

Detection of cytoplasmic effects on production: the influence of number of years of data

A Salehi, JW James *

*Department of Wool and Animal Science, The University of New South Wales,
Sydney 2052, Australia*

(Received 17 January 1995; revised 6 January 1997, accepted 3 March 1997)

Summary – Data were simulated for a single trait determined by additive genetic, cytoplasmic and environmental effects in a sheep flock. The proportions of variance due to genotype and cytoplasm were $h^2 = 0.30$ or 0.60 and $c^2 = 0.05$ or 0.10 , with an extra set in which $h^2 = 0.30$ and $c^2 = 0$. For all five parameter sets, two periods (10 or 20 years) were considered, with the same number of records in each case. Ten replicates of each were run so that 100 independent data sets were obtained. A ten-year data set with half the number of animals was also run, giving 150 analyses in all. Data sets were analysed by derivative-free restricted maximum likelihood, fitting four models, which included additive genotype alone or in combination with either or both an additive genetic maternal effect and a cytoplasmic effect. If the cytoplasmic effect were omitted from the analysis model, heritability and maternal variance were overestimated. If cytoplasmic effects were included, parameter estimates were close to true values. Heritability did not affect the power of tests for cytoplasmic effects, which was higher for $c^2 = 0.10$ than for $c^2 = 0.05$. Detection of cytoplasmic variance was more likely with data spread over 20 rather than over 10 years, and power was reduced with fewer data.

cytoplasmic effect / restricted maximum likelihood / animal model / sheep

Résumé – Détection d'effets cytoplasmiques sur des caractères de production : incidence du nombre des années avec données. On a simulé des données pour un caractère déterminé par les effets génétiques additifs, cytoplasmiques et de milieu, dans un troupeau de moutons. Les parts de variance phénotypique dues aux effets génétiques et cytoplasmiques sont $h^2 = 0,30$ ou $0,60$, et $c^2 = 0,05$ ou $0,10$ respectivement, avec une série supplémentaire où $h^2 = 0,30$ et $c^2 = 0$. Pour les cinq ensembles de paramètres, on considère deux périodes (10 ou 20 ans) avec un nombre égal des données dans les deux cas. On a fait dix répétitions pour chaque situation, et donc 100 fichiers indépendants sont obtenus. Un fichier couvrant 10 années avec un nombre d'animaux réduit de moitié est analysé aussi, soit 150 analyses en tout. Les fichiers sont analysés par la méthode

* Correspondence and reprints

du maximum de vraisemblance restreinte (REML) avec un procédé sans dérivées. Quatre modèles ont été ajustés, incluant, soit le génotype additif seul, soit le génotype additif plus l'un des effets cytoplasmiques ou génotype additive maternel, ou les deux. Quand le modèle d'analyse exclut l'effet cytoplasmique, l'héritabilité et la variance maternelle sont surestimées. Quand le modèle inclut l'effet cytoplasmique, les paramètres estimés approchent leurs valeurs vraies. La puissance des tests des effets cytoplasmiques est plus élevée pour $c^2 = 0,10$ que pour $c^2 = 0,05$, mais elle ne dépend pas de l'héritabilité. La détection d'effets cytoplasmiques est plus probable avec des données réparties sur 20 que sur 10 ans, et la puissance du test est diminuée quand le nombre de données diminue.

effet cytoplasmique / maximum de vraisemblance restreinte / modèle animal / mouton

INTRODUCTION

Although dam and sire contribute equally (with the exception of sex-linked genes in the heterogametic offspring sex) to the genotype of their offspring, it is accepted that progeny performance is affected more by the dam than by the sire (Hohenboken, 1985). A maternal effect may be defined as any influence on a progeny phenotype attributable to the dam, other than nuclear genes. In mammals, milk production, intrauterine environment and parental care are common components of maternal effects, which may be both genetically and environmentally determined.

