Einhaltung des statistischen Risikos 1. Art in unbalancierten gemischten linearen Modellen bei unterschiedlichen Restriktionen der REML-Schätzer und Berechnung deren Schätzfehler

Joachim Spilke
Universität Halle-Wittenberg
Landwirtschaftliche Fakultät
06099 Halle
spilke@landw.uni-halle.de

Hans Peter Piepho
Universität Hohenheim
Fakultät Agrarwissenschaften
70599 Stuttgart
piepho@uni-hohenheim.de

Nobert Mielenz
Universität Halle-Wittenberg
Landwirtschaftliche Fakultät
06099 Halle
mielenz@landw.uni-halle.de

Xiyuan Hu
Universität Halle-Wittenberg
Landwirtschaftliche Fakultät
06099 Halle
hu@landw.uni-halle.de

Zusammenfassung

Basierend auf einer Simulationsstudie wird der Einfluss verschiedener Restriktionen auf die Schätzwerte von Varianzkomponenten (Varianzkomponenten positiv vs. Matrix V positiv definit), Optimierungsalgorithmen (Newton-Raphson vs. Fisher-Scoring) sowie Berechnungen der für die Freiheitsgradapproximationen im statistischen Test erforderlichen Schätzfehler der Varianzkomponenten (Informationsmatrix vs. Hesse-Matrix) auf die Einhaltung des vorgegebenen statistischen Risikos 1. Art für eine Spaltanlage mit unbalancierten Daten geprüft. Für Vergleiche der Stufen der Kleinteilstückfaktoren und der Faktorkombinationen konnte die Einhaltung des nominalen Risikos verbessert werden, nicht jedoch für die Stufen des Grossteilstückfaktors.

Keywords: gemischte Modelle, Unbalanciertheit, Hypothesenprüfung, Risiko 1. Art.

1 Einleitung und Problemstellung

Die Problematik statistischer Tests in gemischten linearen Modellen mit unbalancierten Daten ist lange bekannt und ursächlich mit der Situation verbunden, dass bei den meisten praktischen Anwendungen von unbekannten Varianzkomponenten ausgegangen werden muss (Henderson, 1984, p. 83). Basierend auf den unter Normalverteilung nahezu optimalen Eigenschaften der Schätzer (asymptotische Konsistenz, Erwartungstreue sowie asymptotische Normalverteilung der Schätzwerte und asymptotisch angebbare Fehlervarianzen der Schätzungen) (Searle et al., 1992, p. 254) hat sich insbesondere REML gegenüber verschiedenen Schätzkonzepten durchgesetzt, u.a. gegenüber dem ANOVA-Konzept. Die praktischen Vorteile von REML- gegenüber ANOVA-Schätzern konnten auch in einer Simulationsstudie für landwirtschaftlich bedeutsame Datenstrukturen gezeigt werden (Spilke und Tuchscherer, 2001).
Allerdings ist zu bemerken, dass die Forderungen an einen REML-Schätzer nicht eindeutig sind. So schränkt Searle (1971, p. 418) die mit der Bereitstellung von REML-Schätzern verbundene Maximierung der Likelihoodfunktion streng auf den Parameterraum ein. Das ist gleichbedeutend mit Varianzschätzern größer gleich Null. Diese Forderung ist für populationsgenetische Anwendungen zur Schätzung genetischer Parameter sinnvoll. Für Aufgabenstellungen mit dem Ziel der Schätzung fester Effekte und Hypothesenprüfung ist es jedoch ausreichend, die positive Definitheit der Varianz-Kovarianz-Matrix des Beobachtungsvektors (V) zu sichern. Das bedeutet, dass nicht zwangsläufig nur ausschließlich positive Varianzschätzer akzeptiert werden müssen (vgl. Piepho und Spilke, 1999). Die positive Definitheit der Matrix V besagt definitionsgemäß, dass quadratische Formen b'Vb für beliebige $b \neq 0$ positiv sein müssen. Anschaulich heißt das, dass jede Linearkombination L=b'y des Beobachtungsvektors y eine positive Varianz besitzt. Untersuchungen zeigen, dass es bei Nutzung eines REML-Algorithmus, der nur Schätzwerte größer Null der Varianzkomponenten akzeptiert, zu einer Verzerrung der Varianzkomponentenschätzungen kommen kann. Dieser Sachverhalt prägt sich besonders bei kleinem Datenumfang und einem weiten Verhältnis der Varianzkomponenten aus (Spilke et al., 2003). Es besteht aber die Erwartung, mit einer Modifikation des REML-Algorithmus in der Weise, dass alleiniges Kriterium für die Akzeptanz der Schätzung die positive Definitheit der Matrix V ist, diese Verzerrung zu reduzieren und zu einer Verbesserung der Einhaltung des vorgegebenen statistischen Risikos 1. Art zu kommen.

