Institut für Agrar- und Ernährungswissenschaften
der Naturwissenschaftlichen Fakultät III
der
Martin-Luther-Universität Halle-Wittenberg

Impact of dealing with outliers on variance estimations and genetic evaluations with milk yield test-day models in German Holstein cows

Dissertation

zur Erlangung des akademischen Grades
doctor agriculturarum (Dr. agr.)

vorgelegt von

M.Sc. Agraringenieur
SeyedFaridoddin Rezazadeh Baei
geb. am 11.09.1972 in Babol, Iran

Gutachter:
    Prof. Dr. Hermann H. Swalve
    Prof. Dr. Lutz Schüler
    PD. Dr. Sven König

Verteidigung am: 26.10.2009

Halle/Saale 2009

Impact of dealing with outliers on variance estimations and genetic evaluations with milk yield test-day models in German Holstein cows

Abstract

Due to a large amount of data collected in genetic evaluation systems, it is clear that the presence of errors is nearly inevitable, therefore, outliers or abnormal test-day (TD) yields are common in production data of dairy cattle. Such abnormal measurements may adversely affect genetic evaluation if not accounted for properly. Thus, appropriate quantitative tools have to be devised for detecting and treating outliers. The examination of residuals is one of the earliest methods proposed for detecting outliers. Scaling the residuals by dividing by standard error is an alternative in the situation that they have different variances.

For the purpose of this study, two different treatment schemes, discarding and adjusting of outliers have been implemented prior to variance component estimation using a fixed regression TD model. Each of the two treating schemes were applied to three different regression procedures; simple (SR), robust (RR) and local regressions with both linear (LR1) and quadratic (LR2) polynomial degrees, to remove or to reduce the effects of TD records that are labelled as outliers. In order to do the discarding scheme, various percentage of outlying TD records (0.1, 0.2, 0.4, 0.8, 1.6, 3.2 and 6.4%) which represented the highest values of either absolute residual (ABSR) or absolute standardized residual (ASTAR) have been excluded.

The adjusting procedure consisted of estimating residuals from each three regression estimation procedures, computing the standard deviation for each cow, and restricting the outlier residuals to be within k standard deviations (SD). Thus, a new observation (=k×SD) was created for use in the genetic evaluation.

The objectives were to compare three regression procedures applied to 111,599 TD milk yield records of 11,620 German Holstein cows and to determine the impact of treating outlying milk records on numbers and distributions of outliers, mean of absolute residuals (MAR), mean of squared residuals (MSR), standard deviation of residuals (SDR), variance components, genetic and non-genetic parameters, mean estimated breeding values (MEBV) and rankings of animals. Results were compared to the original dataset with any kind of manipulation. Milk yields in the first lactations were analysed by the model that included the fixed factors of herd-year-season of calving, herd-test-date effect, fixed regression coefficients of the Ali and Schaeffer function (1987) for days in milk (DIM) and random effects of animal and permanent environment.

Preliminary analyses showed both the existence of outliers in the data and dissimilarity between distributions of outliers across regressions and schemes. Relative frequency of outliers indicated that more abnormalities occurred at a specified stage of lactation depending on the regression technique. However, SR and RR showed a different pattern from local regressions (i.e. LR1 and LR2) where the low outliers were less than 50% of total outlying records. By using ASTAR instead of ABSR, the residual terms MAR, MSR and SDR more slowly decreased but ASTAR represented less dependency on DIM and production level.

In all cases, both phenotypic and error variances decreased with increasing number of outliers depending on the method and percentage of treated outliers. Adjusting of outliers less affected the MEBV of animals than discarding them, and cows displayed more fluctuations in MEBV than bulls. Rankings of top one hundred bulls and cows did not differ much in comparison with complete data. According to the results of this research, the RR method seemed to be preferred over the methods SR, LR1 or LR2 because it affects a smaller proportion of treated data. The findings of the current study also suggest that the effectiveness of each regression procedure in detection of outliers is a function of goodness of method-fitting to the data and incidence frequency of outliers during lactation.

Though the results of residual criteria did not show any strong evidence to determine the proper coefficient k for each treating scheme, the model performance was generally improved using both discarding and adjusting methods. However according to the results, especially based on standard deviation of additive genetic and permanent environment parameters, discarding 0.4% to 0.8% and adjusting about 1% to 5% outliers are recommended depending on the regression method. Selection of the best coefficient (k) highly depends on the data set, model, regression strategy and definition of limitations for high and low TD yields whether discarding or adjusting schemes are applied.

Dealing with outlying records contains two main parts. The first is detection of outliers using a suitable model, and the second is treating them through either discarding or adjusting approaches by determining limits for low and high boundaries. If the model does not accurately fit the data or limitations are not correctly defined, some data will incorrectly be defined as outliers as well as some outliers will not be detected.

Acknowledgments

When it comes to the time to write this page, it would not be easy to enumerate all of the support that I have received throughout my research. I am sure that without many of the people, the thesis completion would not have become a reality.

Foremost, I would like to express my sincere gratitude to my advisor Prof. Dr. Hermann H. Swalve for offering this exciting project, for providing data, for the continuous support of my PhD study, and for his patience, motivation, enthusiasm, and immense knowledge. It has been both a pleasure and honour to share with him ideas and meaningful discussions on many subjects related to my research and program of study over last four years.

Besides my advisor, I would like to thank the rest of my thesis committee members: Prof. Dr. Lutz Schüler and PD. Dr. Sven König for their encouragement, insightful comments, and useful questions. Their guidance helped me to improve my thesis and to complete my studies.

My deepest thanks to Ms. Dr. Monika Wensch-Dorendorf for helping me to analyze the data sets under Linux operation system. I would also like to thank Prof. Dr. Strandén for posting his papers from Agrifood Research Center, Finland. This was a great help.

Additionally, I must also thank Dr. Nobert Mielenz for helping me with statistical problems. In particular, I am grateful to Ms. Dr. Renate Schafberg for supporting me when I arrived in Halle\Saale. I will never forget your kindness towards me and my wife.

I want to thank Ms. Heike Braunsdorff, librarian of the Institute, for her help and guidance in finding the needed literatures. I wish to thank Ms. Dr. Kerstin Brügemann who paid attention to me whenever I needed.

This thesis would not be possible without support of my wife, Atefeh who has supported me through the completion of my degree. Thank you for your love, encouragement and providing this opportunity.

Last but by no means least, I especially appreciate Dr. Joachim Wussow whose help and support makes me remember forever.

List of Contents

List of Tables

List of Figures

1. *Introduction*

Breeding programs and management decisions regarding milk production always start from an analysis of lactation yields generated from sets of test-day (TD) records made during lactation period. Since TD milk yield records of dairy cows are the most important ingredient in dairy cattle breeding programs, it is essential to keep them away from contaminations using proper methods.

1.1. Outliers problem with milk yield records

Selection of the best animals for genetic improvement requires the prediction of breeding values which must be deduced from phenotypic data. However, the accuracy of estimating breeding values depends on the quality and the accuracy of the phenotypic and pedigree data available that usually are large in size. The presence of errors in the data is nearly unavoidable and consequently outliers commonly take place in performance data such as TD yields. Therefore, it is difficult to define when a TD record becomes an outlier because there are biological and technical reasons.

The International Committee for Animal Recording (ICAR) recommends that the data entering a country genetic evaluation system should have high quality, irrespective of how quality is defined (Interbull 2001). However, data obtained in a wide range of applications frequently contain atypical observations called outliers (outlying observations). Yang et al. (2004) stated that an outlier or an abnormal TD yields is a 24-h daily milk (fat, protein or other trait) yield which is significantly located outside the range of the rest of the yields made by cows of similar status (i.e. age, stage of lactation, herd or contemporary group). This definition reflects the statistical character of outliers.

Some studies have reported bias of predicted breeding values in genetic evaluation of dairy cattle (Kuhn et al. 1994, Kuhn and Freeman 1995, Lidauer and Mantysaari 1996). Strandén and Gianola (1997) supposed that this bias may be due to outliers that cannot be predicted by the statistical model.

Although the cause for a TD record to be an abnormally low or high milk yield is usually not explainable, few reasons have been mentioned by several authors such as selective reporting or misreporting of the performance records (Bertrand and Wiggans 1998), preferential treatment of a cow (Strandén and Gianola 1997, 1998, 1999), sickness, injury, error in recording (Wiggans et al. 2003), chance, data manipulation, human error (Yang et al. 2004), inappropriate contemporary group formation, animal misidentification (Cardoso et al. 2006), measurement errors, health status, short-term changes in the herd environment and mismanagement of the data (Jamrozik et al. 2007). In every case, regardless of the source of outlying observations, a measured 24-h TD yield may not be representative of either the cow's actual yield or her genetic ability if it is considered as an outlying TD records (Wiggans et al. 2003). This implies that an outlier can be an observation without to be an apparent error caused by above mentioned reasons.

Dealing with outliers is also a difficult task in all scientific areas, but since milk production of dairy cows is influenced by genetic and environmental effects and some kinds of interaction of those effects, e.g. between sire and herd (Suzuki et al. 1998) it would be more complex and consequently, there remains a need for more investigations.

Outliers present some problems in animal breeding analyses. Such data cannot be accommodated in the model leading to bias in the prediction of breeding values (Kuhn et al. 1994, Kuhn and Freeman 1995, Strandén and Gianola 1999). Presence of outliers in the data also results in reduction of accuracy and reliability of estimates through inflating the sum of squares of residuals or prediction error variances for estimated genetic parameters and breeding values and leads to misleading results in genetic evaluations (Yang et al. 2004).

In the last ten years, interest in test day models (TDM) using TD records has increased among dairy breeders because of several advantages (Swalve 1995b, 1998, 2000). In addition to accounting for environmental factors, an advantage of a TDM as fixed regression model is that it may account for differences in the shape of the lactation curve (Bormann et al. 2003). In this context, several mathematical functions have been proposed (Wood 1967, Ali and Schaeffer 1987, Wilmink 1987, Guo and Swalve 1995) differing mainly in the type of regression, in the number of parameters and shapes of lactation curves when milk data are fitted on individual cows (Macciotta et al. 2005). On the other hand, due to the effect of both environmental and genetic factors leading to the random variations in shapes of lactation curves between cows (Rekik and Ben-Gara 2004) a wide range of goodness of fit has been described by researchers (Perochon et al. 1996, Olori et al. 1999, Tekerli et al. 2000). It has been reported that atypical lactations are poorly described by standard models in contrast to typical lactations (Olori et al. 1999). The standard deviations related to means of estimates of parameters for atypical curves are large, suggesting an important heterogeneity of atypical curves (Rekik and Ben Gara 2004). Therefore it is necessary to use suitable models when one attends to detect outlying TD records. It can be achieved by using robust regressions which are robust to the presence of outliers in least-squares models or local smoothing regressions which fit lactation data without the need for a predefined model.

It is concluded from the above discussions that treating outliers by a proper approach which takes into accounts the data departures, may be capable of attenuating the impact of outliers on data analysis. Several procedures are available in order to detect outliers, and each may be applied to prediction residuals generated from the regression models implying that extreme observations can only be determined according to the model used for regression analysis. There is a great deal of doubt about the ways to deal with such abnormal records when estimating lactation yields or breeding values.

One of the earliest proposed methods to detect outliers is the examination of ordinary residuals or some transformations of them. Regardless of which method (excluding or adjusting) is used, these methods are implemented prior or parallel to data analysis. Although a few researchers have studied the impact of detection and correction of outliers (Wiggans et al. 2003, Yang et al. 2004, Jamrozik et al. 2007), there are only a few literary sources which deal with the outliers in the genetic analysis using TDM and analyse the question whether such abnormal recording should be discarded or adjusted.

1.2. The study purposes

What emerges from above considerations and discussions is that the certain cause of an outlying observation can not be determined especially in large data sets that are common in dairy cattle genetic evaluations. Because there are different and controversial opinions how to deal with outliers, in this study some methods will be applied on Holstein data from Northern Germany to analyze the consequences of this data manipulation. The main objectives of the current study are: 1) examination of three regression methods to detect outlying yields, 2) excluding or adjusting the outlying yields when estimating lactation yield and, 3) investigation whether the accuracy of lactation yield estimates is improved by the use of such methods. Results from the procedures will be compared to those from the entire dataset without treatment. Specific objectives included: comparison of fit statistics of different regression models, method comparisons by residual statistics, analysis of distributions of the identified outliers over the lactation period, comparison of rankings of animals (bulls and cows) from different evaluation methods and examination of the influence of using these methods on genetic and non-genetic parameters and their standard errors.

2. *Literature review*

The prediction of breeding values constitutes an integral part of most breeding plans for genetic improvement (Mrode 1996). This prediction is implemented by selection of animals of higher genetic merit than average to be parents of the next generation (Cameron 1997). Merit that is deduced via genetic evaluation, is a linear or nonlinear combination of genetic values for several economical traits. Genetic merit is not observable, so it must be inferred from data (Gianola 2000). Crucial for the accurate prediction of breeding value is the availability of records (Mrode 1996). Linear models are used for evaluation of linear, normally distributed traits such as milk yield (Van Raden 2006) and in the most sciences such as applied statistics data (Pedhazur 1982, Nevitt and Tam 1997).

In the linear model

$$y_i = x_i \beta + \varepsilon_i \qquad \rightarrow \qquad \hat{y}_i = x_i \beta \tag{2.1}$$

where the ε_i is the differences between the observed values of y_i (response variable) and those predicted by the model ($\hat{y}_i$) and x_i is the value of the explanatory variable(s) for the *i*th individual. In general the least-squares approach is adopted to estimate model parameters (β) and residuals (ε). The error term represents the cumulative effect of factors not specified in the model (Rencher and Schaalje 2008).

Since the error ε is unobservable, it must be estimated indirectly using predicted residuals, $\hat{\varepsilon}$ (Zewotir and Galpin 2007). The least-squares regression lines are the best-fitting line to the data when its underlying assumptions are satisfied: a) each of the metric variables is normally distributed, b) the relationships between metric variables are linear, and c) the relationship between response and explanatory variables is homoscedastic. When all three assumptions hold, the least-squares estimators of the β's have minimum variance (Rawlings et al. 1998) otherwise the estimators may be poor and the resulting analysis may provide results of questionable validity (Freund and Littell 2000).

Although violations of the assumptions are often not so severe as to invalidate the analysis (Freund and Littell 2000) these assumptions are frequently violated by the presence of outliers in the data (Nevitt and Tam 1997). Thus, it is useful to examine the data for possible violations and, if violations are found, to employ remedial alternatives (Freund and Littell 2000). In the following pages, more comprehensive details on the outlier topic will be presented.

2.1. Outlier definitions

Though there are various definitions, an outlier is generally considered as a data observation that is far from the pattern for a variable or sample (Stevens 1984). Hawkins (1980) described outliers as observations that deviate so much from other observations and are generated by a different mechanism and may cause doubt about the validity of the results. Wainer (1976) also entered the concept of the fringelier (a marginally extreme value or case), referring to unusual events which occur rarely. Barnett and Lewis (1984) defined an outlier in a set of data as "an observation (or subset of observations) which appears to be inconsistent with that set of data". Likewise, Johnson (1992) considers an outlier as an observation in a data set which appears to be inconsistent with the rest of that data set.

2.2. Sources of outliers

Outliers can arise from several different mechanisms or causes. One of the challenges of working with outliers is that the researcher is rarely sure of the reason for the outlying observations (Evans 1999). Therefore, it is important to study the causes that may be responsible for outliers in a given data set because what should be decided about an outlier is somewhat a function of the derived cause (Osborne and Overbay 2004). Anscombe (1960) classified outliers into two major categories: those arising from errors in the data, and those arising from the inherent variability of the data. Potential sources of outliers are: measurement and recording errors, incorrect distribution assumption, unknown data structure, or inherent variability (Iglewicz and Hoaglin 1993, Barnett and Lewis 1994).