Another possible source of maternal effect is the cytoplasm, especially mitochondrial DNA (mtDNA), which is maternally transmitted (Wagner, 1972). In recent years several studies following that of Bell et al (1985) have presented evidence for the influence of cytoplasm on production in cattle. However, Kennedy (1986) pointed out that careful analysis of data was needed to demonstrate the occurrence of a cytoplasmic effect, and that in particular it was necessary to account for the influence of nuclear genes. An assessment of the importance of fitting an appropriate model was made by Southwood et al (1989). They simulated data with a conventional maternal effect, with a cytoplasmic effect, or with both, in addition to an additive genetic effect. They then analysed the simulated data using models that included only a maternal effect, only a cytoplasmic effect, or both, as well as an additive genetic effect. When the correct model was fitted, the parameter estimates matched the true values, but estimates were biased when an incorrect model was fitted. They simulated a dairy cow population over 60 years, and analysed data from the last 30 years, with about 4 500 records being used.

In Australian Merino sheep it would be unusual to have such a data set, and most data sets would extend over 10–20 years. Over such a period, a separation of cytoplasm and maternal genotype might be more difficult to achieve. We have therefore undertaken studies of the effectiveness of estimating cytoplasmic effects in such populations. Most such populations would be undergoing selection, unlike the random choice of parents simulated by Southwood et al (1989), and so we have simulated populations under selection. Boettcher et al (1996) also used simulation of a dairy cow population with cytoplasmic effects but their interest was in biases introduced through ignoring cytoplasmic effects, and not in the detection of cytoplasmic variance. In this paper we report results for a model that includes an additive genetic effect and a cytoplasmic effect.

MATERIALS AND METHODS

Data were generated by computer simulation of a sheep flock with five age groups of breeding ewes and two age groups of rams. It was assumed that initially all breeding animals had been randomly chosen from the base population, while the juvenile animals were also a random sample of the base population. Thereafter, a fifth of the breeding ewes and a half of the rams were culled each year and replaced by the best available hoggets. The basic simulation used 10 rams and 200 ewes, with data being simulated over 20 or 10 years. Another set of simulations used 20 rams and 400 ewes for mating each year, with data simulated over 10 years.

Each animal had an additive genetic value and a cytoplasmic value. These were added, together with a random environmental deviation, to give the phenotypic value. Except for the base population, an animal's breeding value was the average of its parents' breeding values, plus a segregation effect in which allowance was made for parental inbreeding. An animal's cytoplasmic value was that of its dam. The seed for the random number generator was different for every run of the simulation so that all replicates were independent. The 10-years data sets were simulated separately from the 20-years data sets.

The trait under selection was assumed to have a heritability (h^2) of 0.3 or 0.6, with cytoplasmic variance (c^2) 0.05 or 0.10 as a fraction of phenotypic variance. Cytoplasmic and additive genetic effects were uncorrelated. For a heritability of 0.3, another series of runs with cytoplasmic variance set to zero was carried out, making a total of five models. Ten replicates were simulated for each model. With three data structures described above (population size, number of years) this gave 150 data sets for analysis.

The data were analysed using an animal model restricted maximum likelihood procedure and a derivative-free algorithm (Graser et al, 1987). The DFREML computer package (Meyer, 1989, 1991) was employed to fit four different models to the data. These were models 1, 2, 3 and 7 in DFREML terminology. The models are described in table I.

Table I. Models and parameters used in estimation.

<i>Model</i>	<i>Variance components</i>	<i>No of parameters</i>
1	$\sigma_a^2 \sigma_e^2$	1
2	$\sigma_a^2 \sigma_e^2 \sigma_c^2$	2
3	$\sigma_a^2 \sigma_e^2 \sigma_m^2$	2
7	$\sigma_a^2 \sigma_e^2 \sigma_c^2 \sigma_m^2$	3

a = additive genetic effects; m = maternal genetic effects; c = cytoplasmic genetic effects; e = environmental effects.

