

Estimation of Breeding Values with Heterogeneous Residual Variances by Random Regression Models ^[1]

Çiğdem TAKMA ¹  Yavuz AKBAŞ ¹

^[1] This study was financially supported by The Scientific and Technological Research Council of Turkey (Project No: 109O336)

¹ Ege Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik Anabilim Dalı, TR-35100 Bornova, İzmir - TÜRKİYE

Article Code: KVFD-2015-14364 Received: 10.09.2015 Accepted: 03.11.2015 Published Online: 03.11.2015

Abstract

In this study the third order random regression models (RV1 and RV10) including the fixed, random additive genetic and permanent environmental effects were used. In the RV1 model residual variance was constant and in the RV10 model all test day records were taken as different groups. The predicted error variance was found 6.83 in RV1 model and this variance was changed between 5.30 and 9.19 in RV10 model. Heritability values were estimated 0.12-0.53 for RV1 and 0.04-0.18 for RV10 models. Spearman rank and Kendall rank correlations between estimated breeding values of test day milk yields estimated from RV1 and RV10 models within cows (0.79, 0.61) were found almost same with within sires (0.82 and 0.63). Consequently, these correlations indicate that breeding values estimated from RV1 and RV10 models were highly correlated. Although the association between two set of breeding values estimated from RV1 and RV10 models were high, ranking of cows and sires by breeding values were different for two models. The shift in rank of first 100 cows was found 22% same and 78% different, also the shift in rank of first 50 sires was found 30% same and 70% different on RV1 and RV10 models breeding values rank list. Overall, the RV10 model was found better than RV1 model in estimation of breeding value.

Keywords: Residual variance, Random regression, Breeding value, Turkish Holstein

Şansa Bağlı Regresyon Modelinde Heterojen Hata Varyansları ile Damızlık Değerinin Tahmin Edilmesi

Özet

Bu çalışmada sabit, şansa bağlı genetik ve kalıcı çevre etkilerine sahip üçüncü dereceden şansa bağlı regresyon modelleri (RV1 ve RV10) kullanılmıştır. RV1 modelinde hata varyansı sabit, RV10 modelinde ise denetim günlerinin her biri farklı kabul edilmiştir. Hata varyansı RV1 modelinde 6.83 ve RV10 modelinde 5.30 ile 9.19 arasında tahminlenmiştir. Kalıtım dereceleri RV1 modelinde 0.12-0.53 iken, RV10 modelinde 0.04-0.18 arasında tahminlenmiştir. RV1 ve RV10 modelleri ile tahminlenen denetim günü süt verimi damızlık değerlerine ait Spearman ve Kendall sıra korelasyonları inek (0.79, 0.61) ve boğalarda (0.82, 0.63) birbirine benzer bulunmuştur. Sonuç olarak, her iki modelle elde edilen damızlık değerleri arasındaki ilişkinin yüksek olduğu belirlenmiştir. Sığırların damızlık değerlerinin sıralamasına yönelik korelasyonların yüksek olmasına karşın, sıralama bakımından modeller arasında farklılıklar saptanmıştır. Her iki modele ait damızlık değeri listesinde ilk 100 ineğin sıra değişimleri %22 oranında ortak ve %78 oranında farklı iken, ilk 50 boğanın sıra değişimleri ise %30 oranında ortak ve %70 oranında farklı bulunmuştur. Genel olarak RV10 modeli damızlık değer tahmininde RV1 modelinden daha iyi bulunmuştur.

Anahtar sözcükler: Hata varyansı, Şansa bağlı regresyon, Damızlık değeri, Siyah Alaca

INTRODUCTION

Breeding values for all traits are used to rank and select animals in order to achieve intended genetic improvements. As a matter of fact; breeding values from a random regression model are not equivalent to the breeding values from the more traditional models ^[1]. Last decade therefore, many countries have implemented

random regression models for genetic evaluation of test day records. In random regression models, residual effect is generally assumed constant throughout the trajectory. However, recent studies have shown that the residual variance (RV) is changing over time ^[2,3] because of herd management, weather conditions, lactation number, age at calving, month of calving, days in milk, pregnancy status, medical treatments and milking times etc. ^[4,5].



