

Joachim Spilke und Armin Tuchscherer

# Simulationsuntersuchungen zum Einfluss verschiedener Strategien der Varianzkomponentenschätzung und Hypothesenprüfung auf die statistischen Risiken in gemischten linearen Modellen mit ungleicher Klassenbesetzung

Die vorliegende Arbeit vergleicht für zwei praktisch bedeutsame unbalancierte Versuchspläne die Auswirkungen unterschiedlicher Strategien der Varianzkomponentenschätzung und Approximation der Freiheitsgrade auf die statistischen Risiken erster bzw. zweiter Art.

## 1 Einführung und Problemstellung

Die überwiegende Anzahl in landwirtschaftlichen Versuchen oder Erhebungen gewonnener Beobachtungsdaten können als Realisationen  $y$  einer normalverteilten Zufallsvariable angesehen werden, die durch ein gemischtes lineares Modell zu beschreiben sind. Wenngleich überwiegend Effektschätzung und Hypothesenprüfung über die festen Effekte Ziel der Untersuchungen sind, führt eine realistische Modellierung meist zwangsläufig zur zusätzlichen Berücksichtigung zufälliger Effekte im Sinne von Störgrößen.

Für die Modellierung der Realisationen einer solchen Zufallsvariable bedient man sich meist der folgenden Schreibweise:

$$\underline{y} = \mathbf{X}\underline{\beta} + \mathbf{Z}\underline{u} + \underline{e} \quad (1)$$

mit

$\underline{\beta}, \underline{u}$  = Vektor der unbekanntesten festen bzw. zufälligen Effekte

$\mathbf{X}, \mathbf{Z}$  = bekannte Versuchsplanmatrizen der festen bzw. zufälligen Effekte

$\underline{e}$  = Vektor der unbekanntesten zufälligen Resteffekte.

Weiter soll gelten:

$$\underline{u} \sim N(0, \mathbf{G}), \underline{e} \sim N(0, \mathbf{R})$$

$$E(\underline{y}) = \mathbf{X}\underline{\beta}, \text{Var}(\underline{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \mathbf{V}$$

(HENDERSON 1963, 1975, 1990 p. 1ff).

Falls  $\mathbf{G}$  und  $\mathbf{R}$  bekannt sind, lassen sich „Best Linear Unbiased Estimators“ (BLUE) (d.h., lineare erwartungstreue Schätzungen mit kleinster Varianz) für die festen Effekte in (1) aus

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\hat{\underline{\beta}} = \mathbf{X}'\mathbf{V}^{-1}\underline{y} \quad (2)$$

gewinnen. Ein ausführlicher Beweis findet sich u.a. bei DEMPFLER (1982).

Ein zentrales Problem besteht aber darin, dass die vorteilhaften Eigenschaften der Schätzungen fester Effekte bzw. Vorhersagen zufälliger Effekte, unverzerrte lineare Schätzungen bzw. Vorhersagen mit kleinster Varianz (BLUE bzw. BLUP) zu sein, an bekannte Matrizen  $\mathbf{G}$  und  $\mathbf{R}$  gebunden sind (HENDERSON, 1963). Bei Nutzung von Schätzungen für  $\mathbf{G}$  und  $\mathbf{R}$  bleibt die Erwartungstreue erhalten, die Eigenschaft beste Schätzung bzw. Vorhersage zu sein, geht jedoch verloren. Falls weiterhin die Schätzungen für  $\mathbf{G}$  und  $\mathbf{R}$  sowie die Effektschätzungen bei Nutzung des selben Beobachtungsvektors gewonnen werden, geht auch die Eigenschaft der Linearität verloren. Weiterhin damit im Zusammenhang stehend und von großer Bedeutung für die praktische Nutzung ist, dass

- sich nur bei bekannten  $\mathbf{G}$  und  $\mathbf{R}$  die Fehlervarianzen der festen Effekte exakt angeben lassen

- sowie in vielen Fällen exakte statistische Tests über die festen Effekte existieren (HENDERSON, 1984, p. 83).

Die Voraussetzung bekannter Matrizen  $\mathbf{G}$  und  $\mathbf{R}$  ist bei Anwendungen im Agrarbereich jedoch nie erfüllt.

Entsprechend ist die praktische Anwendbarkeit gemischter Modelle wesentlich daran gebunden wie es gelingt, Schätzwerte für die Varianzparameter mit ausreichend hoher Treff- und Wiederholungsgenauigkeit zur Verfügung zu stellen. Falls unbalancierte Datenstrukturen zu bearbeiten sind, verlieren die im balancierten Fall vorzuziehenden ANOVA-Schätzer ihre vorteilhaften Eigenschaften (Erwartungstreue, kleinste Varianz unter allen quadratischen Schätzfunktionen und bei Normalverteilung kleinste Varianz unter allen erwartungstreuen Schätzfunktionen) mit Ausnahme der Erwartungstreue (AHRENS, 1967; SEARLE, 1971; SPILKE et al., 1998). Dabei wird unter ANOVA das Schätzprinzip verstanden, durch Gleichsetzung mittlerer Abweichungsquadrate mit ihren Erwartungswerten Parameterschätzungen zu gewinnen (SEARLE et al., 1992 p. 34).

Unter den Bedingungen unbalancierter gemischter Modelle gewinnen daher seit längerem ausgearbeitete Konzepte [ML bzw. REML (HARTLEY and RAO; 1967; PATTERSON and THOMPSON, 1971); MINQUE bzw. MIVQUE (RAO, 1971; LAMOTTE, 1973)] praktische Bedeutung. Als Vorteil für REML geben SEARLE et al. (1992, p. 254) die unter Normalverteilung nahezu optimalen Eigenschaften der REML-Schätzer an (asymptotische Konsistenz, Erwartungstreue sowie asymptotische Normalverteilung der Schätzwerte und asymptotisch angebbare Fehlervarianzen der Schätzungen). Ein weiterer Grund für die Bevorzugung von REML gegenüber ML ist die Übereinstimmung von ANOVA und REML im balancierten Fall, falls keine negativen ANOVA-Schätzungen auftreten (ELSTON, 1998). Ebenso ist es ein wichtiger Vorteil, daß auch für unbalancierte gemischte Modelle der Fehler der geschätzten Varianzkomponenten angegeben werden kann. Dieser Sachverhalt wird in entsprechenden Approximationen der Freiheitsgrade der Testgröße ausgenutzt (FAI and CORNELIUS, 1996; KENWARD and ROGER, 1997).

Wenngleich sich SEARLE et al. (1992, p. 254) schließlich für die Bevorzugung von ML und REML gegenüber dem ANOVA-Prinzip aussprechen, fehlen empirische Untersuchungen zu den konkreten Konsequenzen dieser Entscheidung. Daher ist gerade aus Sicht der Anwendung eine Überprüfung der theoretischen Annahmen, insbesondere zwischen ANOVA- und den sich zunehmend durchsetzenden REML-Schätzern für typische unbalancierte Datenstrukturen von besonderer Bedeutung. Entsprechend besteht die Zielstellung dieser Arbeit in erster Linie in einem Vergleich der Auswirkungen dieser Strategien der Varianzkomponentenschätzung auf die Schätzgenauigkeit fester Effekte. Eine solche Untersuchung ist auch deshalb notwendig, da in statistischer Software meist mehrere Schätzverfahren verfügbar sind und für den Nutzer die Konsequenzen der verschiedenen Vorgehensweisen oft nicht überschaubar sind.