2.2.1. Measurement or recording errors

Neither measurement device nor researcher is completely infallible (Evans 1999). At times, errors may be made in the measurement, or due to faulty recording or coding of observations (Iglewicz and Hoaglin 1993) or as a result of an imperfect measurement apparatus (Anscombe 1960). Errors of this nature can often be corrected by returning to the original documents or even the subjects if necessary and possible and entering the correct value (Osborne and Overbay 2004). In these situations researchers may legitimately decide to reject or recode the observations (Anscombe 1960, Hecht 1991, Osborne and Overbay 2004). Iglewicz and Hoaglin (1993), however, advocate the recording of all outliers because if they reoccur in subsequent data collections, it may indicate the need to modify measurement or recording techniques.

2.2.2. Incorrect distributional assumptions

In addition to measurement and recording errors, outliers can occur in data sets due to incorrect distribution assumed for the analysis (Iglewicz and Hoaglin 1993). Douzenis and Rakow (1987) pointed out that the presence of outliers not only may indicate a weakness in the statistical model but may also distort statistics which assume a normally distributed sample. Thus, researchers should consider the assumptions which are made by each model in regression analysis and effectively handle this matter before collecting the data (Evans 1999).

2.2.3. Structure within the data

Another possible cause of outliers is that the data contains a different structure than is accounted for by the sampling method (Jacobs 2001). Iglewicz and Hoaglin (1993) offer an example of data that are presumed to come from random daily samples but actually come from another sample. In this case, the data may need to be investigated more fully before deciding whether to retain, recode, or reject the outlying observations (Evans 1999).

2.2.4. Execution error

Another factor to consider when studying outliers is that they may not be mistakes at all, but indicators that within a sample, such values are possible (Iglewicz and Hoaglin 1993). In such a case, the presence of an outlier can lead the researcher to an important discovery in terms of the potential for what is being studied. Anscombe (1960) pointed out that researchers do not always accomplish what they set out to do.

2.2.5. Inherent variability

Another source of outliers is the natural variation of the observations over the sample and unexpected events during the data generating process (Barnett and Lewis 1994). As the data set becomes larger, the probability for the occurrence of outlying values becomes greater (Evans 1999). An outlying observation may be merely an extreme manifestation of the random variability inherent in the data. If this is true, the values should be retained and processed in the same manner as the other observations in the sample (Grubbs 1969). Due to the deleterious effects that outliers can have, it might be desirable to use a transformation or truncation strategy to keep the observation in the data set and at the same time to minimize the impact on statistical results (Osborne and Overbay 2004).

However, defining the cutoff for what makes an extreme measurement an outlier requires somewhat individual judgment (Jacobs 2001). Not all outliers are illegitimate, and not all illegitimate scores show up as outliers (Barnett and Lewis 1994).

2.3. Classification of outlier detection methods

The methods of outlier detection can be classified into at least two groups a) univariate methods that proposed in earlier works in this field, and b) multivariate methods that have been commonly used in most researches (Ben-Gal 2005). Univariate analysis explores each variable in a data set, separately. Ben-Gal et al. (2003) categorize statistical process control methods by two major criteria: a) methods for independent data (*x* variables) versus methods for dependent data (*y* variable); and b) methods that are model-specific (e.g. least-squares), versus methods that are model-generic (e.g. smoothing local regression). Various outlier detection methods exist and will be briefly discussed presently.

2.4. Univariate outlier detection methods

A univariate method searches the data in *y*-axis or generally one variable to detect outliers. Univariate methods are favoured for their simplicity, but these methods cannot detect observations which violate the correlation structure of the dataset (Franklin and Brodeur 1997). Usual methods of univariate outlier detection have been listed as below:

2.4.1. Scoring of residuals

One method of identifying univariate outliers is simply converting the data values to $\underline{Z}$ scores and screening for high absolute values (Donzenis and Rakow 1987). A $\underline{Z}$ score is the observed value minus the mean, divided by the standard deviation. It represents the number of standard deviations it falls from the mean. However, this rule doesn't always yield the desired result particularly in small data sets (Shoemaker 1999).

2.4.2. Scatter plots

Anscombe and Tukey (1963) encouraged beginning any analysis of residuals by looking at a scatter plot. If outlying samples are present, these will usually show up as points in the score plots lying outside the normal range of variability (Evans 1999, Naes et al. 2002). They provide valuable information not only about the presence of outliers but also the adequacy of the model and/or the validity of its associated assumptions. The patterns of residuals are informative about the type of violation (Serdahl 1996, Zewotir and Galpin 2007). For this purpose, scatter plots of residuals against predictor variables can also help to detect outliers (Larsen and McCleary 1972). In addition, Zewotir and Galpin (2007) recommended using the studentized residuals to the ordinary residuals in the plot, because using them will not be complicated by the non-constant residual variance and will generally be more revealing than the ordinary residuals.

2.4.3. Box Plot

Perhaps the most popular univariate outlier detection technique is the quartile method proposed by Tukey (1977). Since the mean is very sensitive to outliers and to long tails in the distribution, in many situations statisticians instead use the median, which is generally much safer (Hampel et al. 1986). The method is not only robust, but also simple and non-parametric (Franklin and Brodeur 1997). A box plot is a graphical display of data that shows the location, spread, skewness and length of tails the measure of central tendency using median, beneath which 50% of the data points fall (Tukey 1977). An example for a box plot is shown in Figure 2.1.

The box surrounding the median represents the 75th-25th percentiles of the data (Jacobs 2001). The inter-quartile range (IQR) is a measure of the range of data and is the difference between the 25th (Q1) and 75th (Q3) percentiles (Meinert and Norman 1998). This method treats any value greater than the 75th percentile plus 1.5 times the IQR distance, or less than the 25th percentile minus 1.5 times the IQR distance as an outlier (Jacobs 2001). In the US, Bertrand and Wiggans (1998) reported that some edits for outliers based on IQR ranges are imposed for milk yield data from dairy herds sampled by their owners rather than a testing supervisor. Meinert and Norman (1998) classified outliers as possible or probable by IQR (Figure 2.1). Possible outliers are values that lie between 1.5 and 3 IQR ranges from the first quartile (Q1) or the third quartile (Q3). Probable outliers are values falling beyond 3 IQR range units from Q1 and Q3. They used the milk yield outliers detected by box plot method as indicators of accuracy of genetic evaluations of sires.

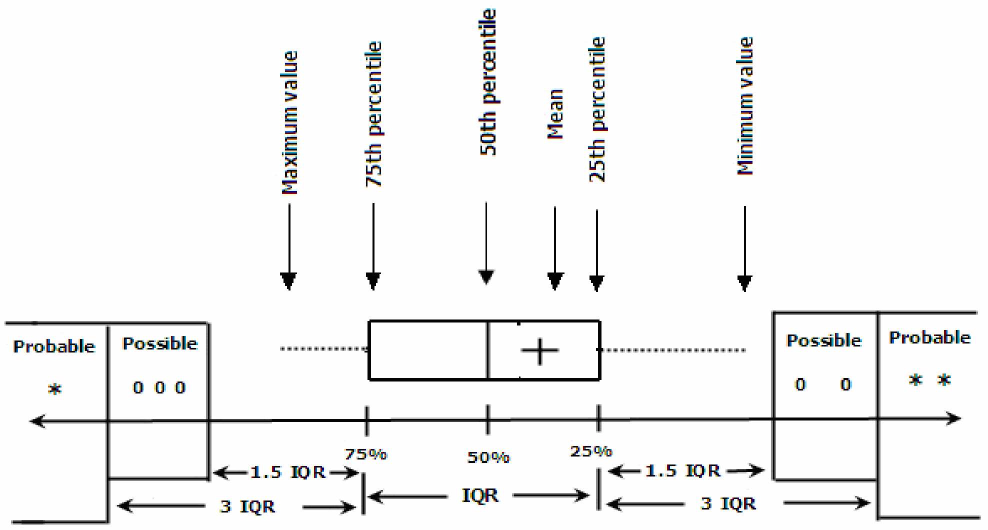

Figure 2.1: An example of a box plot.

2.5. Obtrusive effects of outliers on regression analysis

The aim of a regression analysis is to obtain estimates of the unknown parameters which indicate how a change in one of the independent variables affects the values of the dependent variable. The occurrence of outliers can result in increase in error rates and substantial undesired changes in parameter and statistic estimates when using either parametric or nonparametric analysis (Zimmerman 1994). First, they generally inflate error variance and reduce the power of statistical tests. Second, they can decrease normality if non-randomly distributed, altering the odds of making errors (both Type I and Type II) and in multivariate analyses violate model assumptions. Third, they can severely bias or influence estimates that may be of substantial interest (Osborne and Overbay 2004). Therefore, it is important to identify them prior to modeling and analyzing (Williams et al. 2002, Liu et al. 2004).

2.6. Multivariate outlier detection methods in regression analysis

The classical outlier detection methods are efficient when the data include only one outlier. However, if the data contains more than one outlying observation, the power of these methods decreases drastically, owing to the masking (failing to identify outliers) and swamping (mistaking clean observations for outliers problems). Therefore, a method which avoids these problems is needed (Hadi 1992). In order to decide about the outlier problem in regression analyses, it is also useful to distinguish between x and y outliers. From a practical perspective, three types of outliers are problematic in regression analysis (Rencher and Schaalje 2008).

On the one hand, some observations may fail to be predicted by the model with a reasonable degree of accuracy. The y-outliers are defined as those observations that have a different relationship between y and x variables (Naes et al. 2002). The y-outliers may occur in the case that several populations are mixed in a data set or that some relevant variables have been omitted (Coenders and Saez 2000). On the other hand, some observations are abnormally positioned relative to the majority of x-data (Naes et al. 2002). The presence of such observations in the data set substantially modifies the estimates and weakens the conclusions which may be drawn from the model (Coenders and Saez 2000).

Likewise, it often happens that an observation is an outlier according to the both y- and x-data, simultaneously (Coenders and Saez 2000). Figure 2.2 that was adopted from Rencher and Schaalje (2008), displays these three types of outliers in simple linear regression analysis.

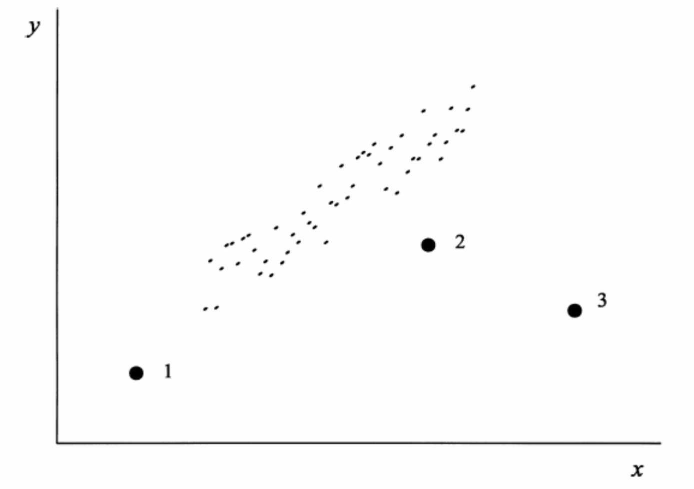

Figure 2.2: Simple linear regression showing three outliers.

The sample 1 is an outlier in *x* direction (leverage point) but is positioned close to the straight line fitted to the rest of the samples. This sample has little effect on the regression equation but it has the potential to be influential. The point 2 in Figure 2.2 is an outlier in *y* direction close to the average in *x*. In this case the regression equation is only slightly influenced by this outlying sample. However, the outlying sample 3 is abnormal both in *x* and in the relation between x and y. This sample has a very strong effect on the regression equation. It is said that the sample is influential. The statistics to measure these types of outliers will be discussed.

2.6.1 The y- Outliers and their detection

Outliers on the dependent variable typically exert greater influence on the parameter estimates and R^2 (Determination Coefficient) than do outliers on the independent variables (Hecht 1991, Serdahl 1996). The outlying data point on *y* pulls the regression line towards itself in an effort to minimize error variance (Serdahl 1996). Hecht (1991) found that analysis of the standardized and studentized residuals were the most effective diagnostic methods for identifying outliers on the *y* axis.

2.6.1.1. Residuals

The most common and traditional tool for detecting *y*-outliers is the residual that for observation *i* is defined as $\varepsilon_i = y_i - \hat{y}_i$. With linear models in Formula (2.1), the error vector ε is unobservable (Rencher and Schaalje 2008). In order to estimate ε for a given sample, the predict residual vector is used as:

$$\hat{\varepsilon} = y - \hat{y} = y - X\beta \qquad (2.2)$$

The prediction errors in a regression model are assumed to be a random variable that are normally and independently distributed with a mean of zero and common variance (Gill 1986). If the model and attendant assumptions are correct, a plot of the residuals versus predicted values or versus *i* (the observation number) should show no systematic pattern (Naes et al. 2002). These plots are therefore useful for checking the model (Rencher and Schaalje 2008). There are no exact rules, but some rules of thumb can be set up for when a residual is to be considered larger than expected (Martens and Naes 1989). Larger y can indicate errors in samples with a different relationship between x and y (Naes et al. 2002). But a popular approach is to detect an observation as an outlier if the magnitude of its residual, ε, exceeds the estimated population standard deviation, SD, multiplied by a designated constant, k, (Anscombe 1960, Anscombe and Tukey 1963, Hecht 1991).

If k is large, then the largest residuals will less likely exceed this product and they will less likely be rejected. If k is small, however, the product of k with SD will be smaller and the observation with the greatest residual will more likely be rejected (Evans 1999). Anscombe (1960) has discussed calculation of k with more detailed information on the. A typical plot of this type is shown in Figure 2.3.

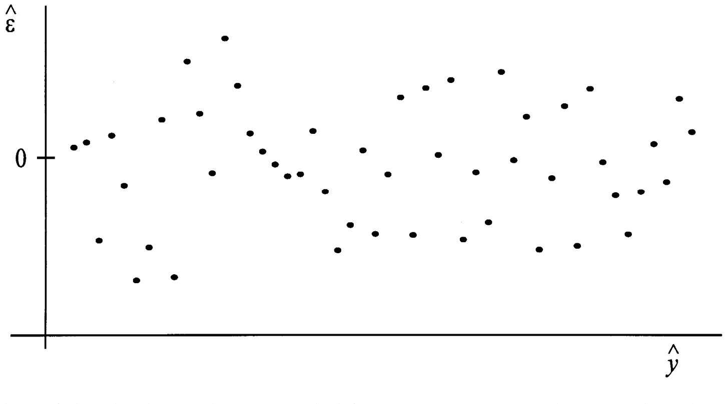

Figure 2.3: Ideal residual plot when model is correct (Rencher and Schaalje 2008).

Using the least-squares estimator $\hat{\beta} = (X'X)^{-1}X'y$, the vector of predicted values can be written as

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy \qquad (2.3)$$

where $H = X(X'X)^{-1}X'$. The $n \times n$ matrix H is called the Hat matrix because it maps y into $\hat{y}$ (Hoaglin and Welsch 1978, Rencher and Schaalje 2008). Using (2.3), the residual vector $\hat{\varepsilon}$ can be expressed in terms of H:

$$\hat{\varepsilon} = y - \hat{y} = y - Hy = (I - H)y = (I - H)\varepsilon \qquad (2.4)$$

2.6.1.2. Standardized residuals

The observed residuals, $\hat{\varepsilon}$, in real data, however, are not independent and do not have common variance as indicated by Formula 2.4 (Rawlings et al. 1998), because they are scale-dependent and their variances depend on the standard errors (Gill 1986).

$$\text{var}(\hat{\varepsilon}_i) = \sigma^2(1 - h_{ii}) \tag{2.5}$$

where h_{ii} is the ith diagonal element of the Hat matrix that the method for its calculation would be presented on the next page. Thus, it is desirable to scale the residuals so that they have the same variance σ^2. There are two common and related methods of scaling (Rencher and Schaalje 2008). For the first method of scaling, Belsley et al. (1980) suggest to standardize each residual that is independent of the residual. Dividing each residual by its SD (standard deviation) gives a standardized residual, denoted with r_i,

$$r_i = \frac{\hat{\varepsilon}_i}{\sigma\sqrt{1 - h_{ii}}} \tag{2.6}$$

where h_{ii} is the ith diagonal element of H and σ is square root of the error variance. All standardized residuals normally distributed with unit variance and mean 0 (Rawlings et al. 1998). In analysis of variance the biggest issue is the distance of an individual from the subgroup. Standardized residuals indicate the distance between any point and the sub-group, and hence are efficient in assisting analysts in examining data for multivariate outliers (Osborne and Overbay 2004).

