey Related Odors, Including A SEM Study Of The Antennal Sensilla," *Environmental Entomology* 28:812-822.

Kackar, R.N. and Harville, D.A. (1984), "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models," *Journal of the American Statistical Association* 79: 853-862

Kenward, M. G. and Roger, J. H. (1997), "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood," *Biometrics* 53: 983-997.

Lehman, E. (1998), *Elements of Large Sample Theory*, New York: Springer-Verlag.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996), *SAS System for Mixed Models.* SAS Institute Inc.: New York.

McBride, J. B. (2000), *Adequacy of Approximations to Distributions of Test Statistics in Complex Mixed Linear Models*, M.S. Project, Brigham Young University.

McCulloch, C. E. and Searle, S. R. (2001), *Generalized, LIear, and Mixed Models*, New York: John Wiley and Sons.

Patterson, H. D. and Thompson, R. (1971), "Recovery of Inter-Block Information When Block Sizes are Unequal," *Biometrika* 58: 545-554.

SAS® version 8, Online Help (1999), SAS Institute Inc, Cary, NC.

Satterthwaite, F. E. (1941), "Synthesis of Variance," *Pyschometrika* 6: 309-316

Verbeke, G. and Molenberghs, G. (1997), *Linear Mixed Models in Practice, A SAS-Oriented Approach*, New York: Springer.

CONTACT INFORMATION

G. Bruce Schaalje
Department of Statistics
Brigham Young University
Provo, UT 84602
Phone: (801) 378-3996
Fax: (801) 378-5722
schaalje@byu.edu

Justin B. McBride
3M Electronic Handling and Protection Division
Bld. A 141-4N-02
6801 River Place Blvd
Austin, TX 78726
Phone: (512) 984-4863
jmcbride@mmm.com

Gilbert W. Fellingham
Department of Statistics
Brigham Young University
Provo, UT 84602
Phone: (801) 378-
Fax: (801) 378-5722
gwf@byu.