Da die Effektschätzung in vielen Fällen mit einer Hypothesenprüfung verbunden ist, erfolgt weiterführend die Bewertung anhand der realisierten statistischen Risiken für Fehler erster und zweiter Art. Es bietet sich an, die Untersuchungen mit einem Vergleich verschiedener Vorgehensweisen zur Approximation der Freiheitsgrade zu verbinden.

## 2 Material und Methode

### 2.1 Untersuchte Versuchspläne

In die Untersuchungen wurden mit der Spalt- und Gitteranlage zwei typische Versuchsanlagen einbezogen. Während im Fall der Gitteranlage ein Beispiel für eine „geplante Unbalanciertheit“ (planned unbalancedness) vorliegt, ist die Spaltanlage vom Grundsatz her eine balancierte Versuchsanlage. Hier wird aber der praktisch wichtige Fall betrachtet, dass durch fehlende Beobachtungen (missing values) aus einer balancierten Versuchsanlage eine ungeplante Unbalanciertheit entsteht (SEARLE et al., 1992 p. 4ff), wobei die „missing completely at random“ Annahme getroffen wird.

Die Beobachtungswerte  $y$  einer Spaltanlage werden als Realisationen einer Zufallsvariable  $\underline{y}$  mit dem folgenden Modell angesehen:

$$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \underline{B}l_k + (\alpha\underline{B}l)_{ik} + e_{ijk} \quad (3)$$

$$i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, r;$$

Dabei bedeuten:

- $\mu$  = allgemeines Mittel,
- $\alpha_i$  = Effekt der  $i$ -ten Stufe des Faktors A (fest)
- $\beta_j$  = Effekt der  $j$ -ten Stufe des Faktors B (fest)
- $(\alpha\beta)_{ij}$  = Wechselwirkungseffekt der  $i$ -ten Stufe und der  $j$ -ten Stufe des Faktors B (fest)
- $\underline{B}l_k$  = Effekt des  $k$ -ten Blocks (zufällig,  $N \sim (0, \sigma_{Bl}^2)$ )
- $(\alpha\underline{B}l)_{ik}$  = Wechselwirkungseffekt der  $i$ -ten Stufe des Faktors A und der  $k$ -ten Stufe des Faktors Block (zufällig,  $N \sim (0, \sigma_{\alpha\underline{B}l}^2)$ )
- $e_{ijk}$  = Resteffekt (zufällig,  $N \sim (0, \sigma_e^2)$ ).

Im vorliegenden Fall wird der für das Versuchswesen häufige Fall betrachtet, dass der Blockeffekt als ein zufälliger Effekt angesehen wird.

Die Simulation basiert auf einer Datenstruktur von  $a=2$ ,  $b=3$  und  $r=4$ . Damit wurde eine unter praktischen Verhältnissen sehr häufige Datenstruktur verwendet. Die unbalancierte Datenstruktur wurde durch Eliminierung von je zwei zufällig ausgewählten Beobachtungen generiert. Damit standen für jede Einzelauswertung 22 Beobachtungswerte zur Verfügung.

Die Beobachtungswerte  $y$  einer Gitteranlage werden als Realisationen einer Zufallsvariable  $\underline{y}$  mit dem folgenden Modell angesehen:

$$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + \underline{B}l_{k(j)} + e_{ijk} \quad (4)$$

$$i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, r ;$$

Dabei bedeuten:

- $\mu$  = allgemeines Mittel
- $\alpha_i$  = Effekt der  $i$ -ten Stufe des Faktors A (fest)
- $\beta_j$  = Effekt der  $j$ -ten Wiederholung (fest)
- $\underline{B}l_k$  = Effekt des  $k$ -ten Blocks innerhalb der  $j$ -ten Wiederholung (zufällig,  $N \sim (0, \sigma_{Bl}^2)$ )
- $e_{ijk}$  = Resteffekt (zufällig,  $N \sim (0, \sigma_e^2)$ ).

Entsprechend der häufig geübten Praxis, bei der Anlage von Gittern die Blöcke innerhalb der Wiederholungen zu randomisieren, wird der Faktor Block als zufällig angesehen.

In 3 Wiederholungen zu je 3 unvollständigen Blöcken mit je 3 Beobachtungen wurden 9 Stufen des Faktors  $i$  und damit je Simulationslauf 27 Beobachtungen erzeugt. Die geplante Unbalanciertheit resultiert also aus der Prüfung von je nur 3 Stufen des Faktors  $i$  je Block. Diese Datenstruktur führt dazu, dass bestimmte Stufen keinmal (z.B.: 1 mit 8, 5 mit 7) oder einmal (z.B.: 1 mit 2, 3 mit 4) gemeinsam in einem Block auftreten. Damit werden die Differenzen zwischen den Effekten der Stufen des Faktors  $i$  mit unterschiedlicher Präzision geschätzt. Der verwendete Versuchsplan ist in Anlage 1 angegeben.

## 2.2 Datensimulation und benutzte Parameter für die Simulation

Die Bearbeitung der in Abschnitt 1 skizzierten Aufgabenstellung erfolgte bei Nutzung der stochastischen Simulation unter Verwendung des Programms SAS (Version 8.1) sowohl für die Datensimulation als auch Datenauswertung. Das genutzte Programmsystem verfügt über geprüfte Zufallszahlengeneratoren (CLARK and WOODWARD, 1992).

Die unterstellten Varianzparameterwerte sind in Tabelle 1 zusammengestellt.

Tab. 1: Genutzte Varianten der Varianzparameter

Varianzkomponente	Parameterwerte					
	Spaltanlage			Gitteranlage		
$\sigma_e^2$	1	1	1	1	1	1
$\sigma_{BI}^2 / \sigma_e^2$	1	0.5	0.1	1	0.5	0.1
$\sigma_{A*BI}^2 / \sigma_e^2$	1	0.5	0.1			

Die für die Simulation verwendeten Verhältnisse der Varianzkomponenten sind von praktischen Erfahrungen abgeleitet und decken einen weiten Bereich ab.

Bei Gültigkeit der Hypothese H1 wurden die in Tabelle 2 zusammengestellten Differenzen der Mittelwerte zugrundegelegt. Dabei wird aus Gründen der durchgängigen Betrachtung der Mittelwertdifferenzen unter der Null- und Alternativhypothese der Vergleich für die Spaltanlage ebenfalls unter H1 betrachtet, obwohl das aus inhaltlicher Sicht bei existierenden Wechselwirkungen zwischen den festen Effekten (wie im vorliegenden Fall simuliert, vgl. Tabelle 2) nicht immer sinnvoll ist. Die Auswertung jeder Simulationsvariante basiert auf 10000 wiederholten Simulationsläufen.