The cytoplasmic model used for simulation was:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_c\mathbf{c} + \mathbf{e}$$

The complete model fitted to the data (model 7) was:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_m\mathbf{m} + \mathbf{Z}_c\mathbf{c} + \mathbf{e}$$

where \mathbf{y} is a $N \times 1$ vector of observations, \mathbf{b} is a vector of fixed effects and \mathbf{X} , \mathbf{Z}_a , \mathbf{Z}_m and \mathbf{Z}_c are incidence matrices relating fixed effects (b), random additive genetic effects (a), maternal genetic effects (m), and cytoplasmic origin effects (c) to \mathbf{y} , respectively, and \mathbf{e} is a vector of random residual effects. The variance-covariance structure among the random effects for the full fitted model was:

$$V(\mathbf{a}) = \mathbf{A}\sigma_a^2 \quad V(\mathbf{m}) = \mathbf{A}\sigma_m^2 \quad V(\mathbf{c}) = \mathbf{I}_c\sigma_c^2 \quad V(\mathbf{e}) = \mathbf{I}_e\sigma_e^2$$

with all covariances zero. Here \mathbf{A} is the additive genetic relationship matrix between animals, σ_a^2 is additive genetic variance, σ_m^2 is maternal genetic variance, σ_c^2 is cytoplasmic variance, σ_e^2 is environmental variance, \mathbf{I}_c is an identity matrix of order equal to the number of cytoplasm origins, and \mathbf{I}_e is an identity matrix of order equal to the number of records. For model 3 it was assumed that $c = 0$, for model 2 that $m = 0$ and for model 1 that both c and m were 0. Fixed effects included in the model were year of record and sex. A single record at hogget age was analysed. For the fitting of the cytoplasmic effect the female in the base population from whom the cytoplasm descended was identified. The convergence criterion chosen for stopping the estimation procedure was a variance of 10^{-7} in the likelihood values for the simplex. Standard errors of parameter estimates were obtained from variation between replicates. The probability of finding significant cytoplasmic effects was assessed by use of likelihood ratio tests of model 7 versus model 3, assuming that -2 times the difference in log likelihoods had approximately a chi-squared distribution with one degree of freedom. This test is an appropriate way to test significance of the changes in the likelihood function when additional terms are included as explained by Rao (1973).

RESULTS

Parameter estimates with their standard errors (calculated from variation among replicates) are presented in tables II, III and IV for different data structures. The results of the log likelihood tests are presented in tables V and VI.

When there was no cytoplasmic effect in the simulated data the estimated heritability was always close to the true value. In the presence of a cytoplasmic effect, heritability was always overestimated when c^2 was omitted from the fitted model. Estimates of h^2 and c^2 with the model fitting additive genetic and cytoplasmic effects (model 2) were all close to the true values for all data sets across all sets of parameters, although a few values differed from the true value by more than two standard errors. With a model that included the maternal genetic effect as the only additional random effect (model 3), estimates of h^2 were higher than for model 2. According to the log likelihood ratio test presented in table V a complex model (model 7) fitting the additive genetic effect, maternal genetic effect and cytoplasmic effects gave significantly better fit than model 3, and $c^2 = 0.10$ gave, as expected, greater power than $c^2 = 0.05$ since a larger χ^2 implies a greater power or probability of rejecting the null hypothesis. Moreover, heritability did not affect

Table II. Average estimates of heritabilities, maternal genetic effects and cytoplasmic effects from REML analyses for a period of 20 years with five different parameter sets.