2.6.2. The x- outliers or leverage points and their detection

Sometimes, even though data points may deviate from the mean, they may not necessarily impact the coefficient of determination or the regression equation (Evans 1999). The outliers on the x axis impact regression statistics, though to a smaller degree than do outliers on the y axis (Hecht 1991). They usually have more effect on the variability of the x scores than they do on the relationship between the variables (Hecht 1991) unless the extreme value of x shows a serious impact on y (Evans 1999).

2.6.2.1 Studentized residuals

A second method of scaling the residuals is the studentized deleted residual introduced by Belsley et al. (1980). For ith observation, it is the residual computed from a regression equation estimated without the ith observation divided by its SD, which is also computed without the ith observation. This prevents the ith observation from influencing its own prediction and from inflating the standard error with which it is being standardized (Coenders and Saez 2000). The result is the Studentized residual, denoted r_i^*:

$$r_i^* = \frac{\varepsilon_i}{\sigma_{(i)}\sqrt{1 - h_{ii}}} \tag{2.7}$$

where the subscript in parentheses indicates that the ith observation has been omitted. As with ε_i and r_i, the r_i^* are not independent of each other (Rawlings et al. 1998). The standardized and Studentized residuals can be obtained from the ordinary residuals without rerunning the regression with the observation omitted (Belsley et al. 1980). Cook and Weisberg (1982) refer to r_i as the studentized residual with internal studentization in contrast to external studentization for r_i^*. The r_i^* are called RSTUDENT by Belsley et al. (1980) and SAS Institute (Rawlings et al. 1998).

2.6.2.2. Hat matrix

The Hat matrix can be helpful for detecting outliers when outliers maintain enough influence on the regression line to make $\hat{y}_i$ close to y_i (Hoaglin and Welsch 1978, Iglewicz and Hoaglin 1993). The leverage of observation i is defined as h_{ii}:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^{n}(x_k - \bar{x})^2} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SS_x} = x_i'(X'X)^{-1}x_i \tag{2.8}$$

where $\bar{x}$ is the mean of the x variable. High leverage points (high values on the diagonal of the Hat matrix) suggest that the corresponding observation may be an outlier, if any h_{ii} be greater than twice the number of the predictors (Hoaglin and Wesch 1978, Evans 1999). The leverages can be plotted against observation number or against other diagnostic tools as residuals (Naes et al. 2002). Hoaglin and Welsch (1978) advocated the use of the Hat matrix followed by an examination of studentized residuals. The Hat matrix offers information about high leverage points, and the studentized residuals allow researchers to identify inconsistent y values (Evans 1999).

2.6.3. The x-y outliers or influence points and their detection

As was mentioned above, some outliers may be very harmful when they affect both regression line and regression parameters. Under these circumstances, Cook's influence measure is the best known and probably the most used one (Cook and Weisberg 1982, Martens and Naes 1989) that makes a contribution to both the residual and the leverage. It can be written as

$$D_i = \frac{r_i^2}{k+1}\left(\frac{h_{ii}}{1-h_{ii}}\right) \tag{2.9}$$

where r_i is the standardized residual and h_{ii} is the ith diagonal element ofthe Hat matrix computed from the full regression and k is the number of parameters in the model. D_i would be large if the standardized residual is large and if the data point is far from the centroid of the x-space (Rawlings et al. 1998). The influence D_i is usually plotted versus ε or observation number that is called an influence plot (Naes et al. 2002).

2.7. Statistical test for presence of outliers

Iglewicz and Hoagin (1993) advocated the inspection and examination of all data for outliers. By testing data for outliers, researchers can avoid making distorted conclusions about data and can yield more robust estimates of parameters (Bacon 1995).

There are several tests for identifying univariate and multivariate outliers, separately. Four common outlier tests for univariate data with normal distributions are the Rosner test, Dixon test, Grubbs test, and the box-plot rule. These techniques are based on hypothesis testing rather the regression methods (Grubbs 1969, Fallon and Spada 1997).

Because multivariate statistics are increasing in popularity with researchers in most scientific fields, the challenge of detecting multivariate outliers demands more attention (Wiggins 2000). The standardized and studentized residuals are often used to identify outliers in analysis based on linear models (Montanari 1995). The observation with the largest absolute studentized residual is usually given special attention and is considered as the observation most likely to be a contaminant.

However, it is important to realize that these residuals will follow a t distribution, implying that (in a large sample) approximately 5% of them will appear extreme. This problem is solved by adjusting the t-test critical value, using a Bonferroni correction. The correction is to replace the usual degree of freedom $\alpha/2$ for critical t with $\alpha\ /2n$. Consequently, this pushes the t-critical into the ends of the distribution, decreasing the chance of declaring a residual to be an outlier (Scott 2003).

The PRESS statistic is another available statistic with the use of regression analysis for detecting the possible presence of outlying observations in the data. The PRESS (prediction sum of squares) is the sum of squares of error (SSE) which was obtained by estimating the equation using all the data without the current (*i*th) observation (Freund and Littell 2000). It is computed as:

$$\sum_{i=1}^{n} \hat{\varepsilon}_{(i)}^{2} = \sum_{i=1}^{n} \left(\frac{\hat{\varepsilon}_i}{1 - h_{ii}} \right)^2 \tag{2.10}$$

Where the (i) indicates that the ith observation was not used in the development of the regression equation. The sum of squares of the PRESS statistic should be compared to the SSE. When it is substantially larger than the SSE, there is reason to suspect that some influential observations or outliers exist (Freund and Littell 2000) because a residual $\hat{\varepsilon}_i$ that corresponds to a large value of h_{ii} contributes more to PRESS (Rencher and Schaalje 2008). However, there are data points that are identified as outliers using some statistics but not by others (Wiggins 2000).

2.8. Deal with outliers: discarding or adjusting

There is a great deal of argue as to what to do with detected outliers (Osborne and Overbay 2004). When an outlier is detected, it is not easy to know exactly what to do about it, but some guidelines can be set up. The first thing to do is always to go back to the lab and look for possible reasons for the sample to be an outlier (Naes et al. 2002). However, in practice, it is not always possible. If there is no explanation for an outlier, the dataset could be analyzed both with and without the outlying observations and then, two results should be compared (Evans 1999, Rencher and Schaalje 2008). If the influence is negligible, it does not matter very much what is actually done with the outlier. Another alternative is to discard the outlier, even though no explanation has been found (Rencher and Schaalje 2008). This will most often, but not always lead to better prediction ability of the predictor (Naes et al. 2002).

Although, it is stated by a few authors that removal of extreme scores produces undesirable consequences, they are in the minority, especially when the outliers are illegitimate (Osborne and Overbay 2004). When the data points are legitimate, some authors (Orr et al. 1991) argue that data are more likely to be representative of the population as a whole if outliers are not removed (Osborne and Overbay 2004). Hecht (1991) asserted that too many researchers advocate the simple rejection of outliers for their treatment but they should not force the data to conform to their preconceived expectations by rejecting detected outlying observations. Adjusting of outliers is a recommended alternative to discarding them. One method of adjusting outliers is the use of transformations (Osborne 2002). By using transformations, extreme scores can be kept in the data set and residual variance can be reduced (Hamilton 1992). However, transformations may not be an appropriate way because of difficulties in its interpretation (Osborne 2002).

Truncation of outliers is one alternative to transformation, wherein extreme values are recoded to the highest or lowest reasonable score (Osborne and Overbay 2004). It could be done using trimming (elimination) or winsorization (truncation) of the data (Iglewicz and Hoaglin 1993). A trimmed mean is defined as the mean calculated by trimming a percentage (usually 5%) of data sets from the top and bottom of data sets. This helps to alleviate the distortion caused by extreme values (Fallon and Spada 1997). A winsorized mean is similar to the trimmed mean but instead of trimming the points, they are set to the lowest or highest value (Sachs 1982, NIST 2003).

Iglewicz and Hoaglin (1993) proposed the operation of winsorization parallel to the trimming. They believe that trimming should be performed on a data set simultaneously, with an equal number of values trimmed from each end of the data set. This demonstrates that winsorization does not discard outliers completely, but rather decreases their distance from the center of the sample (Barnett and Lewis 1984). Instead of elimination, transformations or truncation, researchers sometimes use various robust procedures to protect their data from being distorted by the presence of outliers (Barnett and Lewis 1994, Rencher and Schaalje 2008). However, dealing with outliers needs a model to catch them and results depend on the model fitted to the data (Mayeres et al. 2003).

2.9. Ordinary Least Squares, Robust and Local Regressions

Williams et al. (2002) classified the outlier detection methods between parametric methods and nonparametric methods that are model-free. The classical least squares estimator is widely used in regression analysis because of both the ease of computation and tradition. The most widely employed method for estimating model parameters (β and ε) in least squares is the method of ordinary least squares (Nevitt and Tam 1998). When Ordinary Least Squares (OLS) is used, the outliers can significantly influence the estimates (Lane 2002) even in the presence of only one outlier (Nevitt and Tam 1998). Consequently, when there are outliers in the data, a robust fitting method is necessary. Robust and local regressions are two alternatives in such a case. More details about these specific regressions will be presented:

2.9.1 Ordinary Least Squares method

The Ordinary Least squares (OLS) that was first introduced by Carl Friedrich Gauss around 1794 (Weisberg 2005), is the most common regression estimator. The objective of this method is to

$$Minimize \sum_{i=1}^{n} \varepsilon_i^2$$

where $\varepsilon = y - \hat{y}$ as in 2.2 (Chen 2002). Most outlier detection methods use some measures of distance to evaluate how far away an observation is from the centre of the data. To measure this distance, the sample mean and variance may be used (Franklin and Brodeur 1997). Although this method of identifying outliers is simple, but since the mean is very sensitive to outliers and to long tails in the distribution, even a single regression outlier can totally offset the least squares estimator (Hadi and Simonoff 1993, Iglewicz and Hoaglin 1993, Franklin and Brodeur 1997). Consequently, OLS is not a robust estimator (Chen 2002). One approach to avoid this weakness and sensitivity is utilizing robust scale and location estimators which are inherently resistant to outliers and much safer due to using other statistics such as the median or quartile to mean (Hampel et al. 1986, Franklin and Brodeur 1997).

2.9.2 Robust regression method

This method that gives more robust results in the presence of outliers, also can be used to detect outliers (Chen 2002). However, if no outliers are present and the sample size is large, the robust and classical methods should give similar estimates. If the estimates differ, outliers may be present or the assumed distribution may be incorrect (Olive 2005).

The purpose of robust estimation is to produce an efficient estimator in the presence of outliers, while minimizing bias. This is done by reducing the influence of the outliers on the estimator before it breaks down (Franklin and Brodeur 1997). Hampel (1971, 1974) introduced the concept of the breakdown point, as a measure for the robustness of an estimator against outliers as the smallest percentage of outliers that can cause an estimator to take arbitrary large values. Thus, if an estimator has the larger breakdown point, consequently it is more robust (Ben-Gal 2005).

Contrary to traditional least squares estimators, there is a class of robust regression methods that replace the sum of squared errors with ones less influenced by outliers using trimming or winsorizing simply means (Franklin and Brodeur 1997). The ROBUSTREG procedure in SAS Version 9.1 (SAS 2003) implements the most commonly used robust regression techniques such as Huber M-estimation, high breakdown value estimations (LTS and S), and combinations of these two methods (MM) that have different breakdown measures (Chen 2002).

M-estimation was introduced by Huber (1973), and it is the simplest method among robust regressions both computationally and theoretically (Chen 2002). M-estimators are not robust with respect to *x*-space while at the same time they are still robust with respect to outlying data in the response direction (Rousseeuw and Leory 1987). It means that both OLS and M-estimation methods cannot distinguish good leverage points from bad leverage points. In such cases, high breakdown value estimates (LTS, S and MM) are needed (Chen 2002). Instead of minimizing a sum of squares as in least squares, a Huber-type M-estimator $\hat{\beta}$ of β minimizes a sum of less rapidly increasing functions (ρ) of the residuals (Rousseeuw and Leory 1987):

$$\beta = \sum_{i=1}^{n} \rho\left(\frac{x_i - T}{\sigma}\right)$$

Where for the observation x_i, location estimate T, and scale estimate σ, define the residual, $r_i=(x_i-T)/\sigma$. Typically, the role of this function is to decrease the influence of observations with large residuals. The procedure can also solve this system by using iteratively re-weighted least squares which are robust to outliers (Chen 2002).

As mentioned above, there are many other strategies for robust regressions which search outliers in x or both x and y directions (Chen 2002). More details about these methods can be found in the literature (Yohai 1987, Rousseeuw and Yohai 1984, Rosseeuw and Leroy 1987, Rousseauw and Van Driessen 1998, Venables and Ripley 1999).

The purpose of diagnostics is to find and identify deviations from the assumptions; whereas the purpose of robustness is to safeguard against deviations from the assumptions (Huber 1990). Furthermore, Huber (1990) noted that outlier detection is classified under diagnostics, not robustness. Robustness is making a procedure insensitive to outliers. Though robust estimators are efficient and highly insensitive to unusual values of the dependent variable, one high leverage point can break them completely (Militino and Ugarte 1997).

2.9.3. Local Regression method

Locally weighted regression, or LOESS, is a nonparametric approach to estimate regression functions through multivariate smoothing with a weighted least squares algorithm that locally weights the data (Cleveland and Devlin 1988, Cleveland and Grosse 1991). This procedure is available in SAS (2003) via PROC LOESS. This procedure allows greater flexibility than traditional modeling tools and it performs iterative re-weighting to provide robust fitting when there are outliers in the data (Cohen 1999). The underlying model for local regression is

$$y_i = f(x_i) + \varepsilon_i$$

where for i = 1 to n, y_i and x_i are observations of a response and independent variables. In the above formula, f is the unknown regression function and ε_i is a random error (Cleveland and Grosse 1991, Cohen 1999) and y_i has a constant variance (Cleveland and Loader 1995). For f, it is supposed that the function can be well approximated locally by a parametric polynomials (linear or quadratic) by fitting a regression surface to the data points within a chosen neighbourhood of the certain point x_0 (Cleveland and Loader 1995, Cohen 1999). The radius of each neighbourhood or bandwidth $h(x_0)$ is chosen so that the neighbourhood contains a specified percentage of the data points called the smoothing window $(x-h(x_0), x+h(x_0))$ as demonstrated by Loader (1999). To estimate $f(x)$, only observations within this window are used (Cleveland and Grosse 1991, Loader 1999). The observations are weighted according to the formula:

$$\omega_i(x) = W\left(\frac{x_i - x_0}{h(x_0)}\right)$$

where *W(u)* is a weight function that assigns largest weights to observations close to x_0 (Cleveland et al. 1988). Researcher will usually want to incorporate a weight function, *w(u)*, that gives greater weight to the x_i in the neighbourhood that are close to x_0 and lesser weight to those that are further (Cleveland and Loader 1995). Moreover, the bandwidth $h(x_0)$ has a critical effect on the fit of the local regression. If $h(x_0)$ is too small, insufficient data fall within the smoothing window and a large variance will result. On the other hand, if $h(x_0)$ is too large, the local polynomial may not fit the data well within the smoothing window and the fit will have large bias (Loader 1999). SAS controls the smoothing window through specifying a smoothing parameter in LOESS procedure (SAS 1999).

The smoothing parameter in each local neighbourhood controls the smoothness of the estimated surface (Cohen 1999). A smoothing parameter, α, (between 0 and 1) chooses $h(x_0)$ in such a way that the local neighbourhood always contains a specified number of points (Loader 1999). Of course, as the smoothing parameter α is reduced, the residuals generally get smaller, and show less structure (Loader 1999).