Bezogen auf die SAS-Prozedur Mixed hat die Begrenzung der Varianzkomponenten auf den Parameterraum die Konsequenz, dass jede Schätzung einer Varianzkomponente von Null zu einer Modellreduzierung führt. Das eigentliche „Rechenmodell“ ist damit identisch mit einem Modell ohne den jeweiligen zufälligen Effekt. Damit werden nur Schätzungen größer Null akzeptiert, ansonsten wird der Effekt ignoriert, wenngleich im Ergebnisausdruck ein Schätzwert von Null für die betreffende Varianzkomponente ausgewiesen wird.
Es ist das Ziel dieser Arbeit, mit Hilfe einer Simulationsstudie den Einfluss unterschiedlicher Restriktionen auf die Schätzwerte (Varianzkomponenten positiv vs. Matrix V positiv definit), verschiedener Optimierungsalgorithmen (Newton-Raphson vs. Fisher-Scoring) sowie Berechnungen der für die Freiheitsgradapproximationen im statistischen Test erforderlichen Schätzfehler der Varianzkomponenten (Informationsmatrix vs. Hesse-Matrix) auf die Einhaltung des vorgegebenen statistischen Risikos 1. Art für ausgewählte Kontraste zu prüfen.

2 Material und Methode

2.1 Simuliertes Modell und Datenstruktur

Die Untersuchung bezieht sich auf eine zweifaktorielle Spaltanlage, der wegen ihrer verbreiteten Nutzung eine besondere Bedeutung zukommt. Dabei werden die Beobachtungen y_{ijk} als Realisationen einer Zufallsvariable $\underline{y}_{ijk}$ angesehen, für die das folgende Modell gilt:

$$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \underline{bl}_k + \underline{f}_{ik} + \underline{e}_{ijk} \qquad (1)$$

(i=1,..,a; j=1,...,b; k=1,...,r)

mit

$$\underline{bl}_k \sim N(0,\sigma^2_{bl}); \underline{f}_{ik} \sim N(0,\sigma^2_{RA}); \underline{e}_{ijk} \sim N(0,\sigma^2_{R})$$

Dabei bedeuten:

μ	= allgemeines Mittel
α_i	= fester Effekt der i-ten Stufe des Faktors A
$ß_j$	= fester Effekt der j-ten Stufe des Faktors B
$(\alpha ß)_{ij}$	= fester Interaktionseffekt der i-ten Stufe des Faktors A und der j-ten Stufe des Faktors B
$\underline{bl}_k$	= zufälliger Blockeffekt
$\underline{f}_{ik}$	= zufälliger Grossteilstückfehler
$\underline{e}_{ijk}$	= zufälliger Resteffekt.

Es werden die in praktischen Anwendungen häufigen Datenstrukturen a×b×r = 2×3×3 bzw. 2×3×4 analysiert. Dabei wird der ebenfalls praktisch bedeutsame Fall betrachtet, dass jeweils zwei zufällig ausgewählte Beobachtungen für die Auswertung nicht zur Verfügung stehen. Entsprechend sind 16 bzw. 22 Beobachtungen für die Analyse verfügbar.
Um weiterhin einen weiten Bereich praktischer Gegebenheiten abzudecken, werden die Varianzkomponenten so gewählt, dass sie einem Verhältnis von 1, 0.5 und 0.1 zur Restvarianz entsprechen.
Wichtigstes Bewertungskriterium für den Methodenvergleich ist der Vergleich von nominalen und empirischen Risiko. Entsprechend wurde bei der Planung der Wiederholungsanzahlen das Konfidenzintervall einer Wahrscheinlichkeit zugrunde gelegt. Die für jede Simulationsvariante durchgeführten 10.000 Wiederholungen sichern für einen Wahrscheinlichkeitsparameter von 0.05 eine halbe Intervallbreite von 0.0043 bei einem Konfidenzniveau von 95%, was als eine ausreichende Genauigkeitsvorgabe angesehen wird.
Die Datensimulation entsprechend (1) wurde bei Nutzung des Data Step in der Software SAS (Version 8.02) vorgenommen.

