Genetic analysis of milk yield, fat and protein content in Holstein dairy cows in Iran: Legendre polynomials random regression model applied

Rohullah Abdullahpour1, Mohammad Moradi Shahrbabak2, Ardeshir Nejati-Javaremi2, Rasoul Vaez Torshizi3 and Raphael Mrode4

1Department of Animal Science, Faculty of Agriculture, Science and Research Branch, Islamic Azad University, Tehran, Iran, 2Department of Animal Science, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, 3Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran, 4Scottish Agricultural College, Sir Stephen Watson Building, Penicuik, United Kingdom

Abstract

Data including 219 105 test day records of 22 569 first parity Holstein cows in 56 herds were analysed for milk yield, fat content and protein content. Legendre polynomials were used in a random regression model to explain traits curves, additive genetic and permanent environmental effects along days in milk (DIM). Legendre polynomials of order 3 were used to describe additive genetic effects on the traits. For permanent environmental effects, for milk the order of 6 and for fat and protein content the order of 4 were used. Heterogeneity of residual variance was considered. Restricted maximum likelihood (REML) methodology was used to estimate the required parameters. Variations in genetic, permanent environment and residual effects and heritability changes along DIM were computed and illustrated. Daily heritability estimates averaged as 0.22, 0.14 and 0.23 for milk, fat and protein contents, respectively. Minimum correlations between DIM for additive genetic effects were as 0.49, –0.01 and 0.34 for milk, fat and protein contents. There were higher genetic correlations between test day milk yield and protein content compared to fat content. The genetic trend of milk yield has increased over the years from 1971 to 2005, while the genetic trend for fat and protein content declined.

Keywords: Holstein cows, test day model, fat content, protein content, variance components, genetic trends
Introduction

Genetic evaluation of dairy cattle for longitudinal data can be implemented by means of test day models, which involve analysing the original test day records directly. Some advantages of test day models include accounting more accurately for environmental factors that affect cows at different stages of lactation at the time of test (Jamrozik & Schaeffer 1997, Swalve 2000, Jensen 2001), increases in the accuracy of genetic evaluations, accounting for the number of records per cow and the interval between records (Pool & Meuwissen 1999), decreasing the costs of milk recording by making fewer measurements (Wiggans & Goddard 1997), flexibility, the potential to slightly reduce the generation interval by frequent genetic evaluations (Swalve 2000) and the ability to account for individual differences in the shape of the lactation curves of cows including the persistency of the lactation (Pool & Meuwissen 1999).

There is a curvilinear relationship between test day production yields and DIM. Therefore, it is necessary to use appropriate functions for modelling these traits. The first function used to model the random part of the lactation curve in a random regression model was the curve proposed by Ali and Schaeffer (Pool et al. 2000) and afterwards other parametric families of curves were applied. Another option is the use of polynomial regression models. Legendre polynomials in random regression models have been used in most studies for the genetic analysis of dairy test day records. The greater use of orthogonal Legendre polynomials is due to their beneficial properties. The derived coefficients are very useful for analysing patterns of genetic variation by calculating the eigenfunctions and eigenvalues of covariance function. They behave relatively smooth rather than oscillatory and are therefore suitable on biological grounds (Kirkpatrick et al. 1990). There is no need to adopt restrictive assumptions about the shape of the curve (Kirkpatrick et al. 1990, Mrode 2005) or the dispersion structure (Meyer & Kirkpatrick 2005). Furthermore, they reduce correlations among the estimated coefficients (Schaeffer 2004) and missing records can be predicted more accurately than with the Wilmink curve. Higher orders are estimable when conventional polynomials fail because of better convergence (Pool et al. 2000). In addition, they are easy to apply (Mrode 2005).

