

Simulering för test av multiple trait analys för egenskaper mätta på olika individer

Gunnar Jansson & Öje Danell

Ämnesord: skogsträdsförädling, simulering

SkogForsk – Stiftelsen Skogsbrukets Forskningsinstitut

SkogForsk arbetar för ett långsiktigt, lönsamt skogsbruk på ekologisk grund. Bakom SkogForsk står skogsbolag, skogsägareföreningar, stift, gods, allmänningar, plant-skolor, SkogsMaskinFöretagarna m.fl., som betalar årliga intressentbidrag. Hela skogsbruket bidrar dessutom till finansieringen genom en avgift på virke som avverkas i Sverige. Verksamheten finansieras vidare av staten enligt särskilt avtal och av fonder som ger projektbundet stöd.

SkogForsk arbetar med forskning och utveckling med fokus på fyra centrala frågeställningar: Produktvärde och produktionseffektivitet, Miljöanpassat skogsbruk, Nya organisationsstrukturer samt Skogsodlingsmaterial. På de områden där SkogForsk har särskild kompetens utförs även i stor omfattning uppdrag åt skogsföretag, maskintillverkare och myndigheter.

Serien **Arbetsrapport** dokumenterar långliggande försök samt inventeringar, studier m.m. och distribueras enbart efter särskild beställning.

Forsknings- och försöksresultat från SkogForsk publiceras i följande serier:

SkogForsk-Nytt: Nyheter, sammanfattningar, översikter.

Resultat: Slutsatser och rekommendationer i lättillgänglig form.

Redogörelse: Utförlig redovisning av genomfört forskningsarbete.

Report: Vetenskapligt inriktad serie (på engelska).

Handledningar: Anvisningar för hur olika arbeten lämpligen utförs.

ISSN 1404-305X

Innehåll

Inledning.....	3
Material och metoder.....	4
Modell.....	4
Statistisk och biologisk modell.....	5
Simulering av avkommevärden.....	5
Förutsättningar.....	7
Simuleringsstudie 1.....	7
Simuleringsstudie 2.....	8
Resultat och diskussion	9
Simuleringsstudie 1.....	9
Simuleringsstudie 2.....	10
Slutsats	13
Referenser.....	13

Inledning

Inom kvantitativ genetik och husdjursförädling har under de senaste decennierna skett en mycket snabb utveckling och anpassning av avancerad statistisk skattningsmetodik för genetiska tillämpningar. Motsvarande saknas nästan helt inom skogsträdsförädling.

När det gäller skattning av genetiska korrelationer är metodkraven mycket stora. Svårigheten ligger i att man aldrig kan mäta genotyper eller genetisk variation direkt utan enbart den fenotypvariation som är det samlade resultatet av genetiska effekter och relativt sett betydligt större individuell miljöpåverkan. Skattning av ett genetiskt samband innebär att man ur synliga/mätbara fenotypvärden skall särskilja ett genetiskt samband. Skattningsprincipen bygger på att man via graden av likhet (kovarians) mellan vissa typer av släktingar avseende sambandet mellan egenskaperna identifierar den genetiskt betingade delen av sambandet. Genetiska korrelationer är genom det indirekta skattnings sättet behäftade med stora skattningsfel. Speciellt svåra ur metodsynpunkt är skattningar av genetiska samband mellan egenskaper, som kommer till uttryck på olika individer, t.ex. träd i etträdsparceller kontra träd i flerträdsparceller eller träd från samma familj mätta i olika försök.

Den idag rekommendabla metodiken bygger på "Mixed Model Equations" (MME) och iterativ skattning med s.k. "Restricted Maximum Likelihood"-teknik (REML). Aktuella metodöversikter ges av bl.a. Henderson (1984), Mayer (1990, 1991) och Schaeffer (1986, 1993). REML-tekniken är flexibel och kan relativt lätt anpassas till olika multivariata modeller där designmatrisen varierar mellan variabler. Man kan t.ex. ha olika fixa effekter för de olika egenskaperna. Vidare kan ena egenskapen vara mätt på t.ex. halvsyskonmaterial och den andra på helsyskonmaterial. I regel är kraven såpass speciella i varje enskilt fall att det kräver mer eller mindre skraddarsydda lösningar, bl.a. på grund av att beräkningarna ofta är mycket krävande. Detta innebär att användare inom skogsträdsförädling ofta är hänvisade till egen utveckling av programvara för sina tillämpningar. För den speciella men relativt vanliga situationen där olika egenskaper mäts på olika trädindivider har därför ett FORTRAN-program för skattning av varianser och kovarianser med REML-metodik utarbetats för att användas inom skogsträdsförädling (program TTFSREML, Jansson, 1997). Programmet är gjort för två egenskaper mätta på olika individer och avser helsyskonanalys. För att testa programmet har data med kända varianser och kovarianser simulerats och jämförts med de skattningar som programmet ger.