Tabelle 2: Genutzte Varianten der Mittelwertdifferenzen bei Gültigkeit von H1

Spaltanlage		Gitteranlage	
Parameterkontrast	Wert	Parameterkontrast	Wert
$\mu_1 - \mu_2 = \alpha_1 - \alpha_2 + \frac{1}{b} \left( \sum_{j=1}^b (\alpha\beta)_{1j} - \sum_{j=1}^b (\alpha\beta)_{2j} \right)$	1	$\mu_1 - \mu_2 = \alpha_1 - \alpha_2$	0.5
$\mu_1 - \mu_2 = \beta_1 - \beta_2 + \frac{1}{a} \left( \sum_{i=1}^a (\alpha\beta)_{i1} - \sum_{i=1}^a (\alpha\beta)_{i2} \right)$	1	$\mu_1 - \mu_8 = \alpha_1 - \alpha_8$	0.5
$\mu_{11} - \mu_{12} = \beta_1 - \beta_2 + (\alpha\beta)_{11} - (\alpha\beta)_{12}$	1.5	$\mu_3 - \mu_4 = \alpha_3 - \alpha_4$	1
$\mu_{11} - \mu_{21} = \alpha_1 - \alpha_2 + (\alpha\beta)_{11} - (\alpha\beta)_{21}$	2.0	$\mu_5 - \mu_7 = \alpha_5 - \alpha_7$	1
$\mu_{11} - \mu_{22} = \alpha_1 - \alpha_2 + \beta_1 - \beta_2 + (\alpha\beta)_{11} - (\alpha\beta)_{22}$	2.5		

## 2.3 Schätzung der Varianzkomponenten

Entsprechend der eingangs beschriebenen Aufgabenstellung wird eine Schätzung der Varianzkomponenten nach den Schätzprinzipien ANOVA bzw. REML vorgenommen.

Da vorliegende Untersuchungen sich auf gemischte Modelle beziehen, wird die Varianzkomponentenschätzung nach dem ANOVA-Prinzip nach der als Type 3 bekannten Vorge-

weise vorgenommen (SEARLE, 1987 p. 391ff; DUFNER et. al, 1992 p. 299). Dieses Schätzprinzip baut allein auf der Zerlegung und Interpretation der Summe der Abweichungsquadrate auf und kann daher auch zu negativen Schätzwerten der Varianzkomponenten führen. Durch das Verfahren wird die positive Definitheit der Kovarianz-Matrix der Beobachtungen (Matrix V in Gleichung 1) nicht gesichert. Weiterhin erfolgte eine Maximum-Likelihood-Schätzung der Varianzkomponenten. Dabei wurden sogenannte eingeschränkte Maximum-Likelihood-Schätzer REML (Restricted Maximum Likelihood) entsprechend der Ableitung von PATTERSON and THOMPSON (1971) verwendet. Es wurden nur Schätzwerte der Varianzkomponenten im zulässigen Bereich positiver Werte akzeptiert. Damit ist diese Methode nur asymptotisch erwartungstreu. Insbesondere bei geringen Stichprobenumfängen und im Verhältnis zur Restvarianz kleinen Varianzkomponenten sind verzerrte Schätzungen zu erwarten.

Die angeführten Verfahren sind in der Software SAS Version 8.1(proc mixed) implementiert.

## 2.4 Approximation der Freiheitsgrade der Prüfgröße

Verbunden mit dem Problem praktisch stets unbekannter Kovarianz-Matrizen der geschätzten festen Effekte und der Notwendigkeit der Bereitstellung entsprechender Schätzungen ist weiterhin auch das der Bestimmung der Freiheitsgrade der Prüfgrößen im Signifikanztest. Eine besondere Problematik folgt daraus, dass in gemischten Modellen die Varianz der Prüfgrößen im Allgemeinen nicht durch eine Varianzkomponente bzw. nur ein mittleres Abweichungsquadrat ausgedrückt werden kann. Das gilt insbesondere auch für unbalancierte Datenstrukturen. In diesem Fall stehen mehrere Konzepte zur Bestimmung der Freiheitsgrade zur Verfügung. Eine weite Verbreitung hat das auf SATTERTHWAITE (1941) zurückgehende Konzept gefunden. Die Basisidee dieses Konzeptes besteht in der Ausnutzung der Chi<sup>2</sup>-Verteilung des Quotienten einer Varianzschätzung und ihres Erwartungswertes. Bei Nutzung von REML-Schätzern gilt diese Verteilungsvoraussetzung nur approximativ. Aufbauend auf diesem Ansatz wird in der Software SAS ein Vorschlag von GIESBRECHT and BURNS (1985) umgesetzt. Der Vorschlag basiert auf der approximativen Bestimmung der Varianz der Fehlervarianz der interessierenden Linearkombination fester Effekte und des Erwartungswertes der Fehlervarianz.

Eine weitere in unsere Untersuchung einbezogene Approximation der Freiheitsgrade geht auf KENWARD and ROGER (1997) zurück. Dieser Vorschlag beruht grundsätzlich auch auf der von Satterthwaite vorgeschlagenen Näherung auf Basis Chi<sup>2</sup>-verteilter Zufallsvariable. Die Erweiterung gegenüber GIESBRECHT and BURNS (1985) besteht jedoch darin, die von KACKAR and HARVILLE (1984) beschriebene Unterschätzung der Kovarianz-Matrix der geschätzten festen Effekte durch eine Korrektur zu berücksichtigen, was besonders bei kleinen Stichproben bedeutsam sein kann. Die Nutzung der Vorgehensweise von KENWARD and ROGER (1997) führt wegen der Korrektur bereits bei Tests mit Rang = 1 zu unterschiedlichen Ergebnissen gegenüber dem Satterthwaite-Verfahren. Die untersuchten Approximationen der Freiheitsgrade sind ebenfalls in der Software SAS ab Version 8.0 implementiert.

### 3 Ergebnisse und Diskussion

Die Güte der Varianzkomponentenschätzung wird anhand der Treffgenauigkeit und des mittleren quadratischen Fehlers (MSE) bewertet. Während die Treffgenauigkeit erwartungsgemäß für das verwendete ANOVA-Verfahren sehr gut eingehalten wird, trifft das für REML wie erwartet nicht zu (Tabelle 3). Dabei ist jedoch eine deutliche Abhängigkeit vom Verhältnis der Varianzparameter zur Restvarianz zu beobachten. Besonders bei einem weiten Verhältnis (0.1:0.1:1 bzw. 0.1:1) führt REML zu einer ausgeprägten Überschätzung der Varianzparameter  $\sigma_{Bl}^2$  und  $\sigma_{A*Bl}^2$ . Damit kommt der hohe Anteil von Schätzwerten gleich Null für diese Varianzkomponenten zum Ausdruck. Mit dem Auftreten von Schätzwerten gleich Null kommt es faktisch zu einer Modellreduktion um diesen Effekt und einer Verschiebung dieses für die jeweilige Stichprobe kleinen Varianzanteils (ansonsten würde ein Schätzwert größer Null auftreten) in die Restvarianz bei gleichzeitiger Erhöhung der Restfreiheitsgrade. Das führt zu der beobachteten Unterschätzung der Restvarianz.