2.2 Untersuchte Kontraste fester Effekte

Für alle festen Effekte im Modell (1) werden Werte von Null vorgegeben.
Wir untersuchen die Kontraste mit den folgenden Modelleffekten:

$$\mu_{1.} - \mu_{2.} = \alpha_1 - \alpha_2 + \frac{1}{b}\left(\sum_{j=1}^{b}(\alpha\beta)_{1j} - \sum_{j=1}^{b}(\alpha\beta)_{2j}\right)$$

$$\mu_{.1}-\mu_{.2}=\beta_1-\beta_2+\frac{1}{a}\left(\sum_{i=1}^{a}(\alpha\beta)_{i1}-\sum_{i=1}^{a}(\alpha\beta)_{i2}\right)$$

$$\mu_{11}-\mu_{22}=\alpha_1-\alpha_2+\beta_1-\beta_2+(\alpha\beta)_{11}-(\alpha\beta)_{22}$$

Die Zusammensetzung der untersuchten Kontraste ergibt sich entsprechend Modell (1). Wegen der Vorgabe von Null für alle Effekte ist die Nullhypothese durchgehend gültig.

2.3 Analyse des gemischten Modells

Eine Zusammenstellung aller untersuchten Auswertungsvarianten liefert Tabelle 1.
Für die Datenanalyse steht zunächst die Prozedur Mixed des SAS-Systems (Version 8.2) zur Verfügung. Mit diesem System können REML-Schätzer mit der Bedingung größer Null gewonnen werden (Verfahren REML). Falls auch Schätzwerte kleiner Null akzeptiert werden sollen, ist die Option „NOBOUND" zu verwenden (Verfahren REML-N).
Im Verlauf unserer Untersuchungen zeigte sich aber, dass für einen hohen Anteil simulierter Datensätze keine Schätzungen der festen Effekte als auch deren Standardfehler erzeugt werden (vgl. Abschnitt 3.1). Um dennoch den im Zusammenhang mit der Hypothesenprüfung wichtigen Fall „Varianzkomponenten kleiner Null" zu analysieren, wurde ein Programm bei Nutzung der SAS – Prozedur IML geschrieben (Verfahren REML-IML). Dabei wurden als Startwerte für die Optimierung die geschätzte Varianz des Beobachtungsvektors des jeweiligen Simulationslaufes benutzt (0.8 bzw. 0.5 der geschätzten Varianz als Startwert für $\hat{\sigma}^2_{bl}$ und $\hat{\sigma}^2_{RA}$ bzw. $\hat{\sigma}^2_{R}$). Eine vergleichbare Vorgehensweise wurde ebenfalls bei Nutzung der SAS-Prozedur Mixed realisiert (REML-N-start). In allen übrigen Fällen wurden als Startwerte die Schätzwerte gemäß MIVQUE0 verwendet, wie das standardmäßig in der SAS-Prozedur Mixed vorgesehen ist.
Als Optimierungsverfahren wurde der Newton-Raphson-Algorithmus genutzt, mit Ausnahme der Variante REML-N-score. Diese in Proc Mixed implementierte Vorgehensweise nutzt als Verfahren das Fisher-Scoring sowie zur Bereitstellung der Fehlervarianzen der Varianzkomponenten die Informationsmatrix.
Zum Vergleich der Nutzung von Informations- und Hesse-Matrix (Sorensen and Gianola, 2002 pp 131-139) zur Berechnung der Fehlervarianzen der

Varianzkomponenten wurde in einem weiteren IML-Programm auch die Hesse-Matrix benutzt (REML-IML bzw. REML-IML-obs).

Tabelle 1: Untersuchte Auswertungsvarianten und verwendete Bezeichnungen

<table>
<tr><th></th><th>Optimie-rungs-verfahren</th><th>Startwerte</th><th>Berechnung der Fehler der Varianzen</th><th>Restriktionen an die Schätzwerte</th><th>Bezeichnung</th></tr>
<tr><td rowspan="4">PROC MIXED</td><td rowspan="3">Newton-Raphson</td><td rowspan="2">MIVQUE0</td><td rowspan="3">Hesse-Matrix (observed Hessian)</td><td>positiv</td><td>REML</td></tr>
<tr><td rowspan="3">„Nobound“</td><td>REML-N</td></tr>
<tr><td>Varianz der Beobachtungen</td><td>REML-N-start</td></tr>
<tr><td>Fisher-Scoring</td><td>MIVQUE0</td><td rowspan="2">Informationsmatrix (expected Hessian)</td><td>REML-N-score</td></tr>
<tr><td rowspan="2">PROC IML</td><td rowspan="2">Newton-Raphson</td><td rowspan="2">Varianz der Beobachtungen</td><td rowspan="2">Kovarianzmatrix des Beobachtungs-vektors positiv definite</td><td>REML-IML</td></tr>
<tr><td>Hesse-Matrix (observed Hessian)</td><td>REML-IML-obs</td></tr>
</table>

