Note

Genetic analysis of thoroughbred racing performance in Spain

MD Chico

INRA, CRJ, Station Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas, France

Summary — Variance components of Spanish thoroughbred racing performance are estimated by REML using 2 models. The performance characteristics studied are speed, earnings, and rank for each race. The first model considered (model 1) includes the following fixed effects: jockey; sex; weight; age; distance; race course; and race year. A single race effect is substituted for the last 3 effects of model 1 in the second model 2. Heritability and repeatability estimates are close to zero for speed while they range from 0.07 to 0.17 and from 0.19 to 0.26 respectively for rank and earnings.

genetic analysis / performance / thoroughbred / heritability / repeatability

Résumé — Analyses génétiques de la performance des pur-sang en Espagne. *Les composants de la variance de la performance en course des pur-sang anglais sont estimés par REML suivant 2 modèles. Les caractéristiques de performance étudiées sont la vitesse, les gains et l'ordre d'arrivée pour chaque course. Le premier modèle (1) utilise les effets fixes suivants : jockey, sexe, poids, âge, distance de course, hippodrome et année de course, alors que dans le deuxième modèle (2) un seul effet de course est substitué aux 3 derniers du premier modèle. Les estimations d'héritabilité et de répétabilité sont proches de 0 pour la vitesse alors qu'elles s'étalent entre 0,07 et 0,17 et entre 0,19 et 0,26 respectivement pour les gains et le rang.*

analyses génétiques / performance / pur-sang / héritabilité / répétabilité

INTRODUCTION

Traditionally, a horse's breeding value is estimated through a subjective combination of race performances such as ranking, earnings, and racing time.

The first step in any genetic improvement scheme is to estimate parameters (heri-tability, repeatability, and genetic and phenotypic correlation) of the studied performances. In order to calculate these parameters, we need to estimate the variance of the random factors included in a chosen model. The values obtained determine the choice of an optimal breeding program for genetic improvement.

The most used methods for estimating variance components include Henderson's methods I, II and III (Henderson, 1953), MIVQUE (Rao, 1971), maximum likelihood (Hartley and Rao, 1967), and restricted maximum likelihood (Patterson and Thompson, 1971). Brief descriptions of these methods and their properties are given by Schaeffer (1983), Henderson (1984), and Van Raden (1986). Restricted maximum likelihood (REML) has the most suitable statistical properties for most animal breeding studies, but until recently its use has been severely limited by its computational requirements.

The thoroughbred population in Spain has some interesting characteristics from a genetic standpoint. First, performance data are detailed and complete (cross-data for each race and each horse). Second, its environmental parameters are well known and easily controlled. Third, genetic evaluations have been regularly undertaken using a mixed model since 1986. Despite its small size, this population allows reliable comparisons of various models.

The objective of this study is to compare the variance component estimations obtained with 2 different models. The traits studied are racing speed, earnings, and rank. It also compares the different alternatives to genetic evaluation in the case of thoroughbreds and attempts to clarify certain aspects of the Hill–Cunningham polemic in *Nature* (Gaffney and Cunningham, 1988; Hill, 1988) in complement to Chico (1990).

MATERIALS AND METHODS

Materials

The racing records used in this study are the same ones used in a precedent study (Chico, 1990). Only Grand-Prix races from 1980 to 1989 are used. Grand-Prix are high-prized adult horse races where all participants carry approximately the same weight. Genealogical data is limited to 3 generations, and missing or irrelevant geneaogical information is discarded; the relationship matrix totals 2 705 animals. The number of observations (number of races by number of participants) amounts to 2 596 pertaining to 776 race horses, which were progeny of 315 sires and 597 mares.

Method

The fixed effects included in the mixed model are (in parentheses: number of levels or classes): race (260); jockey (27); sex (2); weight (5); age (4); distance (14); race course (3); and race year (10). Jockeys that have run less than 14 times are grouped in a single class. The 3 variables studied are speed, earnings, and rank in each race. Speed is estimated based on the distance of the race, the winner's run-time, and the distance between horses at arrival (Chico, 1988).