The Iranian Holstein population including an approximate number of one million heads originates from Holstein registered heifers imported from Europe, the United States and Canada during the 1970s and early 1980s. Five centres located at different regions were established for progeny test of young bulls and production semen for artificial insemination. While in 1997 an index including milk yield, fat percentage and herd life was proposed, in 2006 it was recalculated by excluding fat percentage and including fat and protein yields and became the national selection objective (Sadeghi-Sefidmazgi et al. 2012). Each year approximately 80 young bulls are entered into the progeny testing programme of which 12–20 bulls would be selected as proven sires (Dadpasand et al. 2008). While in recent years the contributions of domestic semen has drastically decreased, that of foreign semen for the United States, Canada and European countries were 60 %, 30 % and 10 % in sire of sire pathway and 58 %, 32 % and 10 % in sire of dam pathway with a growing trend for semen.
imported from the US after 2000 (Joezy-Shekalgorabi & Shadparvar 2010). Although national genetic evaluations are routinely carried out twice every year and EBV published on 5 production and 17 linear type traits, long-term genetic and phenotypic trends of the traits might not be clearly illustrated or interpreted (Sadeghi-Sefidmazgi et al. 2012). Up to now, the national genetic evaluation for production traits is carried out using 305 days records by the Animal Breeding Center of Iran. The aim of this study was to apply random regression models using Legendre polynomials to analyse milk yield, fat content and protein content of first lactation records for Holstein dairy cows in Iran.

**Material and methods**

*Data*

Test day data of Iranian Holstein cows were obtained from the Animal Breeding Center (Iran) which is responsible for dairy cattle recording. First lactation records between 5 and 330 days in milk (DIM) from animals having at least one known parent, calved between years 1997 to 2009 at ages ranged from 20 to 36 months, were kept for more edits. Finally the data set included 219 105 test day records on 22 569 cows distributed in 56 herds. The cows had at least 6 tests with the first one recorded before day 36 after parturition and the interval between records was not more than 50 days. The herds were located in 14 different provinces belonging to various distinguished climates in Iran. A summary of the statistics of test day records of traits is shown in Table 1. The pedigree was constructed by successive backward tracing for the ancestors of animals with records. The pedigree included 51 005 animals and 2 365 sires in 18 generations. The arithmetic and harmonic number of progeny per sire were 18.35 and 2.23. The base population contained 466 sires and 6 044 dams.

<table>
<thead>
<tr>
<th>Tests</th>
<th>Number</th>
<th>Milk mean</th>
<th>SD</th>
<th>Fat % mean</th>
<th>SD</th>
<th>Protein % mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20 534</td>
<td>26.11</td>
<td>6.67</td>
<td>3.49</td>
<td>0.98</td>
<td>3.07</td>
<td>0.45</td>
</tr>
<tr>
<td>2</td>
<td>22 251</td>
<td>29.88</td>
<td>7.10</td>
<td>3.09</td>
<td>0.84</td>
<td>2.88</td>
<td>0.38</td>
</tr>
<tr>
<td>3</td>
<td>22 231</td>
<td>30.14</td>
<td>7.22</td>
<td>3.03</td>
<td>0.80</td>
<td>2.91</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>22 177</td>
<td>29.62</td>
<td>7.32</td>
<td>3.06</td>
<td>0.80</td>
<td>2.98</td>
<td>0.38</td>
</tr>
<tr>
<td>5</td>
<td>22 231</td>
<td>28.88</td>
<td>7.39</td>
<td>3.10</td>
<td>0.81</td>
<td>3.04</td>
<td>0.38</td>
</tr>
<tr>
<td>6</td>
<td>22 154</td>
<td>28.02</td>
<td>7.47</td>
<td>3.17</td>
<td>0.82</td>
<td>3.08</td>
<td>0.37</td>
</tr>
<tr>
<td>7</td>
<td>21 621</td>
<td>27.15</td>
<td>7.39</td>
<td>3.22</td>
<td>0.83</td>
<td>3.13</td>
<td>0.37</td>
</tr>
<tr>
<td>8</td>
<td>20 814</td>
<td>26.24</td>
<td>7.29</td>
<td>3.28</td>
<td>0.83</td>
<td>3.17</td>
<td>0.37</td>
</tr>
<tr>
<td>9</td>
<td>19 246</td>
<td>25.09</td>
<td>7.26</td>
<td>3.37</td>
<td>0.84</td>
<td>3.22</td>
<td>0.38</td>
</tr>
<tr>
<td>10</td>
<td>15 196</td>
<td>24.23</td>
<td>7.26</td>
<td>3.43</td>
<td>0.85</td>
<td>3.27</td>
<td>0.39</td>
</tr>
<tr>
<td>11</td>
<td>10 650</td>
<td>23.54</td>
<td>7.29</td>
<td>3.50</td>
<td>0.87</td>
<td>3.31</td>
<td>0.42</td>
</tr>
</tbody>
</table>