Tab. 3: Mittlere Schätzwerte der Varianzkomponenten für REML und TYPE 3 sowie das Verhältnis deren MSE (TYPE 3/REML)

Spaltanlage	Varianzkomponente								
	$\sigma_{Bl}^2$	$\sigma_{A*Bl}^2$	$\sigma_e^2$	$\sigma_{Bl}^2$	$\sigma_{A*Bl}^2$	$\sigma_e^2$	$\sigma_{Bl}^2$	$\sigma_{A*Bl}^2$	$\sigma_e^2$
Parameterwerte	1	1	1	0.5	0.5	1	0.1	0.1	1
Schätzwerte (REML)	1.1 1	0.92	0.9 9	0.5 7	0.47	0.9 7	0.1 4	0.14	0.9 2
Schätzwerte (TYPE 3)	0.9 8	1.03	1.0 1	0.4 9	0.52	1.0 1	0.0 9	0.10	1.0 1
MSE (TYPE3/REML)	1.2 7	1.52	1.0 9	1.3 6	1.80	1.1 7	1.9 7	3.27	1.3 7

Gitteranlage	Varianzkomponente								
	$\sigma_{Bl}^2$	$\sigma_e^2$		$\sigma_{Bl}^2$	$\sigma_e^2$		$\sigma_{Bl}^2$	$\sigma_e^2$	
Parameterwerte	1	1		0.5	1		0.1	1	
Schätzwerte (REML)	0.9 9	1.03		0.5 5	0.97		0.2 1	0.92	
Schätzwerte (TYPE 3)	1.0 1	1.00		0.5 0	1.00		0.1 0	1.00	
MSE (TYPE3/REML)	1.0 6	1.08		1.2 0	1.17		1.6 8	1.37	

Es ist zu betonen, dass die beobachtete Verzerrung auch vom verwendeten Algorithmus zur Gewinnung der REML-Schätzer abhängt. Die Maximierung der Likelihood-Funktion ist in der Software SAS unter der Nebenbedingung implementiert, nur Schätzwerte  $\geq 0$  zuzulassen. Diese beispielsweise bei der Schätzung genetischer Parameter in der Pflanzen- und Tierzucht sinnvolle Restriktion erscheint für den Fall, dass aus gemischten Modellen Schätzungen fester Effekte zu gewinnen sind, nicht zwingend erforderlich. Ausreichend wäre, die positive Definitheit der Matrix V (Gleichung 1) zu sichern, was nicht zwingend Schätzwerte  $\geq 0$  erfordert. Eine solche Vorgehensweise könnte zu einer geringeren Verzerrung der geschätzten Varianzkomponenten führen. Ein wesentlicher Unterschied zwischen den verwendeten Schätzverfahren ergibt sich auch für die MSE der geschätzten

Mittelwertdifferenzen (Tabelle 3). Die TYPE 3-Schätzungen weisen durchgehend höhere MSE auf. Die Erhöhung der MSE weisen eine Abhängigkeit vom Verhältnis der Varianzparameter und der Versuchsanlage auf. Sie sind für die Spaltanlage und weitem Verhältnis der Varianzparameter größer. Beachtet man, dass sich der MSE aus dem Quadrat des Bias und der Fehlervarianz der Schätzung zusammensetzt, bedeuten unsere Simulationsergebnisse, dass die REML-Schätzungen einen deutlich geringeren Schätzfehler aufweisen, da sie trotz ausgeprägter Verzerrung zu einem geringeren MSE führen. Geringere MSE werden auch für die Effektschätzung beobachtet, falls die Varianzkomponenten mit REML gewonnen wurden. Die in Tabelle 4 angeführten Verhältnisse der MSE gelten unabhängig von den bei der Simulation vorgegebenen Effektdifferenzen sowohl bei Gültigkeit der Nullhypothese als auch bei Gültigkeit der Alternativhypothese (Effekte entsprechend Tabelle 2).

Tab. 4: Verhältnis der MSE der Effektschätzungen (TYPE 3/REML)

	Verhältnis der Varianzkomponenten		
	1:1:1	0.5:0.5:1	0.1:0.1:1
<b>Spaltanlage</b>			
$\mu_1 - \mu_2$	1.00	1.06	1.40
$\mu_{.1} - \mu_{.2}$	1.01	1.06	1.42
$\mu_{11} - \mu_{12}$	1.01	1.11	2.12
$\mu_{11} - \mu_{21}$	1.01	1.07	2.05
$\mu_{11} - \mu_{22}$	1.01	1.07	1.20
	1:1	0.5:1	0.1:1
<b>Gitteranlage</b>			
$\mu_1 - \mu_2$	1.02	1.06	1.15
$\mu_1 - \mu_8$	1.03	1.08	1.21
$\mu_3 - \mu_4$	1.03	1.06	1.15
$\mu_5 - \mu_7$	1.04	1.09	1.23

Die Überlegenheit bezüglich des MSE der Effektschätzungen gilt für beide Versuchsanlagen, ist jedoch für die Spaltanlage bei weitem Verhältnis der Varianzkomponenten in Abhängigkeit vom betrachteten Kontrast stärker ausgeprägt. Die MSE sind ausschließlich durch die Fehlervarianz bedingt, da alle Effekte erwartungsgemäß durchgehend unverzerrt geschätzt wurden. Begründet ist das in der Unabhängigkeit der Treffgenauigkeit von den Varianzkomponenten (MIELENZ und SPILKE, 1984). Da einer Effektschätzung meist eine Hypothesenprüfung folgt, sind weiterhin die realisierten statistischen Fehler von Interesse. Die Ergebnisse sind in den Abbildungen 1-4 zusammengestellt. Dabei basiert die Auswertung wiederum auf beiden Schätzmethoden der Varianzkomponenten, aber auch auf den unterschiedlichen Vorgehensweisen zur approximativen Ermittlung der Freiheitsgrade. Wie eingangs erwähnt, sind unter den Bedingungen unbalancierter Datenstrukturen in gemischten Modellen die statistischen Tests im Allgemeinen nur approximativ gültig. Das findet auch in den vorliegenden Ergebnissen seinen Niederschlag, bei denen Verzerrungen des statistischen Fehlers erster Art zu beobachten sind.