Horses 'out of the money' (those that rank below the fifth position) are given artificial earnings equal to half the earnings of the one that ranked immediately ahead of it. Normal distribution of the earnings is then obtained by applying a logarithmic transformation (Langlois, 1975; Chico, 1990).

Ranking is not a normal variable either. To get closer to this requirement, ranking is replaced by a normal score which is the expected value of the rank of a standardized normal variable (Pearson and Hartley, 1972).

Two different methods for the estimation of variance components are used: an EM-type algorithm of Dempster *et al* (1977) to estimate repeatability and a DF-algorithm to estimate the additive genetic variance and the permanent environment variance separately. Two models for fixed effects are applied:

Model 1 includes the effects of jockey, sex, weight, age, distance, race course, and race year.

Model 2 uses a single race effect to replace distance, race course, and year effects.

This single race effect represents the interaction of race distance, race course and year, as well as the effect of the environment in the considered race. The second model should yield smaller estimates of residual variance because it better accounts for race environment.

The REML method developed by Patterson and Thompson (1971) associated with the EM-type algorithm of Dempster *et al* (1977) is

used to estimate performance repeatability, only considering the horse effect.

The analytical model is defined as follows:

$$y = Xb + Zu + e$$

where:

y is the vector of observations for speed, earning, or rank;

b is the vector of unknown fixed effects;

u is the vector of random horses effects representing both genetic and permanent environmental effects associated with a horse;

e is the vector of random residual effects;

X is the known incidence matrix for fixed effects; and

Z is the known incidence matrix for random effects.

Henderson's mixed-model equations (MME) were used to predict animals' value.

$$\begin{vmatrix} X'X & X'Z \\ Z'X & Z'Z + a_1I \end{vmatrix} \begin{vmatrix} b \\ u \end{vmatrix} = \begin{vmatrix} X'Y \\ Z'Y \end{vmatrix}$$

where if $u = a + p$:

$a_1 = (1 - r)/r = s_e^2/(s_a^2 + s_p^2)$;

$r = (s_a^2 + s_p^2)/s_y^2 = $ repeatability; $s_y^2 = s_a^2 + s_e^2 + s_p^2$;

$a = $ genetic value;

$p = $ permanent effect;

$s_a^2 = $ variance of genetic value;

s_e^2: variance of error;

$s_p^2 = $ variance of permanent effect; and

$s_y^2 = $ total variance.

REML method with a derivative-free algorithm (DFREML)

The use of a derivative-free REML algorithm was suggested by Grasser *et al* (1987) to estimate variance components for a single trait analyzed with an animal model. The basic principle is the maximization of the log likelikood of a series of contrasts of the data, without calculating the lst and 2nd derivatives of this log likelikood.

The linear model of analysis, an individual animal model (IAM), is used to obtain an additive genetic value not only for animals with records but also for any parent included in the relationship matrix. It also gives a BLU estimate for the fixed effect included in the model (Meyer, 1988). The programs of Meyer (1988) are used for this algorithm.

The analytical model is now:

$$y = Xb + Z_u u + Z_p p + e$$

where:

y, **X**, **b** and **e** are as before;

$\mathbf{Z_u}$ and $\mathbf{Z_p}$ are known incidences matrices;

u is the vector of random additive genetic value effects; and

p is the vector of random permanent environment effect.

It was assumed that the random vectors have zero mean and that

$$\text{var} \begin{vmatrix} Y \\ u \\ p \\ e \end{vmatrix} = \begin{vmatrix} V & 0 & 0 & 0 \\ 0 & G & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & R \end{vmatrix}$$

where:

$$V(Y) = Z\,G\,Z' + Z\,P\,Z' + R;$$

G and **R** are non-singular matrices;

$G = \text{var}(u) = A \cdot s_u^2$;

A is the full relationship matrix with all animals from the pedigree;

$P = Is_p^2$; s_p^2 is the permanent environment variance;

$R = Is_e^2$; and s_e^2 is the error variance.