An Example of local fitting for the four smoothing parameters 0.2, 0.4, 0.6 and 0.8 adopted from Loader (1999) is shown in the Figure 2.3. Clearly, α =0.2 produces a much sharper fit than α =0.8 which is highly smoothed, since it does not fit the data well (Loader 1999). Thus the maximum smoothing parameter of 1 essentially fits a function for the entire data set, which is obviously the smoothest yet poorest fitting curve, while small values, such as 0.1 may actually fit the curve to all data points (Freund and Littell 2000). There are several strategies to select the smoothing parameter. A simple one is to examine plots of the fit residuals versus the predictor variable and to choose the largest smoothing parameter that shows no clearly pattern in the fit residuals (Cohen 1999).

Like the smoothing parameter, the degree of the local polynomial affects the bias and variance. A high polynomial degree can always provide a better prediction with less bias than a low polynomial degree (Cleveland and Loader 1995). Figure 2.4 displays, for example, including local constant, linear, quadratic and cubic fits for the ethanol dataset adopted from Loader (1999). Though the local constant fit in Figure 2.4 is sharp and somewhat bias, it doesn’t fit the data well at the left boundary. Both the boundary bias and sharpness were reduced by local linear fit but this is rarely adequate. The local quadratic and local cubic fits produce better performance containing less sharpness and better fit to data the but increase variance especially at boundaries that can be a problem (Loader 1999).

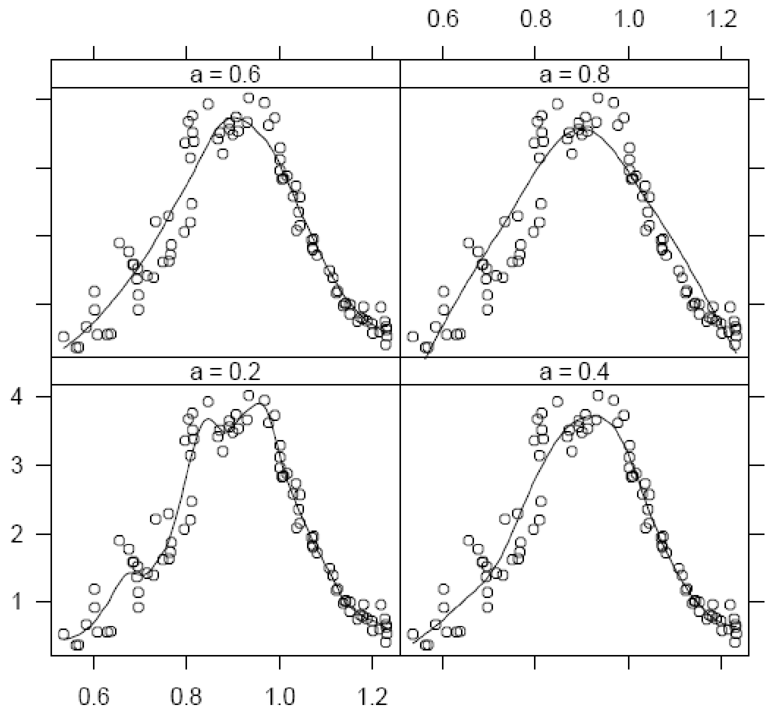

Figure 2.4: Local fitting at four different smoothing parameter: α = 0.2, 0.4, 0.6 and 0.8 (from Loader (1999)).

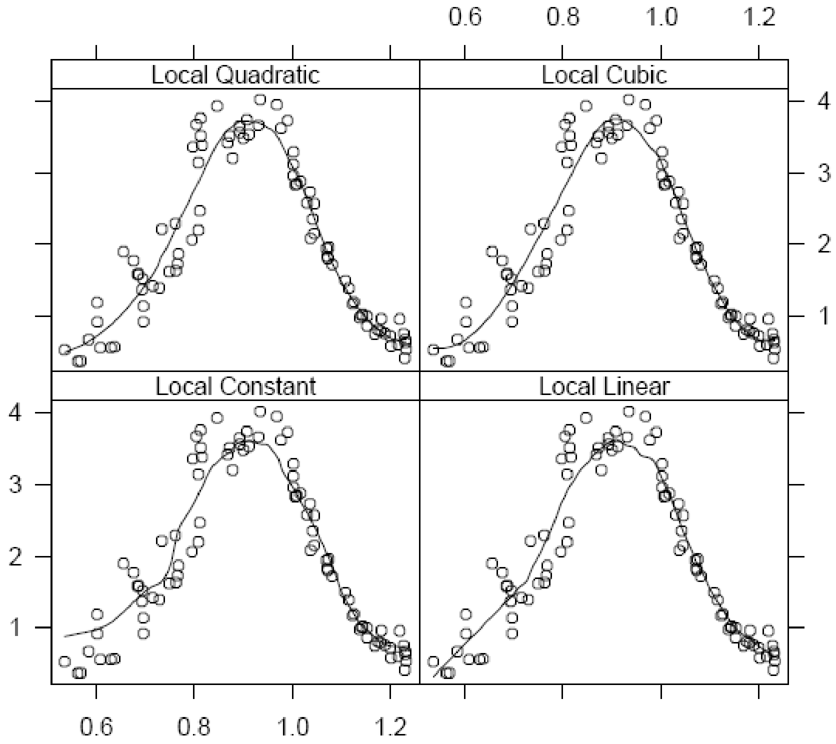

Figure 2.5: Effect of changing the polynomial degree on model fit (from Loader (1999)).

3. *Materials and methods*

3.1. Data

The sample is a subset of the data used in Brügemann (2008). It contains data from four cattle breeding service centers in Germany.

The final set of test data included a total of 111,599 TD records of 11,620 first lactation Holstein cows which calved between 1997 and 2002 (black-white=11576, red-white=44). These data come from the six largest dairy farms situated in north-eastern of Germany. Days in milk (DIM) were restricted to those between 5 and 330. Only records corresponding to 2 times a day milking and collected under the supervision of a milk recording expert were used. Each TD record contained data completely measured on daily (24-h) milk yields (in kg) with only two sampled milking and with sampling interval of approximately 4 weeks.

Four seasons of calving (January-March, April-June, July-September and October-December) were defined. To be included in the data set, each cow was required to have at least nine test-day yields per lactation. Herd-test-dates were required to have at least 9 test-day records. There were 113 herd-year-season and 357 herd-test-date (HTD) contemporary groups in the final data. The final pedigree file consisted of 58315 animals, the ancestors. Table 3.1 presents the structure of data and pedigree files. In this context, Table 3.2 presents the mean and the standard deviation of milk yields and days in milk by test numbers.

Table 3.1: Structure of the data.

Item	**No.**
First lactation with ≥ 9 TD records	11,620
TD records no.	111,599
HYS levels no.	113
HTD levels no.	357
Number of inbred animals	14,889
Average and maximum inbreeding coefficient%	1.97, 26.56
Mean, minimum and maximum of age at calving in month	27.58, 20, 40
Herds no.	6
Average, minimum and maximum TD records per herd, no.	18,560, 7,147, 37,455

Table 3.2: Mean and standard deviation (SD) of TD yields and days in milk by test numbers.[1]

Test No.	Test-day milk yields		Days in milk	
	Mean	SD	Mean	SD
1	24.648	5.735	22.437	12.336
2	27.486	5.258	54.880	14.304
3	27.000	5.247	88.165	16.159
4	25.680	5.169	121.010	17.209
5	24.284	5.142	154.227	17.904
6	22.831	5.088	187.668	18.110
7	21.462	5.103	220.618	17.953
8	19.803	5.154	253.549	17.764
9	17.550	5.624	285.992	17.175
10	16.544	5.629	311.244	13.997
11	16.480	5.695	318.645	6.805
Mean	22.983	6.337	164.197	92.476

1- The number for TD 1-9 was 11620 and it was 6385 and 634 for TD 10 and 11 respectively.

3.2. Methods

In order to achieve the objectives that have been initially determined, it necessary to pursue the following strategies.

3.2.1. Tests to determine the existence of outliers

Prior to apply for any detection or treatment of abnormal recordings, it is necessary to perform some tests in determining the existence of outliers. To do this, PRESS residuals and statistic, standardized residuals and studentized residuals with both usual degree of freedom for t-statistic and one corrected by Bonferroni adjustment were used.

3.2.2. Methods to detect outliers

As earlier has been explained, dealing with outliers requires a model to be fitted to the data using a regression procedure. The selected regression procedures were the methods; Simple Regression (SR), Robust Regression (RR) and Local Regression (LR) with first degree (LR1) and second degree (LR2) of polynomial function with two smoothing parameter (0.5 and 0.6). These regression procedures were utilized to detect the abnormal TD records as well as to check and compare whether regression improves the criterions as more as possible.

All regression methods SR, RR and LR are carried out by the Windows version 9.1.3 of Statistical Analysis Software (SAS) using Proc REG, Proc ROBUSTREG and Proc LOESS procedures respectively and the produced outputs were used for further utilizing. Robust estimation includes many different procedures. In the present study, one such method that uses the M-estimator (Huber 1973) for the robust regression method was applied. For the RR method, the residual was specified as convergence criterion and the maximum number of iterations during the parameter estimation being limited to 1000. For all local smoothing methods, only one iteration was considered, although the locally re-weighting methods requiring more iterations and yielding more stable results, are robust to the influences of outliers. However, in this case it was not possible to be handled by the LOESS procedure (at least for some lactations). The RR method was used due to its useful characters as to be resistant versus outlying observations and LR owing to its desirable properties for situations in which one does not know a suitable parametric form of the regression surface.

3.2.3. Schemes to treat outliers

In order to deal with outlying TD observations before implementation of genetic evaluations, two treating schemes, discarding and adjusting, were considered and each of them was applied to three regression procedures.

The discarding scheme excluding 0.1, 0.2, 0.4, 0.8, 1.6, 3.2 and 6.4 percent of outlying TD yields with the highest values for either absolute residual (ABSR) or absolute standardized residual (ASTAR) were employed. The marginal residuals for each observation were defined as y_i-$\hat{y}_i$ where $\hat{y} = X\beta$. The ASTAR of individual recordings was calculated as the residuals (R) divided by standard division of the residual (STDR).

With the adjusting scheme, a TD record was adjusted if its R was beyond k times of its STDR. In this case, these records (y_i^*) were replaced by a value as following:

$$y_i^* = \begin{cases} y_i & \text{if } \left| r_i \right| \le (k \times STDR) \\ \hat{y}_i - k \times STDR & \text{if } r_i < -(k \times STDR) \\ \hat{y}_i + k \times STDR & \text{if } r_i > (k \times STDR) \end{cases}$$

where STDR is the standard deviation of residuals estimated by regression method for a specific lactation of an animal in a given TD. Also, k is a constant chosen by the user and is designed so that approximately between 0.1 to less than 11.5 percents of records are labelled as outlier (see Table 4.10). Outliers for each method were defined as observations that were located outside of specified limits. In general, this method is similar to the method adopted by Yang et al. (2004) except that they computed the standard errors for each contemporary groups instead of each observation point. The model to capture the outlying data by both SR and RR procedures included four DIM covariates (X_2, X_3, X_4 and X_5) of Ali and Schaeffer's (1987) function and LR procedure used only one covariate, X_5 (it comes following) in the model. Subsequently, the data obtained after treating data specified as outlying records by both discarding and adjusting schemes were used in order to compare the impact of using these methods on model residual criterions, genetic parameters and estimated breeding values including the full pedigree.

In addition, variance components and parameters as well as breeding values were obtained by a regular analysis with complete data. Outliers were quantified by number and proportion (relative to the total number of observations), average, standard deviation, and minimum and maximum values of differences between y^*_i and y_i. Outliers were also characterized by number of outliers per cow and lactation.

3.3. Model

A single trait repeatability animal model which uses fixed regression model (FRM) for genetic analysis of TD milk yields was defined with following components:

$$y_{ijkl} = HYS_i + HTD_j + \sum_{m=1}^{5} b_m X_m + a_k + pe_k + e_{ijkl}$$

Where:

y_{ijkl} is a TD milk yield observation of cow *k*,
HYS_i is fixed effect of *i*th herd-year-season of calving,
HTD_j is fixed effect of *j*th herd-test-date,
a_k is additive genetic random effect of *k*th cow,
pe_k is permanent environment random effect of *k*th cow,
b_m are *m*th fixed regression coefficients of age at calving and Ali and Schaeffer's function
X_1 = age at calving in months,
X_2 = DIM/330,
X_3 = ln(330/DIM),
$X_4 = (\mathrm{DIM}/330)^2$ and
$X_5 = (\ln(330/\mathrm{DIM}))^2$.

This model is similar to the both repeatability models used by Swalve (1995a) or similar to a combination of two models TY1 and TY3 of Ptak and Schaeffer (1993), except that the regression coefficients are not nested within age-season effects. The inclusion of factors was based on the literature and a preliminary analysis of test day data. Effects that had a major influence on the shape of the lactation curve were herd, age at calving and year and season of calving. Month of production is automatically included in HTD effect. The main non-genetic effects in the TDM are the herd-test-date (HTD) effect and fixed regressions to account for the average shape of lactation curves (Strabel et al. 2005). All fixed effects are included as cross-classified effects, not as fixed regressions.

In the FRM which initially was proposed by Ptak and Schaeffer (1993) and Reents et al. (1995) for modelling dairy TD data, it is assumed that additive genetic and permanent environmental effects (variance components) are constant throughout a lactation. This implies equal correlations between any two tests, regardless how far away they are from each other (Jensen 2001). Thus, residual variance was assumed to be constant across lactation. Furthermore, (co)variances between residuals for records were assumed to be zero in this model.

Although, such assumptions result in parsimonious parameterisation, they do not seem to make a good use of TD information and do not reflect phenotypic observations well (Liu et al. 1998) due to assuming that individuals have a constant additive genetic effect throughout the course of lactation and do not allow for differences in lactation curve shapes between animals (Szyda and Liu 1999). Advantages and disadvantages of FRM were well discussed in literature (e.g. Swalve 2000, Jensen 2001). Here, it is assumed that the shape of the lactation curve is not influenced by random genetic and permanent environmental effects. However, it was adopted due to the computation limitations and simplicity.

The analyses to compute variance components were implemented by VCE package V5.1.0 (Kovac and Groeneveld 2002). For estimating the model parameters, the conjugate gradient algorithm (default option) was used through setting the sparse matrix algorithm as solver (Groeneveld 2003). The estimations were stopped when the Log likelihood changed by less than 0.0001 in the last two successive iterations or until the number of iterations reached 1000 whichever occurred first. However, the preliminary analysis showed that thirty iterations were sufficient to reach convergence.

For the purpose of estimating breeding values by PEST package V4.2.3 (Groeneveld et al. 1990), the solver IOC (iteration on coefficient) was chosen. The estimated variance components obtained by VCE package were used as values of the specified parameters for PEST. For PEST, the parameters to stop computations were 1×10^{-7} (convergence criterion) and 200,000 (maximum number of iteration) respectively. Both packages were run under a Linux operation system. The same pedigree data were used in all analyses.

In matrix notation, the model can be written as:

$$y = Xb + Za + Wp + e \tag{3.1}$$

where y is the vector of TD milk yields; *b* is the vector of fixed effects being herd-year-season of calving (HYS) subclass and herd-test-date (HTD) subclass and fixed regression coefficients (X_1 to X_5); *a* is the vector of animal additive genetic effects; *p* is the vector of animal permanent environmental effects; and *e* are the random residual effects. *X*, *Z*, and *W* are matrices relating observations to factors *b*, *a* and *p* in the model, respectively.

The corresponding Henderson's mixed model equations (Henderson 1984) for analysis are:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} & Z'R^{-1}W \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W+P^{-1} \end{pmatrix} . \begin{pmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \end{pmatrix} \tag{3.2}$$

Assume that:

$$\begin{pmatrix} a \\ p \\ e \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} G & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & R \end{pmatrix}\right]$$

$$E\begin{pmatrix} y \\ a \\ p \\ e \end{pmatrix} = \begin{pmatrix} Xb \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

where $Var(a) = G = A\sigma_a^2$, $Var(p) = P = I_p\sigma_p^2$ and $R = I_n\sigma_e^2$. For this case, the matrix R is diagonal and has the constant residual variance. Therefore, multiplying (3.2) throughout by the scalar residual variance (σ_e^2) results in a equivalent equation:

$$\begin{pmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}.\alpha_1 & Z'W \\ W'X & W'Z & W'W + P^{-1}.\alpha_2 \end{pmatrix} . \begin{pmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \\ W'y \end{pmatrix} \tag{3.3}$$

with $\alpha_1 = \sigma_e^2 / \sigma_a^2$, $\alpha_2 = \sigma_e^2 / \sigma_p^2$.

where $Var(y) = ZGZ' + WPW' + R$, A = additive genetic relationship matrix among the animals, *I* = identity matrix and *R* = residual matrix, σ^2_a= variance of the additive genetic effect, σ^2_p= variance of the permanent environmental effect and σ^2_e= variance of the residual effect that were assumed to be constant over lactation.

