U. Coskun et al. Large Animal Review 2020; 26: 349-352 349

Comparison of repeatable and random regression models for genetic parameter estimation on Thoroughbreds

UMIT COSKUN[1], HASAN ÖNDER[1]* AND SAMET HASAN ABACI[1]

[1]Ondokuz Mayis University, Faculty of Agriculture, Department of Animal Science, 55139, Samsun, Turkey

SUMMARY

Introduction - The combination of multiple factors (track, year, ages, etc.) is effective in achieving the maximum level of race performance in Thoroughbreds.

Aim - The aim of this study is to estimate and compare genetic parameters on the number of race success characteristics in Thoroughbreds with random regression models (RRM) and repeatable animal models (RAM) with a different number of repetitions. It was also aimed to investigate which number of observation points would be sufficient for genetic parameter estimation for Thoroughbred.

Materials and methods - As data, 111312 test day race completion time (sec) records of 13625 Thoroughbreds raced taken from the Jockey Club of Turkey between 2005 and 2016 were used. Competition performances were compared with different measurements using the same repeatability model. Variance components of Thoroughbreds were obtained by using RRM and RAM using DFREML and WOMBAT package, respectively.

Results and discussion - When AIC and BIC values were examined, it was observed that the values in RRM were lower than RAM method for ten races. According to Akaike Weights results, while the fifth race shows 35.56% better fit than the fourth race in the model. The AW values of other number of races showed less superiority; thus, 5th and 6th races can be preferred over its competitors in terms of Kullback-Leibler discrepancy.

Conclusion - Our results showed that number of race of five were sufficient to estimate genetic parameters for Thoroughbred horse. Also, RRM method can be preferred compared to RAM method.

KEY WORDS

Heritability, Random regression, REML, Repeatable animal models, Thoroughbreds.

INTRODUCTION

The contributions of horse races are made in every community and culture from the 17th century to the present day and have been improved by selection of horseracing performance for about 300 years. While horses are bought and sold at very high prices, high-performing horses in the races are making significant gains to their owners[1, 2].

The Thoroughbred horse breed was established in England in the early 1700s based on crosses between stallions of Arabian origin and indigenous mares[3]. The Thoroughbred horse breed is used in a variety of sports such as running, jumping and hunting. In equestrian sports, Thoroughbred horses have a very important place for flat racing for speed at middle - distances between 1400 m and 2400 m all around the world[4]. Thoroughbred racing industry and its professionals are now even more capable of selecting the most appropriate horses for breeding and racing[5]. The statistical power of these studies has also been restricted by the statistical methods and computational resources, making it difficult to include all necessary environmental factors in the analyses and potentially biasing estimates[6]. There are many types of methods for predicting genetic parameters, such as the animal model or the sire model restricted maximum likelihood (REML), best linear unbiased prediction (BLUP), random regression (RR), and Bayesian methods[1, 2, 7-11]. To use in dairy cattle breeding for the analysis of test day production records random regression models (RRM) has been introduced by Henderson (1982) and Laird and Ware (1982) as mentioned by Schaeffer[11]. During the last 20 years, longitudinal data analyzed by RRM are used extensively on many economical traits at different animal breeding scenarios[1, 10, 12-14]. Due to the improved modelling of variances and genetic parameters, RRM is a good choice for modelling the traits which are measured repeatedly for each individual, but change gradually and continually in time[8, 15].

This study aims to estimate and compare genetic parameters on the number of race success Thoroughbreds with random regression models (RRM) using L(3,3) legendre polynomials and repeatable animal model with restricted maximum likelihood (REML) method. Gomez et al[8] used the L(2,2) model, while Buxadera and da Mota[12] used the L(3,2) model. In our study, we used the L(3,3) model with fixed residual variance, which is used extensively in other species[10,13]. It was also aimed to investigate which number of observation points would be sufficient for genetic parameter estimation for Thoroughbred.

Corresponding Author:
Hasan Onder (hasanonder@gmail.com).

MATERIALS AND METHODS

Materials

As data, 111312 test day race completion time (sec) records of 13625 Thoroughbreds taken from the Jockey Club of Turkey between 2005 and 2016 were used after exclusion. Only animals with two to ten repeated measurements per animal were used. Some animals whose parents were not reached and the animals that were recorded for less than two races were excluded from the study (number of excluded records were 3361). The average percentage of known parental grandparent was 21.4% and maternal grandparent was 43.7%. The average inbreeding was calculated as 0.07. The race tracks type, the race year and the age of the horse were fixed, and the race tracks distance were taken as covariates. Sample size for the number of races were 13625, 12880, 12126, 11383, 10674, 10051, 9510, 8973 and 8464 from two races to ten races, respectively. In the model where variance elements are estimated included the fixed effects; race tracks (3 types), the race year (12 years) and the age (9 different ages) of the horses and the race tracks distance (16 different distances from 800m to 2400m with increasing 100m) were taken as covariates.