Huvudsyftet med simuleringarna i denna rapport har varit att jämföra de skattningar av parametrar som FORTRAN-programmet ger med simulerade data genererade utifrån kända varianser och kovarianser. Ett andra syfte var att visa hur simuleringar kan användas för att skatta medelfel till genetiska korrelationer.

Material och metoder

Simuleringarna har gjorts för en situation med två egenskaper mätta på olika individer och avser helsyskon.

Modell

I matrisform har modellen följande utseende:

$$\begin{bmatrix} \mathbf{y} \\ \tilde{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{X}} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \tilde{\mathbf{b}} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{Z}}_1 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \tilde{\mathbf{p}} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{Z}}_2 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \tilde{\mathbf{f}} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \tilde{\mathbf{e}} \end{bmatrix}$$

där

- $\mathbf{y}$ och $\tilde{\mathbf{y}}$ = observationsvektor för de två egenskaperna
- $\mathbf{X}$ och $\tilde{\mathbf{X}}$ = designmatris för fixa blockeffekter
- $\mathbf{b}$ och $\tilde{\mathbf{b}}$ = vektor med fixa blockeffekter
- $\mathbf{Z}_1$ och $\tilde{\mathbf{Z}}_1$ = designmatris för slumpmässiga föräldraeffekter
- $\mathbf{p}$ och $\tilde{\mathbf{p}}$ = vektor för slumpmässiga föräldraeffekter
- $\mathbf{Z}_2$ och $\tilde{\mathbf{Z}}_2$ = designmatris för slumpmässiga familjeeffekter
- $\mathbf{f}$ och $\tilde{\mathbf{f}}$ = vektor med slumpmässiga familjeeffekter
- $\mathbf{e}$ och $\tilde{\mathbf{e}}$ = vektor med slumpmässiga residualer

Förväntade medelvärden var

$$\begin{aligned} E \begin{bmatrix} \mathbf{y} \\ \tilde{\mathbf{y}} \end{bmatrix} &= \begin{bmatrix} \mathbf{X}\mathbf{b} \\ \tilde{\mathbf{X}}\tilde{\mathbf{b}} \end{bmatrix} \\ E \begin{bmatrix} \mathbf{p} \\ \tilde{\mathbf{p}} \end{bmatrix} &= E \begin{bmatrix} \mathbf{f} \\ \tilde{\mathbf{f}} \end{bmatrix} = E \begin{bmatrix} \mathbf{e} \\ \tilde{\mathbf{e}} \end{bmatrix} = \mathbf{0} \end{aligned}$$

Varians-kovariansmatriserna är (föräldrarna antas vara obesläktade)

$$\text{Var} \begin{bmatrix} \mathbf{p} \\ \tilde{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_{11} & \mathbf{g}_{12} \\ \mathbf{g}_{21} & \mathbf{g}_{22} \end{bmatrix} \text{ med invers } \begin{bmatrix} \mathbf{g}^{11} & \mathbf{g}^{12} \\ \mathbf{g}^{21} & \mathbf{g}^{22} \end{bmatrix} \text{ och}$$

$$\mathbf{g}_{11} = \mathbf{I}\sigma_{p_1}^2, \mathbf{g}_{22} = \mathbf{I}\sigma_{p_2}^2, \mathbf{g}_{12} = \mathbf{g}_{21} = \mathbf{I}\sigma_{p_1 p_2}$$

$$\text{Var} \begin{bmatrix} \mathbf{f} \\ \tilde{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_{11} & \mathbf{f}_{12} \\ \mathbf{f}_{21} & \mathbf{f}_{22} \end{bmatrix} \text{ med invers } \begin{bmatrix} \mathbf{f}^{11} & \mathbf{f}^{12} \\ \mathbf{f}^{21} & \mathbf{f}^{22} \end{bmatrix} \text{ och}$$

$$\mathbf{f}_{11} = \mathbf{I}\sigma_{f_1}^2, \mathbf{f}_{22} = \mathbf{I}\sigma_{f_2}^2, \mathbf{f}_{12} = \mathbf{f}_{21} = \mathbf{I}\sigma_{f_1 f_2}$$

$$\text{Var} \begin{bmatrix} \mathbf{e} \\ \tilde{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{r}_{22} \end{bmatrix} \text{ med invers } \begin{bmatrix} \mathbf{r}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{r}^{22} \end{bmatrix} \text{ och } \mathbf{e}_{11} = \mathbf{I}\sigma_{e_1}^2, \mathbf{e}_{22} = \mathbf{I}\sigma_{e_2}^2$$

