Journal of Dairy Research

cambridge.org/dar

Using quantile regression for fitting lactation curve in dairy cows

Hossein Naeemipour Younesi[1,2], Mohammad Mahdi Shariati[1], Saeed Zerehdaran[1], Mehdi Jabbari Nooghabi[3] and Peter Løvendahl[4]

[1]Department of Animal Science, Ferdowsi University of Mashhad, 91779 Mashhad, Iran; [2]Department of Animal Science, University of Birjand, 97191 Birjand, Iran; [3]Department of Statistics, Ferdowsi University of Mashhad, 91779 Mashhad, Iran and [4]Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark

Research Article

Cite this article: Naeemipour Younesi H, Shariati MM, Zerehdaran S, Jabbari Nooghabi M and Løvendahl P (2019). Using quantile regression for fitting lactation curve in dairy cows. *Journal of Dairy Research* **86**, 19–24. https://doi.org/10.1017/S0022029919000013

Received: 23 August 2018
Revised: 26 October 2018
Accepted: 23 December 2018
First published online: 7 February 2019

Keywords:
Lactation curve; milk production traits; quantile regression; somatic cell score

Author for correspondence:
Mohammad Mahdi Shariati,
Email: mm.shariati@um.ac.ir

Abstract

The main objective of this study was to compare the performance of different 'nonlinear quantile regression' models evaluated at the τth quantile (0·25, 0·50, and 0·75) of milk production traits and somatic cell score (SCS) in Iranian Holstein dairy cows. Data were collected by the Animal Breeding Center of Iran from 1991 to 2011, comprising 101 051 monthly milk production traits and SCS records of 13 977 cows in 183 herds. Incomplete gamma (Wood), exponential (Wilmink), Dijkstra and polynomial (Ali & Schaeffer) functions were implemented in the quantile regression. Residual mean square, Akaike information criterion and log-likelihood from different models and quantiles indicated that in the same quantile, the best models were Wilmink for milk yield, Dijkstra for fat percentage and Ali & Schaeffer for protein percentage. Over all models the best model fit occurred at quantile 0·50 for milk yield, fat and protein percentage, whereas, for SCS the 0·25th quantile was best. The best model to describe SCS was Dijkstra at quantiles 0·25 and 0·50, and Ali & Schaeffer at quantile 0·75. Wood function had the worst performance amongst all traits. Quantile regression is specifically appropriate for SCS which has a mixed multimodal distribution.

Introduction

Mathematical models used in fitting longitudinal traits, such as lactation and growth, contribute toward better management decisions (Leclerc et al. 2008; Løvendahl & Chagunda, 2011). There is a long history of implementing mathematical models to explain the specific shape of the lactation curve (Wood, 1967; Cobby & Le Du, 1978; Adediran et al. 2012). The lactation curve has been fitted using either nonlinear Wood (Wood, 1967), Wilmink (Wilmink, 1987), or polynomial (Ali & Schaeffer, 1987) models. There is no single model for the lactation curve that is best in all aspects. From the statistical point of view, one modeling challenge is to capture the rapid increase in milk production before the peak yield (Wilmink, 1987). In physiological applications the ability to predict average energy balance and metabolic diseases is a major goal (Friggens et al. 2007; Friggens & Løvendahl, 2008; Sundrum, 2015).

All longitudinal models applied in fitting the lactation curve assume that there is a common shape of lactation shared by all animals; i.e. the mean curve. 'The mean curve' represents the quantile 0·50 under normal distribution assumptions. In this study, a new approach is chosen to fit different lactation curves evaluated at different production levels using quantile regression (QR). QR as a useful supplement to the least squares regression was introduced by Koenker & Bassett (1978), in economics. Conceptually, a quantile p corresponds to a production level that divides producing animals into two groups with either proportion p (higher production) or $1-p$ (lower production). This model is a flexible tool that allows a specific lactation curve at any production level of the trait. The quantile in which lactation curve is of interest can be pre-determined regarding economic and management criteria. Therefore, using QR can be more appropriate, for instance, where the pattern of lactation differs between high and low producing cows. A more insightful example is the level of somatic cell score (SCS) in mastitic cows which is not distributed as SCS in healthy cows (Rupp & Boichard, 1999; Nash et al. 2000).