Model	Parameters		Parameter estimates		
	h^2	c^2	h^2 (se)	m^2 (se)	c^2 (se)
1	0.30	0.00	0.3026(0.0097)		
	0.30	0.05	0.3767(0.0132) ^b		
	0.30	0.10	0.4251(0.0088) ^b		
	0.60	0.05	0.6398(0.0080) ^b		
	0.60	0.10	0.6988(0.0094) ^b		
2	0.30	0.00	0.2932(0.0128)		0.0056(0.0024) ^a
	0.30	0.05	0.2975(0.0115)		0.0556(0.0093)
	0.30	0.10	0.2883(0.0080)		0.1009(0.0076)
	0.60	0.05	0.5916(0.0091)		0.0454(0.0048)
	0.60	0.10	0.6021(0.0148)		0.0986(0.0078)
3	0.30	0.00	0.2770(0.0129)	0.0116(0.0045) ^a	
	0.30	0.05	0.3042(0.0140)	0.0482(0.0112) ^b	
	0.30	0.10	0.3287(0.0177)	0.0711(0.0093) ^b	
	0.60	0.05	0.6128(0.0098)	0.0260(0.0039) ^b	
	0.60	0.10	0.6424(0.0121) ^b	0.0536(0.0085) ^b	
7	0.30	0.00	0.2785(0.0124)	0.0066(0.0047)	0.0030(0.0016)
	0.30	0.05	0.2809(0.0116)	0.0101(0.0042) ^a	0.0521(0.0091)
	0.30	0.10	0.2876(0.0118)	0.0031(0.0017)	0.1053(0.0107)
	0.60	0.05	0.5883(0.0100)	0.0026(0.0019)	0.0433(0.0048)
	0.60	0.10	0.5736(0.0119)	0.0044(0.0021)	0.1089(0.0095)

For comparison of estimate with true value assuming approximate normality. ^a $P < 0.05$; ^b $P < 0.01$; h^2 = heritability; m^2 = maternal genetic variance; c^2 = cytoplasmic variance.

the ability to detect a cytoplasmic effect. The average χ^2 values were higher for the larger data sets, but among the larger data sets, they were higher for 20 years rather than 10 years. Furthermore, table VI, in which the numbers of significant χ^2 values for replicates for each data set are presented, also indicates that the 20-year period was slightly more likely to show significance than a 10-year period, even with the same amount of data. In table V it can be seen that the average χ^2 value for 20 years is about twice that for 10 years with an equal amount of data, and thus with smaller cytoplasmic contributions the difference in power would be more noticeable.

DISCUSSION

Overestimation of heritability by model 1 across all sets of parameters and periods except for $c^2 = 0$, indicated that with an analysis model lacking cytoplasmic effects

Table III. Average estimates of heritabilities, maternal genetic effects and cytoplasmic effects from REML analyses for a period of 10 years with an equal amount of data given five different parameter sets.

<i>Model</i>	<i>Parameters</i>		<i>Parameter estimates</i>		
	h^2	c^2	h^2 (<i>se</i>)	m^2 (<i>se</i>)	c^2 (<i>se</i>)
1	0.30	0.00	0.2911(0.0073)		
	0.30	0.05	0.3666(0.0123) ^b		
	0.30	0.10	0.4085(0.0076) ^b		
	0.60	0.05	0.6357(0.0067) ^b		
	0.60	0.10	0.6864(0.0061) ^b		
2	0.30	0.00	0.2838(0.0065) ^a		0.0058(0.0028)
	0.30	0.05	0.3105(0.0147)		0.0495(0.0053)
	0.30	0.10	0.3036(0.0092)		0.0962(0.0074)
	0.60	0.05	0.5894(0.0080)		0.0524(0.0047)
	0.60	0.10	0.5710(0.0223)		0.1099(0.0036) ^a
3	0.30	0.00	0.2827(0.0094)	0.0055(0.0032)	
	0.30	0.05	0.3147(0.0089)	0.0445(0.0043) ^b	
	0.30	0.10	0.3252(0.0101) ^a	0.0763(0.0107) ^b	
	0.60	0.05	0.6044(0.0067)	0.0316(0.0057) ^b	
	0.60	0.10	0.6253(0.0098) ^a	0.0744(0.0080) ^b	
7	0.30	0.00	0.2831(0.0093)	0.0040(0.0031)	0.0008(0.0003) ^a
	0.30	0.05	0.2987(0.0096)	0.0075(0.0030) ^a	0.0509(0.0030)
	0.30	0.10	0.2984(0.0092)	0.0097(0.0049)	0.0877(0.0072)
	0.60	0.05	0.5766(0.0099)	0.0040(0.0017) ^a	0.0527(0.0058)
	0.60	0.10	0.5760(0.0095)	0.0107(0.0048)	0.1079(0.0068)

