THE IMPORTANCE OF DIFFERENT TRAITS IN GENETIC IMPROVEMENT OF TROTTERS

Th. Árnason
IHBC AB, Knubbo, S-740 41 Morgongåva, Sweden.

SUMMARY

The application of various traits in the genetic evaluation of trotters is reviewed. Many racing performance traits exhibit skew distributions. The use of Box and Cox power transformation is illustrated on trotting data. For improved distribution of racing time, the following transformation is proposed: log(racing time - x), where x is a localization parameter. Problems encountered with genetic evaluation of racing performance are addressed.

INTRODUCTION

Harness racing of trotters is a popular sport in many countries. Breeders of trotters are showing a growing interest in the application of modern knowledge and methodology of animal breeding as an aid in improving racing capacity by means of genetic selection. A crucial problem is to establish appropriate criteria for racing performance of trotters. Large files of accumulated racing competition results are often available and this has stimulated interest in the use of such data as a basis for genetic evaluations. Application of various racing variables has been attempted. Most commonly, earnings have been used as a measure of racing success, but other complementary traits, such as racing time, rank at finish, earnings per start, number of starts, etc. have been utilized as criteria for racing success. (For review articles, see: Katona,1985; Tolley et al.,1985.)

BREEDING GOALS

In most general terms it can be stated that the objective of the breeders of trotters is to supply harness sport racing with horses that are capable of providing exciting races and that are attractive as a betting object and as a public entertainment. Speed and the explosive expression of athletic power is certainly important for the popularity of horse racing. Since only successful racing horses earn big money in tough competition with contemporaries, breeders compete with each other in breeding and raising winners. At the same time the racing business is becoming increasingly internationalized and inter-population competition is attracting significant attention. Establishment of breeding organizations has resulted in formal definitions of breeding goals and the establishment of breeding schemes in several countries.

The verbal descriptions of breeding goals for trotters often include terms such as: racing speed, ability to win, temperament, conformation, precocity, endurance, health, and regularity of gait. Surely the true aggregate genotype involves a host of anatomical, physiological, endocrinological and mental components that affect the frequency, the stride length, rhythm and efficiency in the diagonal (lateral for pacers) movements of the legs, at the right moments during races, in response to the behavior of the competitors and the signs from the driver.

Until the complexity of the underlying biological elements is better understood in a remote future, a reasonable composite trait of racing success must be defined. In monetary terms the aggregate genotype could simply be expressed as a strictly monotone increasing function of earnings over a given age span, as the cost of input is practically invariant.

MEASURES OF PERFORMANCE

From the filed competition results available in the chief trotting sport countries, the following measures of racing performance can be extracted:

Earnings: Money earnings have been suggested as a logical measure of racing performance for trotters (Minkema,1975; Langlois,1983). One problem with the use of the earning variables is the extremely skew distributions of phenotypic records. Apart from a cluster of zero earners in the left end of the frequency distribution, logarithmic or root transformations provide reasonable approximations to normal distribution of records (Arnason, et al., 1982, Klemetsdal, 1989). Langlois (1983) showed that log transformations of earnings were optimum, provided that the total purse of races represented a log-linear function of the level of the race. Deviations from this assumption might explain why power (root) transformations seem preferable to the log transformation of earnings in some datasets.

Another significant problem is how to treat earning variables of non-starters, i.e. horses that never enter a race course and therefore do not provide any direct measure of their racing capacity. Obviously the validity of decisions whether to treat non-starters as missing values or to assign them zero values depends on the true association between this preselection criterion and racing performance.

Klemetsdal (1992) has suggested that the inflationary effect on variance in earnings should be eliminated by standardizing data according to time periods before use in genetic evaluations.