$$\text{Cov} \left(\begin{bmatrix} \mathbf{p} \\ \tilde{\mathbf{p}} \end{bmatrix}, \begin{bmatrix} \mathbf{f} \\ \tilde{\mathbf{f}} \end{bmatrix} \right) = \text{Cov} \left(\begin{bmatrix} \mathbf{p} \\ \tilde{\mathbf{p}} \end{bmatrix}, \begin{bmatrix} \mathbf{e} \\ \tilde{\mathbf{e}} \end{bmatrix} \right) = \text{Cov} \left(\begin{bmatrix} \mathbf{f} \\ \tilde{\mathbf{f}} \end{bmatrix}, \begin{bmatrix} \mathbf{e} \\ \tilde{\mathbf{e}} \end{bmatrix} \right) = \mathbf{0}.$$

Statistisk och biologisk modell

Den biologiska modellen utgör grunden för simuleringarna. Utifrån genetiska varianser och kovarianser genereras data som svarar mot den statistiska modellen. Den statistiska modellen, $y_{jj'k} = \mu + p_j + p_{j'} + f_{jj'} + e_{jj'k}$, leder till följande sammansättning av den totala fenotypiska variansen

$$\sigma_t^2 = \sigma_p^2 + \sigma_f^2 + \sigma_e^2$$

där

$$\sigma_t^2 = \text{fenotypisk varians}$$

$$\sigma_p^2 = \text{föräldravarians}$$

$$\sigma_f^2 = \text{familjevariens}$$

$$\sigma_e^2 = \text{residualvariens}$$

Den genetiska modellen, $P = \mu + A + D + E$, ger på motsvarande sätt

$$\sigma_P^2 = \sigma_A^2 + \sigma_D^2 + \sigma_E^2$$

där

$$\sigma_P^2 = \text{fenotypisk varians}$$

$$\sigma_A^2 = \text{additiv föräldravarians}$$

$$\sigma_D^2 = \text{dominansvariens}$$

$$\sigma_E^2 = \text{miljövariens}$$

Sambandet mellan de statistiska och genetiska varianstermerna är

$$\sigma_p^2 = \frac{1}{4}\sigma_A^2$$

$$\sigma_f^2 = \frac{1}{4}\sigma_D^2$$

$$\sigma_e^2 = \frac{1}{2}\sigma_A^2 + \frac{3}{4}\sigma_D^2 + \sigma_E^2$$

Simulering av avkommevärden

För varje individ genereras en avkommeobservation med ett y-värde för den simulerade egenskapen. För att detta skall vara möjligt väljs först:

h_1^2 heritabilitet för egenskap 1

h_2^2 heritabilitet för egenskap 2

d_1^2 dominansvariens i förhållande till fenotypisk varians för egenskap 1

d_2^2 dominansvariens i förhållande till fenotypisk varians för egenskap 2

r_A additiv genetisk korrelationskoefficient

r_D dominanskorrelation

r_E miljömässig korrelation, i detta fall förutsatt vara 0

$\sigma_{P_1}^2$ fenotypisk varians för egenskap 1

$\sigma_{P_2}^2$ fenotypisk varians för egenskap 2
 b_{1j} för $i = 1, K \ n_1$ fix blockeffekt för egenskap 1
 b_{2j} för $i = 1, K \ n_2$ fix blockeffekt för egenskap 2

Utifrån dessa förutsättningar beräknas additiv genetisk varians (σ_A^2) dominansvariens (σ_D^2) och miljövariens (σ_E^2):

$$\begin{aligned}
 \sigma_{A_1}^2 &= h_1^2 \sigma_{P_1}^2 \\
 \sigma_{A_2}^2 &= h_2^2 \sigma_{P_2}^2 \\
 \sigma_{D_1}^2 &= d_1^2 \sigma_{P_1}^2 \\
 \sigma_{D_2}^2 &= d_2^2 \sigma_{P_2}^2 \\
 \sigma_{E_1}^2 &= \sigma_{P_1}^2 - \sigma_{A_1}^2 - \sigma_{D_1}^2 \\
 \sigma_{E_2}^2 &= \sigma_{P_2}^2 - \sigma_{A_2}^2 - \sigma_{D_2}^2
 \end{aligned}$$

Tredje steget blir sedan att generera data. Detta följer nedanstående arbetsgång.