İletişim (Correspondence)



+90 232 3112714



cigdem.takma@ege.edu.tr

Homogeneous residual variance assumption leads to lower or higher impact on the evaluation for different parts of the lactation [6]. Therefore, the information coming from each part of the lactation where the residual variance is actually larger than the assumed homogeneous value will have lower weight than it really has [7]. In addition, Olori et al. [8] conveyed that constant residual variance assumption causes residual variances to be underestimated and heritability values to be overestimated in early stage of lactation. Instead of constant residual assumption, lactation can be divided into different classes with assuming homogeneity of residual variance within the classes and heterogeneity between them [7-9]. This approach is easy to implement and useful to provide the information on the expected pattern of residual variance which changes over lactation.

There are few studies on heterogeneity of residual variance in random regression model for the estimation of genetic parameters for test day milk records [7-10]. However, there is not enough research in which examines different residual variance structures on test day milk yield estimated breeding values (EBV) by random regression models. Fujii and Suzuki [11] reported genetic parameters and EBV's for milk yield using random regression models under homogeneous and heterogeneous residual variances, larger permanent environmental variances than additive genetic variances and there was no difference between EBV's from homogeneous and heterogeneous residual variance models.

In this study, it is aimed to compare effect of homogeneity and heterogeneity residual variance on breeding values of test day milk records for Turkish Holsteins using random regression model.

MATERIAL and METHODS

Data

Test day milk records obtained from different farms who are the members of Isparta province Cattle Breeders Association in Turkey were the material of this study. First lactation test day milk yields were collected at monthly periods (TD1-TD10) from 2001 through 2011. Last test day milk records less than 5 kg were excluded and records used from monthly milk between 5th days and 307th days. Age at first calving of cows was also limited between 20 and 51 months. In final data set total of 43206 test day milk records from 6085 Turkish Holstein cows in 248 herds were analyzed. The descriptive statistics of the final data set were given in Table 1.

Method

Third order random regression model using legendre polynomial was preferred due to the best fit [10,12-17] as follows:

Table 1. Structure of the final data

Tablo 1. Veri setinin yapısı

Item	Statistics
Number of records	43206
Mean of milk yields per TD	18.78
Number of herds	248
Number of herd-year-season level	3809
Number of Animals with records	6085
Number of Sires with progeny records	667
Number of Dams with progeny records	4241

$$y_{ijk} = HYS_i + \sum_{m=1}^5 \beta_m X_m(t_{jk}) + \sum_{m=1}^3 \alpha_{jm} \phi_m(t_{jk}) + \sum_{m=1}^3 p_{jm} \phi_m(t_{jk}) + e_{ijk}$$

where y_{ijk} is the k^{th} test day milk yield of the cow j at i^{th} herd-year-season, HYS_i is the i^{th} herd-year-season effect, β_m is the m^{th} fixed regression coefficients associated with the m^{th} covariate, t_{jk} is the k^{th} test day of the cow j , $X_m(t_{jk})$ is the m^{th} covariates (X_1 : Age at first calving, $c=305$, $X_2=DIM/c$, $X_3=(X_2)^2$, $X_4=\ln(c/DIM)$, $X_5=(X_4)^2$ depending on $DIM=t$ evaluated at t_{jk}), α_{jm} is the m^{th} additive genetic random regression coefficients for cow j , p_{jm} is the m^{th} permanent environmental random regression coefficients for cow j , ϕ_m is the m^{th} polynomial and e_{ijk} is the random residual effect with $e_{ijk} \sim N(0, \sigma^2 e_{ijk})$. In this model homogeneity and heterogeneity residual variance assumptions were tested. In RV1 model, the residual variance was assumed constant throughout lactation. On the contrary to RV1, residual variance was assumed different for each test day in RV10 model.