Speed: Trotting speed is obviously closely related to success in races, as the winner of a race is the one maintaining the highest average speed throughout the course. Average trotting speed in a race is expressed as the racing time per unit of distance (km or mile). Racing time has frequently been used as a measure of racing performance in trotters (Rönningen, 1975; Katona and Osterkorn, 1977; Ojala and Van Vleck, 1981; Tolley et al., 1983). Availability of racing time records is obviously restricted to horses that have actually competed. Most authors have adopted normal distribution of best racing time records. In reality, however a slight positive skewness of the distribution is apparent. Log transformation of racing time records ($\log_{10}$ (best racing time per km in sec. - 1 min)) gives a somewhat closer fit to the normal distribution in data on Swedish standardbred trotters. The improvement obtained in the present data was far less pronounced though, than found in an earlier study (Arnason et al., 1982). The achievement from the use of log-transformed racing time is probably trivial for simpler genetic evaluation procedures. However, the transformation is more appealing for prediction of genetic changes from selection, and is possibly preferable in animal model applications with deep pedigrees and when available records span a long period. Certainly a linear trend in improvement of racing time is not to be expected, while a diminishing rate of progress seems more plausible. One second's improvement in racing time, from 1.15.0 to 1.14.0 corresponds to 0.0300 units on the log linear scale, while the difference between 1.22.0 and 1.21.0 is worth one-third less after the log transformation (0.0202 units). The asymptotic levelling off of racing times observed in thoroughbreds also contradicts the use of a linear measure of racing speed (Cunningham, 1976; James, 1990). Presumably racing time records should generally be transformed by log(racing time - x), where x is a localization parameter with an optimum value close to (at least one time unit below) the ultimate asymptotic level for speed for that particular gait. The current world record for trotters corresponds to about 75% of the gallop speed of the best thoroughbreds. Apparently a plateau has not yet been reached for trotting speed.

Rank at finish: Several authors have tried to devise linear measures of racing performance in trotters by functions of ranks of several first horses at the finish (e.g. Arnason et al., 1982; Katona, 1985; Ojala, 1987). Transformations may improve the distributional properties of such measures.

Alternatively, Tavernier (1990,1991) has proposed that series of ranks in equine competitions should be dealt with by order statistics. One of the advantages of using order statistics is the avoidance of the undesirable effects of asymmetrical distributions and of outliers. At the same time, distributions of functions of the order statistics are often derivable. Tavernier has indeed derived the posterior density of an underlying variable related to rank which enables Bayesian estimators of genetic parameters as well as genetic values to be computed by iterative methods.

Health, endurance: Even though health and endurance are assumed to be important traits in the breeding of trotters, direct measures are not readily available in filed form. In genetic evaluations of Swedish trotters, the number of starts (races) is included as a variable of minor importance (Arnason and Svendsen, 1991). The idea is to get some measure of endurance, since good health status is a prerequisite for a frequently raced horse. The problem is that the argument cannot be reversed, i.e. few starts do not necessarily indicate injuries or other health problems. Assuredly a few competitions in races of high class (stakes) will wear a horse as much as many races of mediocre competition.

Of course it should not be overlooked that any racing performance trait is an indirect measure of health and soundness.

Start status:

In any trotter population a proportion of horses never enter a race course. In Sweden, this proportion accounts for about 40% of standardbred trotters. How these absent horses should be treated in statistical analysis and genetic evaluation of trotting performance traits is by no means clear. The measure of start status (also called start frequency) is truly a discrete variable, where value 1 represent a started horse and value 0 a non-starter. The reasons why a certain proportion of horses do not race probably differ between populations and even between time periods. Certainly it is of fundamental importance to know the true association between this preselection criterion and racing performance. Klemetsdal (1992) has shown that such preselection process can result in substantial underestimation of genetic trend, thus giving biased estimates of breeding values.

Other traits:

Disqualifications, precocity, flyer and stayer ability are traits that can be extracted from the filed records of racing performance available from the trotting sport organizations. Interest in the application of these variables has not been very keen, as yet.

Samples of radiological examinations of the legs of trotters are sometimes used as indicators of defects or lesions in limbs. In many countries radiological inspection is already used in connection with approval of stallions. Radiological assessment and muscle biopsy as well as other biochemical and physiological markers for racing performance and health will probably be applied in the future.

Data transformation

Concern over the distributional properties of racing variables is bound up with what the observations are intended for. If the purpose is a simple ranking of animals according to performance, there is no great cause for worry. If, however, the intention is to use the variables in BLUP animal model methods, then non-linearity and heterogeneity of variances might introduce serious bias and impair the efficiency of selection. The power transformation technique of Box and Cox (1964) provides means for seeking the optimum transformation of variables. The use of this technique on trotting data is illustrated below.

The data comprise accumulated 3-5-year-old racing records of 17,545 Swedish standardbred trotters born 1981-88. The following linear model was used to describe the variation in five racing performance traits:

$$y_{ijk} = f_i + s_j + e_{ijk} \qquad (1)$$

where y_{ijk} is the transformed observation; f_i is the fixed effect of i^{th} sex/birth-year subclass; s_j is the random effect of sire (j=1,...,449); and e_{ijk} is the random error. Meyer's (1987) REML-PK programs were used to estimate sums of squares and heritabilities from the intra-class correlation of half-sibs.