The MME (Henderson, 1973) are:

$$\begin{vmatrix} X'X & X'Z & X'Z \\ & Z'Z + a_1A^{-1} & Z'Z \\ & & Z'Z + a_2I \end{vmatrix} \begin{vmatrix} b \\ u \\ p \end{vmatrix} = \begin{vmatrix} X'Y \\ Z'Y \\ Z'Y \end{vmatrix}$$

where:

$a_1 = (1 - r) / h^2$; $a_2 = (1 - r) / (r - h^2)$;

$h^2 = s_a^2 / s_y^2 = $ heritability.

RESULTS AND ANALYSIS

Table I shows the repeatability estimated by REML-EM and DFREML methods using

Table I. Estimates of heritability (h^2) and repeatability (r) using REML–EM and REML–DF methods with models 1 and 2.

Trait	Model	REML–EM	REML–DF	
		r	h^2	r
Speed	1	0.00	0.01	0.01
	2	0.02	0.00	0.01
Log (earnings)	1	0.23	0.09	0.23
	2	0.25	0.08	0.25
Rank	1	0.21	0.07	0.21
	2	0.25	0.10	0.26

the 2 models proposed. The iterative REML method always converges for all variables in both algorithms.

This table clearly shows that repeatability estimates for all 3 traits are slightly higher in model 2 than in model 1. This indicates a better data correction due to considering race effects; in addition to race course, distance and year, it includes racing characteristics. The repeatability of speed is very low in all methods and models. The repeatability of ranking and logarithm of earnings reach the same levels in all 3 methods with a small advantage in the second model (similar results were obtained by Chico et al (1990)).

The difference between heritability and repeatability (ignoring dominance effects and epistasis) is the proportion of total variance attributable to permanent environmental effects. Permanent environmental effects on racing performance may include such factors as owner, trainer, early nutrition, and injury.

The speed heritability estimates obtained are close to zero. Model 1 values are comparable to those obtained by Chico (1990). This is also confirmed by the work of Langlois (1975, 1978) who obtains heritability values below 0.10 as Buttram et al (1988).

Heritability estimates are in the order of magnitude of 0.1 for the logarithm of earnings and slightly larger for rank, whatever model is used. These results confirm those previously obtained by Chico (1988, 1990) with smaller values and smaller standard error.

The estimates of residual variances for each trait by model 2 are smaller than those from model 1 indicating a better ability of model 2 to account for variation in these racing traits. The earnings and rank heritability estimates valuation in this study are also equivalent to those obtained by Field and Cunningham (1976), Langlois and Chico (1989), Katona and Distl (1988), Klemestsdal (1990), and Arnason et al (1988).

This analysis allows us to conclude that speed is not the proper criterion for selection program for Spanish thoroughbred population. Heritability and repeatability are close to zero. However, earnings and rank are similar traits because estimates of the genetic correlation between these 2 traits are close to 1 in both models and their heritability and repeatability values are similar. Ranking appears to be the most important thoroughbred performance evaluation criterion. This fully justifies Tavernier's proposition (1990, 1991).

CONCLUSION

Thoroughbred race performances are moderately heritable and repeatable in terms of earning and rank and are independent of chronometric measurement. It is therefore not surprising that chronometric performances show little improvement over time (Gaffney and Cunningham, 1988; Hill, 1988). Speed is not a good selection criterion for thoroughbreds in Spain. On the other hand, earning and rank are good criteria for genetic improvement programs thanks to heritability, repeatability as well as genetic and phenotypic correlations.

Model 2 shows a better ability to account for variation in racing traits thanks to smaller estimates of residual variances. Heritability and repeatability are also slightly higher in model 2. Model 2 would be preferable to apply in future work.