Further, QR is extremely robust against outliers and heterogeneous residuals (Buchinsky, 1995). QR is being used in many research areas such as medicine (Wei et al. 2006; Beyerlein, 2014), economics (Chernozhukov & Hansen, 2006) and genetics (Huang & Lin, 2007; Briollais & Durrieu, 2014). Different functions can be accommodated in the QR framework based on the shape of a curve. In this paper, four widely used functions in animal science including Wood (Wood, 1967), Wilmink (Wilmink, 1987), Dijkstra (Dijkstra et al. 1997) and Ali & Schaeffer (Ali & Schaeffer, 1987) were examined.

© Hannah Dairy Research Foundation 2019

CAMBRIDGE UNIVERSITY PRESS

Applying QR can lead to making precise management decisions for each level of production, whereas regular linear models do not provide this information. The aim of this study was to introduce QR in modeling the lactation curve and to evaluate the goodness of fit of different functions in QR framework at different quantiles of milk production traits and SCS. In this study quantiles 0·25, 0·50 and 0·75 were chosen in order to estimate and compare lactation curves between low, medium and high producing cows.

Materials and methods

Data were recorded on 13 977 first lactation dairy cows in 183 herds from 1991 to 2011 by the Animal Breeding Center of Iran. The average number of test day records of each cow was 7·91 with a minimum of three and maximum of 10 records. The average daily milk yield, fat and protein percentage and SCS were 30·16 kg, 3·35, 3·06 and 4·47, respectively. All herds had three times milking program. Records with DIM <5 and >300 were removed. Age at first calving was between 20 and 40 months. The natural logarithm transformation of somatic cell count (SCC) was performed to decrease the deviation from normality (Ali & Shook, 1980; Ødegård et al. 2004). Some descriptive statistics of the data have been presented in Supplementary Table S1.

Four mathematical functions were used in the QR model to describe the lactation curve. In the QR theory there is no assumption underlying the distribution of error terms but randomness. Ignoring error terms, the systematic part of the functions under study were:

$$\text{Wood (WD)} \quad y_{\tau t} = at^b e^{-ct} \tag{1}$$

$$\text{Wilmink (WIL)} \quad y_{\tau t} = a + b \times e^{-0.046t} + ct \tag{2}$$

$$\text{Ali \& Schaeffer (AS)} \quad y_{\tau t} = a + b\left(\frac{t}{305}\right) + c\left(\frac{t}{305}\right)^2 + d\left[\ln\left(\frac{305}{t}\right)\right] + g\left[\ln\left(\frac{305}{t}\right)\right]^2 \tag{3}$$

$$\text{Dijkstra (DJ)} \quad y_{\tau t} = ae^{[b(1-e-ct)/(c-dt)]} \tag{4}$$

where $y_{\tau t}$ represents the trait; a, b, c, d and g are the parameters that define the scale and shape of the lactation curve; t is DIM. For milk yield, the Wood model is a gamma function, in which parameter a approximates the initial milk yield after calving, b is the inclining slope parameter up to the peak yield, and c is the declining slope parameter (Wood, 1967). In Wilmink model, parameter a is the level of production, b is the initial raise to the peak and c is the decreasing rate after the peak. The factor k was set to 0·046 (Wilmink, 1987; Macciotta et al. 2005) and is associated to the time of peak yield. The Ali & Schaeffer model was a second order regression of yields on days in lactation and on natural logarithm of 305-days yield divided by the day in lactation (Ali & Schaeffer, 1987). Dijkstra et al. (1997) proposed a mechanistic model with four parameters that describe the mammary growth pattern (cell proliferation and apoptosis) of mammals during pregnancy and lactation. The parameters of all functions were estimated using the 'nlrq' function of the 'quantreg' package in R (Koenker, 2017).