Sammanlagt föräldrabidrag i egenskap 1:

$$p_{11} + p_{12} = \sqrt{\frac{1}{4} \sigma_{A_1}^2} \cdot s_k + \sqrt{\frac{1}{4} \sigma_{A_1}^2} \cdot s_{k'} = \sqrt{\frac{1}{4} \sigma_{A_1}^2} \cdot (s_k + s_{k'})$$

där p_{11} och p_{12} är effekt av förälder 1, respektive 2 och s_k och $s_{k'}$ är pseudo-slumptal för föräldereffekt 1 respektive föräldereffekt 2 och är genererade för $N(0, 1)$.

Sammanlagt föräldrabidrag i egenskap 2:

$$p_{21} + p_{22} = \sqrt{r_A^2 \frac{1}{4} \sigma_{A_2}^2} \cdot s_k + \sqrt{(1-r_A^2) \frac{1}{4} \sigma_{A_2}^2} \cdot s_{k+1} + \sqrt{r_A^2 \frac{1}{4} \sigma_{A_2}^2} \cdot s_{k'} + \sqrt{(1-r_A^2) \frac{1}{4} \sigma_{A_2}^2} \cdot s_{k'+1}$$

där p_{21} och p_{22} är effekt av förälder 1, respektive 2 och k respektive k' är samma slumptal som i föregående uttryck.

Familjebidraget i egenskap 1 (f_1):

$$f_1 = \sqrt{\frac{1}{4} \sigma_{D_1}^2} \cdot s_{k+2}$$

Familjebidraget i egenskap 2 (f_2):

$$f_2 = \sqrt{r_D^2 \frac{1}{4} \sigma_{D_2}^2} \cdot s_{k+2} + \sqrt{(1-r_D^2) \frac{1}{4} \sigma_{D_2}^2} \cdot s_{k+3}$$

Avkommeobservation i egenskap 1:

$$y_1 = b_{1i} + p_{11} + p_{12} + f_1 + \sqrt{\frac{1}{2}\sigma_{A_1}^2} \cdot s_{k+4} + \sqrt{\frac{3}{4}\sigma_{D_1}^2} \cdot s_{k+5} + \sqrt{\sigma_{E_1}^2} \cdot s_{k+6}$$

där

$$\sqrt{\frac{1}{2}\sigma_{A_1}^2} \cdot s_{k+4} + \sqrt{\frac{3}{4}\sigma_{D_1}^2} \cdot s_{k+5} + \sqrt{\sigma_{E_1}^2} \cdot s_{k+6}$$

utgör den slumpmässiga residualen i den statistiska modellen.

Avkommeobservation i egenskap 2:

$$y_2 = b_{2i} + p_{21} + p_{22} + f_2 + \sqrt{\frac{1}{2}\sigma_{A_2}^2} \cdot s_{k+7} + \sqrt{\frac{3}{4}\sigma_{D_2}^2} \cdot s_{k+8} + \sqrt{\sigma_{E_2}^2} \cdot s_{k+9}$$

med motsvarande residualterm sammansatt av de tre sista termerna.

Förutsättningar

I tabell 1 redovisas förutsättningarna för de två simuleringstudierna.

Tabell 1.
Förutsättningar för simuleringarna

	Simuleringsstudie 1	Simuleringsstudie 2
Antal föräldrar	30	60
Antal block	20	20
Antal avkommor per familj och block	2	2
Antal familjer	125	275
Antal simuleringar	25	30

Simuleringsstudie 1

I några simuleringar har 25 % av observationerna slumpmässigt tagits bort för att simulera en överlevnad på 75 %. Simuleringar har gjorts dels med ett partiellt diallelt korsningsschema (figur 1), dels med ett faktoriellt korsningsschema (figur 2).

I några fall har simuleringarna baserats på faktiska datadesigner och den slutgiltiga MME-lösningen och variansskattningen. Detta innebär att MME-lösningens fixa effekter för respektive egenskap utnyttjats. Till dessa har sedan slumpgenererade värden för föräldrar, familj och residual adderats. Modellen för att generera avkommeobservation blir i detta fall för egenskap 1:

$$y_{1ij'k} = b_{1i} + \sigma_p s_{1j} + \sigma_p s_{1j'} + \sigma_f s_{1jj'} + \sigma_e s_{1ij'k}$$

och för egenskap 2:

$$y_{2ij'k} = b_{2i} + r_A \sigma_p s_{1j} + \sqrt{(1-r_A^2)} \sigma_p s_{2j} + r_A \sigma_p s_{1j'} + \sqrt{(1-r_A^2)} \sigma_p s_{2j'} + r_D \sigma_f s_{1jj'} + \sqrt{(1-r_D^2)} \sigma_f s_{2jj'} + \sigma_e s_{2ij'k}$$

där

$y_{tijj'k}$ är observation k på förälderkombination jj' i block i och egenskap t och alla s är genererade normalfördelade pseudo-slumptal $N(0, 1)$.