Methods

Repeatable animal model approach using sire and dam pedigree information was used. The following model was used to apply the RRM[1].

$$Y_{ijkl} = RT_i + RY_j + HA_k + \sum_{m=1}^{K_B} \beta_m \, x_{(m)}(t_{ij}) + \sum_{m=1}^{K_A} a_{jm} \, \varphi_m(t_{ij}) + \sum_{m=1}^{K_P} P_{jm} \, \varphi_m(t_{ij}) + e_{ijkl}$$

here; Y_{ijkl}: observation value of the horse l at race track ith at race year of jth and age of kth, RTi: ith race tracks (sand, grass, artificial grass); RY_j: jth race year (2005-2016); HA_k: kth horse age (2-18); β_m: mth fixed regression coefficients for horse j; t_{ij}: ith test day of the horse j; $x_{(m)}\ (t_{ij})$: mth covariates (race tracks distance: from 800 to 2400 (with 100m increase); α_{jm}: mth additive genetic random regression coefficients for horse j; P_{jm}: mth permanent environmental random regression coefficients for horse j; φ_m: mth polynomial evaluated for the race t_{ij}; K_B, K_A and K_P are the order of fitted fixed, random additive and random permanent regression coefficients; e_{ijkl}: random residual effect for y_{ijkl}.

In RRM, the (co) variances of the random regression coefficients and the heritability values (additive genetic effect / phenotypic effect) were estimated by the REML method. The estimates were obtained using the AI-REML algorithm, thus avoiding the problems of possible local maximum estimates "without derivative". Third-order Legendre polynomials have been used to define the structure of (co)variance between observations of the same individual.

The individual repeatable animal model used to estimate the components of variance and heritability are presented below:

$$Y_{ijkl} = \mu + RT_i + RY_j + HA_k + b\left(X_{ijkl} - \bar{X}\right) + a_l + P_l + e_{ijkl}$$

here; Y_{ijkl}: observation value of the horse l at race track ith at race year of jth and age of kth, μ: population mean, RT_i: the effect of ith race track, RY_j: the effect of jth race year, HA_k: the effect of kth horse age, b: scalar coefficient for race completion time, X_{ijkl}: race time for animal l in ijtth sub group, $\bar{X}$ mean of the race completion time for the population, a_l: is the random animal effect of Thoroughbreds l.,: is the random permanent environmental effect of Thoroughbreds l. and e_{ijkl}: random error term $e_{ijkl} \sim N(0, \sigma_e^2)$.

In RAM, heritability (h^2) was calculated from genetic parameters (additive genetic effect / phenotypic effect) obtained by REML method.

To compare the models, Akaike information criterion (AIC), Bayesian information criterion (BIC) and Log likelihood values were used[12]. The DXMMR option running under the DFREML program for the RRM and the WOMBAT package program for the repeatable animal models were used.

RESULTS

The values of the variance components obtained by the RRM and RAM methods are given in Table 1. According to the findings obtained with the RAM method, the additive genetic effect value varied between 3.651 and 4.066, while it varied between 4.014 and 5.165 with the RRM method. When the permanent environmental effect values are examined, it varied between 2.310 and 2.899 according to the RAM methods, while it varied between 2.209 and 7.045 in the RRM method. When the methods were examined according to phenotypic effects, values between 13.938 and 17.267 were observed in RAM methods, while values were found between 14.776 and 20.015 in RRM method.

DISCUSSION

Estimated heritability values were similar with the results of Buxadera and da Mota[12], but higher than the estimates of da Gama, Borquis[7]. Orhan and Kaygisiz[16] found the heritability of race completion time as 0.05 which was the lowest estimates of the trait, this may have occurred from their sample size which had 3184 races. Faria, Maiorano[14] found the heritability of race completion time from 0.45 to 0.56, which is higher than the results of this study. This difference may have occurred as a result of the race track distances they used. The highest heritability was observed in 5th and 6th races which shows that the 5th and 6th races were enough to comment on the horse's racing performance.

When the AIC and BIC values were examined in Table 4, it was observed that the values of the RRM method were lower than those of the RAM method for 10 race. Therefore, RRM method can be preferred instead of RAM method. Many researchers recommended the use of RRM[8, 11, 12, 17-21].