DXMRR option of the DFREML statistical package [18] was used for fitting of the models. The goodness of fits for the models with different error variances were examined using as Akaike's information criterion-AIC [19]. This likelihood based criterion has been calculated as: $AIC = -2 * \text{LogL} + 2 * p$ where p denotes the number of parameters estimated. The model which gives the lowest AIC values was chosen as the better approximating model [20,21]. Furthermore, two error structures were compared by Likelihood ratio test-LRT [22]. LRT for model i and j was $LRT_{ij} = 2 * (\text{LogL}_i - \text{LogL}_j)$. In the LRT, the Log Likelihood (LogL) differences were tested using Chi-square test with the degree of freedom determined as the number of the parameter differences between the models [23]. Spearman rank and Kendall rank correlations between EBV of test day milk yields obtained from RV1 and RV10 models were also calculated for cows and sires. Shifts in rank of animals according to EBV were determined to show the changes in ranks of cows and sires between models.

RESULTS

The comparison of RV1 and RV10 models were presented in Table 2. The differences in LogL values of the

RV1 and RV10 models were found significant ($P < 0.05$). A decrease in AIC values with increased numbers of parameters was noticed between models (Table 2).

The residual, additive genetic, permanent environmental variances and heritability estimates of test day milk yields from the RV1 and RV10 models were summarized in Table 3. Two models had different tendency for estimation of variance components. The residual variance was 6.83 from the RV1 model and changed between 5.30 and 9.19 from RV10 model. In RV10 model, the predicted residual variances were lower at the beginning of the lactation but higher at end of the lactation. Estimates of additive genetic variances were varied from 1.83 to 14.28 for RV1 and 0.91 to 3.77 for RV10 models. Changes of the permanent environmental variances were found between 5.24 to 8.94 for RV1 and 6.39 to 14.45 for RV10 models. Heritability estimates of test day milk yields from the RV1 and RV10 models were ranged from 0.12 to 0.53 and 0.04 to 0.18, respectively (Table 3).

The Spearman rank correlations between EBV's from RV1 and RV10 models were statistically significant ($P < 0.01$) and found 0.79 for cows and 0.82 for sires. The difference between the probabilities in the same and different orders of animal according to EBV with Kendall rank correlations were also statistically significant ($P < 0.01$) and determined 0.61 for cows and 0.63 for sires.

Mean and standard deviations for test day milk yield EBV's from RV1 and RV10 models were illustrated in Table 4. Number and percent of cows and sires were detected to indicate the rank changes of Turkish Holsteins sorted

by their EBV's from RV1 and RV10 models. Moreover, largest rank shift of cows and sires was also determined according to various top lists EBV's from both models rank list in Table 4.

The mean of EBV's for cows ranged from 8.98 to 6.79 for RV1 model and changed from 1.05 to 1.38 for RV10 model. The mean of EBV's of sires changed from 5.34 to 3.35 and from 1.03 to 0.63 for RV1 and RV10 models, respectively. Mean of EBV's were decreased when increased number of animals in top lists as expected. On the other hand, when only the first 10 cows are considered, there are 2 cows (20%) on both lists. The percentages of cows on both lists were 30% when the first 20 and 50 cows are considered. When only the first 100 cows are considered, there are 22 cows (22%) same and 78% different on RV1 and RV10 models EBV's rank list. However, the first 5, 10, 25 and 50 sires are considered, the percentages of sires on both lists were $\geq 25\%$. 50% sires (5 sires) are on both lists when the first 10 sires are considered. Largest shifts in rank for the first 10, 20, 50 and 100 cows, and for the first 5, 10, 20 and 50 sires are presented based on the ranking for milk yield EBV's. As shown in Table 4, when the first 10 cows are considered, the cow in the 4th rank on RV1 list appeared in the 9th rank on RV10 list, which was the largest rank shift in this group. When the first 5 sires are considered, the sire in the 4th rank on RV1 list appeared in the 3th rank on RV10 list. Considering largest rank shifts, the order for RV10 is generally higher than the orders for RV1. The largest rank shifts for only first 10 cows and 50 sires showed opposite direction. This shows each two model are different in estimation of breeding value.