*Statistical models*

Single-trait random regression models using orthogonal Legendre polynomials were used to fit the data. Some criteria were used to select models from several fitted ones of random regression using Legendre polynomials, natural cubic splines and B-splines functions.
(Abdullahpour, unpublished). The model equation for analyses was as follows:

$$y_{tijk} = F_i + \sum_{k=0}^{n} \Phi_{jtk} \beta_k + \sum_{k=0}^{3} \Phi_{jtk} u_{jk} + \sum_{k=0}^{op} \Phi_{jtk} p_{ejk} + e_{tijk}$$

(1)

where $y_{tijk}$ is the test day observation, $F_i$ represents fixed effects in the model including herd-test date, milking frequency, calving year-month and age of cow at calving as covariate, $\beta_k$ are fixed regression coefficients, $u_{jk}$ and $p_{ejk}$ are the $k$-th random regression coefficients for the $j$-th animal for additive genetic and permanent environmental effects respectively, $\Phi_{jtk}$ is the $k$-th Legendre polynomial for the standardized test day record of cow $j$ made on day $t$, $n$ is the order of Legendre polynomials for fitting traits curve for the whole population which was 8 for milk and 9 for fat content and protein content, $op$ is the order of Legendre polynomials for permanent environmental effect equal to 6, 4 and 4 for milk, fat content and protein content respectively and $e_{tijk}$ is the random residual effect. To account for heterogeneity in residual effect along the lactation trajectory, DIM were partitioned into 11 equal segments of about 30 days each and an independent residual variance structure was assumed.

Analyses were done using ASREML software (Gilmour et al. 2009) to estimate covariance components of random regression coefficients by REML methodology. Variances along DIM for permanent environmental effect were estimated similar to variances for additive genetic effect using $\hat{G} = \hat{\phi}_k \hat{\phi}$, in which $\hat{G}$ is the additive genetic (co)variance matrix for DIM with dimension $t \times t$ ($t$ is the number of days in milk), $\phi$ is the matrix of Legendre polynomials for standardized DIM with dimension of $t \times k$ ($k$ is the order of Legendre polynomials for this random effect) and $\hat{K}$ is the estimated (co)variance matrix for random regression coefficients which is a $k \times k$ matrix.

Standard errors of (co)variances were obtained from variances of (co)variances as explained by Fischer et al. (2004) for additive genetic variances:

$$\text{var(vec}(\hat{G})) = (\phi \otimes \phi) \text{var(vec}(\hat{K}))(\phi \otimes \phi)$$

(2)

In which $\text{var(vec}(\hat{G}))$ is a ($t \times t$) by ($t \times t$) (co)variance matrix, vec($\hat{G}$) is a ($t \times t$)×1 vector containing the columns of $\hat{G}$ matrix, vec($\hat{K}$) is a ($k \times k$)×1 vector containing columns of matrix $\hat{K}$ and $\text{var(vec}(\hat{K}))$ can be approximated from the appropriate elements of the inverse of the average information matrix in a REML procedure and $\otimes$ denotes the direct product. The variance for phenotype (co)variance matrix is built as:

$$\text{var(vec}(\hat{P})) = \text{var(vec}(\hat{G})) + \text{var(vec}(\hat{E})) + 2\text{cov(vec}(\hat{G}), \text{vec}(\hat{E}))$$

(3)

And for heritability estimates:

$$\text{var}(\hat{h}^2) = \text{var}(\hat{g}_{i,i} / \hat{p}_{i,i}) \approx (\hat{p}_{i,i}^2 \text{var}(\hat{g}_{i,i}) + \hat{g}_{i,i}^2 \text{var}(\hat{p}_{i,i}) - 2\hat{g}_{i,i} \hat{p}_{i,i} \text{cov}(\hat{g}_{i,i}, \hat{p}_{i,i})) / \hat{p}_{i,i}^4$$

(4)

where $\hat{g}_{i,i}$ and $\hat{p}_{i,i}$ are elements of $\hat{G}$ and $\hat{P}$, var($\hat{g}_{i,i}$), var($\hat{p}_{i,i}$) and cov($\hat{g}_{i,i}, \hat{p}_{i,i}$) are variance and covariance of genetic and phenotypic variance at time $i$.