To determine the adequate number of race for genetic parameter estimations, value of the heritability and Akaike Weights (AW), in the AIC sense that minimizes the Kullback-Leibler discrepancy, Wagenmakers and Farrell[22] interpreted together. According to AW results, while the fifth race shows 35.56% better fit than the fourth race in the model, the sixth race shows 23.64% better fit than the fifth model. The AW values of other number of races showed less superiority; thus, 5th and 6th races can be preferred over its competitors in terms of Kullback-Leibler discrepancy. And also heritability values were found as higher at these number of races.

CONCLUSIONS

Within the conditions of this research, it became evident that

Table 1 - Values of variance components obtained by RRM and RAM methods.

Methods	Number of races	Additive Genetic Effect	Permanent Environmental Effect	Residual Effect	Phenotypic Effect
RAM	2	3.651 ± 0.253	2.617 ± 0.201	7.669 ± 0.093	13.938 ± 0.149
	3	3.991 ± 0.240	2.310 ± 0.178	8.387 ± 0.073	14.688 ± 0.140
	4	3.961 ± 0.232	2.485 ± 0.170	9.045 ± 0.065	15.491 ± 0.137
	5	3.969 ± 0.228	2.592 ± 0.165	9.236 ± 0.058	15.797 ± 0.133
	6	4.066 ± 0.227	2.605 ± 0.163	9.525 ± 0.055	16.197 ± 0.133
	7	4.059 ± 0.226	2.684 ± 0.161	9.813 ± 0.052	16.556 ± 0.132
	8	4.015 ± 0.224	2.754 ± 0.160	10.066 ± 0.050	16.835 ± 0.130
	9	4.007 ± 0.224	2.837 ± 0.160	10.246 ± 0.049	17.091 ± 0.130
	10	4.016 ± 0.224	2.899 ± 0.160	10.351 ± 0.047	17.267 ± 0.131
RRM	2	4.014 ± 0.234	2.209 ± 0.164	8.543 ± 0.044	14.766 ± 0.134
	3	4.449 ± 0.231	2.645 ± 0.163	8.543 ± 0.044	15.637 ± 0.136
	4	4.799 ± 0.261	3.231 ± 0.187	8.543 ± 0.044	16.573 ± 0.155
	5	5.042 ± 0.300	3.715 ± 0.213	8.543 ± 0.044	17.300 ± 0.177
	6	5.165 ± 0.334	4.006 ± 0.233	8.543 ± 0.044	17.714 ± 0.193
	7	5.161 ± 0.361	4.174 ± 0.247	8.543 ± 0.044	17.878 ± 0.205
	8	5.031 ± 0.383	4.449 ± 0.263	8.543 ± 0.044	18.023 ± 0.218
	9	4.782 ± 0.409	5.222 ± 0.299	8.543 ± 0.044	18.547 ± 0.243
	10	4.427 ± 0.454	7.045 ± 0.379	8.543 ± 0.044	20.015 ± 0.303

Heritability estimates for RAM and RRM on the number of races were given in Table 2. The values of heritability estimates were moderate and close to each other, RRM estimates were higher when compared to RAM estimates. Additive genetic correlations were nearly two times higher than phenotypic correlations (Table 3) estimated by RRM among the number of race. This finding was an evidence that genetic factors have greater influence than the environmental factors.

Table 2 - Heritability estimates for RAM and RRM on number of races.

Number of races	RAM		RRM	
	Heritability	Permanent Environmental Effect Ratio	Heritability	Permanent Environmental Effect Ratio
2	0.262 ± 0.017	0.188 ± 0.015	0.284 ± 0.014	0.150 ± 0.011
3	0.272 ± 0.015	0.157 ± 0.013	0.289 ± 0.013	0.169 ± 0.011
4	0.256 ± 0.014	0.160 ± 0.011	0.291 ± 0.014	0.195 ± 0.012
5	0.251 ± 0.013	0.164 ± 0.011	0.291 ± 0.015	0.215 ± 0.013
6	0.251 ± 0.013	0.161 ± 0.010	0.288 ± 0.016	0.226 ± 0.013
7	0.245 ± 0.012	0.162 ± 0.010	0.279 ± 0.017	0.233 ± 0.014
8	0.238 ± 0.012	0.164 ± 0.010	0.257 ± 0.017	0.247 ± 0.015
9	0.234 ± 0.012	0.166 ± 0.010	0.221 ± 0.018	0.282 ± 0.016
10	0.233 ± 0.012	0.168 ± 0.010	0.271 ± 0.019	0.352 ± 0.018

Heritability values estimated by RRM had increasing trend until the number of six races and then it tended to decrease; however, the heritability values estimated by RAM had a decreasing trend after the second race.