Table 2. LogL and AIC values for RV1 and RV10 models

Tablo 2. RV1 ve RV10 modellerinde LogL ve AIC değerleri

Models	Number of Parameters	AIC	Log Likelihood Values	LRT
RV1	13	140325	-70150	-
RV10	22	140143	-70043	107*

AIC: Akaike's information criterion; LRT: Likelihood ratio test; * LRT values between RV1 and RV10 models are significant ($P < 0.05$)

Table 3. Estimations of additive genetic (G), permanent environmental (PE), residual variances (RV) and heritability estimates (h^2) of test day milk yields from the RV1 and RV10 models

Tablo 3. RV1 ve RV10 modellerinde denetim günü süt verimleri için eklemeli genetik, kalıcı çevre, hata varyansları ve kalıtım derecesi (h^2) tahminleri

Model	Parameters	Test Day									
		1	2	3	4	5	6	7	8	9	10
RV1	G	5.81	3.47	2.40	1.97	1.83	1.94	2.55	4.20	7.73	14.28
	PE	8.94	7.40	6.96	6.98	7.00	6.78	6.27	5.64	5.24	5.64
	RV	6.8									
	h^2	0.27	0.20	0.15	0.12	0.12	0.12	0.16	0.25	0.39	0.53
RV10	G	3.77	3.14	2.65	2.17	1.74	1.40	1.16	1.00	0.92	0.91
	PE	12.17	8.49	7.65	6.80	6.39	6.63	7.60	9.25	11.56	14.45
	RV	5.30	7.88	6.26	5.50	7.46	6.84	5.94	6.82	7.86	9.19
	h^2	0.18	0.16	0.16	0.15	0.11	0.09	0.08	0.06	0.05	0.04

Table 4. Shifts in rank of cows and sires ranked by EBV's from RV1 and RV10 models**Tablo 4.** Sığırların RV1 ve RV10 modelleri ile tahminlenen damızlık değerleri sıralamasındaki değişimler

Animals Considered	X±S for EBV's		Number of Animals on Both Lists	% of Animals on Both Lists	Largest Rank Shift	
	RV1 Model	RV10 Model			RV1 Model	RV10 Model
Cows						
first 10	8.98±0.59	1.38±0.08	2	20	4	9
first 20	8.44±0.69	1.31 ±0.06	6	30	17	13
first 50	7.59±0.86	1.17±0.13	15	30	50	22
first 100	6.79±1.02	1.05±0.15	22	22	94	8
Sires						
first 5	5.34±0.81	1.03±0.13	2	40	4	3
first 10	4.76±0.82	0.94±0.13	5	50	10	5
first 20	4.16±0.85	0.81±0.16	5	25	16	2
first 50	3.35±0.87	0.63±0.18	15	30	47	76

DISCUSSION

In this study constant residual variance (RV1) and heterogeneous residual variances (RV10) for each test day periods were compared. The RV1 model was differ from RV10 model for the goodness of fit test. The RV10 model was fit better than RV1 model due to lower AIC value. The predicted residual variances were lower at beginning and had fluctuating tendency in middle but higher at the end of the lactation under RV10 model. Fujii and Suzuki^[11] observed similar tendency for beginning and end of lactation but had generally lower estimates of residual variances. Estimates of Olori et al.^[8] are not similar to those in this study with higher estimates at beginning and lower estimates at the end of the lactation. Moreover, Olori et al.^[10] had highest estimates at the beginning of lactation and relatively constant in mid lactation with lower values than this study. The fluctuating estimates of residual variances might clarify with number of test day records used and the models in the analysis.