3.4. Comparisons

The evaluation of the model fit was based on checking the residuals (R, ABSR and ASTAR) and the correlation between TD yields and their predicted values. Moreover, the proportion of identified outlying recordings over lactation and versus production level by various regressions will be checked. The total number of detected outliers, the percentage, no. of cows with at least one outlying TD recording, percentage of low outliers and percentage of lactations with a single outlying TD for each data set produced by methods will be examined. In each case (treating method), the total number of lactations with one (TD1), two (TD2) and by eight (TD8) outlier records for whole lactations will be calculated.

It is very important to note that the residuals, $\hat{e}_i$, estimated by PEST package are conditional meaning that the effects of the factors included in the model on observations were removed by $\hat{e}_i = y_i - (\mathrm{x}_i'\hat{\mathrm{b}} + \mathrm{z}_i'\hat{\mathrm{a}} + \mathrm{w}_i'\hat{\mathrm{p}})$ where y_i is the corresponding observation.

Average of milk yield production (MPA), mean of absolute residuals (MAR), mean of squares of residuals (MSR) and standard deviation of residual (SDR) will be compared. In addition, results of genetic evaluations, additive genetic (V_a), permanent environmental (V_c) and residual (V_e) variances, heritability (h^2), permanent variance ratio (c^2) and error variance ratio (e^2) and; their standard errors (SE (h^2), SE (c^2) and SE (e^2)) were estimated.

Comparisons are also made of the top 100 bulls (with 50 or more daughters having records) and top 100 cows having records, on number of top 100 bulls and cows in common with the different analyses, average of estimated breeding values (MEBV), and Spearman rank correlations. Correlations between the complete data with no treatment and those from the methods will be also estimated for all animals.

4. Results

4.1. Tests for determining the existence of outliers

As explained before, it is useful to examine data for possible outliers and to employ remedial alternatives, although practically most datasets from every field of science includes one or more outlying observation(s). In this study, testing was done through applying absolute standardized residual (ASTAR), absolute PRESS residual (APRESSR), PRESS statistic (PRESS) and studentized residual (STUR). Residuals from a regression analysis will generally be dependently (unequally) distributed, even if outliers are not present in the data. Therefore, it is advisable to divide the residuals by their standard deviations. Similarity between PRESS statistic and the residual sum of squares (SSE) relevant to a specified lactation, suggests that there are no outliers or influential observations in the data. The advantage of using the studentized residuals over ordinary residuals is that they have a two-sided symmetric student distribution. Belsley et al. (1980) pointed out that studentized residuals have a t-distribution with n-p-1 degrees of freedom, where n and p are the number of observations and model parameters, respectively. This means that one can test the significance of any single studentized residual using a t-table.

Table 4.1 presents the percentage of detected outliers by STAR and PRESS residuals which are greater than 2 and 3 in absolute value, respectively. Moreover, it shows the identified outliers by studentized residuals in two statistical levels of $\alpha=0.05$ and $\alpha=0.01$ with both the usual alpha for two-side test ($\alpha/2$) and the alpha corrected by Bonferroni method ($\alpha/2n$). It also shows the comparison of the average of PRESS to SSE for all lactations.

Table 4.2 shows the summary statistics for the three residual terms PRESS, SSE and PRESS/SSE that were produced by SR (simple regression) analysis and their absolute values. However, these residuals will be needed to explain the observed ranges of outlier tests which were applied to the simple regression.

Standardized residuals (STAR) are useful in detecting abnormal observations or outliers. The good thing about STAR is that they quantify how large the residuals are in standard deviation (SD) units, and therefore, can easily be used to identify the outliers. In general, any observation with an ASTAR greater than two (that includes about 95% of them) should be considered for further investigation, although such observations are not necessarily outliers (Birkes and Dodge 1993). Shiffler (1988) showed that the magnitude of a standardized residual in a univariate data set could never exceed $(n - 1)/\sqrt{n}$ where n is the number of observations (Gray and William 1994). It means that with available data, range of ASTAR would be 2.67, 2.85 and 3.01 for 9, 10 and 11 tests per lactation, respectively. However, 1.7% data points having ASTAR greater than two (the specified threshold for being a potential outlier) were detected (Table 4.2).

Another useful test to recognize outlying TD records is examining studentized residuals (STUR). The studentized residual for a given data point not only depends on the ordinary residual but also on the size of the mean square error and the leverage h_{ii}. The STUR is the standardized residual calculated without the current observation. It is clear that increasing the significance level of a statistical test will decrease the probability of obtaining a significant error. As Table 4.1 shows, STUR detected 8.44% and 2.82% at significant levels of 5% and 1%, respectively, versus 1.85% and 0.64% whereas their alphas were corrected for degree of freedom using Bonferroni adjustment. Since the Bonferroni criterion is relatively conservative, Williams et al. (1999) provided an alternative to this inequality that results in drooping more outlier observations.

The PRESS residual and PRESS statistic, behaved somewhat different from the last two ones. The predicted residual for the observation *i* is called PRESS residual (PRESSR) that was explained in section 1.7 and is the numerator of PRESS statistics. The sum of squares of PRESSR is called the PRESS statistics (predicted residual sum of squares). Furthermore, an extreme value is often considered to be an outlier if an outlier test flags the predicted residual greater than 3.0 in absolute value. Approximately 34% data points provided a APRESSR value greater than the value 3.0 (Table 4.2) which implies a potential difficulty in identifying outlying data using this type of residual criterion. Though PRESS in most cases is larger than SSE but the general idea is that it should not be much larger in contrast to what is given in Table 4.2. As a rule, if the PRESS statistic for an individual lactation is considerably larger than its SSE this indicates the existence of outliers in that particular lactation. Clearly, a lower PRESS statistic suggests a better prediction performed by the model if the aim is to compare different models.

About 53.5% of the data were identified as outliers by the PRESS statistics. On average, PRESS is about 750 times greater than SSE, probably due to its sensitivity to dropping TD records in a lactation stage prior to the peak yield. Looking at PRESS and its absolute value (APRESSR) presented in Table 4.2, one can find the high average value and large standard deviation (SD) of PRESSR and APRESSR. Figure 4.1 shows the high variability of the average of PRESSR for the first month of lactation period. It means that the measure of the PRESS statistic is seriously affected by eliminating (see equation 2.10 on page 14) the TD records relating to the first 45 days of lactation before peak milk yield.

Silvestre et al. (2006) studied the different shapes of the seven lactation curve functions based on TD records from varying sample schemes including the intervals from calving to first test (8, 30, 60, and 90 day) and two intervals between tests (4 and 8 week). When time intervals from calving to first test were equal to 60-d or more were analyzed, the SD of error increased and the correlation (R) between true and estimated yields decreased. Their results for both 4 and 8 week test schemes show poor fit for the lactations groups having the first test within 60 and 90 days after calving. It is interesting that a poorer results were observed for the lactation groups with DIM=60 and 4 weeks interval containing 9 records per lactation compared to the group with DIM=30 and 8 weeks interval including only 5 tests per lactation. This fact obviously indicates that the reduction in R was a consequence of the increased interval from calving to the first test (60 or 90 DIM). Thus, the efficiency of most lactation models strongly depends on the interval from calving to the first test (Silvestre et al. 2006).

From the results it may be concluded that the PRESSR and PRESS statistics may not be useful outlier tests for the data with few TD records per lactation as herein. It is proved again by taking the average of PRESS statistics for the lactations with 9, 10 and 11 test records, separately. The number of TD records decreased from 15149 in the lactations with 9 records to 14688 and 6862 in lactations with 10 and 11 records, respectively. A possible cause of this decrease is that a increase in the number of milk tests will raise the chance of occurring more than one milk tests in the first 45 days in milk after calving.

Therefore, the results from the tests underline the presence of outliers in the data set regardless of which outlier test is used. Below, other aspects of the fit of the regression models as well as the relationship between frequencies of the detected outliers and lactation stage and/or TD yield values will be explained.

Table 4.1: Percentage of outliers detected by ASTAR, STUR and PRESSR and PRESS[1].

Test	Percentage of detected outliers
ASTAR >2	1.71
STUR $t_{95\%}$	8.44
STUR $t_{95\%b}$	1.85
STUR $t_{99\%}$	2.82
STUR $t_{99\%b}$	0.64
PRESSR >3	34.4
PRESS > (100×SSE)	53.5

1- In this table the limitation for a point to be an outlier using ASTAR was a value greater than 2, and using PRESSR limitation was a value greater than 3. Using STUR a point was an outlier if its critical t-value was larger than the t-value in the t-table for levels 5% and 1% with ($t_{95\%b}$ and $t_{99\%b}$) and without Bonferroni correction ($t_{95\%}$ and $t_{99\%}$). Using PRESS statistics, the range for a particular lactation having outlier(s) was a PRESS greater than 100×SSE.

Table 4.2: Mean, SD, Min and Max values for various residuals and statistics resulting from SR^1.

Type of Residual	Mean	SD	Min	Max
R	0.0000	1.457	-16.05	17.149
STDR	1.1746	0.8615	0.0007	8.6572
STAR	0.0002	1.027	-2.429	2.4139
STUR	-0.0314	1.806	-51.08	63.495
PRESSR	1.2393	38.79	-1916.0	1666.4
ABSR	0.9819	1.076	0.000	17.149
ASTAR	0.8609	0.559	0.000	2.4294
ASTUR	1.1294	1.409	0.000	63.495
APRESSR	8.9917	37.76	0.000	1916.0
SSE^2	20.380	27.51	0.0550	443.752
PRESS^2	14469	88876	0.76404	3672788
Ratio PRESS/SSE^2	755.15	3660.9	1.9457	133212

1- In this table R stands for residual, STDR for standard deviation of residual, STAR for standardized residual, STUR for studentized residual, PRESSR for predicted residual sum of squares and ABSR, ASTAR, ASTUR and APRESS stand for absolute values of R, STAR, STUR and PRESS, respectively. Moreover, SSE stands for sum of squared errors, PRESS stands for predicted residual sum of squares and Ratio stands for the proportion PRESS/SSE.

2- These three criteria (SSE, PRESS and Ratio) were applied to the individual lactations.

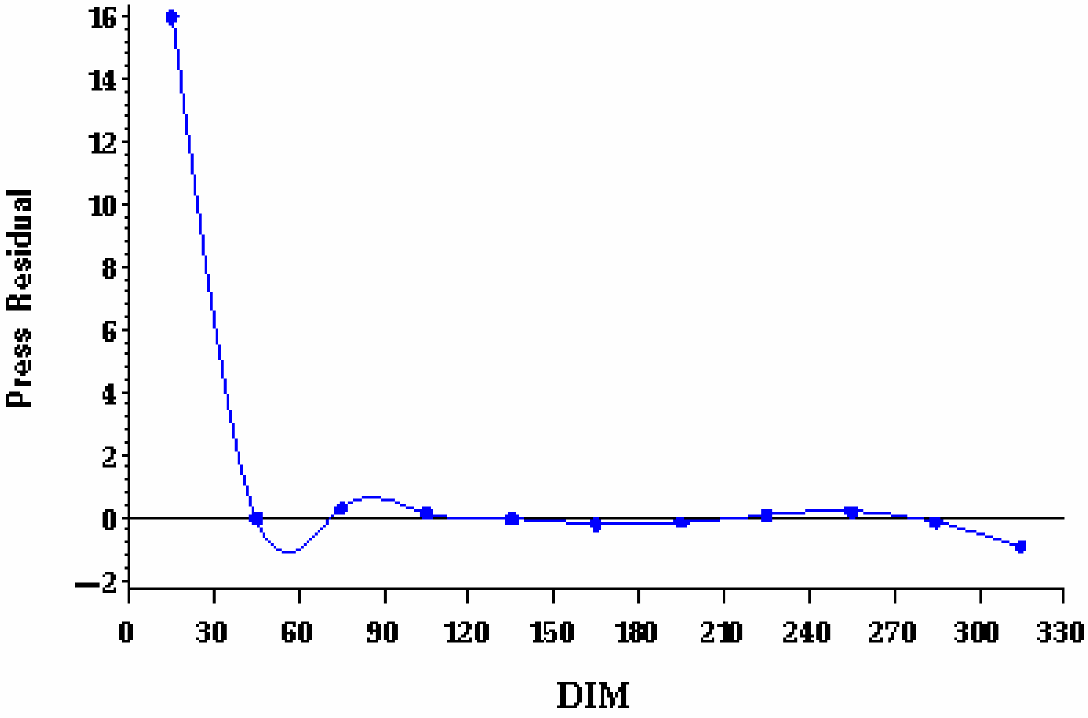

Figure 4.1: Variation of PRESS residual versus days in milk (DIM) resulting from SR model.

4.2. Checking the fit of the models

Figures 4.2, 4.3 and 4.4 show the patterns of variations for predicted yields, residual terms and correlations between predicted and observed milk yields over the lactation period, respectively. Despite a high variance at the beginning and the end of lactation, both SR and RR models fitted well to the TD data points for these parts.

On the other hand, the curves fitted by the local regressions LR1 and LR2 with both smoothing parameters 0.5 and 0.6 (LR1-0.5, LR1-0.6, LR2-0.5 and LR2-0.6) matched the average of the actual data (MPA) from 5 to 75 DIM sufficiently, and afterward they gradually moved away from the MPA by end of lactation. The LR regressions underestimated the predicted values between 180 and 270 DIM and overestimated it later, indicating that these models were unable to fit the data accurately within this phase of the lactation course.

As it is observable from Figure 4.2, around DIM 270 up to 300, the slope of the MPA curve changed slightly. This figure also shows that local smoothing regressions performed in a slightly different way compared with the rest of methods for the latest months when milk yield was again significantly increased. While the predicted residuals for the LR1 and LR2 models followed a pattern in which they had the same sign for several successive months, inconsistent patterns of deviations about zero were found for the SR and RR models indicating a more satisfactory description of the lactation data by these models (Figure 4.3). The average of absolute standardized residuals followed a similar pattern for all regressions excluding SR which remained nearly stable during lactation. Late lactations (later than month 6) were particularly difficult to predict with any of the LR models.

Table 4.3 shows the correlations between TD milk records and other statistics such as predicted values (PRED), residuals (R), standard deviation of residuals (STDR) and standardized residuals (STAR) for all regression procedures. The estimated correlations between predicted and observed milk yields were high for all models, but the observed milk yields showed an intermediate correlation with the residuals (15-28) with an exception for LR2-0.5 (=9). The existence of these correlations again indicates that higher prediction residuals generally are associated with a larger observed yields and vice versa. This table demonstrates that local regressions show slightly higher correlation of TD records with their predicted values compared to SR and RR, though this difference is negligible. On the other hand, increasing the smoothing parameter from 0.5 to 0.6 caused a decrease in PRED as an insufficient performance. In addition, existence of a higher correlation between milk yields and their residuals (R) than standardized residuals (STAR) implies that a larger (or smaller)

TD record has a higher (or lower) chance to be an outlier if it is detected by ABSR. However, standardizing the residuals somewhat alleviated this correlation by 5% (see Table 4.3).

Figure 4.4 illustrates with more details the pattern of correlation between the original and predicted TD records by each regression procedure in various parts of the lactation period. While the predicted values generated from the local regressions were highly correlated with the true values (Table 4.2), these models (except LR2-0.5 method) suffer from a lack of fit (i.e. under- and over-estimation of the predictions) at late lactation (Figure 4.2).

The narrow range of correlations and their high values (around 1.0 and 0.93) between predicted and observed milk yields over the lactation for all models suggest that the predicted values were more highly associated with the magnitude of milk yields than the stages of lactation. Furthermore, local regressions with the smaller smoothing parameter 0.5 predicted the original points better than 0.6 did.