Konvergens ansågs uppnådd när den relativa förändringen för all varians- och kovariansskattningar från en iteration till nästa var mindre än 0,00001 (i några fall 0,0001).

	Förälder																														
	1	2	3	4	5	6	7	8	9			23	24	25	26	27	28	29	30												
1		x	x	x	x	x																									
2			x	x	x	x	x																								
3				x	x	x	x	x																							
4					x	x	x	x	x																						
5																															
.							.																								
.								.																							
.									.																						
.										.																					
22												x	x	x	x	x															
23													x	x	x	x	x														
24														x	x	x	x	x													
25															x	x	x	x	x												x

Figur 1.

Partiellt diallelt korsningsschema som har använts i huvuddelen av simuleringarna. x markerar de korsningar som ingår.

	Förälder				
	1	2	3	4	5
6	x	x	x	x	x
7	x	x	x	x	x
8	x	x	x	x	x
9	x	x	x	x	x
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
27	x	x	x	x	x
28	x	x	x	x	x
29	x	x	x	x	x
30	x	x	x	x	x

Figur 2.

Faktoriellt korsningsschema som använts i några av simuleringarna. x markerar de korsningar som ingår.

Simuleringsstudie 2

Simuleringsstudie 2 skiljer sig från studie 1 genom att simuleringarna har genomförts med ett större datamaterial (tabell 1). Ett partiellt diallelt korsningsschema med 5 korsningar per förälder utnyttjades. I denna studie redovisas även exempel på hur simuleringar kan utnyttjas för att skatta medelfel till genetiska korrelationer.

Resultat och diskussion

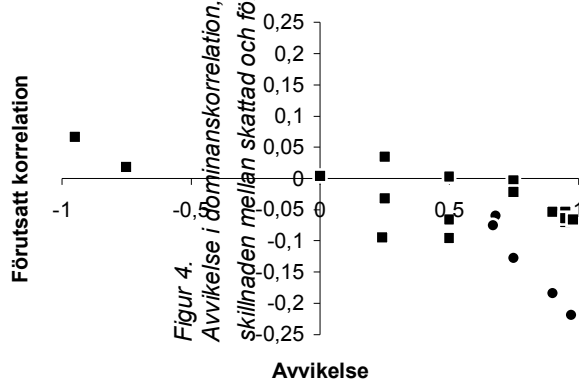
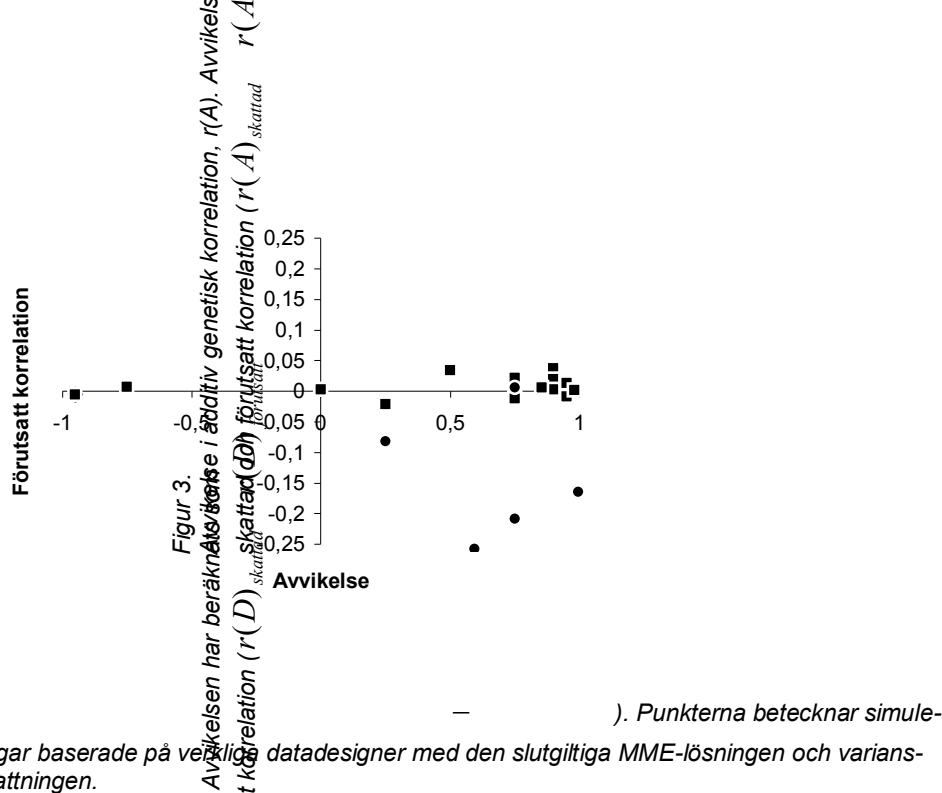
Simuleringsstudie 1

Tabell 2 visar en sammanställning av resultaten av simuleringarna. I figur 3 och 4 visas avvikelserna mellan skattad och förutsatt korrelation för den additiva korrelationen respektive dominans korrelationen.