The models were fitted to data and compared using root mean square error (RMSE), Akaike's information criterion (AIC) and the log-likelihood.

RMSE is a kind of generalized standard deviation and was calculated as follows

$$\text{RMSE} = \frac{\text{RSS}}{\sqrt{n-p-1}} \tag{5}$$

where, RSS is the square root of residual sum of squares, n is the number of observation and p is the number of parameters in the model,

AIC was calculated using the following equation (Akaike, 1974)

$$\text{AIC} = -2\log(L) + 2p, \tag{6}$$

Therein L is the maximized value of the likelihood function for the estimated model and p is the number of estimated parameters.

Smaller values of RMSE, AIC and higher values of the log-likelihood ($\log(L)$) indicate a better fit when comparing models.

Quantile regression

The QR method is able to find more effects than classical regression and tests whether there is a change in the τth quantile of response variable for any certain $\tau \in (0, 1)$ instead of localizing on the variation in the mean of response variable (Koenker & Bassett, 1978). When the conditional distributions of response variable are non-Gaussian, the mean might not be the best summary, and any variation in distributions may not be detected. In QR, for fixed $\tau \in (0, 1)$ the following model can be determined

$$Q_\tau(y_i|\ \beta,\ x_i) = f(x_i, \beta_\tau) \tag{7}$$

where x_i is the vector of covariates and β_τ denotes a vector of unknown parameters on the τth (conditional) response quantile and is estimated such that proportion τ of the data having records below that quantile. In our analyses, covariate x_i was DIM, model f was either WD, WIL, AS or DJ and parameter vector β_τ was consisted of parameters of the models, namely, a, b, c, d, g. The model is defined as

$$y_i = f(x_i, \beta_\tau) + \varepsilon_i, \tag{8}$$

where ε_i is random error term which can take any distribution. The parameters β_τ can be estimated by solving the following

$$\hat{\beta}(\tau) = {}^{\arg\min}_{\beta \in R} \sum_{i=1}^{n} \rho_\tau(y_i - f(x_i, \beta_\tau)) \tag{9}$$

$$\rho_\tau(x) = x(\theta - I(x < 0) = \begin{cases} \tau_x & \text{if} \quad x \geq 0 \\ (\tau - 1)x & \text{if} \quad x < 0 \end{cases} \tag{10}$$

where, the QR loss function (10) is an asymmetric absolute loss

https://doi.org/10.1017/S0022029919000013 Published online by Cambridge University Press

Table 1. Goodness of fit statistics of different models in the same quantile

		Trait								
		Milk			Fatp			Prop		
Function	Quantile	AIC	Log *L*	RMSE	AIC	Log *L*	RMSE	AIC	Log *L*	RMSE
Wood	0·25	688 867	−344 430	0·02	255 097	−127 545	0·003	71 461	−35 727	0·001
Wilmink		**688 769**	−344 382	0·02	254 996	−127 495	0·003	70 451	−35 222	0·001
Ali & Schaeffer		688 790	−344 390	0·02	255 004	−127 497	0·003	**69 094**	−34 542	0·001
Dijkstra		688 808	−344 400	0·02	254 989	−127 490	0·003	69 247	−34 620	0·001
Wood	0·50	668 036	−334 015	0·02	245 801	−122 898	0·003	67 246	−33 620	0·001
Wilmink		667 823	−333 909	0·02	245 639	−122 816	0·003	66 126	−33 060	0·001
Ali & Schaeffer		667 912	−333 951	0·02	245 661	−122 825	0·003	**64 408**	−32 199	0·001
Dijkstra		667 856	−333 924	0·02	245 627	−122 809	0·003	65 527	−32 760	0·001
Wood	0·75	675 903	−337 949	0·02	264 906	−132 450	0·003	95 101	−47 548	0·001
Wilmink		**675 631**	−337 813	0·02	264 698	−132 346	0·003	94 191	−47 093	0·001
Ali & Schaeffer		675 718	−337 854	0·02	264 744	−132 367	0·003	**92 736**	−46 363	0·001
Dijkstra		675 652	−337 822	0·02	264 686	−132 339	0·003	92 808	−46 400	0·001