It can be seen that RV1 model was appeared with higher genetic variance estimates at the beginning and end of lactation, but lower at the middle of the lactation. Genetic variances for RV10 model follows lower estimates at the both side of the lactation than RV1. In our study, estimates of additive genetic variances (1.83-14.28 for RV1 and 0.91-3.77 for RV10) were lower than the estimates of Rekaya et al.^[9]. These estimates were higher than the estimates obtained from Fujii and Suzuki^[11] for the early and late part of lactation. The differences in genetic variations were the reason of genetic differences among cows since variance components for test day milk yields are change according to lactation stages, parity, year of calving and generally populations etc. Also inconsistent environmental factors in early and late stages of lactation may partly explain the discrepancy in estimates in our study.

Furthermore, permanent environmental variances (6.39-14.45) had higher values at the beginning and end of lactation, but lower at the middle of the lactation under RV10 model. The estimates of permanent environmental variances (5.24-8.94) tended to slightly increase at the beginning and decrease at the end of lactation under RV1 model. Olori et al.^[8], Rekaya et al.^[9], Fujii and Suzuki^[11] have also found similar findings but reported higher estimates than our results.

The heritability estimates from RV1 model (0.39-0.15) were found lower at middle of the lactation and higher early and late part of the lactation because of lower environmental but higher genetic variances. Heritability estimates from RV10 model (0.3-0.11) were found lower for early and late part of the lactation as expected due to lower genetic variances at these stages. When compared these results to Olori et al.^[8] and Olori et al.^[10] have obtained opposite trend and higher estimates because of higher estimates of genetic variances. Heritability estimates only reflect the proportion of genetic variance of test day milk yields, but they cannot say anything about causes of variances of the test day milk yields. Constant residual variance assumption caused lower permanent environmental and higher genetic variance estimates and therefore may be induced over estimated heritability values at late of lactation stages (TD9 and TD10) in our study.

In this study, spearman rank correlations that clarify correlations of cows and sires EBV's ranking for RV1 and RV10 models were found high and significant. The difference between the probability of same order EBV's of two models and probability of different orders EBV's of two models were also high and significant. These correlations explain that, estimated breeding values of animals from RV1 and RV10 models are highly correlated. Moreover, ranking of sires was less affected than ranking of cows with heterogeneous residual variance model. When the

differences of largest rank shifts for models are examined, it can be seen that more extreme changes in ranks occurred for the both lists. A possible explanation of this finding could be reason of lower accuracy for cows EBV's. The drastic changes in ranks of EBV's in the cow and sire lists and also different model lists seemed to may be associated with differences on lactation curves that deviated from the standard lactation curve.

Consequently, our results indicate that type of residual variance assumption in random regression models might have significant effect on the variance components at any stages of the lactation. This effect is also available in EBV's as on variance components. In this study, it can be seen that heterogeneous residual variance assumption is more effective than constant residual assumption. Therefore, heterogeneous residual variance effect will be more informative for detecting reliable breeding values.

ACKNOWLEDGEMENT

The authors would like to thank the Cattle Breeders Association of Turkey for providing data.