Die positive Definitheit der Matrix V in den IML-Programmen (REML-IML, REML-IML-obs) wurde durch die Forderung nach ausschließlich positiven Eigenwerten abgesichert. Diese Anforderung wurde über einen Strafterm bei der Maximierung der Likelihood-Funktion berücksichtigt. Der Strafterm lieferte nur bei negativ von Null verschiedenen Eigenwerten einen negativen Beitrag zur Likelihoodfunktion.
Bei der Hypothesenprüfung wurde mit zwei Approximationen für die Freiheitsgrade gearbeitet. Die Satterthwaite-Methode (Giesbrecht and Burns,1985; Fai and Cornelius, 1996), verfügbar bei Nutzung der Option METHOD=SATTERTH in der Prozedur Mixed und die Methode von Kenward and Roger (1997), verfügbar bei Nutzung der Option METHOD=KENWARDROGER. Die Differenz zwischen Satterthwaite-Methode und der Kenward-Roger-Methode besteht darin, dass bei letztgenannter Methode eine Korrektur der Standardfehler der festen Effekte nach Kackar and Harville (1981) erfolgt.

2.4 Untersuchte Maßzahlen

Zur Charakterisierung der Auswertungsmethoden verwenden wir die folgenden Maßzahlen:
Bias der geschätzten Varianzkomponenten,
Anteil nicht abgeschlossener Rechnerläufe,
Abweichung der mittleren geschätzten (Mittel aus den Schätzungen der Standardfehler je Simulationslauf) und empirischen Standardfehler (geschätzte Standardabweichung aus den Effektschätzungen je Simulationslauf),
Empirischer Fehler 1. Art als Anteil beobachteter Ablehnungen der Nullhypothese bei einem nominalem Fehler α= 5%.

3 Ergebnisse der Untersuchungen

3.1 Schätzung der Varianzkomponenten

Die in den Tabellen 2 und 3 zusammengestellten Ergebnisse der Varianzkomponentenschätzung zeigen einen deutlichen Bias der Schätzwerte für die Variante REML und untermauern damit die Motivation für die vorliegende Untersuchung.
Für die beiden Datenstrukturen ist der Bias der Varianzkomponentenschätzungen bei REML-IML gegenüber REML beträchtlich reduziert. Das betrifft insbesondere die Schätzwerte für σ^2_{bl} und σ^2_{RA} für die Datenstruktur 2×3×3. Dabei ist die Verzerrung der Varianzschätzer bei einem weiten Verhältnis der Varianzkomponenten i.A. stärker ausgeprägt.
Der deutliche Unterschied zwischen den untersuchten REML-Schätzern wird erklärlich wenn man beachtet, dass für die Datenstruktur 2×3×3 bei Nutzung der Methode REML der Prozedur Mixed und einem Varianzverhältnis 0.1:0.1:1 in mehr als der Hälfte der Fälle für σ^2_{bl} und σ^2_{RA} Schätzwerte von Null auftreten, für das Varianzverhältnis 1:1:1 beträgt der Anteil noch 1/3 bzw. 1/4. Die Begrenzung der Varianzkomponenten auf den Parameterraum hat die Konsequenz, dass jede Schätzung einer Varianzkomponente von Null zu einer Modellreduzierung führt. Das eigentliche „Rechenmodell“ ist damit identisch mit einem Modell ohne den jeweiligen zufälligen Effekt. Damit werden nur Schätzungen größer Null akzeptiert, ansonsten wird der Effekt

ignoriert, wenngleich im Ergebnisausdruck ein Schätzwert von Null für die betreffende Varianzkomponente ausgewiesen wird.