Breeding values were predicted using $\text{EBV}_k = \phi \hat{u}_k$, in which $\text{EBV}_k$ is a vector containing estimated breeding values for animal $k$ for the range of DIM used in constructing the $\phi$ matrix and $\hat{u}_k$ is the vector of random regression coefficients estimated for animal $k$. 
Results

The lactation curves revealed that the peak of milk yield occurred after the 40th DIM when the percentages of fat and protein were smallest. The average of fat content was significantly more than protein content in the first and last one third of the lactation but in the second one third, they became closer to each other. Variance and coefficient of variation for protein percentage were much lower than those for milk and fat percentage. Variance for protein content did not show specific changes dependent to lactation stage, but that for fat content decreased sharply pre-peak milk and after that increased slightly towards the end of lactation, while milk yield variance had a reverse trend along the lactation.

Estimates of variances, covariances and correlations among the random regression coefficients for milk yield, fat content and protein percentage are shown in Table 2. Random coefficients for permanent environmental effects showed greater variation in case of milk yield and smaller variation for fat and protein contents than those for additive genetic effects.

<table>
<thead>
<tr>
<th>Tests</th>
<th>Mean Milk</th>
<th>Mean SD</th>
<th>Mean Fat%</th>
<th>Mean SD</th>
<th>Mean Pro%</th>
<th>Mean SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>22.49</td>
<td>±0.60</td>
<td>0.133</td>
<td>±0.223</td>
<td>0.074</td>
<td>±0.248</td>
</tr>
<tr>
<td>1</td>
<td>1.334</td>
<td>4.49</td>
<td>±0.11</td>
<td>0.061</td>
<td>-0.234</td>
<td>-0.444</td>
</tr>
<tr>
<td>2</td>
<td>-1.426</td>
<td>0.176</td>
<td>1.824</td>
<td>±0.05</td>
<td>-0.026</td>
<td>-0.333</td>
</tr>
<tr>
<td>3</td>
<td>0.329</td>
<td>-0.464</td>
<td>0.873</td>
<td>±0.03</td>
<td>-0.03</td>
<td>-0.385</td>
</tr>
<tr>
<td>4</td>
<td>-0.813</td>
<td>0.064</td>
<td>-0.311</td>
<td>-0.019</td>
<td>0.476</td>
<td>±0.02</td>
</tr>
<tr>
<td>5</td>
<td>0.455</td>
<td>-0.172</td>
<td>0.037</td>
<td>-0.21</td>
<td>0.036</td>
<td>0.338</td>
</tr>
<tr>
<td>6</td>
<td>-0.423</td>
<td>0.122</td>
<td>0.092</td>
<td>-0.043</td>
<td>-0.117</td>
<td>-0.012</td>
</tr>
</tbody>
</table>

Means and SD for DIM, milk yield, Fat% and Pro% for monthly tests.

Daily additive genetic, permanent environmental and residual variances for the traits varied across lactation days (Figure 1).
The genetic variance declined sharply for fat content but only moderately for protein content at the early lactation. However, it increased after days for peak of milk yield for the three traits. Permanent environmental variation for fat and protein contents declined immediately in early lactation and increased rapidly afterwards towards the end of lactation. It later increased from mid-lactation until the end of parity. The residual variances for traits were highest at the beginning of lactation. Then residual variances decreased until mid-lactation followed by an increase towards the end of parity. The increase for the case of protein content happened steadily in the second half part of lactation. Residual variances for milk yield were in the range of variances for other random effects in the model but for fat and protein contents residual variances were higher than variances from other sources across DIM. The ratio of residual variance to phenotypic variance was highest for fat content and smallest for milk yield during all DIM.