Fatp, fat percentage; Prop, protein percentage.

function that allocates weights τ and $(\tau - 1)$ to the positive and negative deviations. $L1$-norm estimator or least absolute deviation (LAD) estimator is obtained by taking $\tau = 1/2$ and is often chosen as a replacement to least squares estimators. It performs better in the attendance of heavy-tailed distributions and in the upper or lower quantiles of the response variable, especially in situations where data include outliers.

Results and discussion

Goodness of fit statistics for milk production traits from different QR models are presented in Table 1. In a given quantile, the RMSE between models was similar and could not be used for model comparison. In terms of AIC, across all models, the 0·50th quantile (**Q50**) was best in describing milk yield data followed by the Q75 and Q25. For fat and protein percentage Q50 had the lowest AIC, followed by the Q25 and Q75 (Table 1). Based on AIC and log *L*, across quantiles, functions WIL, DJ and AS performed best in analyzing milk yield, fat and protein percentage, respectively (Table 1). WD function was consistently poor for all milk production traits in different quantiles. Goodness of fit statistics for SCS with different QR models have been presented in Table 2. In terms of AIC, among quantiles across all models the Q25 was the best followed by the Q50 and Q75. The results indicated that WD function had the worst fit of lactation curve for SCS, whereas, DJ and AS functions were best based upon AIC (Table 2).

To our knowledge, there is no report on the performance of QR models on analysis of lactation curve, but our findings can be compared with other studies based on random regression models. Boujenane (2013) showed that based on RMSE, WD, AS and WIL models were similar in fitting lactation curve of Moroccan Holstein dairy cows. Quinn et al. (2005) reported that the AS model gave the best fit to the lactation curve of Irish cows, compared with the WIL model. The results obtained in this study were different to those of Olori et al. (1999), who found that AS and WIL provided residuals error smaller than WD function. Elahi Torshizi et al. (2011) reported that amongst models (WD, WIL, AS and DJ), based on RMSE the AS model was the best model and WIL the worst model for milk yield trait. In fitting WD, WIL, AS and DJ models to lactation curve of Australian Holstein dairy cows data, Adediran et al. (2012) showed that AS and DJ performed best and worst, respectively. In fitting average test day milk yield, Adediran et al. (2012) reported that regression with DJ function had the lowest Bayesian Information Criterion (BIC) compared to WD, WIL and AS functions. Dematawewa et al. (2007) analyzed extended lactation curve for milk, fat and protein yield in US Holsteins and found that WIL was best function based on BIC but WD, WIL and DJ produced similar RMSE. Some studies have reported that WD model is more appropriate in fitting lactation curve (Papajcsik & Bodero, 1988; Boujenane, 2013; Ferreira et al. 2015). Pakdel et al. (2010) applied six nonlinear models in different lactations to describe SCS and reported that at first and second lactations AS function was best, whereas in later lactations Morant & Gnanasakthy (**MG**) function was the best model. In the study of Rodriguez-Zas et al. (2000), AS model had the lowest MSE and the performance of WD model was worst and based on log Land AIC, MG and AS models were best, respectively.