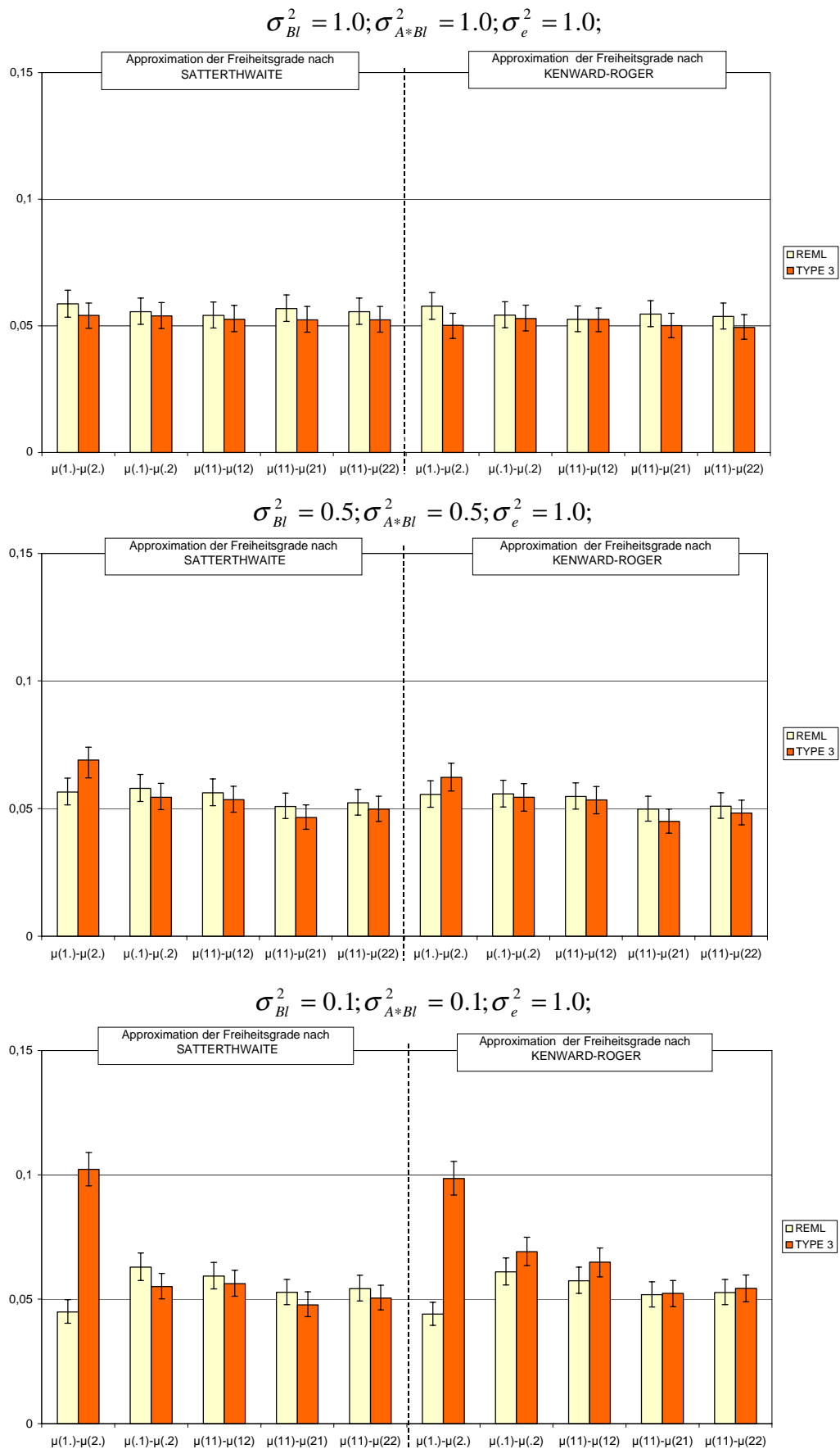


Abb. 1 Realisierte Fehler erster Art und deren Konfidenzintervall ( $P = 0.95$ ) bei Gültigkeit der Nullhypothese für die Spaltanlage (vorgegebenes Risiko erster Art  $\alpha = 0.05$ )

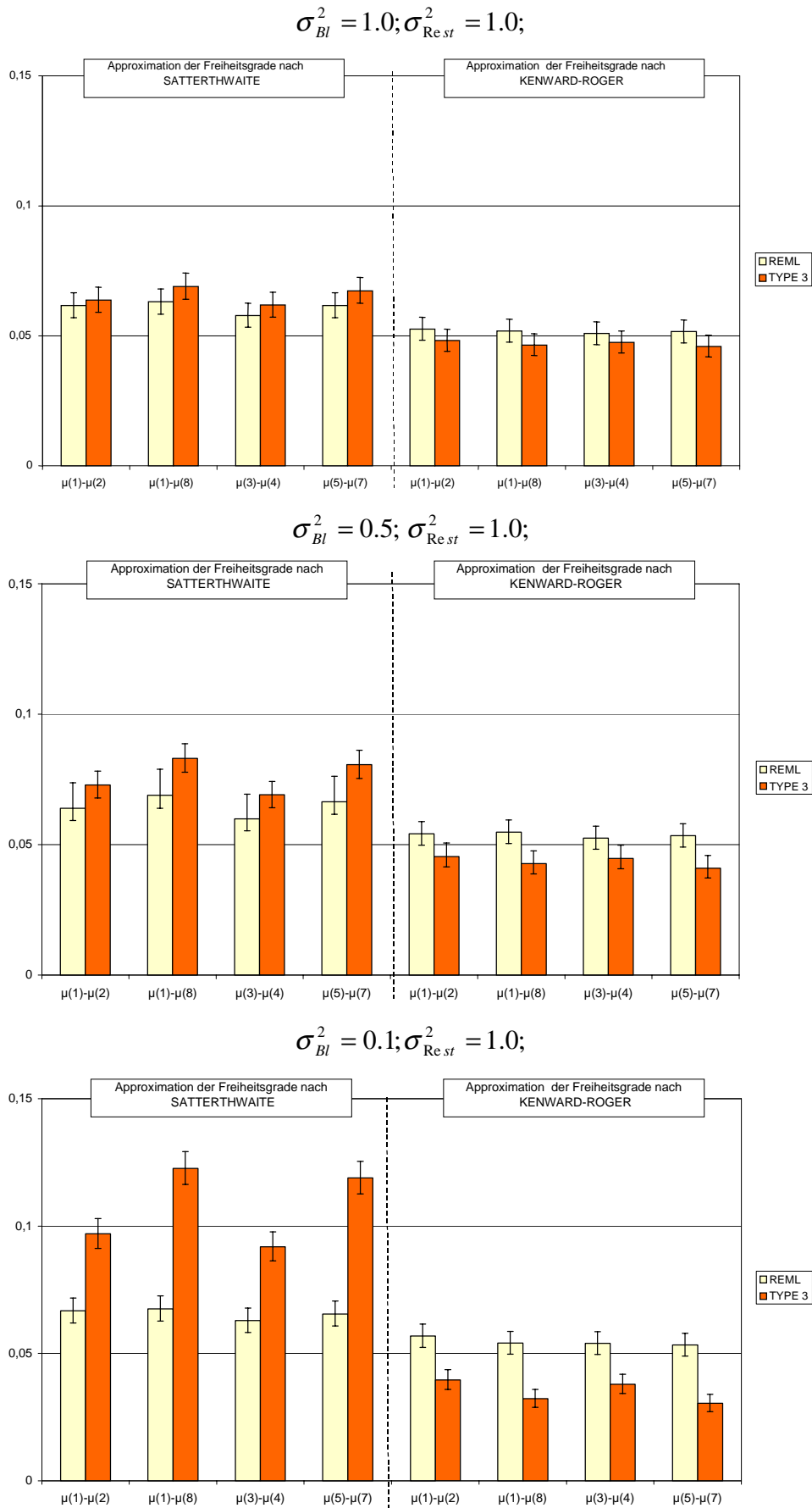


Abb. 2 Realisierte Fehler erster Art und deren Konfidenzintervall ( $P = 0.95$ ) bei Gültigkeit der Nullhypothese für die Gitteranlage (vorgegebenes Risiko erster Art  $\alpha = 0.05$ )