Tabell 2.
Resultat av simuleringarna.

Förutsättningar								Skattad	
r(A)	r(D)	h_1^2	h_2^2	Antal obs. trait 1	Antal obs. trait 2	Design	Överlevnad %	r(A)	r(D)
-0,95	-0,95	0,20	0,20	5 000	5 000	Part	100	-0,955	-0,884
-0,75	-0,75	0,20	0,20	5 000	5 000	Part	100	-0,743	-0,732
0,00	0,00	0,20	0,20	5 000	5 000	Part	100	0,003	0,004
0,25	0,25	0,20	0,20	5 000	5 000	Part	100	0,229	0,284
0,25	0,25	0,20	0,20	5 000	5 000	Part	100	0,168	0,218
0,50	0,50	0,20	0,20	5 000	5 000	Part	100	0,534	0,503
0,75	0,75	0,20	0,20	5 000	5 000	Part	100	0,738	0,742
0,75	0,75	0,05	0,70	5 000	500	Fakt	100	0,738	0,622
0,75	0,75	0,20	0,20	5 000	5 000	Part	100	0,772	0,746
0,75	0,75	0,20	0,20	5 000	5 000	Part	75	0,756	0,728
0,90	0,90	0,20	0,20	5 000	5 000	Part	100	0,911	0,846
0,90	0,50	0,05	0,70	5 000	500	Part	100	0,937	0,404
0,90	0,50	0,05	0,70	5 000	500	Fakt	100	0,920	0,434
0,90	0,90	0,20	0,20	1 250	5 000	Part	75	0,903	0,716
0,95	0,95	0,20	0,20	5 000	5 000	Part	100	0,964	0,880
0,95	0,95	0,20	0,20	5 000	5 000	Fakt	100	0,941	0,896
0,95	0,95	0,20	0,20	5 000	5 000	Part	100	0,963	0,886
0,98	0,98	0,20	0,20	5 000	5 000	Part	100	0,982	0,914
0,9941	0,6804	0,0458	0,2034	1 180	220	Part	Verklig	0,830	0,620
0,7513	0,6701	0,0070	0,2446	2 961	220	Part	Verklig	0,543	0,595
0,8536	0,9731	0,2117	0,6590	3 857	245	Fakt	Verklig	0,860	0,754
0,5945	0,2406	0,1484	0,6002	218	245	Fakt	Verklig	0,337	0,146
0,8533	0,9310	0,2133	0,5295	3 857	245	Fakt	Verklig	0,833	0,761