Table 4.3: Correlations between TD records and residual statistics for various regressions[1].

Statistics	SR	RR	LR-0.5	LR2-0.5	LR1-0.6	LR2-0.6
PRED	97.322	97.306	98.625	99.663	97.429	98.926
R	22.987	23.136	20.458	9.182	27.930	15.500
STDR	3.328	10.273	15.218	10.503	20.026	15.354
STAR	19.317	20.174	15.264	5.597	22.335	10.663

1- In this table all correlations have been presented in percent.

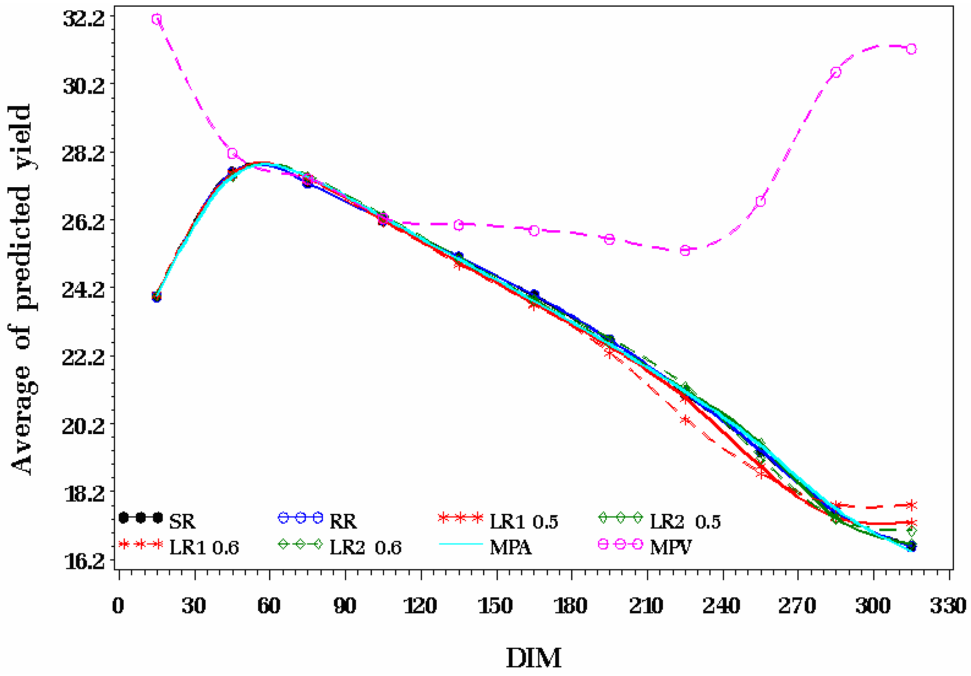

Figure 4.2: The plots of original data (MPA), its variance (MPV) and predicted values versus days in milk (DIM) for various regressions.

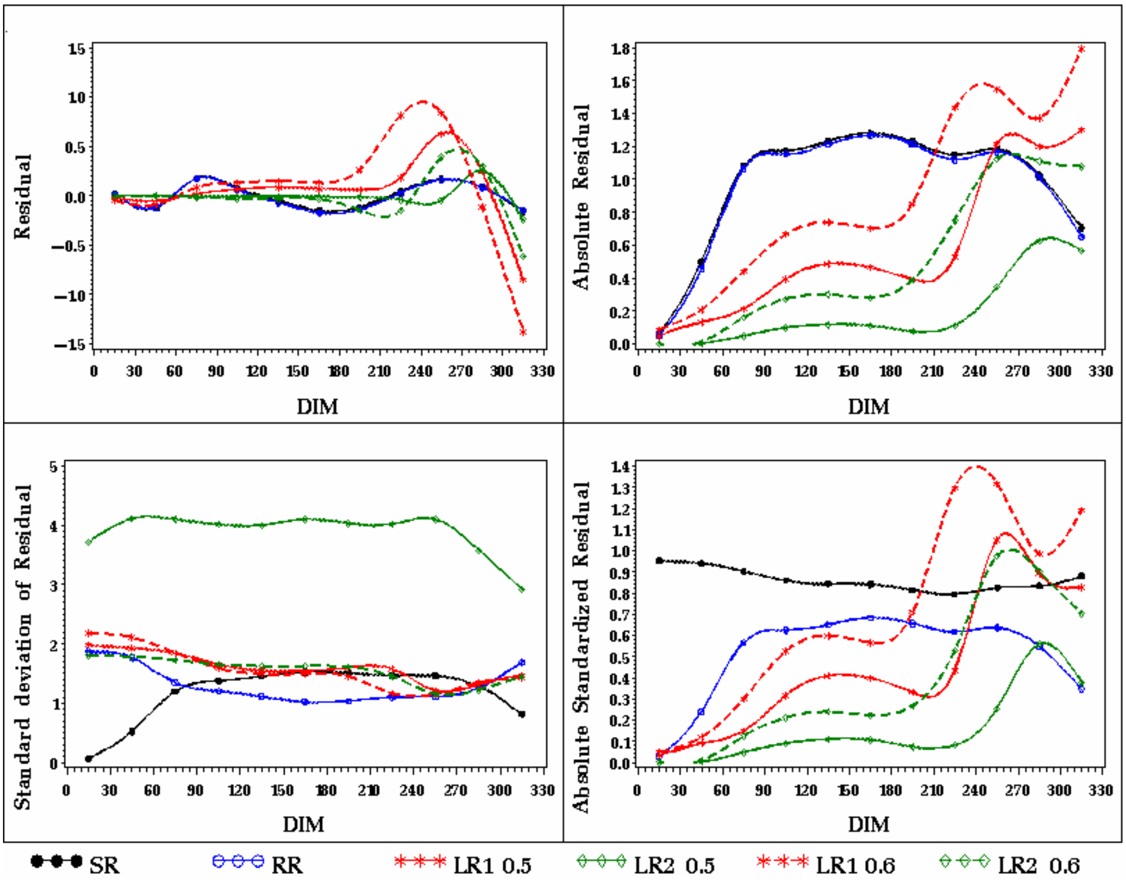

Figure 4.3: The plots of average residual, absolute residual, standard deviation of residual and absolute standardized residuals versus days in milk (DIM) for various regressions.

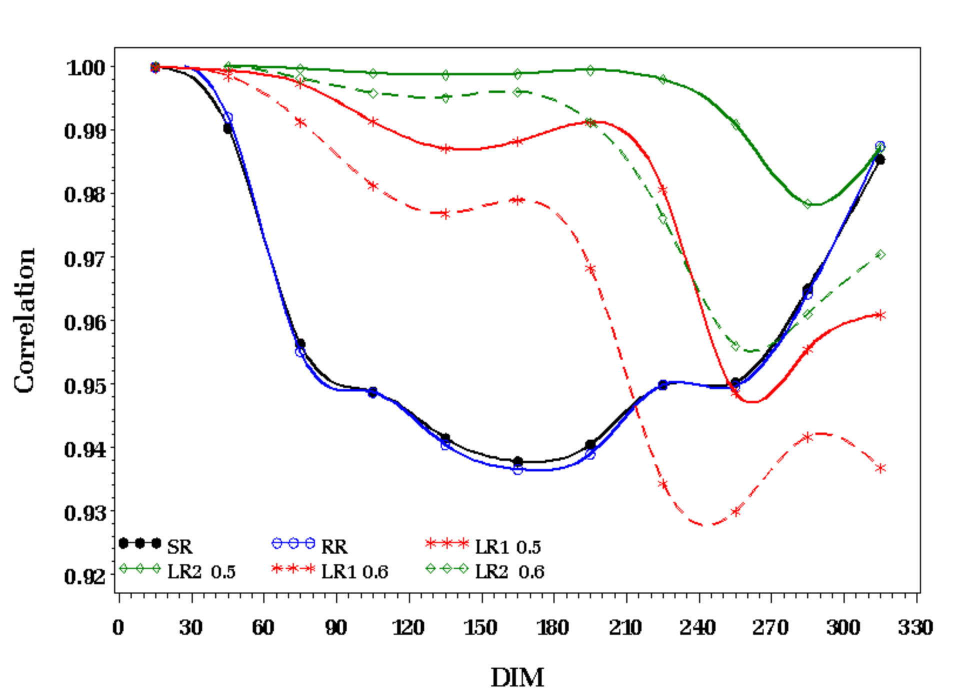

Figure 4.4: Correlation between TD yields and predicted values versus DIM for various regressions.

4.3. The distribution frequency of outliers

Although homogeneity of variance (homoscedasticity) is assumed in the model, the data should be tested for it. Several significance tests are available using absolute values of the residuals as Modified Levene’s test which is less sensitive to the normality assumption (NIST 2003). The significant F-test value of 173.01 ($p<0.0001$) for TD size and value of 26.70 ($p<0.0001$) for DIM are evidence that variances are not homogeneous (heteroscedasticity) between eight different production levels classified according to their (daily) milk kilograms (<10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40 and >40) and lactation stages (DIM from 5 to 330 with 30-days interval). A number of remedies for this heterogeneity exist including deleting outliers (if their removal is essential), using weights by weighted regression, transformations, nonparametric and robust procedures. All Tukey's Studentized Range Tests that were applied to the pairwise comparisons between different production levels and lactation stages, were significantly different ($p<0.0001$) with exception for DIM levels 45 vs. 75 and 15 vs. 165. According to Mayeres et al. (2003), the distribution of outliers should be independent of the value of production (TD milk yield) or the stage of lactation (DIM). Tables 4.4 and 4.5, and Figure 4.5 demonstrate the distribution pattern of the detected outliers (in percent) during the lactation course using all three regression strategies, SR, RR and LR (with linear and quadratic degrees as well as two smoothing parameters 0.5 and 0.6) and for both schemes. In these tables, the relative frequencies are based on the average of all coefficients k.

The outliers detected by ABSR (Figure 4.5a) for SR and RR were well distributed between peak and end of lactation. However both local regressions (especially for LR2) were somewhat different from SR and RR, a high portion of the outlying records occurred in the last three months of lactation. This could be a consequence of the goodness of the fit of the models over the lactation. Overall regressions (except RR), switching from ASTAR to ABSR slightly changed the incidence pattern of the abnormal recordings (Figure 4.5b). The observed difference between ABSR and ASTAR for the relative frequencies of detected outliers confirms that the predicted residuals do not have a constant variance over the lactation period. Moreover, Figure 4.5c shows the distribution of adjusted outlying observations over the lactation period. It is very obvious that the distributions for SR and RR are quite similar to those for discarding by the use of ASTAR, but two local regressions show a slight difference in distribution of outliers between adjusting scheme and discarding using ASTAR (Figure 4.5). Using SR, percentage of OTD yields at the peak was somewhat higher than other methods, which may be due to a smaller variance at this stage of lactation (see Figure 4.2). The interesting thing with RR is that the frequency of outliers did not arise in the early and in the late stages of lactation, confirming that the RR method has a sufficient fit for the data sets which resulted from both elimination and substitution of outlying observations.

In contrast to SR and RR procedures, when an adjusting scheme was applied to the data by both local regressions, the relative frequency of occurrence of outlying observations highly increased before the end of lactation (similar to discarding scheme). However, local regressions, especially when applying a larger smoothing parameter (0.6), did not fit the data accurately after DIM=150. Consequently, for larger smoothing parameters the distribution of outliers remains somewhat unsatisfactory. Although, both local regressions, particularly LR2 with the smaller smoothing parameter, matched the data well (see PRED in Table 4.3) but all of them suffer from the lack of uniform distribution of outliers over the lactation period.

Similar to DIM, it is necessary to examine the regression methods to find a relationship between size of a TD yield and the probability to be an outlier. However, there is an expectation that the number of outliers for the high (or low) production animals remains low and for the animals with moderate production remains more in agreement with their frequency. The findings shown in Tables 4.6 and 4.7 and Figure 4.6 seem to confirm this theory for both discarding (ABSR and ASTAR) and adjusting schemes. There is a small difference between ABSR and ASTAR in catching the outlying points, so ASTAR captures the smallest and the largest TD records with less probability than ABSR (Figure 4.6a and 4.6b). In other words, it captures the data points located near the center of the data, because they are more.

In the Table 4.6, the frequency (Freq) of TD values (in percent) in the category of 20-25 kg for ASTAR, for example, is 30.26 where its frequency for SR and LR1-0.5 is 24.56 and 33.04, respectively. Scaling of frequencies for the category >40 by converting them to the corresponding frequencies for the category of 20-25 kg resulted in 0.22 (24.56×0.27/30.26) and 0.29 (33.04×0.27/30.26) rather 0.65 and 0.38 for SR and LR1-0.5. The corresponding standardized quantities of the category <10 for SR and LR1-0.5 are 1.88 and 2.53, but the observed frequencies (Freq) are 8.28 and 3.98, respectively. These differences imply that despite standardizing the residuals, ASTAR still catches the TD records with the small and large values more than expected. The relatively insignificant value of the differences indicates that after standardizing the residuals, heterogeneous variance corresponding to TD magnitude and lactation stage were greatly taken into account. The incidence pattern for the distribution of the adjusted records by size classes is similar to that for ASTAR (Figure 4.6c). Despite dissimilarity between two strategies (discarding and adjusting), both methods captured smaller TD records as outliers more frequently than the bigger ones. This is more likely due to their higher count.

Since LR with the smoothing parameter =0.5 (for LR1 and LR2) described the data much better than 0.6, and since it provided a more consistent pattern for outliers over lactation, it is chosen as local regression candidates to examine the impact of outlier treatment on the results of genetic evaluations and consequently on ranking the top cows and bulls.

Table 4.4: Distribution of outliers detected (in percent) using the discarding scheme (ABSR or ASTAR) during lactation (DIM) for different regression methods.

DIM	Freq	ABSR						ASTAR					
		SR	RR	LR1-0.5	LR2-0.5	LR1-0.6	LR2-0.6	SR	RR	LR1-0.5	LR2-0.5	LR1-0.6	LR2-0.6
15	8.14	0.00	0.00	0.00	0.00	0.00	0.00	11.75	0.00	0.00	0.00	0.00	0.00
45	9.52	0.44	0.31	0.00	0.00	0.00	0.00	13.00	0.31	0.00	0.00	0.00	0.00
75	9.41	10.80	10.89	0.21	0.16	0.62	0.17	12.27	9.81	0.12	0.23	0.19	0.07
105	9.38	10.69	10.26	1.51	0.73	2.04	0.55	8.99	7.39	1.79	0.52	1.24	0.16
135	9.35	15.35	15.32	4.15	1.15	3.69	1.10	8.29	11.58	3.58	0.67	1.94	0.20
165	9.28	15.91	15.75	3.66	0.72	3.26	0.44	8.31	20.29	2.69	0.61	1.22	0.06
195	9.38	16.17	16.64	2.06	0.21	5.94	3.51	9.62	19.45	1.67	0.29	9.29	0.01
225	9.45	10.21	10.63	9.29	2.22	18.42	10.93	6.75	9.86	20.67	0.13	50.99	3.43
255	9.52	10.95	10.63	32.12	16.75	22.22	33.20	6.39	10.87	59.74	9.85	29.68	70.29
285	9.36	8.63	8.95	23.12	57.28	15.03	30.51	6.95	9.99	7.96	85.34	2.65	24.48
315	7.21	0.85	0.61	23.88	20.78	28.78	19.59	7.69	0.47	1.79	2.36	2.81	1.31

Table 4.5: Distribution of outliers detected (in percent) using the adjusting scheme during lactation (DIM) for different regression methods.