QR can provide useful information on the shape and variation of lactation curves at different levels of production. We did not find similar QR analysis of lactation curve to compare the results. Ferreira et al. (2015) split the data to four sets according to the quantiles Q25, Q50 and Q75 and then employed a regression with several nonlinear functions to each of the four datasets. They showed that for low and medium producing cows WD function fitted best to the data and for high producing cows DJ had the best model fit. The problem with splitting the data is that only a fraction of the data is used in the analysis and the rest are kept out, whereas, in QR analysis parameters are estimated using a loss function based on whole data irrespective of which quantile is being considered (Koenker & Bassett, 1978).

https://doi.org/10.1017/S0022029919000013 Published online by Cambridge University Press

Table 2. Goodness of fit statistics of different models for all quantile for somatic cell score

Function	Quantile	AIC	Log *L*	RMSE
Wood	0·25	355 718	−177 856	0·006
Wilmink		**355 598**	−177 796	0·006
Ali & Schaeffer		355 598	−177 794	0·006
Dijkstra		**355 571**	**−177 781**	0·006
Wood	0·50	371 729	−185 862	0·005
Wilmink		**371 660**	−185 827	0·005
Ali & Schaeffer		371 647	−185 818	0·005
Dijkstra		**371 637**	**−185 815**	0·005
Wood	0·75	397 960	−198 977	0·006
Wilmink		**397 969**	−198 981	0·006
Ali & Schaeffer		**397 914**	**−198 952**	0·006
Dijkstra		397 930	−198 961	0·006

The estimated parameters of WD, WIL, AS and DJ models at different quantiles for milk yield, fat and protein percentage and SCS have been shown in Tables S1, S2, S3 and S4 of the supplementary file. For milk yield all parameters at the three quantiles were significantly different from zero ($P < 0{\cdot}01$), except in AS function, where parameter *g* at Q25 and Q50 was not significantly different from zero ($P > 0{\cdot}05$). Over quantiles, parameter *a* increased and *b* decreased in WD and WIL functions. Parameter *c* was quite constant in the WD and WIL functions. The trend of parameters in AS function over quantiles was quadratic such that the highest absolute values were observed in Q50 followed by Q75 and Q25. In DJ function, parameter *a* was increasing with quantiles and parameters *b* and *c* were constant (Supplementary Table S1). Scott et al. (1996) reported that the WD function has the tendency to overestimate milk yield prior to the peak yield and at the final stage of lactation. Elahi Torshizi et al. (2011) reported that the WD function underestimates milk yield in the days 5 to 100 of lactation.

For fat percentage all parameters at the three quantiles were significant ($P < 0{\cdot}05$), except, parameter *g* of AS at Q50, which was not statistically different from zero ($P > 0{\cdot}05$). Parameter *a* in both WD and WIL functions increased over quantiles, whereas, parameter *c* did not differ among quantiles. Parameter *b* was increasing by quantile in WIL, but constant in WD function (Supplementary Table S2). There was an irregular trend for parameters of AS function over quantiles. In DJ function parameters *a* and *d* increased and parameters *c* and *b* were constant over quantiles (Supplementary Table S2).

For protein percentage all parameters at different quantiles were significant ($P < 0{\cdot}01$), except, in AS function, where, parameter *b* at Q50 and Q75 and parameter *c* at Q75 were not significantly different from zero ($P > 0{\cdot}05$). In general, standard errors of the parameters were much smaller than the ones of fat percentage (Supplementary Table S3). Therefore, the estimated parameters are more reliable. For protein percentage parameter *a*, that models the initial production level was increasing in all functions. It was a general pattern that parameter *c* in WIL and WD functions were always constant that shows declining (for milk yield) or increasing (for fat and protein percentage) slope is not changed by production level (Supplementary Tables S1–S3).

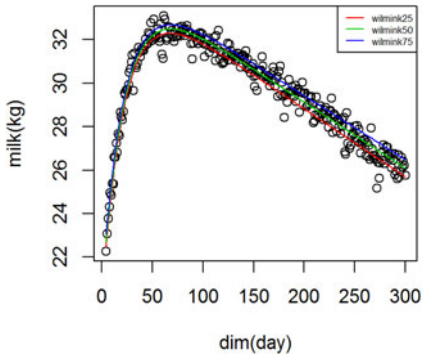


Fig. 1. Wilmink lactation curves for milk yield across quantiles.