Tabelle 2: Bias (%) der geschätzten Varianzkomponenten
(Spaltanlage 2x3x3, zwei fehlende Beobachtungen, 10,000 Simulationsläufe)

	1:1:1			0.5:0.5:1			0.1:0.1:1		
	σ^2_{bl}	σ^2_{RA}	σ^2_{R}	σ^2_{bl}	σ^2_{RA}	σ^2_{R}	σ^2_{bl}	σ^2_{RA}	σ^2_{R}
REML	15,8	-10,2	-4,4	20,4	-5,1	-6,8	66,2	60,0	-12,1
REML-N*	-2,4	1,6	3,1	-3,3	2,7	2,6	-8,4	5,7	2,3
REML-N-start*	-4,4	4,0	1,0	-3,4	5,6	1,0	24,0	17,0	0,9
REML-N-score*	50,4	56,6	1,1	24,7	35,2	1,3	-4,1	24,0	2,0
REML-IML	-4,4	4,3	-0,9	-5,3	6,2	-2,1	-5,4	27,6	-6,0
REML-IML-obs*	-3,6	10,0	-1,3	-4,8	17,3	-2,7	-8,8	72,4	-5,5

* Verfahren mit unvollständig abgeschlossenen Läufen (vgl. Tab. 4)

Tabelle 3: Bias (%) der geschätzten Varianzkomponenten
(Spaltanlage 2x3x3, zwei fehlende Beobachtungen, 10,000 Simulationsläufe)

	1:1:1			0.5:0.5:1			0.1:0.1:1		
	σ^2_{bl}	σ^2_{RA}	σ^2_{R}	σ^2_{bl}	σ^2_{RA}	σ^2_{R}	σ^2_{bl}	σ^2_{RA}	σ^2_{R}
REML	10,7	-7,9	-1,7	14,1	-5,9	-3,4	45,6	40,3	-8,3
REML-N*	-1,9	2,3	1,6	-2,0	2,3	1,8	-4,5	-0,2	2,2
REML-N-start*	-2,4	2,9	1,4	-2,8	2,6	1,7	-5,0	0,9	1,9
REML-N-score*	75,7	74,4	-0,5	15,0	17,6	0,7	-4,1	6,5	1,9
REML-IML	-2,4	2,8	1,0	-2,9	2,9	1,0	-4,2	4,0	0,2
REML-IML-obs*	-2,5	3,7	0,9	-3,3	4,6	1,1	-5,2	10,5	0,6

* Verfahren mit unvollständig abgeschlossenen Läufen (vgl. Tab. 4)

Die beobachteten Verzerrungen der weiterhin untersuchten Verfahren liegen teilweise zwar unter denen von REML-IML. Für eine praktische Nutzung dieser Vorgehensweisen gibt es jedoch erhebliche Hemmnisse, da mit einem teilweise hohen Anteil Simulationsläufe ohne Effektschätzung und Hypothesenprüfung vorliegen (Tabelle 4).

Tabelle 4: Anteil nicht abgeschlossener Berechnungen für die untersuchten Auswertungsmethoden, Varianzverhältnisse und Datenstrukturen

Auswertungs-methode	Varianzverhältnis und Datenstruktur					
	1:1:1		0.5:0.5:1		0.1:0.1:1	
	2×3×3	2×3×4	2×3×3	2×3×4	2×3×3	2×3×4
REML-N	14,3	2,9	17,6	3,6	27,4	5,5
REML-N-start	6,6	0,1	9,6	0,2	18,0	0,3
REML-N-score	62,2	54,5	54,9	43,8	46,2	25,3
REML-IML-obs	4,8	0,8	7,5	1,2	14,0	2,7

Die Gründe für nicht abgeschlossene Berechnungen sind unterschiedlich. Bei Nutzung von REML-N bestehen sie zu etwa 1/3 der Fälle in nicht erreichter Konvergenz (Ausgabe: „`WARNING: Did not converge`“), zu etwa 2/3 der Fälle erfolgt die Akzeptanz einer Lösung mit nicht positiv definiter Matrix V, deren weitere Verarbeitung dann zu Problemen führt (Ausgabe: „`WARNING: Stopped because of infinite likelihood`“). Das konnte bei Prüfung der bereitgestellten Schätzwerte der Varianzkomponenten in einem eigenen IML-Programm gezeigt werden. Die etwa gleiche Situation ist bei Nutzung von REML-N-score zu beobachten. Bei REML-N-start treten ausschließlich Probleme bei der Konvergenz auf.
Bei Nutzung von REML-IML-obs ist der Grund der unvollständigen Abarbeitung, dass die Sicherung einer positiven Definitheit der Matrix V nicht damit verbunden ist, dass diese Eigenschaft zwangsläufig auch für die Hesse-Matrix gilt.
Wegen der teilweise nur unvollständigen Abarbeitung ist eine vergleichende Einschätzung des empirischen statistischen Risikos 1. Art aller in Tabelle 1 angeführten Varianten nicht möglich. Gerade für die Datensätze, für die sich eine Varianzkomponentenschätzung schwierig gestaltet, werden keine Schätzwerte für die Effekte und deren Fehler bereit gestellt. Daher ist in diesen Fällen auch keine Hypothesenprüfung möglich. Entsprechend beschränken wir uns bei der weiteren Darstellung auf die Varianten, bei denen für alle Simulationsläufe vollständige Ergebnisse vorliegen (REML, REML-IML).