Table 3 - Additive genetic (below diagonal) and phenotypic (above diagonal) correlations estimated by RRM among the number of race.

	2	3	4	5	6	7	8	9	10
2		0.428	0.424	0.414	0.400	0.379	0.350	0.309	0.254
3	0.998		0.464	0.463	0.454	0.434	0.403	0.354	0.288
4	0.992	0.998		0.492	0.488	0.472	0.442	0.392	0.321
5	0.985	0.994	0.999		0.509	0.499	0.473	0.427	0.358
6	0.975	0.988	0.995	0.999		0.515	0.498	0.460	0.400
7	0.964	0.979	0.989	0.995	0.999		0.516	0.493	0.446
8	0.950	0.968	0.981	0.990	0.995	0.999		0.521	0.495
9	0.933	0.954	0.970	0.981	0.989	0.995	0.999		0.540
10	0.911	0.936	0.954	0.968	0.979	0.987	0.994	0.998	

While the estimated additive genetic correlations were found to be high, phenotypic correlations were found to be too low. The maximum phenotypic correlation was obtained as 0.54, show medium relation between the number of races 9 and 10. Results of the comparison criteria for RRM and RAM methods are given in Table 4. Considering the data of 10 races at the same time in both methods, AIC and BIC values obtained from the RRM method were found to be lower than the RAM method. However, considering the smaller number of races, the AIC and BIC values of the RAM methods were found to be lower.

Table 4 - Comparison criteria for RRM and RAM methods.

Method	Number of parameters	Number of records	Log Likelihood	AIC	BIC
RRM	13	111312	-192780.0	385586.0	385711.1
RAM-2	3	27250	-47472.9	94951.7	94976.4
RAM-3	3	40131	-70004.6	140015.2	140041.0
RAM-4	3	52257	-91902.8	183811.6	183838.2
RAM-5	3	63640	-111711.3	223428.5	223455.7
RAM-6	3	74314	-130719.1	261444.2	261471.8
RAM-7	3	84365	-148852.5	297710.9	297738.9
RAM-8	3	93875	-166077.4	332160.8	332189.1
RAM-9	3	102848	-182250.6	364507.2	364535.8
RAM-10	3	111312	-197283.7	394573.5	394602.3

the heritability estimates were higher and more reliable estimated by RRM than RAM method. Use of RRM brings an opportunity to successfully select stallions which could be selected as sires for the next generation because the random regression procedure provides further information for the selection process. Thus, the use of random regression as a tool in the evaluation of race horse performance can be highly recommended. Also, our results showed that 5 races were sufficient to estimate genetic parameters for Thoroughbred horse because the sample size was high enough to state that expression.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

ACKNOWLEDGEMENTS

The authors wish to thank the Jockey Club of Turkey for assistance. This study was a short summary of first author's MSc thesis.