Figure 2 shows heritability changes of traits across lactation trajectory. Generally, heritability increased toward DIM but there were a slight and sharp decrease respectively for fat content and milk yield at the beginning of lactation and a slight decrease at the end of lactation for protein content. The average of daily heritability estimates for milk, fat and protein content were 0.22, 0.14 and 0.23, respectively. Heritabilities computed for 305 day lactation were also the same, which were computed over days 5 to 305 and are similar to the average daily heritabilities.
Figure 3 shows the additive genetic and permanent environmental correlations between DIM for three traits. The correlations between adjacent days were approximately 1 for all traits. Minimum correlation coefficients between DIM for additive genetic effect were as 0.49, −0.01 and 0.34, and for permanent environmental effect were 0.1, −0.1 and 0.12 for milk yield, fat content and protein content, respectively.
Figure 4 shows the genetic trends based on 305 day and daily breeding values of milk, fat and protein contents from 1971 to 2005. The genetic changes in traits are lower in early lactation compared to days afterwards.

Discussion

Estimates of variances, covariances and correlations among the random regression coefficients for additive genetic and permanent environmental effects on milk correspond closely to the estimates of other researchers. López-Romero & Carabaño (2003) and Bignardi et al. (2009) reported similar values for milk coefficients using a similar model. Togashi et al. (2008) estimated similar correlations between intercept, second and third additive genetic random coefficients for milk.

The trends for estimated daily variance components of additive genetic and permanent environmental effects along lactation are consistent with the findings of other researchers.
such as De Roos et al. (2004) and Hammami et al. (2008). These trends indicate that genetic variation of milk tests were lowest around the days of peak and were higher on average in the second half part of lactation and a very high variation for the effects of the permanent environment among cows at the onset of milking which declined rapidly. Highest residual variances for traits at the beginning of lactation might be for the fact that the cows are in a more unstable condition due to the effect of parturition. A slight increase in residual variance appearing for milk and fat content at the end of lactation might be for the reason that the cows were forced to stop milking. The ratio of residual variance to phenotypic variance of traits might indicate that the model of analysis was more suitable for milk yield than for fat and protein contents, which had a higher proportion of residual variances. This phenomenon was salient for fat content. It could be that there are other critical factors influencing fat content which the model did not account for. This could be due to problems of preferential treatment in the farm in which cows are fed according to their level of milk production. Since milk is negatively correlated with the percentages, it could be a reason of the higher residual variances for the contents.

The heritability estimates of daily fat content also were smaller than of milk and protein content at all DIM, apart from in early lactation. Some researchers showed daily fat content heritabilities are greater than those of daily milk (Druet et al. 2005, Penasa et al. 2010). While Druet et al. (2005) estimated daily fat content heritabilities to be more than of daily protein content for most of the DIM, Penasa et al. (2010) estimated them lower than of daily protein content. Liu et al. (2000), Jakobsen et al. (2002), Druet et al. (2005), Silvestre et al. (2005), Bohmanova et al. (2008) and Hammami et al. (2008) who focused on fat and protein yield, reported daily heritability estimates of fat yield being smaller than of milk and protein. This phenomenon might be related to some differences in genetic basis of these traits. Milk fat products are derived from metabolites, whereas protein products are all gene derivate products. Therefore, protein production requires more direct genetic control (Lopez-Villalobos 2012). Fat percentage may be controlled by relatively fewer genes each with larger effects, whereas protein percentage may be controlled by far more genes each with smaller effects (Lopez-Villalobos 2012).

Razmkabir (2011) estimated heritability ranges for daily milk yield as 0.11 to 0.25, for daily fat yield as 0.05 to 0.11 and for daily protein yield as 0.06 to 0.17 along the first lactation of Holstein cattle in Iran and the heritability estimates for fat and protein content in our study were nearly double to those for fat and protein yield with similar trend and this is in agreement with Druet et al. (2005) and a review study by Lopez-Villalobos (2012) in which he indicated that heritabilities of percentage traits are at least twice as high as those for yield. Shadparvar & Yazdanshenas (2005) estimates of daily fat percentage heritability for the first lactation of Holstein cows in Iran, ranged from 0.038 to 0.094 which nearly increased across DIM. Our estimates of daily heritability for fat content were in contrast to most other studies, in the way that they were much lower. Some researchers reported much higher values of heritability for daily fat and protein content (Druet et al. 2005, Penasa et al. 2010).