3.2 Standardfehler der untersuchten Kontraste

Weiterhin in engem Zusammenhang mit der Hypothesenprüfung sind die Standardfehler der festen Effekte. Entsprechend soll in diesem Abschnitt vergleichend untersucht werden, inwiefern die Vorgehensweisen REML und

REML-IML zu einer Beeinflussung der Standardfehler führen und welche Wirksamkeit die bei Kenward and Roger (1997) dargestellte Korrektur gemäß Kackar and Harville (1981) aufweist. Eine Einschätzung der Güte der geschätzten Standardfehler und der Korrektur kann bei Nutzung der berechneten Standardfehler aus den 10.000 Schätzwerten je Simulationslauf vorgenommen werden.

Die in Tabelle 5 dargestellten Ergebnisse verdeutlichen, dass die geschätzten Standardfehler in jedem Fall eine Unterschätzung gegenüber dem empirischen Standardfehler aufweisen. Dabei ist die Unterschätzung der auf Basis von REML-IML berechneten unkorrigierten Standardfehler nicht durchgehend kleiner. Das ist jedoch für die korrigierten Standardfehler der Fall. Hier sind die Abweichungen der Standardfehler gegenüber REML meist beträchtlich geringer.

Tabelle 5: Abweichung der mittleren geschätzten vom empirischen Standardfehler der Kontraste (%) (zwei fehlende Beobachtungen, 10000 Simulationsläufe)

		ohne Korrektur			mit Korrektur nach Kenward-Roger		
		1:1:1	0.5:0.5:1	0.1:0.1:1	1:1:1	0.5:0.5:1	0.1:0.1:1
Spaltanlage 2×3×3							
$\hat{\mu}_{1.} - \hat{\mu}_{2.}$	REML	-6,3	-5,0	-4,1	-5,6	-3,9	-3,1
	REML-IML	-1,7	-3,0	-6,0	-0,3	-0,6	-1,4
$\hat{\mu}_{.1} - \hat{\mu}_{.2}$	REML	-4,0	-5,2	-10,2	-1,9	-3,0	-8,8
	REML-IML	-6,7	-8,1	-10,3	-1,4	-1,6	-2,4
$\hat{\mu}_{11} - \hat{\mu}_{22}$	REML	-5,4	-5,0	-7,1	-4,1	-3,3	-6,1
	REML-IML	-3,8	-5,9	-9,3	-0,6	-1,0	-1,6
Spaltanlage 2×3×4							
$\hat{\mu}_{1.} - \hat{\mu}_{2.}$	REML	-3,9	-3,1	0,7	-3,8	-2,6	-1,4
	REML-IML	-0,4	-0,7	-2,6	-0,1	-0,2	-0,9
$\hat{\mu}_{.1} - \hat{\mu}_{.2}$	REML	-2,0	-2,8	-4,8	-0,9	-1,8	-3,9
	REML-IML	-2,0	-2,3	-3,6	-0,5	-0,5	-1,5
$\hat{\mu}_{11} - \hat{\mu}_{22}$	REML	-3,5	-3,3	-3,1	-2,9	-2,8	-2,3
	REML-IML	-1,0	-1,8	-3,9	-0,2	-0,4	-2,0

3.3 Ergebnisse der Signifikanzprüfung

Zur besseren Interpretation der Ergebnisse der Signifikanzprüfung sind die mittleren Freiheitsgrade für die Prüfgröße in Tabelle 6 dargestellt. Um die Höhe der geschätzten Freiheitsgrade bei Nutzung von REML und REML-IML bewerten zu können, sind die entsprechenden Freiheitsgrade für die balancierten Daten ebenfalls angeführt. Da der Datenumfang bei Unbalanciertheit kleiner als bei der entsprechenden Balanciertheit ist, sind entsprechend auch geringere Freiheitsgrade zu erwarten. Die Freiheitsgrade aus REML-IML ergeben gerade diese Tendenz, während das für die Freiheitsgrade aus REML nicht zutrifft. Erklärbar werden diese Ergebnisse jedoch wenn man beachtet, dass bei REML mit einem hohen Anteil eine Modellreduzierung vorgenommen wird (vgl. Abschnitt 3.1). Die damit verbundene Erhöhung der Restfreiheitsgrade wirkt sich besonders deutlich auf den Kontrast $\mu_{1.} - \mu_{2.}$ aus.