References

1. Abaci S. H, Coskun U., Onder H. (2019). Comparison of Some random regression models for racing performances of the British racing horses in Turkey. BSJ Agri., 2 (1):6-9.
2. Tozaki T., Miyake T., Kikuchi M., Kakoi H., Hirota K., Kusano K., Ishikawa Y., Nomura M., Kushiro A., Nagata S. (2019). Heritability estimates of fractures in Japanese Thoroughbred racehorses using a non linear model. J. Anim. Breed. Genet. 136 (3): 199-204. http://doi.org/10.1111/jbg.12387
3. Binns M., Boehler D., Bailey E., Lear T., Cardwell J., Lambert, D. (2012). Inbreeding in the Thoroughbred horse. Anim. Genet., 43 (3): 340-342. doi:10.1111/j.1365-2052.2011.02259.x
4. Paksoy Y., Ünal N. (2019). Multivariate analysis of morphometry effect on race performance in Thoroughbred horses. R. Bras. Zootec., 48. e20180030. doi:10.1590/rbz4820180030
5. Velie B., Hamilton N., Wade C. (2015a). Heritability of racing performance in the Australian Thoroughbred racing population. Anim Genet., 46 (1): 23-29. doi:10.1111/age.12234
6. Velie B., Hamilton N., Wade C. (2015b). Performance selection for Thoroughbreds racing in Hong Kong. Equine Vet. J., 47 (1): 43-47. doi:10.1111/evj.12233
7. da Gama M. P. M., Borquis R. R. A., de Araújo Neto F. R., de Oliveira H. N., Fernandes G. M., da Mota, M. D. S. (2016). Genetic parameters for racing performance of Thoroughbred horses using Bayesian linear and Thurstonian models. J. Equine Vet. Sci., 42, 39-43. doi:10.1016/j.jevs.2016.03.02
8. Gómez M., Molina A., Menendez-Buxadera A., Valera M. (2011). Estimation of genetic parameters for the annual earnings at different race distances in young and adult Trotter Horses using a Random Regression Model. Livest. Sci., 137 (1-3): 87-94. doi:10.1016/j.livsci.2010.10.004
9. Gómez M., Valera M., Molina A. (2010). Genetic analysis of racing performance of trotter horses in Spain. Livest. Sci., 127(2-3): 197-204. doi:10.1016/j.livsci.2009.09.012.
10. Onder H., Sen U., Takma C., Ocak S., Abaci S. H. (2015). Genetic parameter estimates for growth traits in Saanen kids. Kafkas Univ. Vet. Fak. Derg., 21 (6): 799-804. doi: 10.9775/kvfd.2015.13407
11. Schaeffer L. (2004). Application of random regression models in animal breeding. Livest. Prod. Sci., 86 (1-3): 35-45. doi:10.1016/S0301-6226(03)00151-9
12. Buxadera A. M., da Mota, M. D. S. (2008). Variance component estimations for race performance of thoroughbred horses in Brazil by random regression model. Livest. Sci., 117 (2-3): 298-307. doi:10.1016/j.livsci.2007.12.027
13. Cruz A., Menéndez-Buxadera A., Gutiérrez G., Morante R., Burgos A., Gutiérrez J. P. (2020). Genetic (co)variance across age of fiber diameter and standard deviation in Huacaya alpacas, estimated by repeatability, multi-trait and random regression models. Livest. Sci., 231: 103863. doi:10.1016/j.livsci.2019.103863
14. Faria R. A. S., Maiorano A. M., Correia L. E. C. d. S., Santana Jr, M. L., Silva, J. A. I. I. V. (2019). Time class for racing performance of the Quarter Horse: Genetic parameters and trends using Bayesian and multivariate threshold models. Livest. Sci., 225: 116-122. doi:10.1016/j.livsci.2019.05.013
15. Meyer K. (2004). Scope for a random regression model in genetic evaluation of beef cattle for growth. Livest. Prod. Sci, 86(1-3): 69-83. doi:10.1016/S0301-6226(03)00142-8
16. Orhan H., Kaygisiz, A. (2010). Genetic and Environmental parameters effecting racing performance of Turk-Arabian Horses raised at Anatolian state farm. Asian J. Anim. Vet. Adv., 5, 112-119. doi:10.3923/ajava.2010.112.119
17. Bartolomé E., Menéndez Buxadera A., Molina A., Valera, M. (2018). Plasticity effect of rider-horse interaction on genetic evaluations for Show Jumping discipline in sport horses. Anim. Breed. Genet., 135 (2): 138-148. doi:10.1111/jbg.12315
18. Bussiman, F. d. O., Perez, B., Ventura, R. V., Peixoto, M., Curi, R., Balieiro, J. C. d. C. (2018). Pedigree analysis and inbreeding effects over morphological traits in Campolina horse population. Animal, 12(11): 2246-2255. doi:10.1017/S175173111800023X
19. Gómez M., Menendez-Buxadera A., Valera M., Molina A. (2015). A reaction norm model approach to estimate the genetic effect of temperature on sportive performance of trotter horses. J. Anim. Breed. Genet., 132(3): 256-267. doi:10.1111/jbg.12118
20. Kim N. Y., Son J. K., Cho I. C., Shin S. M., Park S. H., Seong P. N., Woo J. H., Park N. G., Park H. B. (2018). Estimation of genetic parameters for temperament in Jeju crossbred horses. Asian-Australas J. Anim. Sci., 31(8): 1098-1102 doi:10.5713/ajas.17.025
21. Posta J., Komlósi I., Mihók S. (2009). Breeding value estimation in the Hungarian Sport Horse population. Vet. J., 181 (1): 19-23. doi: 10.1016/j.tvjl.2009.03.006
22. Wagenmakers E. J., Farrell S. (2004). AIC model selection using Akaike weights. Psychon. Bull. Rev., 11 (1): 192-196. doi:10.3758/BF03206482