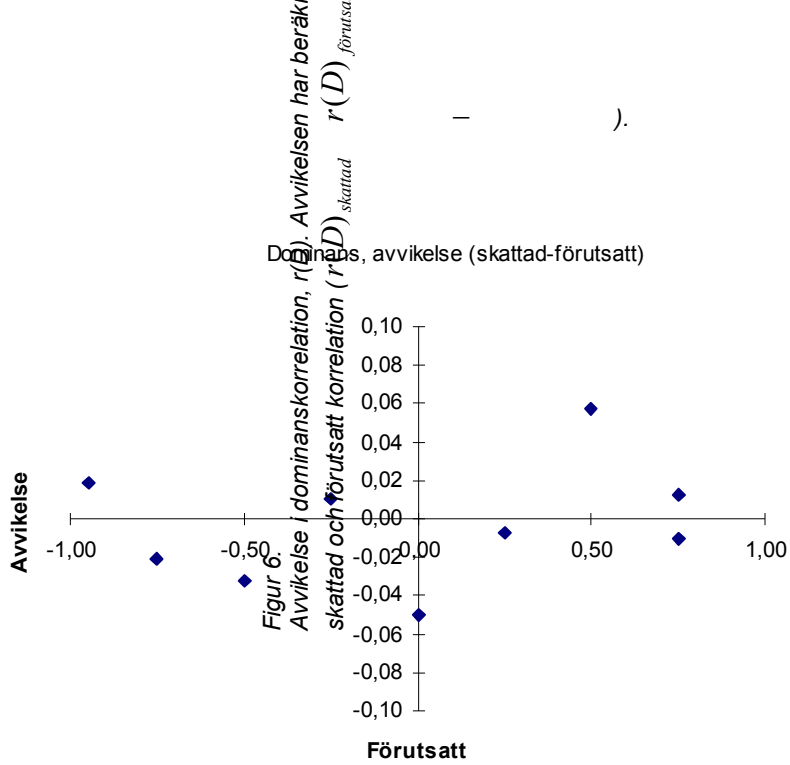
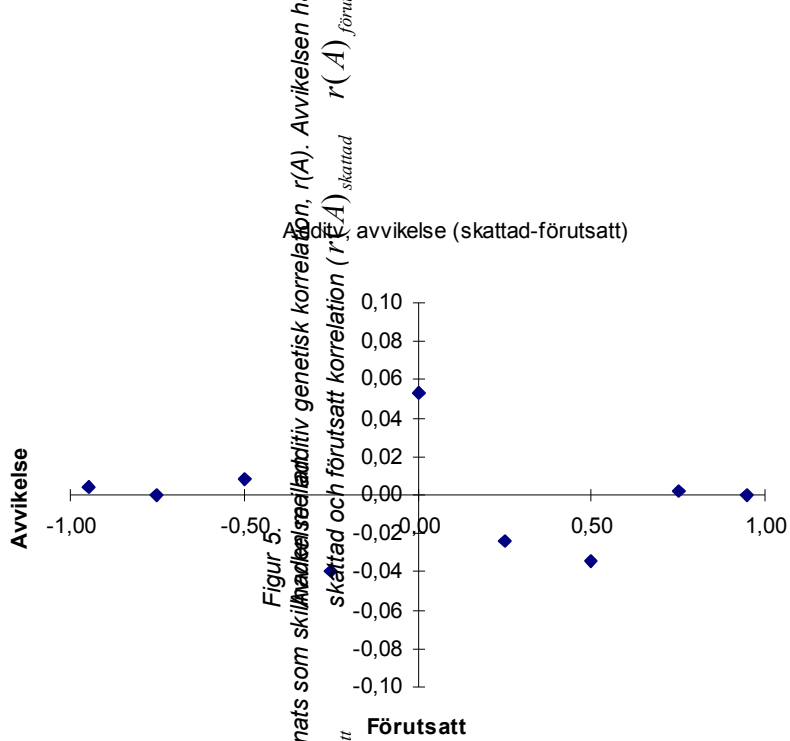
Resultaten visar på god överensstämmelse mellan de förutsatta parametrarna och de efter simuleringarna skattade parametrarna med beaktande av de medelfel som de är behäftade med. Simuleringar baserade på data hämtade ur verkliga material med den slutgiltiga MME-lösningen och variansskattningen har den största avvikelserna mellan förutsatt och skattad korrelation och uppvisar en viss tendens till underskattning jämfört med den förutsatta genetiska korrelationen. Detta kan förklaras av låg heritabilitet och litet datamaterial.



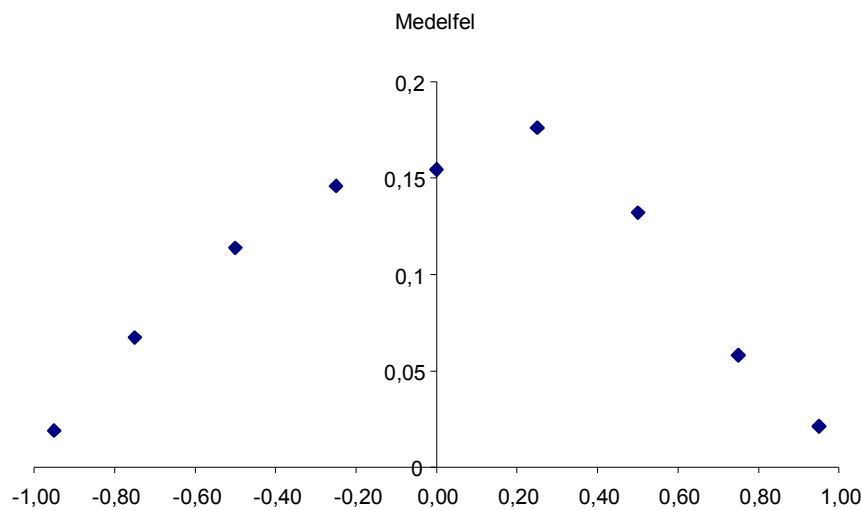
Simuleringsstudie 2

I den andra simuleringsstudien användes 60 föräldrar och 275 familjer, d.v.s. ett betydligt större material än i Simuleringsstudie 1. Ett partiellt diallelt korsningsschema användes där varje förälder ingick i 5 korsningar. I figur 5 och 6 visas avvikelsen mellan skattad och förutsatt korrelation för den additiva korrelationen respektive dominans korrelationen.