A normalized transformed variate, $z^{(\lambda)}$, was formed,

$$z^{(\lambda)} = (x^{\lambda} - 1)/(\lambda \dot{g}_{(x)}{}^{\lambda-1}) \qquad (2)$$

where x is the original untransformed observation and $\dot{g}_{(x)}$ is the geometric mean of the original observations. Since the original distributions were positively skewed, only positive power values λ ranging from 0.1 to 1.0 were employed.

The empirical maximum log-likelihood L_{max} (λ) was determined as:

$$L_{max}(\lambda) = -(n/2)\ln(S_e(\lambda)/n) \qquad (3)$$

where n is the number of observations, and $S_e(\lambda)$ is the residual sum of squares from analysis of the normalized variable (2) using the linear model (1). The resulting L_{max} values, together with skewness, kurtosis and heritabilities for different λ values indicated that the optimum transformations for this data are as follows: for number of starts the 5^{th} root; for percentage placed 1^{st} to 3^{rd} the optimum λ value is about 0.8; for earnings per start, a square root transformation seems almost appropriate; and for accumulated earnings, the 4^{th} root transformation seems valid. Log transformation of best racing time is apparently preferable to power transformation.

EVALUATION OF BREEDING VALUES FOR RACING PERFORMANCE

Studies on heritability of the racing performance traits based on earning, speed, or rank functions, have with few exceptions yielded estimates in the range 0.2 to 0.4 (for a review, see Tolley et al., 1985; some more recent results are reported in EAAP Publication no. 42, 1989). Heritability of racing performance up to the age of 5 years has generally been estimated higher than the heritability of life-time performance. Heritability of number of starts has been estimated to 0.1 and heritability estimates of start status have been obtained in the order of 0.1 to 0.2, by linear methods. (Klemetsdal, 1989; Arnason et al., 1989).

Racing measures based on time, earnings and ranks have been found to be mutually highly genetically correlated, both within and between age classes. The phenotypic correlations are considerably lower indicating quite distinct environmental influence on these traits. Genetic correlation between number of starts and the other racing variables seems fairly weak. No estimates on the association between start status and the other traits were found in literature. Recent estimates in Scandinavian trotters by the method of Janss and Foulley (1993) indicate close genetic correlations in the current data (Arnason and Klemetsdal, unpublished results). The nature and the consequences of these findings require further investigation.

The observed genetic variability in racing performance has encouraged the implementation of mixed-model methodology (BLUP) for genetic evaluation of trotters. The acceptance of animal models for genetic evaluation of horses is predominant. The mixed linear models used to describe data on racing performance of trotters include fixed systematic environmental effects that are known to affect the variation in the traits and which are readily registered in the files. Examples of such effects are sex, age, year, birth year, country of birth, season of birth, etc. for accumulated racing results. Models including the results of individual races may introduce such additional fixed effects as racetrack, racing distance, starting method, track condition, etc. Tavernier (1989) introduced random maternal-herd effects in the model as applied to French trotters.

The inclusion of environmental impact of trainers and drivers on trotting performance are encumbered with certain technical difficulties. In the first place the horses frequently change trainers and drivers, which causes difficulties when evaluations are based on accumulated racing results. Secondly, a more general problem is how to treat these effects statistically. Logically the effects of trainers and drivers are random by their nature. The main problem is that zero covariance between these effects and breeding values of the horses is inconceivable. Treating these effects as fixed is not correct but may be a reasonable alternative when the degree of connectedness is satisfactory, and may yield 'unbiased' but not 'best' (minimum mean squared error) predictions of breeding values. A thorough investigation of this problem is needed.

Proper modelling of competition traits is not always easy. Since the genetic level of its competitors acts as an environmental effect on an individual horse, great care should be taken to model the data in such a way that covariance between vectors of breeding values and residual effects (or other random effects) is eliminated. For instance the inclusion of birth-year as a fixed effect in the statistical model instead of age-year effects could be a way to reduce bias when the main competition is among contemporaries (Arnason and Svendsen, 1991). Tolley et al. (1989) illustrated how a system of interdependent equations (borrowed from econometrics) may be used to eliminate covariance between genetic values for speed and the error term in a model which includes pace and purse of individual races. An extension of this approach into mixed-model framework might be worth considering.

A general question can be raised about the validity of the assumption of normally distributed additive genetic values of speed, or indeed other measures of racing performance in horse populations. It is a well established fact that asymmetric mutational effects and non-linear heritabilities are common for reproductive fitness traits(Mackay, 1990; Frankham , 1990). Speed has undoubtedly been a very strong component of fitness in equine evolutionary history. Non-linear heritabilities of racing speed at the gallop and possibly other gaits, might therefore be expected. The effects of such phenomena on genetic trend and ranking by the BLUP method are not clear. Presumably, ranking will be most erroneous for individuals deviating widely from the mean. As the main interest lies in correct ranking of the very best animals for selection purposes, this problem may not be trivial.