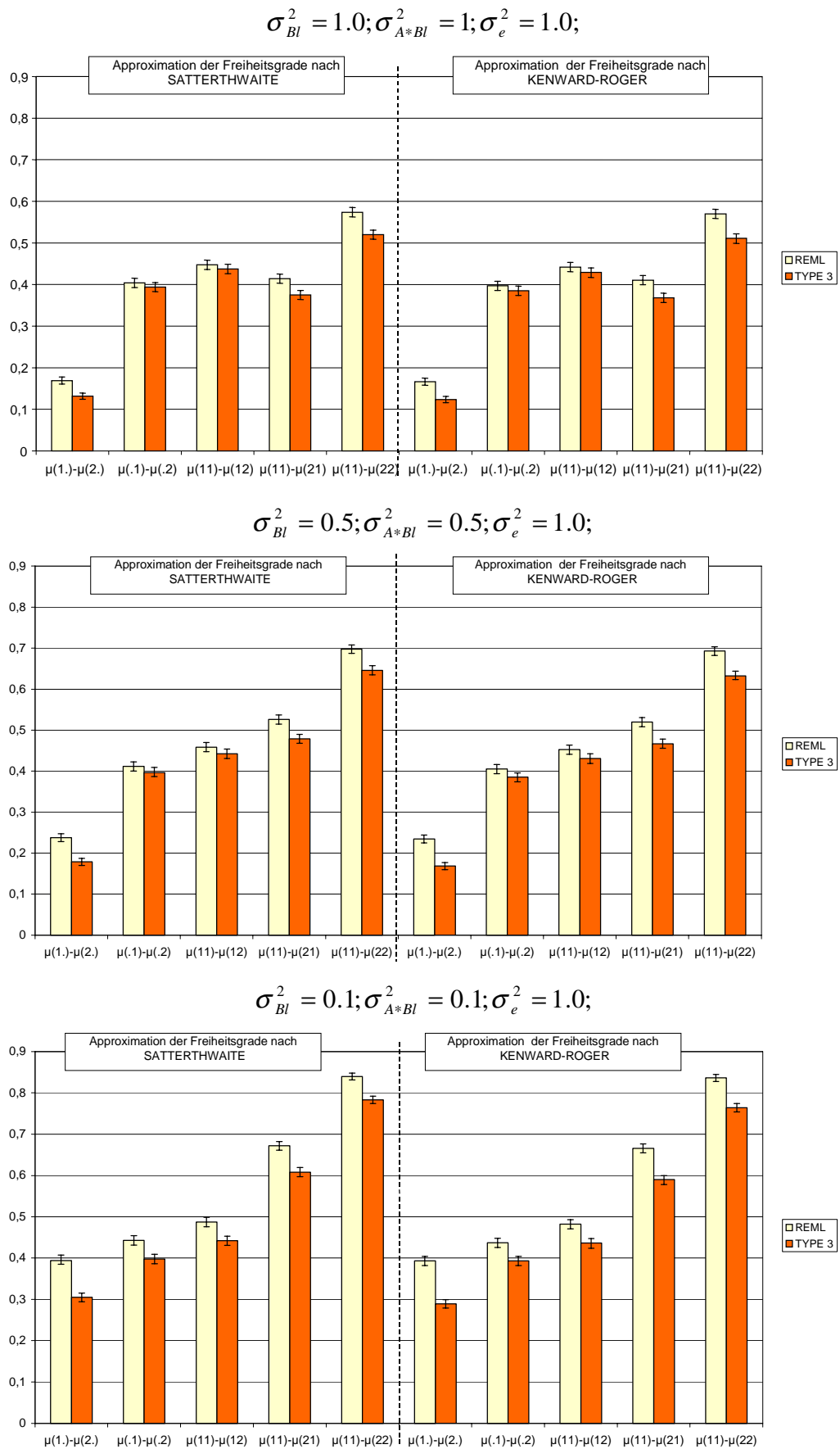


Abb. 3: Realisierte Power und deren Konfidenzintervall ( $P = 0.95$ ) bei Gültigkeit der Alternativhypothese für die Spaltanlage