REFERENCES

Arnason T, Bendroth M, Philipsson J, Henriksson K, Darenius A (1988) Genetic evaluation of Swedish trotters. In: Proc EAAP-Symp Commission on Horse Production. Helsinky, Finland, June 27–July 1, EAAP Publication, N° 42, 106-130

Buttram ST, Wilson DE, Willham RL (1988) Genetics of racing performance in the American quarter horses. III. Estimation of variance components. J Anim Sci 66, 2808-2816

Chico MD (1988) Influencia de los factores ambientales en los resultados de carreras, aplicación de la metodología de modelos mixtos (modelo animal) en el pura sangre inglés. Doctoral Thesis ETSIA, Universidad politécnica de Madrid, Spain

Chico MD (1990) Estimation par un REML des composantes de la variance des résultats en courses des pur-sang en Espagne. In: Abstracts of 41st Annual Meeting of the EAAP Toulouse, France, July 9-12 1990, Study Commission of Horse Production, 2, 394-395

Chico MD, Alenda R, Jurado JJ, Alonso A, De La Mata I (1990) Cátalogo n° 2 de Valoración Genética del Pura Sangre Inglés en España, Publicado por el Ministerio de Agricultura, Pesca y Alimentación, Servicio de Extensión Agraria

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. JR Soc Stat B 39, 1-38

Field JK, Cunningham EP (1976) A further study of the inheritance of racing performance in thoroughbred horses. J Hered 67, 247-248

Gaffney B, Cunningham EP (1988) Estimation of genetic trend in racing performance in thoroughbred horses. Nature 332, 722-724

Grasser 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, 1363-1370

Hartley HO, Rao JNK (1967) Maximum likelihood estimation for the mixed model analysis of variance model. Biometrika 54, 1363-1370

Henderson CR (1953) Estimation of variance and covariance components. Biometrics 9, 226-252

Henderson CR (1973) Sire evaluation and genetic trend. Proc Anim Bred and Genet Symp in Honor of JL Lush, ASAS/ADSA, Champaign, IL, 10-41

Henderson CR (1984) ANOVA, MIVQUE, REML and algorithms for estimation of variances and covariances, Dept of Anim Sci, Cornell University, Ithaca, NY

Hill WG (1988) Why aren't horses faster? Nature 332

Katona O, Distl O (1988) Sire evaluation in German trotter (standard-bred) population. In: Proc EAAP Symp Commission on Horse Prod, Helsinky, Finland, June 27–July 1, EAAP Publication, N° 42, 55-61

Klemetsdal G (1990) Breeding performance in horses – a review. In: Proc 4th World Congress on Genetics applied to Livestock Production, Edinburgh, Scotland, July 23-27, XVI, 184-193

Langlois B (1975) Analyse statistique et génétique des gains des pur-sang anglais de trois ans dans les courses plates françaises. Ann Génét Sél Anim 7, 387-408

Langlois B (1978) Aspects spécifiques de l'amélioration génétique des chevaux de pur-sang. Féd Europ Zootech (Comm Cheval) 29e Congrès, Stockholm, Sweden

Langlois B, Chico MD (1989) Études françaises sur l'estimation de la valeur génétique des chevaux de pur-sang. In: Abstracts 40th Ann Meeting EAAP, Dublin, Ireland, August 27-31, Study Commission of Horse Production, 2, 331-332

Meyer K (1988) DFREML a set of programs to estimate variance components for animal models with several random effects using a derivative-free algorithm. Smet Sel Evol 21, 317-340

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554

Pearson ES, Hartley HO (1972) Biometrika Tables for Statisticians. Vol 2, 27-35, Cambridge University Press, UK

RAO CR (1971) Minimum variance quadratic unbiased estimation of variance components. J Multivar Anal 1, 455

Schaeffer LR (1983) Notes on linear model theory, best linear unbiased prediction, and variances components estimation. Dept Anim Poult Sci Univ of Guelph, Ontario, Canada

Tavernier A (1990) Estimation of breeding value of jumping horses from their rank. Livest Prod Sci 26, 277-290

Tavernier A (1991) Genetic evaluation of horses based on ranks in competitions. Genet Sel Evol 23, 159-173

Van-Raden PM (1986) Computational strategies for estimation of variance components. Ph D Dissertation, IOWA State Univ, Ammens, USA