For SCS all parameters at different quantiles were significant ($P < 0{\cdot}01$), except in AS function, where, parameter *g* at Q25 and Q50 and parameter *d* at Q50 were not significantly different from zero ($P > 0{\cdot}05$). Parameter *a*, the initial level of SCS, showed an increasing trend over quantiles, that was similar to other studied traits (Supplementary Table S4).

Inadequacy of WD function in fitting lactation curve of SCS has been reported by Pakdel et al. (2010) and Rodriguez-Zas et al. (2000) in random regression framework.

From management and breeding point of view, the aim is not only fitting the best model, but also understanding the details of lactation at specific levels of production. QR is a tool to disentangle different shapes of lactation curves among different production levels. The best model fit was at Q50 for milk production traits. It is a natural finding as milk production traits are always assumed as normal traits and it is expected that estimation of the lactation curve based on the 'mean' in linear models or 'median' in quantile regression produce similar curves. Nonetheless, SCS is a different trait that is not perfectly normal even after log transformation. For SCS best model fit occurred at Q25 and the curves at different quantiles were further apart. This is due to the mixture distribution underlying somatic cell count of healthy cows, subclinical and clinical mastitic cows (Madsen et al. 2008). QR has the advantage that it provides more information at different production levels which can lead to making management or selection decisions better.

Figure 1 shows WIL lactation curves at three quantiles for milk yield that performed best among all models. For fat and protein percentage best models across quantiles were DJ and AS, respectively (Figs. 2 and 3). For each trait a specific model performed best and the ranking among models was not altered at different quantiles. It can be observed visually and from model fit criteria in Table 1 that milk yield and fat percentage were fitted adequately, such that the peak, lactation curve extremes, and persistency were captured by all models with negligible differences. For protein percentage all models except AS and DJ failed to capture the high protein percentage in first days and its rapid decline afterward. The differences between lowest and highest AIC of four models at Q50 for milk yield, fat and protein percentage were 378, 174 and 2836, respectively, that shows the comparable performance of the models for milk yield and fat percentage but not for protein percentage.

Figure 4 Shows lactation curves based on best models at each quantile for SCS. The variation in SCS was higher compared to milk production. In genetic analysis, SCS is analyzed as a proxy for mastitis because these are correlated traits (Heringstad et al.

https://doi.org/10.1017/S0022029919000013 Published online by Cambridge University Press

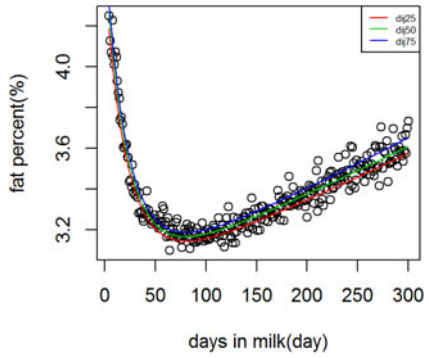


Fig. 2. Dijkstra lactation curves for fat percentage across quantiles.

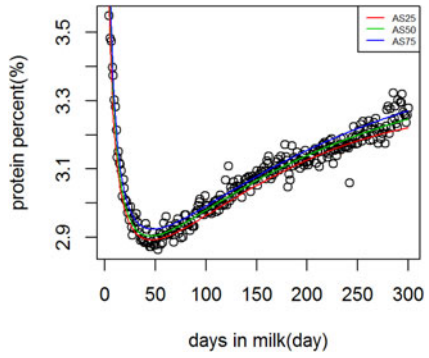


Fig. 3. Ali & Schaeffer lactation curves for protein percentage across quantiles.