The estimated heritabilities for test day milk yield, fat and protein content were lower than most reported heritabilities for 305 day lactation records of these traits. While in Pander et al. (1992), the estimates for lactation concentrations of fat and protein were as 0.63 and 0.47, respectively and their estimates for individual test day records of these traits were lower and
ranged from 0.11 to 0.48 and 0.21 to 0.43, respectively, Jamrozik & Schaeffer (1997) showed higher heritability estimates for milk traits from test day models. Sahebhonar et al. (2011) estimated heritability of aggregate lactation fat and protein percentages for the Holstein population in Iran higher than of our estimates. Nilforooshan et al. (2008) and Toghiani et al. (2010) reported a 305 day fat percentage heritability greater than of our estimates. Shadparvar & Yazdanshenas (2005) estimated 305 day fat percentage heritability as 0.26 which were much higher than of their test day estimates. Razmkabir et al. (2009) reported much higher heritability for aggregate lactation fat and protein yield than estimates of Razmkabir et al. (2011) for test day yields. These evidences indicate that test day models resulted in lower heritability of milk fat and protein for the Holstein cows in Iran. One reason for this condition could come from the elimination of intra-animal phenotypic variation by aggregating several test day records to a single 305 day record. This component of variation could vastly increase residual variance of a test day model, if effects in model of analysis would not account for it. For this reason, in circumstances of high diversity of climates, environmental changes, management and feeding systems like in Iran, about traits like milk fat, for which an animal is highly sensitive to these factors, a test day model might result in much greater residual variance and hence lower heritability. This problem may be reduced by providing new sources of information into the model of analysis like diet quality, which could better account for intra-animal variation or dealing with herds more similar regarding to their condition and management. Ahrabi et al. (2005) using data from one herd, reported significantly higher heritabilities for daily fat yield compared to Razmkabir (2011) who studied a great number of herds within several provinces.

The estimates of genetic correlations between tests indicate that the breeding values of milk and protein content in far DIM are relatively closer compared to those of fat content and there were relatively higher correlations for permanent environmental effects for milk yield compared with fat and protein contents. Some researchers showed similar additive genetic and permanent environmental correlation coefficients between DIM for milk yield (Pool & Meuwissen 2000, Jakobsen et al. 2002, Costa et al. 2008, Bignardi et al. 2009). Jakobsen et al. (2002) reported additive genetic correlations between DIM for fat and protein yield similar to those obtained here for fat and protein contents.

The genetic trends demonstrated a positive genetic change for milk and negative changes for fat and protein contents (Figure 4). However, the negative genetic trends for fat and protein contents were not seen after the year 2002. High genetic change for milk curve could be attributed to the higher additive genetic variance compared to residual variance in milk and as a consequence of selection mainly for milk in Iran during the past years. Thus in addition to the higher genetic variance, the availability of national genetic evaluations for milk yield has accelerated genetic improvement in milk yield in recent years. Sahebhonar et al. (2011) reported genetic correlations between 305 day measure of milk, fat and protein contents as $-0.579 \pm 0.013$ and $-0.618 \pm 0.016$, respectively and a genetic correlation between 305 day measure of fat and protein contents as $0.697 \pm 0.014$. These negative genetic correlations might explain the opposite direction for genetic trends for contents compared to milk, indicating that selection for milk yield resulted in correlated negative response in fat and protein contents. The lower genetic change for the traits at the onset stage of lactation could be attributed to the lower genetic variation and heritability for these traits in early lactation.
Acknowledgements
The authors would like to thank the three anonymous reviewers for their valuable comments and suggestions.

References
Costa CN, de Melo CMR, Packer IU, de Freitas AF, Teixeira NM, Cobuci JA (2008) Genetic parameters for test day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. R Bras Zootec 37, 602-608