For comparison of estimate with true value assuming approximate normality. ^a $P < 0.05$; ^b $P < 0.01$; h^2 = heritability; m^2 = maternal genetic variance; c^2 = cytoplasmic variance.

some of these effects are then attributed to the additive genetic effect. When c^2 is accounted for by fitting model 2, c^2 estimates are not different from the true values. Higher estimates of heritability with model 3 compared to model 2, showed that with model 3 some of the cytoplasmic effects are then attributed to the additive genetic effect and most attributed to a maternal effect. Southwood et al (1989) stated that if an appropriate animal model is applied to the data, variance components can correctly be partitioned among additive direct, additive maternal and cytoplasmic effects. Model 7, including the additive genetic effect, maternal genetic effect and cytoplasmic effect was used to demonstrate cytoplasmic variation, distinct from maternal genetic effects, since in real data as distinct from our simulated data we can not know that there is no maternal effect other than that of the cytoplasm. The ability to detect a cytoplasmic variance component does not seem to depend on the amount of additive genetic variance. The importance of the number of years is clearly illustrated in tables V and VI. Within the same period

Table IV. Average estimates of heritabilities, maternal genetic effects and cytoplasmic effects from REML analyses for a period of 10 years with half the amount of data given five different parameter sets.

<i>Model</i>	<i>Parameters</i>		<i>Parameter estimates</i>		
	h^2	c^2	h^2 (<i>se</i>)	m^2 (<i>se</i>)	c^2 (<i>se</i>)
1	0.30	0.00	0.3099(0.0136)		
	0.30	0.05	0.3859(0.0158) ^b		
	0.30	0.10	0.4331(0.0141) ^b		
	0.60	0.05	0.6587(0.0093) ^b		
	0.60	0.10	0.6663(0.0149) ^b		
2	0.30	0.00	0.3001(0.0120)		0.0052(0.0022) ^a
	0.30	0.05	0.3079(0.0224)		0.0486(0.0088)
	0.30	0.10	0.2990(0.0223)		0.0947(0.0073)
	0.60	0.05	0.5851(0.0164)		0.0629(0.0097)
	0.60	0.10	0.5756(0.0197)		0.0772(0.0116)
3	0.30	0.00	0.2798(0.0196)	0.0035(0.0020)	
	0.30	0.05	0.2989(0.0212)	0.0438(0.0114) ^b	
	0.30	0.10	0.3276(0.0249)	0.0809(0.0083) ^b	
	0.60	0.05	0.6161(0.0226)	0.0362(0.0101) ^b	
	0.60	0.10	0.6144(0.0246)	0.0531(0.0116) ^b	
7	0.30	0.00	0.2710(0.0188)	0.0034(0.0019)	0.0042(0.0023)
	0.30	0.05	0.2706(0.0266)	0.0131(0.0066)	0.0463(0.0091)
	0.30	0.10	0.2922(0.0227)	0.0196(0.0061)	0.0881(0.0095)
	0.60	0.05	0.5625(0.0205)	0.0119(0.0067)	0.0594(0.0089)
	0.60	0.10	0.5604(0.0314)	0.0212(0.0106)	0.0769(0.0131)

For comparison of estimate with true value assuming approximate normality. ^a $P < 0.05$;
^b $P < 0.01$; h^2 = heritability; m^2 = maternal genetic variance; c^2 = cytoplasmic variance.

of 10 years, power was lower when the amount of data was halved. However, for a given amount of data, the power is higher if the data is spread over 20 rather than 10 years. Thus data extending over a number of years is preferable, but if enough records are available, it is worth analysing data from a short period.