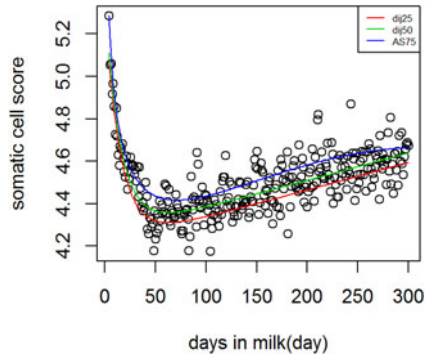


Fig. 4. Dijkstra and Ali & Schaeffer lactation curves for somatic cell score across different quantiles.

2000). In this study QR was used to separate the population based on three levels of SCS production, that is not necessarily an indicator of healthy cows *vs.* mastitic cows. However, it is possible to study different shapes of SCS during lactation among cows that are very low in SCS and susceptible cows to mastitis.

This study aimed at introducing QR in fitting lactation curve and growth pattern. Software packages used in animal breeding estimate these curves and predict genetic merit of animals in a linear mixed model analysis (Madsen & Jensen, 2008; Gilmour et al. 2015). Combining QR and high dimensional random genetic effects in a single mixed model QR analysis is the next development.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0022029919000013.

Acknowledgments. The authors are grateful to the Animal Breeding Center of Iran for providing the data.

Financial support. None.

Competing interest. The authors declare they have no competing interest.

References

Adediran S, Ratkowsky D, Donaghy D and Malau-Aduli A (2012) Comparative evaluation of a new lactation curve model for pasture-based Holstein-Friesian dairy cows. *Journal of Dairy Science* **95**, 5344–5356.

Akaike H (1974) A new look at the statistical model identification. *IEEE transactions on Automatic Control* **19**, 716–723.

Ali T and Schaeffer L (1987) Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science* **67**, 637–644.

Ali A and Shook G (1980) An optimum transformation for somatic cell concentration in Milk1. *Journal of Dairy Science* **63**, 487–490.

Beyerlein A (2014) Quantile regression—opportunities and challenges from a user's perspective. *American Journal of Epidemiology* **180**, 330–331.

Boujenane I (2013) Comparison of different lactation curve models to describe lactation curve in Moroccan Holstein-Friesian dairy cows. *Iranian Journal of Applied Animal Science* **3**, 817–822.

Briollais L and Durrieu G (2014) Application of quantile regression to recent genetic and-omic studies. *Human Genetics* **133**, 951–966.

Buchinsky M (1995) Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. *Journal of Econometrics* **68**, 303–338.

Chernozhukov V and Hansen C (2006) Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* **132**, 491–525.

Cobby J and Le Du Y (1978) On fitting curves to lactation data. *Animal Science* **26**, 127–133.

Dematawewa C, Pearson R and VanRaden P (2007) Modeling extended lactations of Holsteins. *Journal of Dairy Science* **90**, 3924–3936.

Dijkstra J, France J, Dhanoa M, Maas J, Hanigan M, Rook A and Beever D (1997) A model to describe growth patterns of the mammary gland during pregnancy and lactation. *Journal of Dairy Science* **80**, 2340–2354.

Elahi Torshizi M, Aslamenejad A, Nassiri M and Farhangfar H (2011) Comparison and evaluation of mathematical lactation curve functions of Iranian primiparous Holsteins. *South African Journal of Animal Science* **41**, 104–115.

Ferreira AG, Henrique DS, Vieira RA, Maeda EM and Valotto AA (2015) Fitting mathematical models to lactation curves from Holstein cows in the southwestern region of the state of Parana, Brazil. *Anais da Academia Brasileira de Ciências* **87**, 503–517.

Friggens NC and Løvendahl P (2008) The potential of on-farm fertility profiles: In-line progesterone and activity measurements. In Fertility in dairy cows: bridging the gaps. In Royal M.D., Friggens N.C. and Smith R.F. (eds), *British Society of Animal Science*. UK, Cambridge: Cambridge University Press, pp. 72–78.

Friggens N, Ridder C and Løvendahl P (2007) On the use of milk composition measures to predict the energy balance of dairy cows. *Journal of Dairy Science* **90**, 5453–5467.