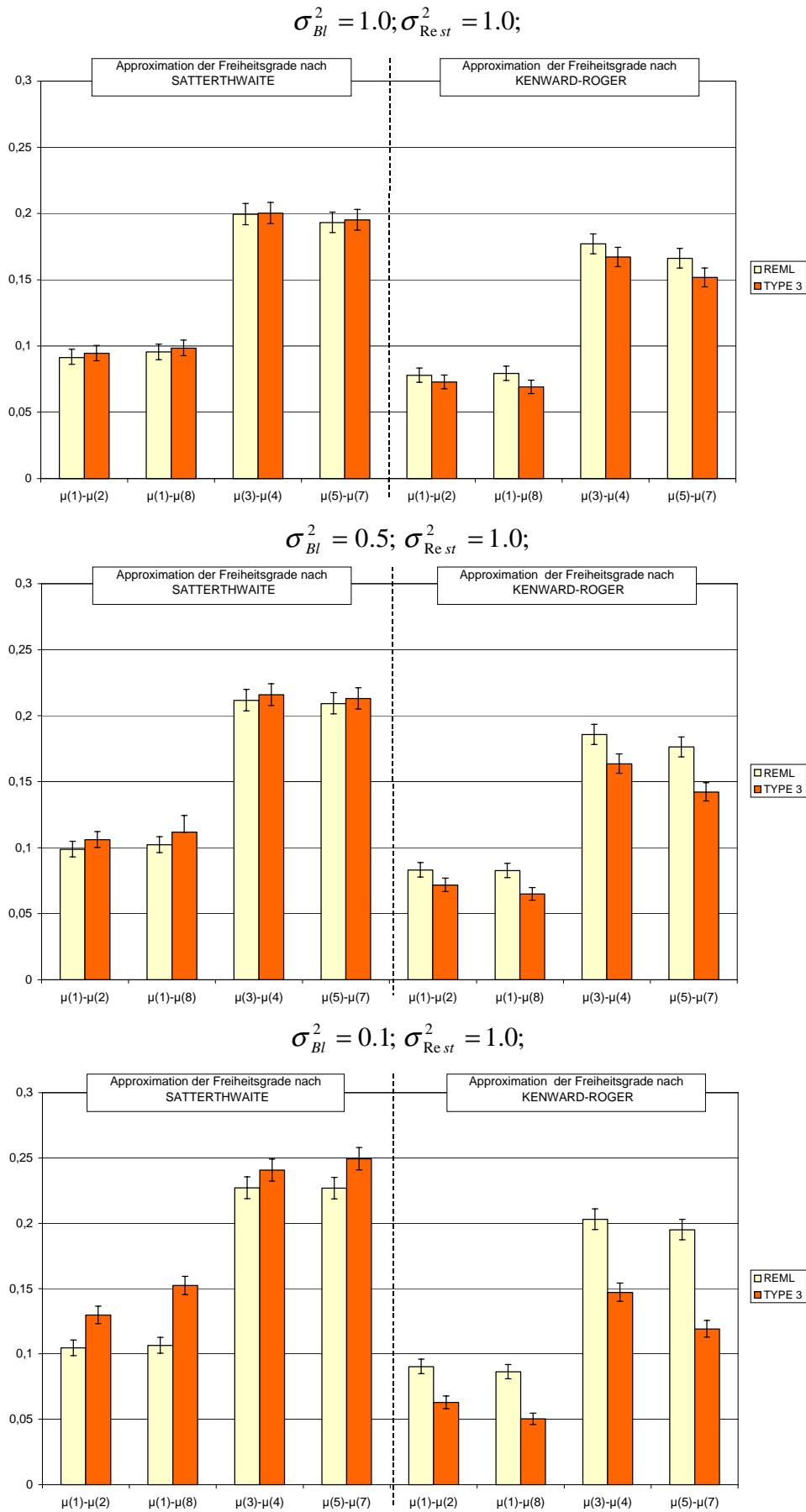


Abb. 4: Realisierte Power und deren Konfidenzintervall ( $P = 0.95$ ) bei Gültigkeit der Alternativhypothese für die Gitteranlage

Zunächst soll auf die bei Nutzung von REML-Schätzern gewonnenen Ergebnisse eingegangen werden. Unter den Bedingungen, dass die simulierten Effekte der Nullhypothese entsprechen, d.h. keine Parameterdifferenzen vorliegen, ist für die Approximation nach Satterthwaite stets ein höherer Fehler erster Art zu beobachten. Die in jedem Fall zu beobachtende Reduzierung bei Nutzung der Vorgehensweise nach Kenward-Roger wird verständlich, wenn man beachtet, dass diese Vorgehensweise eine Korrektur der Unterschätzung der geschätzten Kovarianz-Matrix der festen Effekte beinhaltet.

Während bei einem engen Varianzverhältnis das vorgegebene statistische Risiko für Spalt- und Gitteranlage gut eingehalten wird, treten bei einem weiten Varianzverhältnis leichte Überschätzungen auch bei Nutzung der Approximation nach Kenward-Roger im Fall der Spaltanlage auf. Ebenfalls anhand von Simulationsuntersuchungen wurde gezeigt, dass für den angewendeten REML-Algorithmus die Korrektur der Unterschätzung der Fehlervarianz nie vollständig ist und bei einem weiten Verhältnis der Varianzkomponenten noch eine bedeutsame Unterschätzung bestehen bleibt (SPILKE et. al., 2001).

Das Schätzprinzip ANOVA sollte für die vorliegenden Modelle grundsätzlich nicht angewendet werden. Während für ein enges Verhältnis der Varianzkomponenten bei Nutzung der TYPE 3-Schätzer das statistische Risiko noch ebenfalls gut eingehalten wird, wird die Problematik der Anwendung dieser Methode bei einem weiten Verhältnis der Varianzkomponenten deutlich. So ist die Verzerrung des statistischen Risikos bei Verwendung von TYPE 3 extrem (z.B. 0.09 für den Vergleich  $(\mu_1 - \mu_2)$  bei  $\sigma_{Bl}^2 = 0.1; \sigma_{Res}^2 = 1.0$ ). Eine genauere Analyse dieses Ergebnisses zeigt, dass etwa die Hälfte der Ablehnungen der Nullhypothese auf Varianzkomponenten beruhen, die zu einer nicht positiv definiten Matrix V führen. Unter diesen Bedingungen wird ein Schätzfehler gleich Null ausgegeben. Die dafür von der Software formal berechnete Überschreitungswahrscheinlichkeit führt stets zu Ablehnung von  $H_0$  und damit zu der beobachteten hohen Ablehnungsrate.

Die für REML-Schätzer entwickelte Approximation der Freiheitsgrade nach Kenward-Roger wird in der genutzten Software für TYPE 3 formal durch Nutzung entsprechender Schätzwerte umgesetzt (WOLFINGER, 2000). Die in der Kenward-Roger-Approximation enthaltene Korrektur führt für Fälle mit nicht positiv definiten Matrix V zu teilweise völlig überhöhten Schätzfehlern und damit zu einer geringeren Ablehnung der Nullhypothese gegenüber der Satterthwaite-Approximation.

Die beschriebene Problematik tritt in jedem Fall bei der Nutzung der TYPE 3-Schätzer auf, wird jedoch erst bei einem weiten Verhältnis der Varianzkomponenten und dem damit verbundenen hohen Anteil negativer Schätzwerte deutlich. Gleichermaßen gilt das auch für die Gitteranlage und erklärt die hohe Überschreitung des nominalen statistischen Risikos erster Art bei Nutzung der Vorgehensweise von Satterthwaite und die ausgeprägte Unterschreitung für Kenward-Roger bei  $\sigma_{Bl}^2 = 0.1; \sigma_{Res}^2 = 1.0$ .

Unter den Bedingungen der Gültigkeit der Alternativhypothese zeigen sich für die Power der untersuchten Tests die gleichen Verhältnisse (Abbildung 3 bis 4). Dabei sind die

Ergebnisse für TYPE 3 ebenfalls dargestellt, obwohl beziehungsweise auf die obigen Aussagen nur die bei Nutzung von REML die nominale Fehlerwahrscheinlichkeit erster Art eingehalten wird. Die Trennschärfe der Vorgehensweise nach Satterthwaite ist durchgehend geringer, was bei Beachtung der beschriebenen Korrektur (Vergrößerung) der Fehlervarianzen der festen Effekte nach Kenward-Roger verständlich ist.

Für die untersuchten Datenstrukturen, Versuchspläne und Verhältnisse der Varianzkomponenten kann daher bei Nutzung der Kombination REML und Kenward-Roger auch unter den Bedingungen ungleicher Klassenbesetzung der Fehler erster Art gut eingehalten werden. Die beobachteten Verzerrungen sind gering und stets kleiner als bei Nutzung der Approximation nach Satterthwaite. Die Verwendung von TYPE 3 zur Varianzkomponentenschätzung führt zu deutlich höheren MSE der Effektschätzung und zu einer teilweise ausgeprägten Verzerrung des vorgegebenen statistischen Risikos erster Art.

Für praktische Anwendungen sollte daher als Methode der Varianzkomponentenschätzung REML und für die Approximation der Freiheitsgrade die von KENWARD and ROGER (1997) vorgeschlagene Vorgehensweise angewendet werden. Es ist zu erwarten, dass diese Aussage auch für hier nicht untersuchte Datenstrukturen und Versuchspläne mit geringem Stichprobenumfang gilt.