The major difference between 10 years and 20 years of data for the detection of cytoplasmic effects is the difference in number of generations in which the cytoplasm may become independent from nuclear genes of the founder females. The distributions, averaged over all simulations, of cytoplasmic generation numbers are shown in table VII. Distributions were checked for different simulation models and were virtually identical in all cases except for the difference between the number of years. With 10-year data, less than 20% of cytoplasmic generation numbers are separated from the founders by more than two generations, but for 20-year data the value is greater than 50%.

Table V. Average χ^2 values obtained from log likelihood ratio tests of model 3 against model 7 for different sets of parameters and data sets.

<i>Parameters</i>		<i>Periods</i>		
h^2	c^2	<i>20 years</i>	<i>10 years equal data</i>	<i>10 years half data</i>
0.30	0.0	0.22	0.07	0.21
0.30	0.05	23.59	8.97	4.20
0.30	0.10	43.26	20.21	9.43
0.60	0.05	16.21	9.38	5.36
0.60	0.10	47.29	24.74	7.69

Table VI. Number of replicates for which log likelihood ratio tests of model 3 against model 7 were significant at levels of 5 and 1% for different sets of parameters and data sets.

<i>Parameters</i>		<i>Periods</i>					
h^2	c^2	<i>20 years</i>		<i>10 years equal data</i>		<i>10 years half data</i>	
		<i>No of χ^2 values significant</i>					
		a	b	a	b	a	b
0.30	0.0	—	—	—	—	—	—
0.30	0.05	8	8	8	7	4	4
0.30	0.10	10	10	10	9	10	5
0.60	0.05	10	9	9	4	7	5
0.60	0.10	10	10	10	10	8	6

^a $P < 0.05$; ^b $P < 0.01$.

Table VII. Distribution (in percentages) of number of cytoplasmic generations for 10 and 20 years of data.

<i>Period</i>	<i>Generations</i>						
	0	1	2	3	4	5	>5
20 years	10.4	18.0	17.3	17.3	16.8	13.2	7.0
10 years	18.8	32.5	30.3	15.8	2.5	0.1	0

In this study there were no non-cytoplasmic maternal effects to complicate the simulation. Further work is in progress to investigate cases where a genetic maternal effect occurs.

ACKNOWLEDGMENT

The senior author is grateful for a scholarship from the Ministry of Culture and Higher Education of Iran for the PhD study during which this study was conducted. Grateful acknowledgment is also made to Dr K Meyer for providing the DFREML program for analysing the data.

REFERENCES

- Bell BR, McDaniel BT, Robison OW (1985) Effects of cytoplasmic inheritance on production traits in dairy cattle. *J Dairy Sci* 68, 2038–2051
- Boettcher PJ, Kuhn MT, Freeman AE (1996) Impact of cytoplasmic inheritance on genetic evaluations. *J Dairy Sci* 79, 663–675
- Graser HU, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J Anim Sci* 64, 1362–1370
- Hohenboken WD (1985) Maternal effect. In: *World Animal Science. General and quantitative genetics* (AB Chapman, ed), Elsevier, Amsterdam, 135–149
- Kennedy BW (1986) A further look at evidence for cytoplasmic inheritance for production traits in dairy cattle. *J Dairy Sci* 69, 3100–3105
- Meyer K (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet Sel Evol* 21, 317–340
- Meyer K (1991) DFREML program to estimate variance components by restricted maximum likelihood using a derivative free-algorithm. *User Notes*, Version 2.0, University of New England
- Rao CR (1973) *Linear Statistical Inference and its Applications*. J Wiley and Sons, New York, 417–420
- Southwood OI, Kennedy BW, Meyer K, Gibson JP (1989) Estimation of additive maternal and cytoplasmic genetic variances in animal models. *J Dairy Sci* 72, 3006–3012
- Wagner RP (1972) The role of maternal effects in animal breeding. II. Mitochondria and animal inheritance. *J Anim Sci* 35, 1280–1287