DIM	Freq	SR	RR	LR1-0.5	LR2-0.5	LR1-0.6	LR2-0.6
15	8.14	11.29	0.00	0.00	0.00	0.00	0.00
45	9.52	12.57	0.08	0.00	0.00	0.00	0.00
75	9.41	11.37	6.21	0.30	0.79	0.41	0.20
105	9.38	9.42	5.90	2.96	1.88	2.44	0.45
135	9.35	8.89	12.15	5.71	2.34	3.40	0.56
165	9.28	8.18	21.95	4.85	2.03	2.38	0.21
195	9.38	8.17	19.51	3.05	1.03	9.56	0.11
225	9.45	6.67	9.97	15.70	0.82	40.69	6.22
255	9.52	7.14	12.55	49.88	13.08	29.79	54.59
285	9.36	8.06	11.29	12.52	70.03	4.53	33.88
315	7.21	8.23	0.38	5.01	8.01	6.80	3.78

Table 4.6: Distribution of outliers detected (in percent) using the discarding scheme (ABSR or ASTAR) per various TD milk yield levels for different regression methods.

TD value	Freq	ABSR						ASTAR					
		SR	RR	LR1-0.5	LR2-0.5	LR1-0.6	LR2-0.6	SR	RR	LR1-0.5	LR2-0.5	LR1-0.6	LR2-0.6
10 <	2.32	14.88	15.17	23.74	11.98	33.08	18.69	8.28	6.66	3.98	3.19	5.73	4.32
10-15	8.06	19.39	19.92	15.19	15.71	16.33	15.14	13.88	16.44	11.04	19.24	11.22	14.17
15-20	20.77	16.19	16.58	16.13	26.98	10.95	18.98	20.02	24.82	27.35	35.86	22.09	30.97
20-25	30.26	17.60	17.63	22.85	27.48	15.49	25.54	24.56	24.88	33.04	29.81	30.19	34.17
25-30	25.13	13.23	12.10	13.34	13.32	13.75	14.68	18.69	14.80	16.97	10.20	20.87	13.25
30-35	10.98	9.94	9.61	6.20	3.96	6.99	5.77	9.17	7.32	6.12	1.18	7.71	2.80
35-40	2.21	4.50	4.69	0.93	0.47	1.34	0.65	4.75	4.14	1.12	0.50	1.44	0.29
> 40	0.27	4.27	4.31	1.63	0.10	2.08	0.55	0.65	0.95	0.38	0.01	0.75	0.04

Table 4.7: Distribution of outliers detected (in percent) using the adjusting scheme per various TD milk yield levels for different regression methods.

TD value	Freq	SR	RR	LR1-0.5	LR2-0.5	LR1-0.6	LR2-0.6
10 <	2.32	6.31	3.98	3.98	3.78	5.26	3.81
10-15	8.06	12.06	12.88	12.24	18.58	11.62	15.47
15-20	20.77	20.20	24.68	28.03	35.11	24.03	32.78
20-25	30.26	26.35	29.68	31.39	28.67	30.79	31.79
25-30	25.13	20.95	18.02	17.16	10.98	19.59	12.61
30-35	10.98	10.28	8.31	6.03	2.41	7.24	3.09
35-40	2.21	3.03	1.88	0.94	0.42	1.13	0.40
> 40	0.27	0.81	0.56	0.24	0.04	0.35	0.05

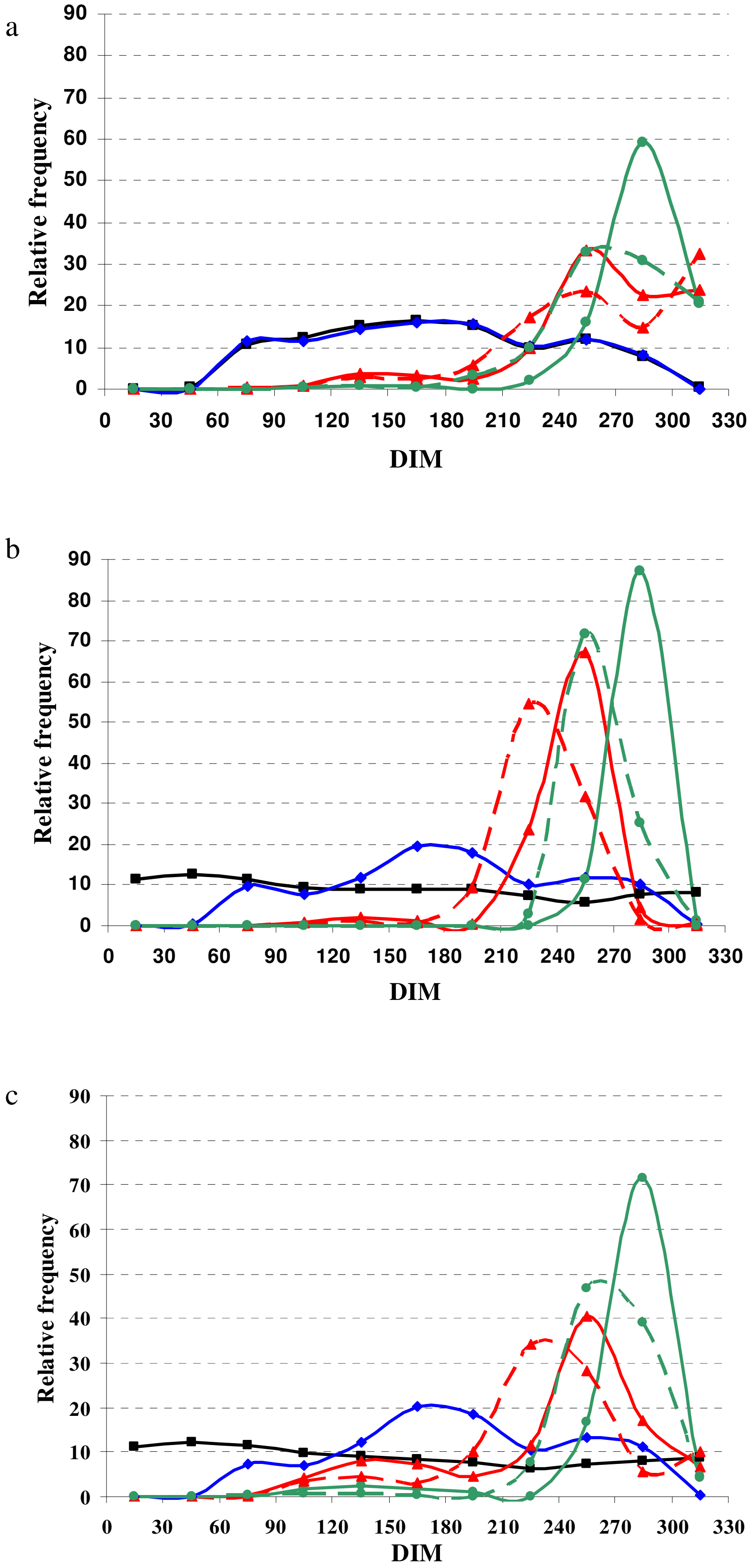

Figure 4.5: Plots of relative frequency of outliers for discarding scheme with both residual criteria ABSR (a) and ASTAR (b) and adjusting (c) scheme versus days in milk (DIM) for different regression methods. The symbols are for SR —■—, RR —◆—, LR1-0.5 —▲—, LR1-0.6 –▲–, LR2-0.5 —●— and LR2-0.6 –●–.

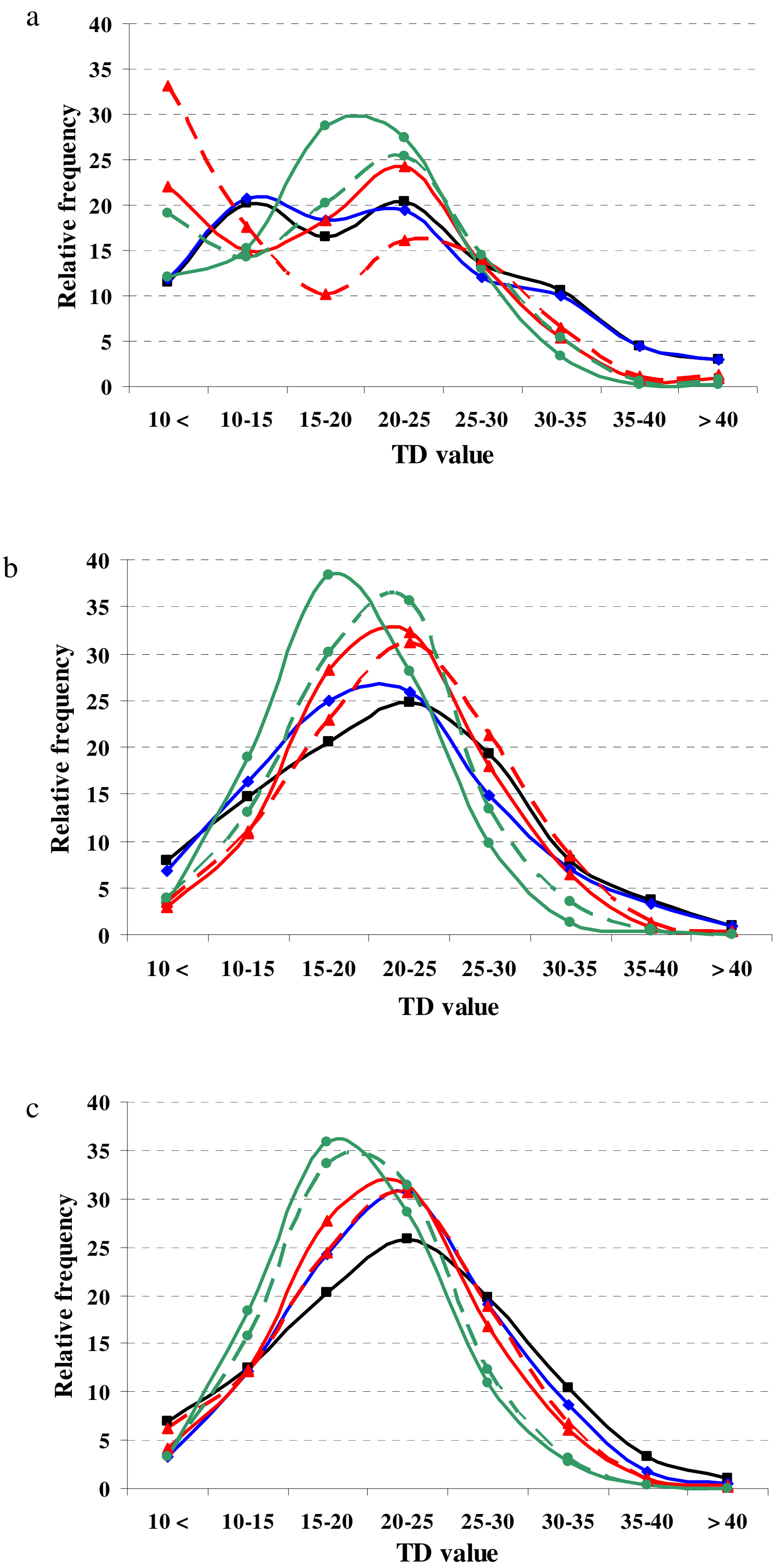

Figure 4.6: Plots of relative frequency of outliers for discarding scheme with both residual criteria ABSR (a) and ASTAR (b) and adjusting (c) scheme versus TD size levels (TD value) for different regression methods. The symbols are for SR —■—, RR —◆—, LR1-0.5 —▲—, LR1-0.6 –▲–, LR2-0.5 —●— and LR2-0.6 –●–.

4.4. Counting outliers per lactations

Tables 4.8 and 4.9 give the distribution of lactations (total number in all lactations) including outlying test-days (OTD) in various count classes of OTD number for the discarding approaches (ABSR and ASTAR). Table 4.10 presents the same information for the adjusting methods. It is remarkable that very few lactations with more than one abnormal record are identified.

Among the regression methods, SR had the highest proportion of lactations with several possible outliers by eight TD records (TD8) that indicates its inability to be independent of production levels. Due to this problem, using ABSR can result in elimination of the records that may not be an outlier or those generated from cows with low or high genetic potential leading to inaccurate selection results. An alternative to this is using ASTAR instead of ABSR. As it is clear from Table 4.9, using ASTAR greatly relieved this problem by standardizing the ordinary residuals with maximum four OTD per lactation. When the adjusting method was applied to the data, the proportion of lactations with a single outlier increased up to four OTD per lactation by correcting about 10% TD records (Table 4.10). This positive outcome for the adjusting case, however, may be due to using both residual and its standard deviation having correlation with the magnitude of TD records.

In the case of discarding of OTD, the frequency of the low outliers for SR and RR was larger than the frequency of high outliers. However, local regressions (LR1 and LR2) show a different pattern where the low outliers were less than 50% out of total OTD records (Table 4.8 and 4.9). Similar outcomes were obtained by adjusting the outlying candidates. These findings strongly emphasize that the proportion of low outliers depends on the regression method and the percentage of identified outliers as well as the fit of the model. Totally, RR showed the greatest percentage of lactations with a single OTD (up to 100%) and in contrast, SR had the greatest proportions of lactations with two or more outliers whether discarding or adjusting schemes were applied to the data.

For all regressions and schemes, however, along with increasing the proportion of detected outlying records, the frequency of lactations with two or more extreme yields increased. This occurs as a consequence of increasing the number of outliers and increasing the probability of observing two or more outliers within a specific lactation. By contrary, an increase in the proportion of detected outliers was associated with a decrease in the incidence of lactations with only a single abnormal record. However, current results do not provide information regarding relative frequency of low outliers in various stages of lactation that may be different for each type of regression procedure.

Table 4.8: The number of outlying TD (TD no.), no. of lactations with at least one outlying TD (Lactation), percentage of low outliers (Low), percentage of lactations with a single outlying TD (Single) and the number of lactations with one to eight outlying TD (TD1, TD2,..,TD8), by various percentages (Percent) for the discarding scheme using ABSR applied to different regressions.

Analysis	**Percent**	**TD no.**	**Lactation**	**Low**	**Single**	**TD 1**	**TD 2**	**TD 3**	**TD 4**	**TD 5**	**TD 6**	**TD 7**	**TD 8**
Full data	0	0	0	0	0	0	0	0	0	0	0	0	0
SR	0.1	111	102	69.37	91.18	93	9	0	0	0	0	0	0
SR	0.2	223	198	71.30	87.37	173	25	0	0	0	0	0	0
SR	0.4	446	385	70.18	84.94	327	55	3	0	0	0	0	0
SR	0.8	892	716	64.57	78.91	565	127	23	1	0	0	0	0
SR	1.6	1785	1306	60.22	71.98	940	279	64	20	3	0	0	0
SR	3.2	3571	2311	55.47	62.09	1435	611	180	59	19	6	1	0
SR	6.4	7142	3982	51.74	49.60	1975	1231	521	169	57	23	5	1
RR	0.1	111	104	69.37	93.27	97	7	0	0	0	0	0	0
RR	0.2	223	212	73.09	94.81	201	11	0	0	0	0	0	0
RR	0.4	446	411	72.87	91.48	376	35	0	0	0	0	0	0
RR	0.8	892	779	67.49	86.65	675	95	9	0	0	0	0	0
RR	1.6	1785	1428	62.75	79.48	1135	241	42	8	2	0	0	0
RR	3.2	3571	2551	57.41	70.64	1802	560	131	39	14	5	0	0
RR	6.4	7142	4351	53.08	57.48	2501	1208	434	141	46	18	3	0
LR1	0.1	111	96	53.15	85.42	82	13	1	0	0	0	0	0
LR1	0.2	223	183	49.78	79.78	146	34	3	0	0	0	0	0
LR1	0.4	446	353	50.22	75.35	266	81	6	0	0	0	0	0
LR1	0.8	892	664	46.30	69.43	461	178	25	0	0	0	0	0
LR1	1.6	1785	1239	44.99	61.26	759	417	60	3	0	0	0	0
LR1	3.2	3571	2333	43.21	56.49	1318	813	184	16	1	1	0	0
LR1	6.4	7142	4141	42.30	45.83	1898	1619	516	87	16	5	0	0
LR2	0.1	111	94	36.94	84.04	79	13	2	0	0	0	0	0
LR2	0.2	223	185	32.29	81.62	151	30	4	0	0	0	0	0
LR2	0.4	446	364	32.96	79.95	291	64	9	0	0	0	0	0
LR2	0.8	892	671	35.09	71.98	483	155	33	0	0	0	0	0
LR2	1.6	1785	1191	39.72	59.70	711	368	111	0	1	0	0	0
LR2	3.2	3571	2103	44.67	49.74	1046	668	373	11	4	1	0	0
LR2	6.4	7142	3525	50.78	36.06	1271	1062	1065	89	32	6	0	0

Table 4.9: The number of outlying TD (TD no.), no. of lactations with at least one outlying TD (Lactation), percentage of low outliers (Low), percentage of lactations with a single outlying TD (Single) and the number of lactations with one to eight outlying TD (TD1, TD2,..,TD8), by various percentages (Percent) for the discarding scheme using ASTAR applied to different regressions.