Tabelle 6: Mittlere Freiheitsgrade der Prüfgröße bei Unbalanciertheit (zwei fehlende Beobachtungen) bzw. Balanciertheit (Spaltanlage, 10000 Simulationsläufe)

		a×b×r = 2×3×3			a×b×r = 2×3×4		
		1:1:1	0.5:0.5:1	0.1:0.1:1	1:1:1	0.5:0.5:1	0.1:0.1:1
$\hat{\mu}_{1.} - \hat{\mu}_{2.}$	REML	4,19	5,04	6,75	5,40	6,71	9,79
	REML-IML	1,77	1,68	1,53	2,89	2,82	2,71
	bei Balanciertheit	2	2	2	3	3	3
$\hat{\mu}_{.1} - \hat{\mu}_{.2}$	REML	7,24	7,59	8,38	10,80	11,30	12,60
	REML-IML	6,52	6,69	6,95	10,49	10,68	11,03
	bei Balanciertheit	8	8	8	12	12	12
$\hat{\mu}_{11} - \hat{\mu}_{22}$	REML	6,48	7,39	8,76	9,25	10,80	13,30
	REML-IML	5,09	5,70	6,50	7,78	9,06	10,82
	bei Balanciertheit	5,61	6,51	7,73	8,01	9,58	11,80

Als komplexe Kriterien zum Vergleich der Verfahren zur Schätzung der Standardfehler und Freiheitsgrade ist das empirische statistische Risiko 1. Art zu betrachten. Für die in Tabelle 7 zusammengestellten Ergebnisse ist kennzeichnend, dass unabhängig von der betrachteten Methode der Varianzkomponentenschätzung bei Nutzung der Korrektur die Anpassung an das nominale Risiko 1. Art von $\alpha = 0.05$ deutlich besser ist. Durch Nutzung von

REML-IML ist eine bedeutsame Verbesserung der Einhaltung nominalen Risikos 1. Art gegenüber REML für die Kontraste $\mu_{.1} - \mu_{.2}$ und $\mu_{11} - \mu_{22}$ zu beobachten.

Tabelle 7: Empirisches statistisches Risiko 1. Art für eine Spaltanlage (zwei fehlende Beobachtungen, nominal α = 0.05, 10000 Simulationsläufe)

		ohne Korrektur der Fehlervarianz der festen Effekte			mit Korrektur der Fehlervarianz der festen Effekte		
		1:1:1	0.5:0.5:1	0.1:0.1:1	1:1:1	0.5:0.5:1	0.1:0.1:1
Spaltanlage 2×3×3							
$\hat{\mu}_{1.} - \hat{\mu}_{2.}$	REML	0,0707	0,0626	0,0470	0,0691	0,0600	0,0456
	REML-IML	0,0171	0,0140	0,0123	0,0138	0,0109	0,0094
$\hat{\mu}_{.1} - \hat{\mu}_{.2}$	REML	0,0562	0,0596	0,0665	0,0538	0,0567	0,0618
	REML-IML	0,0578	0,0618	0,0677	0,0496	0,0508	0,0542
$\hat{\mu}_{11} - \hat{\mu}_{22}$	REML	0,0625	0,0565	0,0582	0,0600	0,0537	0,0562
	REML-IML	0,0628	0,0611	0,0657	0,0533	0,0514	0,0503
Spaltanlage 2×3×4							
$\hat{\mu}_{1.} - \hat{\mu}_{2.}$	REML	0,0620	0,0618	0,0470	0,0608	0,0606	0,0460
	REML-IML	0,0437	0,0421	0,0393	0,0400	0,0365	0,0332
$\hat{\mu}_{.1} - \hat{\mu}_{.2}$	REML	0,0525	0,0541	0,0596	0,0506	0,0528	0,0586
	REML-IML	0,0551	0,0550	0,0595	0,0522	0,0523	0,0562
$\hat{\mu}_{11} - \hat{\mu}_{22}$	REML	0,0556	0,0523	0,0543	0,0538	0,051	0,0527
	REML-IML	0,0549	0,0524	0,0555	0,0513	0,0493	0,0510