Gilmour A, Gogel B, Cullis B, Welham S & Thompson R (2015) *ASReml User Guide Release 4·1 Structural Specification*. Hemel Hempstead: VSN International Ltd.

Heringstad B, Klemetsdal G and Ruane J (2000) Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. *Livestock Production Science* **64**, 95–106.

Huang B and Lin DY (2007) Efficient association mapping of quantitative trait loci with selective genotyping. *The American Journal of Human Genetics* **80**, 567–576.

Koenker R (2017) quantreg: Quantile Regression. R package version 5.33. Available at: http://CRAN.R-project.org/package=quantreg

Koenker R and Bassett G (1978) Regression quantiles. *Econometrica: Journal of the Econometric Society* **46**, 33–50.

Leclerc H, Duclos D, Barbat A, Druet T and Ducrocq V (2008) Environmental effects on lactation curves included in a test-day model genetic evaluation. *Animal: An International Journal of Animal Bioscience* **2**, 344–353.

Løvendahl P and Chagunda M (2011) Covariance among milking frequency, milk yield, and milk composition from automatically milked cows. *Journal of Dairy Science* **94**, 5381–5392.

Macciotta NPP, Vicario D and Cappio-Borlino A (2005) Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *Journal of Dairy Science* **88**, 1178–1191.

Madsen P & Jensen J (2008) A user's guide to DMU: a package for analysing multivariate mixed models, version 6, release 4. Danish Institute of Agricultural Sciences, Tjele, Denmark.

Madsen P, Shariati MM and Ødegård J (2008) Genetic analysis of somatic cell score in Danish Holsteins using a liability-normal mixture model. *Journal of Dairy Science* **91**, 4355–4364.

Nash D, Rogers G, Cooper J, Hargrove G, Keown JF and Hansen L (2000) Heritability of clinical mastitis incidence and relationships with sire transmitting abilities for somatic cell score, udder type traits, productive life, and protein yield. *Journal of Dairy Science* **83**, 2350–2360.

Ødegård J, Heringstad B and Klemetsdal G (2004) Bivariate genetic analysis of clinical mastitis and somatic cell count in Norwegian dairy cattle. *Journal of Dairy Science* **87**, 3515–3517.

Olori V, Brotherstone S, Hill W and McGuirk B (1999) Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livestock Production Science* **58**, 55–63.

Pakdel A, Heydaritabar M and Nejati Javaremi A (2010) The feasibility of nonlinear models to describe the milk somatic cell score of Iranian holstein cows throughout different lactation periods. *Iranian Journal of Animal Science* **41**, 185–192 (in Persian).

Papajcsik I and Bodero J (1988) Modelling lactation curves of Friesian cows in a subtropical climate. *Animal Science* **47**, 201–207.

Quinn N, Killen L and Buckley F (2005) Empirical algebraic modelling of lactation curves using Irish data. *Irish Journal of Agricultural and Food Research* **44**, 1–13.

Rodriguez-Zas SL, Gianola D and Shook GE (2000) Evaluation of models for somatic cell score lactation patterns in Holsteins. *Livestock Production Science* **67**, 19–30.

Rupp R and Boichard D (1999) Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. *Journal of Dairy Science* **82**, 2198–2204.

Scott T, Yandell B, Zepeda L, Shaver R and Smith T (1996) Use of lactation curves for analysis of milk production data. *Journal of Dairy Science* **79**, 1885–1894.

Sundrum A (2015) Metabolic disorders in the transition period indicate that the dairy cows' ability to adapt is overstressed. *Animals* **5**, 978–1020.

Wei Y, Pere A, Koenker R and He X (2006) Quantile regression methods for reference growth charts. *Statistics in Medicine* **25**, 1369–1382.

Wilmink J (1987) Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science* **16**, 335–348.

Wood P (1967) Algebraic model of the lactation curve in cattle. *Nature* **216**, 164–165.

https://doi.org/10.1017/S0022029919000013 Published online by Cambridge University Press