Anlage 1: Versuchsplan der simulierten Gitteranlage

Wiederholung	Block innerhalb Wiederholung	Stufen des Faktors A
1	1	1, 2, 3
	2	4, 5, 6
	3	7, 8, 9
2	1	8, 2, 5
	2	4, 1, 7
	3	3, 6, 9
3	1	2, 7, 6
	2	1, 5, 9
	3	4, 8, 3

4. Literatur

CLARK, M. R., Woodward, D. E. (1992): Generating random numbers with BASE SAS Software. Observations 1 (3) Third Quarter SAS Institute Inc.  
 DEMPFFLE, L. (1982): Zuchtwertschätzung beim Rind mit einer ausführlichen Darstellung der BLUP-Methode. Fortschritte der Tierzucht und Züchtungsbiologie.H3; Verlag Paul Parey Hamburg und Berlin.  
 DUFNER, J., Jensen, U., Schumacher, E. (1992): Statistik mit SAS. B. G. Teubner Stuttgart.  
 ELSTON, D. A. (1998): Estimation of denominator degrees of freedom of F-distributions for accessing Wald statistics for fixed-effect factors in unbalanced mixed models. Biometrics 54: 1085-1096.  
 FAI, A. H. T., 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.  
 GIESBRECHT, F. G., 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.

- HARTLEY, H. O., Rao, C. R. (1967): Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54: 93 - 108.
- HENDERSON, C. R. JR, Henderson, C. R. (1979): Analysis of covariance in mixed models with unequal subclass numbers. *Communications in Statistics, Series A* 8: 751-787.
- HENDERSON, C. R. JR. (1982): Analysis of covariance in the mixed model: higher-level, nonhomogeneous and random regressions. *Biometrics* 38: 623 - 640.
- HENDERSON, C. R. (1963): Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding*, NAS-NRC Publ. No.982, 141 - 163.
- HENDERSON, C. R. (1975): Best linear estimation and prediction under a selection model. *Biometrics* 31: 423 - 447.
- HENDERSON, C. R., (1984): Application of linear models in animal breeding. University of Guelph.
- HENDERSON, C. R. (1990): Statistical method in animal improvement: historical overview. In: *Advances in Statistical Methods for Genetic Improvement of Livestock*. Springer Verlag, New York.
- KACKAR, A. N., Harville, D. A. (1981): Unbiasedness of two-stage estimation and precision procedures for mixed linear models. *Communications in Statistics, Series A* 10: 1249-1261.
- KENWARD, M. G., Roger, J. H. (1997): Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53: 983-997.
- LAMOTTE, L. R. (1973): Quadratic estimation of variance components. *Biometrics* 29: 311-330.
- MIELENZ, N., Spilke, J. (1984): Die Auswertung von Kombinationseignungsprüfungen bei besonderer Berücksichtigung der Methode BLUP. *Archives of Animal Breeding* 27: 311 - 324.
- PATTERSON, H. D., Thompson, R. (1971): Recovery of interblock information when block sizes are unequal. *Biometrika* 58: 545-554.
- RAO C. R. (1971): Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* 1: 445-456.
- SATTERTHWAITE, F. E. (194): Synthesis of variance. *Psychometrika* 6: 309-316.
- SEARLE, S. R. (1971): *Linear Models*. John Wiley and Sons, Inc. New York.
- SEARLE, S. R. (1987): *Linear models for unbalanced data*. John Wiley and Sons, Inc. New York.
- SEARLE, S. R. Casella, G.; McCulloch, C. E., (1992): *Variance Components*. John Wiley and Sons, Inc. New York.
- SEARLE, S. R. (1971):. Topics in variance component estimation. *Biometrics* 27: 1 - 76.
- SPIKKE, J., Dörfel, H.; Warnstorff, K. (1998): Negative Varianzkomponentenschätzungen in zweifaktoriellen Versuchs-Konsequenzen für die Modellwahl? Sitzung der AG Landwirtschaftliches Versuchswesen der Deutschen Region der Internationalen Biometrischen Gesellschaft, 16./17. März, Mainz.
- SPIKKE, J., Piepho, H. P., Hu, X. (2001): Simulationsuntersuchungen zur Hypothesenprüfung über feste Effekte bei verschiedenen Freiheitsgradapproximationen in gemischten linearen Modellen bei kleinen und unbalancierten Stichproben. 5. Konferenz der SAS-Anwender in Forschung und Entwicklung 8./9. März 2001, Universität Hohenheim.
- WOLFINGER, R. (2000): persönliche Mitteilung.

**Simulationsuntersuchungen zum Einfluss verschiedener Strategien der Varianzkomponentenschätzung und Hypothesenprüfung auf die statistischen Risiken in gemischten linearen Modellen mit ungleicher Klassenbesetzung** (J. Spilke, A. Tuchscherer)

## Zusammenfassung

*Bei Nutzung der Monte-Carlo Simulation wurden zwei Strategien der Varianzkomponentenschätzung (TYPE 3 und REML) sowie zwei Vorgehensweisen zur Approximation der Freiheitsgrade (Satterthwaite und Kenward-Roger) unter den Bedingungen unbalancierter Daten (Spaltanlage mit fehlenden Beobachtungen und Gitteranlage) verglichen. Beide Methoden zur Varianzkomponentenschätzung ermöglichen unverzerrte Schätzungen für die festen Effekte. Bedeutsame Unterschiede ergeben sich jedoch für die Effektschätzung und Hypothesenprüfung, indem TYPE 3 zu einem höheren MSE und extremen Abweichungen vom nominalen statistischen Risiko erster Art führt. Die Kombination REML-Kenward-Roger führt zur besten Einhaltung des nominalen statistischen Risikos erster Art.*

**Stichworte:** Gemischte lineare Modelle, unbalancierte Daten, Varianzkomponentenschätzung, Hypothesenprüfung, Approximation der Freiheitsgrade

**Comparison of different strategies of variance component estimation and degrees of freedom approximation on the statistical errors in mixed linear models for unbalanced data** (J. Spilke, A. Tuchscherer)

## Summary

*Based on a Monte carlo study the aim of the paper is to compare different strategies of variance component estimation (TYPE 3 and REML) as well as degrees of freedom approximation (Satterthwaite and Kenward-Roger) in a unbalanced data situation. An incomplete block design (lattice design) and a split-plot-design (with missing values) were investigated. Both methods of variance component estimation lead to unbiased fixed effect estimates. By contrast, the statistical errors of the first and second kind differ considerably in dependence of the used procedure to estimate the unknown variance components and approximate the degrees of freedom. TYPE 3 estimates leads to bigger MSE of the estimated fixed effects and extreme departures from the error of first kind. The combination REML-Kenward-Roger delivers the best control nominal of the statistical error of the first kind.*

**Key words:** Mixed linear models, unbalanced data, variance component estimation, hypothesis testing, degrees of freedom approximation

*Prof. Dr. habil. Joachim Spilke, Martin-Luther-Universität Halle-Wittenberg, Landwirtschaftliche Fakultät, AG Biometrie und Agrarinformatik, Tel. : 0345 5522690, Fax: 0345 55 22128, Email: spilke@landw.uni-halle.de, 06099 Halle.*

*Dr. Armin Tuchscherer, Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Forschungsbereich Genetik und Biometrie, Email: atuchs@fjbn-dummerstorf.de, Tel. : 038208 68904, Fax: 038208 68605, W.-Stahl-Allee 2, 18196 Dummerstorf.*