REFERENCES

- Speidel SE:** Random regression models for the prediction of days to Finish in beef cattle. In partial fulfillment of the requirements for the Degree of Doctor of Philosophy. Colorado State University, Department of Animal Sciences, Fort Collins, Colorado. 2011.
- Jaffrezic F, White IMS, Thompson R, Hill WG:** A link function approach to model heterogeneity of residual variances over time in lactation curve analyses. *J Dairy Sci*, 83, 1089-1093, 2000. DOI: 10.3168/jds.S0022-0302(00)74973-3
- Takma Ç, Akbaş Y:** Heterogeneity of residual variances of test day milk yields estimated by random regression model in Turkish Holsteins. *J Anim Vet Adv*, 8 (4): 782-787, 2009.
- Perochon L, Coulon JB, Lescourret F:** Modelling lactation curves of dairy cows with emphasis on individual variability. *Anim Sci*, 63, 189-200, 1996. DOI: 10.1017/S1357729800014740
- Swalve HH:** Theoretical basis and computational methods for different test day genetic evaluation methods. *J Dairy Sci*, 83, 1115-1124, 2000. DOI: 10.3168/jds.S0022-0302(00)74977-0
- Rekaya R, Carabaño MJ, Toro MA:** Assessment of heterogeneity of residual variances using change point techniques. *Genet Sel Evol*, 32, 383-394, 2000. DOI: 10.1186/1297-9686-32-4-383
- Lo'pez-Romero P, Rekaya R, Carabaño MJ:** Assessment of homogeneity vs. heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. *J Dairy Sci*, 86, 3374-3385, 2003. DOI: 10.3168/jds.S0022-0302(03)73941-1
- Olori VE, Hill WG, Brotherstone S:** The structure of the residual error variance of test day milk yield in random regression models. *Interbull Bulletin*, 20, 103-108, 1999.
- Rekaya R, Carabaño MJ, Toro MA:** Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. *Livest Prod Sci*, 57, 203-217, 1999. DOI: 10.1016/S0301-6226(98)00181-X
- Olori VE, Hill WG, Mcguirk BJ, Brotherstone S:** Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. *Livest Prod Sci*, 61, 53-63, 1999. DOI: 10.1016/S0301-6226(99)00052-4
- Fujii C, Suzuki M:** Comparison of homogeneity and heterogeneity of residual variance using random regression test-day models for first lactation Japanese Holstein cows. *J Anim Sci*, 77, 28-32, 2006. DOI: 10.1111/j.1740-0929.2006.00316.x
- Van Der Werf JHJ, Goddard ME, Meyer K:** The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J Dairy Sci*, 81, 3300-3308, 1998. DOI: 10.3168/jds.S0022-0302(98)75895-3
- Pool MH, Meuwissen THE:** Prediction of daily milk yields from a limited number of test days using test day models. *J Dairy Sci*, 82, 1555-1564, 1999. DOI: 10.3168/jds.S0022-0302(99)75383-X
- Veerkamp RF, Thompson R:** A covariance function for feed intake, live weight, and milk yield estimated using a random regression model. *J Dairy Sci*, 82, 1565-1573, 1999. DOI: 10.3168/jds.S0022-0302(99)75384-1
- Kettunen A, Mäntysaari EA, Pösö J:** Estimation of genetic parameters for daily milk yield of Primiparous Ayrshire Cows by random regression test day models. *Livest Prod Sci*, 66, 251-261, 2000. DOI: 10.1016/S0301-6226(00)00166-4
- Strabel T, Szyda J, Ptak E, Jamrozik J:** Comparison of random regression test-day models for production traits of dairy cattle in Poland. *Interbull Bulletin*, 31, 197-200, 2003.
- Takma Ç, Akbaş Y:** Variance components and genetic parameter estimates using random regression models on test day milk yields of Holstein Friesians. *Kafkas Univ Vet Fak Derg*, 15, 547-551, 2009. DOI: 10.9775/kvfd.2009.055-A
- Meyer K:** DFREML 3.0a Program Package and User Notes. Animal Genetics and Breeding Unit, University New England, Armidale, New South Wales, Australia, 1998.
- Akaike H:** Information theory as an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary, 1973.
- Apiolaza LA, Gilmour AR, Garrick DJ:** Variance modeling of longitudinal height data from a *Pinus radiata* progeny test. *Can J For Res*, 30, 645-654, 2000. DOI: 10.1139/x99-246
- Lo'pez-Romero P, Carabaño MJ:** Comparing alternative random regression models to analyse first-lactation daily milk yield data in Holstein-Friesian cattle. *Livest Prod Sci*, 82, 81-96, 2003. DOI:10.1016/S0301-6226(03)00003-4
- Rao CR:** Linear Statistical Inference and Its Applications. 2nd ed., John Wiley&Sons, New York, 1973.
- Huelsenbeck JP, Bull JJ:** A likelihood ratio test to detect conflicting phylogenetic signal. *Syst Biol*, 45, 92-98, 1996. DOI: 10.1093/sysbio/45.1.92