Das gilt nicht für den Kontrast $\mu_{1.} - \mu_{2.}$. Hier ist auffällig, dass die Verzerrung des statistischen Risikos 1. Art bei allen Vorgehensweisen groß ist. Im Fall der Nutzung der Schätzwerte von REML kommt es durch die damit verbundene Modellreduzierung zu einer Erhöhung der Restfreiheitsgrade. Sowohl dieser Sachverhalt als auch die trotz Korrektur noch unterschätzten Standardfehler führen zu einer Überschreitung des nominalen statistischen Risikos. Im Fall der Nutzung von Schätzwerten der Vorgehensweise REML-IML sind zwar die Abweichungen der berechneten Freiheitsgrade von dem zu erwartenden Wert geringer, dennoch wird das nominale Risiko deutlich unterschritten. Das lässt sich darauf zurückführen, dass bei den für diesen Test geringen Freiheitsgraden bereits eine kleine Unterschätzung dramatische Auswirkungen hat und zu einer schlechten Einhaltung des nominalen Risikos 1. Art führt.

4 Schlussfolgerungen

Die dargestellten Ergebnisse zeigen, dass die Nutzung des Schätzprinzips REML bei unterschiedlichen Bedingungen an die Schätzer zu einer erwartet deutlichen Beeinflussung der Schätzwerte der Varianzkomponenten führt. Mit der alleinigen Bedingung an die Schätzwerte, eine positiv definite Matrix V zu sichern, kann die Verzerrung der Schätzwerte der Varianzkomponenten deutlich reduziert werden.
Ebenso werden aber auch die Ergebnisse der Signifikanzprüfung beeinflusst. Im Fall der untersuchten Datenstrukturen einer Spaltanlage und Verhältnissen der Varianzkomponenten kann bei Nutzung von REML-IML und der Korrektur gemäß Kackar and Harville (1981) für die Kontraste $\mu_{.1} - \mu_{.2}$ sowie $\mu_{11} - \mu_{22}$ das nominale Risiko 1. Art gut eingehalten werden. Vor allem bei der Datenstruktur 2×3×3 kann gegenüber REML eine deutliche Verbesserung beobachtet werden. Diese Verbesserung durch Nutzung des Schätzprinzips REML und der Restriktion V positiv definit sollte ausgenutzt werden.
Dabei ist jedoch zu beachten, dass die Verbesserung an die gleichzeitige Nutzung der Korrektur der Standardfehler der festen Effekte gebunden ist.
Wenn für den Kontrast $\mu_{1.} - \mu_{2.}$ auch bei Nutzung von REML-IML noch eine bedeutsame Verzerrung auftritt, so drückt sich damit der Sachverhalt aus, dass in diesem Fall die Prüfgröße nur mit einer hohen Unsicherheit geschätzt werden kann.

Literatur

[1] 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. J. Statist. Comput. Simul., 54, 363-378.

[2] 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, 477-486.

[3] Henderson, C.R. (1984). Application of linear models in animal breeding, Guelph: University of Guelph.

[4] Kackar, A.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and precision procedures for mixed linear models. Communications in Statistics, Series A 10, 1249-1261.

[5] Kenward, M.G., and Roger, J.H. (1997)., Small Sample Inference for fixed effects from restricted maximum likelihood. Biometrics, 53, 983-997.

[6] Piepho, H.-P., und Spilke, J. (1999). Anmerkungen zur Analyse balancierter gemischter Modelle mit der SAS-Prozedur MIXED. Zeitschrift für Agrarinformatik, 7, 39-46.

[7] Satterthwaite, F.E. (1941). Synthesis of variance. Psychometrika, 6, 309-316.

[8] SAS Institute Inc. (1999). SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute Inc.

[9] Searle, S.R. (1971). Linear models. New York: Wiley.

[10] Searle, S.R. Casella, G. and Mc Culloch, C.E. (1992). Variance component.s, New York: Wiley.

[11] Sorensen, D. and Gianola, G. (2002)., Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics, New York: Springer.

[12] Spilke, J. and Tuchscherer, A. (2001). Simulationsuntersuchungen zum Einfluss verschiedener Strategien der Varianzkomponentenschätzung und Hypothesenprüfung auf die statistischen Risiken in gemischten linearen Modellen mit ungleicher Klassenbesetzung. Zeitschrift für Agrarinformatik, 4, 66-75.

[13] Spilke, J., Hu, X. and Piepho, H.-P. (2003). A simulation study on tests of hypotheses for fixed effects in mixed models for blocked experiments with missing data. (unveröffentlichtes Manuskript).