GENETIC CHANGES

An intensive selection of stallions on the basis of phenotypic racing performance has been practiced in many trotter populations for quite a long time. During the last decade, increasingly sophisticated selection indices have been taking over as selection criteria for racing performance. Phenotypic trend in racing time within the Swedish standardbred trotter population is illustrated in Figure 1. The improvement observed is of course attributed to both genetic and environmental changes. Environmental changes include enhanced training methods, as well as improved tracks, harness and sulkies. A large but sometimes overlooked part of the 'environmental' improvement is however advanced genetic capacity of the competitors, which boosts the racing speed. According to the BLUP results, where the model hopefully allows the correct separation of environmental and genetic changes, genetic progress accounts for 60% of the phenotypic trend. The estimated annual genetic progress in racing performance traits of Swedish standardbred trotters corresponds to about 5% of the phenotypic standard deviation. A corresponding figure of 3.6% in French trotters was reported by Tavernier (1989). Minkema (1981) estimated the genetic progress in Dutch trotters to exceed 5% of the phenotypic standard deviation.

REFERENCES

ARNASON, TH., DARENIUS, A. and PHILIPSSON, J. (1982) Livest. Prod. Sci. 8:557-565.
ARNASON, TH., BENDROTH, M., PHILIPSSON, J. and DARENIUS, A. (1989) EAAP Publication No.42.,106-130.
ARNASON, TH. and SVENDSEN, M. (1991) 42nd Annual Meeting of EAAP, Berlin.
BOX, G.E.P. and COX, D.R. (1964) J.R.Statist.Soc.B26:211-243.
CUNNINGHAM, E.P. (1976) In: Proc. Int. Symp. on Genetics and Horse Breeding. Royal Dublin Society, 17-18 Sept. 1975.
FRANKHAM, R. (1990) Proc. 4th World Congr. Genet. Appl. Livestock Prod. 13:238-239.
JAMES, J.W. (1990) Proc. 4th World Congr. Genet. Appl. Livestock Prod. 16:221-224.

JANSS, L.L.G. and FOULLEY, J.L. (1993) Livest. Prod. Sci. 33:183-198.
KATONA, Ö and OSTERKORN (1977) Züchtungskunde 49:185-192.
KATONA, Ö. (1985) 36th Annual Meeting of EAAP, Halkidiki.
KLEMETSDAL, G. (1989) EAAP Publication No.42.,95-105.
KLEMETSDAL, G. (1990) Proc. 4th World Congr. Genet. Appl. Livestock Prod. 16:184-193.
KLEMETSDAL, G. (1992) Acta Agric. Scand.,Sect. A. Animal Sci.42:226-231.
LANGLOIS, B. (1983) 38th Annual Meeting of EAAP, Madrid.
MACKAY, T.F.C. (1990) Proc. 4th World Congr. Genet. Appl. Livestock Prod. 13:219-228.
MEYER, K. (1987) Stencil from a workshop, Orange, NSW, Australia, March 2-6, 1987.
MINKEMA, D. (1975) Ann. Génét. Sél. anim. 7:99-121.
MINKEMA, D. (1981) Ann. Génét. Sél. anim. 13:245-254.
OJALA, M.J. (1987) J. Anim. Sci. 64:117-125.
OJALA, M.J. and VAN VLECK, L.D. (1981) J. Anim. Sci. 53:611-619.
RÖNNINGEN, K. (1975) Z. Tierzüchtg. Züchtgsbiol. 92:164-175.
TAVERNIER, A. (1989) EAAP Publication No.42.,41-54.
TAVERNIER, A. (1990) Proc. 4th World Congr. Genet. Appl. Livestock Prod. 16:194-197.
TAVERNIER, A. (1991) Genet. Sel. Evol. 23:159-173.
TOLLEY, E.A.; NOTTER, D.R. and MARLOWE, T.J. (1983) J. Anim. Sci. 56:1294-1305.
TOLLEY, E.A.; NOTTER, D.R. and MARLOWE, T.J. (1985) Animal Breeding Abstracts, 53:163-185.
TOLLEY, E.A.; NOTTER, D.R. and MARLOWE, T.J. (1989) EAAP Publication No.42.,131-141.

Figure 1.

Phen. trend in racing time