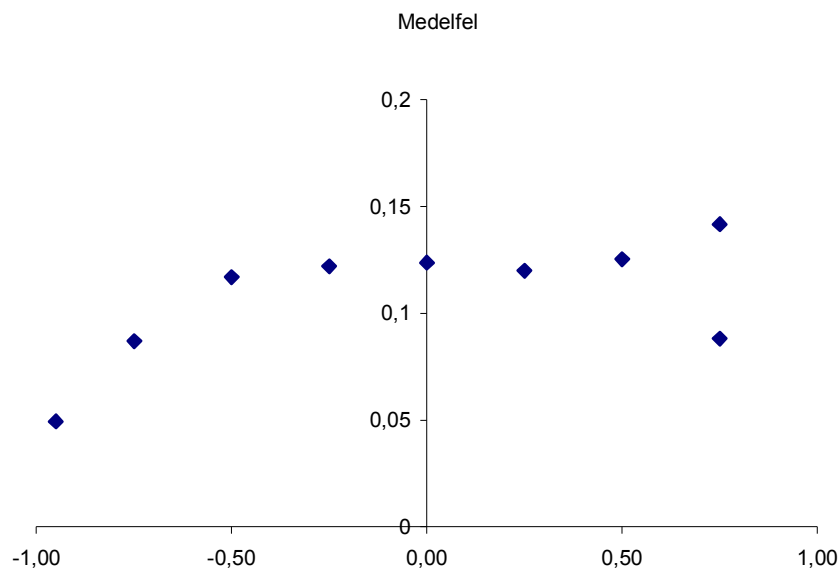
Resultaten visar på god överensstämmelse mellan de förutsatta parametrarna och de efter simuleringarna skattade parametrarna (med beaktande av de relativt stora fel de är behäftade med) över hela skalan från korrelationer på -1 till +1.



REML enligt EM algoritmen ger inga felskattningar till varianskomponenter eller därur skattade parametrar. För att få felskattningar kan simuleringar användas. Tillvägagångssättet är då att först skatta effekter och varianser ur data och sedan utnyttja dem i simuleringarna. Till dessa ”kända” blockeffekter adderas genetiska och miljömässiga effekter från normalfördelade pseudoslumpstal med de varianser som tidigare skattats från data. Dessa simuleringar upprepas ett antal gånger för varje dataset. Medelfelet till skattningarna erhålls sedan som standardavvikelsen till en (ko)varianskomponent eller parameter. I figur 7 visas exempel på simulerat medelfel för additiv genetisk korrelation. Felet är minst när korrelationen är nära -1 eller +1 och störst när korrelationen ligger nära 0. I figur 8 visas motsvarande för dominanskorrelationen.



Figur 7.
Medelfel vid olika additiv genetisk korrelation skattad från simuleringarna.



Figur 8.
Medelfel vid olika dominanskorrelation skattad från simuleringarna.

Slutsats

Huvudsyftet med simuleringstudien var att utifrån kända varianser och kovarianser simulera data för att jämföra med de skattningar som FORTRAN-programmet ger. Varken avvikelser i additiv genetisk korrelation eller dominans korrelation visar på några större systematiska avvikelser.

Den använda REML-proceduren ger inga felskattningar till varianskomponenter eller därur skattade parametrar. I rapporten ges exempel på hur simuleringar kan användas för att ta fram felskattningar till genetiska korrelationer.

Referenser

- Henderson, C. R. 1984. Applications of linear models in animal breeding. Guelph, University of Guelph. 423 s.
- Jansson, G. 1997. TTFSREML (Two Trait Full Sib REstricted Maximum Likelihood). Dataprogram för skattning av varianser och avelsvärden i avkomme försök med diallell eller faktoriell korsningsplan och med två egenskaper mätta på olika individer. (SkogForsk), Uppsala.
- Mayer, K. 1990. Present status of knowledge about statistical procedures and algorithms to estimate variance and covariance components. Proc. 4th World congress on Genetics applied to Livestock Production, Edinburgh 23-27 July, 1990, Vol XIII, 407–418.
- Mayer, K. 1991. Estimating variances and covariances for multiple animal models by restricted maximum likelihood. Genet. Sel. Evol. 51, 67–83.
- Schaeffer, L. R., Wilton, J. W. & Thompson, R. 1978. Simultaneous estimation of variance and covariance components from multitrait mixed model equations. Biometrics 34: 199–208.
- Schaeffer, L.R. 1993. Variance component estimation methods. Kompendium. University of Guelph, Guelph, Ontario.