Analysis	**Percent**	**TD no.**	**Lactation**	**Low**	**Single**	**TD 1**	**TD 2**	**TD 3**	**TD 4**	**TD 5**	**TD 6**	**TD 7**	**TD 8**
Full data	0	0	0	0	0	0	0	0	0	0	0	0	0
SR	0.1	111	93	65.77	86.02	80	8	5	0	0	0	0	0
SR	0.2	223	199	61.88	90.45	180	14	5	0	0	0	0	0
SR	0.4	446	396	62.56	89.14	353	36	7	0	0	0	0	0
SR	0.8	892	778	62.00	87.02	677	88	13	0	0	0	0	0
SR	1.6	1785	1453	59.89	81.21	1180	214	59	0	0	0	0	0
SR	3.2	3571	2862	57.27	80.12	2293	429	140	0	0	0	0	0
SR	6.4	7142	5344	56.13	74.31	3971	949	423	1	0	0	0	0
RR	0.1	111	111	69.37	100.00	111	0	0	0	0	0	0	0
RR	0.2	223	223	68.16	100.00	223	0	0	0	0	0	0	0
RR	0.4	446	446	68.39	100.00	446	0	0	0	0	0	0	0
RR	0.8	892	892	65.92	100.00	892	0	0	0	0	0	0	0
RR	1.6	1785	1785	61.40	100.00	1785	0	0	0	0	0	0	0
RR	3.2	3571	3567	58.67	99.89	3563	4	0	0	0	0	0	0
RR	6.4	7142	6963	55.61	97.43	6784	179	0	0	0	0	0	0
LR1	0.1	111	111	23.42	100.0	111	0	0	0	0	0	0	0
LR1	0.2	223	223	24.22	100.0	223	0	0	0	0	0	0	0
LR1	0.4	446	446	23.32	100.0	446	0	0	0	0	0	0	0
LR1	0.8	892	892	24.33	100.0	892	0	0	0	0	0	0	0
LR1	1.6	1785	1785	25.66	100.0	1785	0	0	0	0	0	0	0
LR1	3.2	3571	3489	28.42	97.65	3407	82	0	0	0	0	0	0
LR1	6.4	7142	5989	33.24	80.80	4839	1147	3	0	0	0	0	0
LR2	0.1	111	111	13.51	100.0	111	0	0	0	0	0	0	0
LR2	0.2	223	223	21.08	100.0	223	0	0	0	0	0	0	0
LR2	0.4	446	446	21.97	100.0	446	0	0	0	0	0	0	0
LR2	0.8	892	892	20.52	100.0	892	0	0	0	0	0	0	0
LR2	1.6	1785	1785	22.30	100.0	1785	0	0	0	0	0	0	0
LR2	3.2	3571	3571	25.93	100.0	3571	0	0	0	0	0	0	0
LR2	6.4	7142	6284	38.07	86.35	5426	858	0	0	0	0	0	0

Table 4.10: The number of outlying TD (TD no.), no. of lactations with at least one outlying TD (Lactation), percentage of low outliers (Low), percentage of lactations with a single outlying TD (Single) and the number of lactations with one to eight outlying TD (TD1, TD2,..,TD4), by various percentage (Percent) for the adjusting scheme applied to different regressions and coefficients (k).

Analysis	**Level**	**k**	**Percent**	**TD no.**	**Lactation**	**Low**	**Single**	**TD 1**	**TD 2**	**TD 3**	**TD 4**
Full data		0	0	0	0	0	0	0	0	0	0
SR	1	2.10	0.78	871	761	62.00	87.254	664	84	13	0
SR	2	2.05	1.22	1365	1140	60.66	83.070	947	161	32	0
SR	3	2.00	1.71	1905	1544	59.69	80.894	1249	229	66	0
SR	4	1.95	2.62	2927	2375	58.46	80.884	1921	356	98	0
SR	5	1.90	3.82	4260	3369	56.88	78.807	2655	537	177	0
SR	6	1.85	5.1	5686	4385	56.68	76.534	3356	757	272	0
SR	7	1.80	6.46	7206	5387	56.09	74.290	4002	953	430	2
SR	8	1.75	7.94	8861	6406	55.61	71.823	4601	1168	624	13
SR	9	1.70	9.46	10559	7401	55.12	69.288	5128	1422	817	34
SR	10	1.65	11.04	12321	8268	54.36	64.998	5374	1810	1009	75
RR	1	3.4	0.73	820	820	65.73	100.00	820	0	0	0
RR	2	3.2	1.13	1258	1256	64.79	99.841	1254	2	0	0
RR	3	3.0	1.72	1924	1921	62.16	99.844	1918	3	0	0
RR	4	2.8	2.62	2924	2899	61.18	99.138	2874	25	0	0
RR	5	2.6	3.78	4224	4137	58.93	97.897	4050	87	0	0
RR	6	2.4	5.39	6016	5754	56.80	95.464	5493	260	1	0
RR	7	2.3	6.37	7114	6665	56.41	93.308	6219	443	3	0
RR	8	2.2	7.5	8371	7618	55.72	90.194	6871	741	6	0
RR	9	2.1	8.75	9760	8539	54.98	85.877	7333	1191	15	0
RR	10	2.0	10.12	11299	9387	54.08	79.994	7509	1845	32	1
LR1	1	2.6	0.37	410	410	23.41	100.00	410	0	0	0
LR1	2	2.4	0.74	827	827	24.79	100.00	827	0	0	0
LR1	3	2.2	1.23	1378	1378	25.25	100.00	1378	0	0	0
LR1	4	2.0	2.05	2285	2283	27.31	99.912	2281	2	0	0
LR1	5	1.8	3.28	3665	3570	28.81	97.339	3475	95	0	0
LR1	6	1.7	4.05	4525	4290	30.19	94.522	4055	235	0	0
LR1	7	1.6	5.07	5661	5120	31.69	89.434	4579	541	0	0
LR1	8	1.5	6.49	7241	6033	33.38	80.043	4829	1200	4	0
LR1	9	1.4	8.61	9607	7318	34.62	69.596	5093	2161	64	0
LR1	10	1.3	11.4	12722	9066	34.66	62.784	5692	3093	280	1
LR2	1	1.4	0.13	140	140	17.14	100.00	140	0	0	0
LR2	2	1.3	1.64	1827	1827	22.33	100.00	1827	0	0	0
LR2	3	1.2	3.33	3713	3713	25.94	100.00	3713	0	0	0
LR2	4	1.1	4.04	4508	4508	27.99	100.00	4508	0	0	0
LR2	5	1.0	4.50	5020	5020	29.70	100.00	5020	0	0	0
LR2	6	0.9	5.00	5576	5547	31.73	99.477	5518	29	0	0
LR2	7	0.8	5.57	6211	6012	34.47	96.690	5813	199	0	0
LR2	8	0.7	6.51	7260	6300	38.50	84.762	5340	960	0	0
LR2	9	0.6	8.37	9343	6382	45.01	54.967	3508	2787	87	0
LR2	10	0.5	11.13	12419	6385	50.17	15.129	966	4824	575	20

4.5. Impact on residual criteria

Tables 4.11 through 4.14 and Figures 4.7 and 4.8 illustrate the milk production average (MPA), mean of absolute residuals (MAR), mean of squared residuals (MSR) and standard deviation of residuals (SDR) obtained from the full data which individual records were deleted and corrected using different estimation procedures (i.e. SR, RR, LR1 and LR2). Perhaps it is again proper to state that henceforth, LR1 and LR2 imply first and second order of local regressions with a smoothing parameter =0.5.

While differences among regressions were small, MPA for both ABSR and ASTAR showed an increasing trend with increasing discarding percentage. But for SR and RR analyses, the increase in MPA alleviated after k=0.8 in contrast to LR1 and LR2 which continuously increased (Tables 4.11 and 4.12 and Figure 4.7). The increasing trend of MPA indicates that the sum of absolute deviations from the average of milk production of the whole dataset (=22.98 kg) for low outliers is larger than that for high outliers. The variation in MPA depends on both the number and the distance of low and high outliers from the average of the data. Since the average of all observations for the low outliers is greater than that for the high outliers, consequently discarding them leads to an increase in the milk average (MPA). This is expected for SR and RR because the incidence of abnormal records was fairly equally distributed over the lactation course. However, an increase in the percentage of discarded outlying data for SR and RR caused a decrease in the ratio of low outliers to total outliers (Table 4.8, 4.9 and 4.10) and consequently the increase of MPA stopped. The reduction in MPA after discarding 1.6% of outliers with SR and RR significantly implies that the low outliers are less deviated from their predicted values compared to the high outliers although the amount of low outliers is greater than that of high outliers. Therefore, the total effect of discarding the lower outliers on MPA will be smaller than the effect of the higher outliers that leads to an increase in MPA.

Abnormal low and high records for both local regressions, however, were accumulated in the last stage of lactation when milk production decreases and eventually stops. In this stage the milk yield average is generally lower than the total average (=22.98 kg), consequently their elimination will lead to an increase in the production average of the sample (MPA). In contrast to SR and RR, the pattern of MPA for LR1 and LR2 using the adjusting method showed a linear decreasing trend with increasing the outlier number (Table 4.13 and 4.14 and Figure 4.8). This finding points to the existence of a relative difference between the two schemes of treatment of the outlying data. In the adjusting case, the abnormal records are not discarded; rather their large deviations are reduced by winsorization or reducing the weight of outliers resulting in maintenance of the relative impact of their original yields on the total average. For LR1 and LR2, this finally resulted in a MPA towards a lower level due to higher incidence of high abnormalities in the end of lactation where OTD yields on average are smaller than the data average. For SR and RR, this does not lead to a slight change in MPA owing to relatively equal distribution of outliers over the lactation.

All methods performed better than data with no treatment in terms of MAR, MSR and SDR for both discarding and adjusting strategies. The original data without treatment always gave the higher residual criterions than both ABSR and ASTAR. However, some degree of dissimilarity is seen in the pattern of three residual terms between ABSR and ASTAR, so that those for ASTAR decreased more slowly compared with ABSR. The main cause for this finding is that the farthest data points (on both high and low sides) are less probable to be detected as outliers by ASTAR than ABSR, likely due to a lower correlation between milk yield records and their ASTAR compared to ABSR. The decreasing trend of the residual terms for SR and RR is quite similar (in both ABSR and ASTAR cases) in contrast to LR1 and LR2 methods. Among all regressions, LR2 contrary to RR, presented the smallest difference in residual terms from original data at a given discarding percentage, as a result of an inconsistent distribution of the detected outliers over the lactation.

It seems that k=0.4 is a good starting point to discard extreme values, because it included a high portion of total reduction in all three residual terms by removing a small number of data points. It is observable from Figure 4.7 that MSR and SDR curves are slightly sharper than those for MAR. It implies that there are few observations with very large differences from their expected values, consequently squaring these residuals boosts their impacts on MSR. Moreover, it is notable that in order to calculate the true total residual differences between each given analysis and the original data, one should multiply MAR and MSR by the number of TD records used by that analysis.

As was explained earlier, elimination of outliers might introduce bias, therefore it was decided to use a residual limit. The corresponding results have been presented in Tables 4.13 and 4.14 and Figure 4.8. The residual measures for data with adjusted records show a decreasing linear pattern. However, the observed difference for the pattern of residual terms between discarding and adjusting schemes again emphasizes that there is a small portion of large residuals (detected by the adopted model) in the data which should be noticed when trying to deal with the outliers. Similar to the discarding approach, for the adjusting case also the greatest decrease in residual terms was found by the RR method, besides a moderate reduction by SR and LR1. It is necessary to state that contrary to the discarding scheme, the adjusting level does not imply the exact percentage of the corrected records due to using different coefficients (i.e. k) which determine the final percentage of the affected TD records for each regression strategy. Therefore, in order to find the exact percentage of corrected TD yields for the adjusting case, one should be referred to Table 4.10.

Table 4.11: Milk production average (MPA), mean of absolute residuals (MAR), mean of squared residuals (MSR), standard deviation of residuals (SDR), variance components (V_a, V_{pe}, V_p and V_e) and their ratios (h^2, c^2 and e^2) and standard errors (SE) and mean log likelihood (MLL) from different regressions for discarding scheme using ABSR.

Analysis	**Percent**	**MPA**	**MAR**	**MSR**	**SDR**	V_{pe}	V_a	V_e	V_p	c^2	h^2	e^2	**SE(c^2)**	**SE(h^2)**	**SE(e^2)**	**MLL**
Full data	0	22.98	2.133	8.462	2.909	7.399	7.154	9.402	23.95	0.30889	0.29864	0.39248	0.01403	0.01582	0.00315	0.1477
SR	0.1	22.99	2.119	8.274	2.877	7.414	7.186	9.196	23.80	0.31156	0.30198	0.38646	0.01405	0.01583	0.00308	0.1326
SR	0.2	22.99	2.110	8.186	2.861	7.454	7.170	9.100	23.72	0.31419	0.30222	0.38359	0.01398	0.01570	0.00301	0.1253
SR	0.4	23.00	2.096	8.043	2.836	7.518	7.126	8.945	23.59	0.31872	0.30209	0.37919	0.01422	0.01598	0.00307	0.1138
SR	0.8	23.00	2.077	7.877	2.807	7.537	7.160	8.765	23.46	0.32125	0.30517	0.37359	0.01431	0.01608	0.00304	0.1004
SR	1.6	23.00	2.051	7.664	2.768	7.564	7.171	8.537	23.27	0.32503	0.30814	0.36684	0.01436	0.01607	0.00297	0.0856
SR	3.2	23.00	2.016	7.429	2.726	7.527	7.248	8.291	23.07	0.32633	0.31421	0.35946	0.01466	0.01641	0.00297	0.0731
SR	6.4	22.98	1.972	7.151	2.674	7.570	7.197	8.012	22.78	0.33230	0.31600	0.35170	-	-	-	0.0661
RR	0.1	22.99	2.118	8.271	2.876	7.415	7.181	9.192	23.79	0.31171	0.30188	0.38642	0.01402	0.01578	0.00307	0.1323
RR	0.2	22.99	2.109	8.176	2.859	7.452	7.178	9.089	23.72	0.31419	0.30262	0.38319	0.01400	0.01572	0.00300	0.1244
RR	0.4	23.00	2.095	8.038	2.835	7.505	7.150	8.938	23.59	0.31809	0.30305	0.37886	0.01367	0.01544	0.00303	0.1132
RR	0.8	23.00	2.075	7.861	2.804	7.522	7.186	8.747	23.45	0.32071	0.30637	0.37292	0.01431	0.01612	0.00303	0.0989
RR	1.6	23.01	2.048	7.651	2.766	7.577	7.159	8.522	23.26	0.32577	0.30781	0.36642	0.01429	0.01597	0.00294	0.0846
RR	3.2	23.00	2.014	7.414	2.723	7.528	7.245	8.274	23.05	0.32665	0.31434	0.35901	0.01469	0.01643	0.00297	0.0707
RR	6.4	22.99	1.972	7.155	2.675	7.535	7.212	8.017	22.76	0.33101	0.31682	0.35218	0.01490	0.01661	0.00299	0.0663
LR1	0.1	22.99	2.123	8.347	2.889	7.405	7.184	9.276	23.87	0.31028	0.30102	0.38870	0.01407	0.01585	0.00316	0.1394
LR1	0.2	22.99	2.117	8.284	2.878	7.411	7.201	9.208	23.82	0.31114	0.30230	0.38656	0.01407	0.01585	0.00310	0.1352
LR1	0.4	23.01	2.107	8.186	2.861	7.415	7.250	9.102	23.77	0.31198	0.30505	0.38298	0.01388	0.01557	0.00301	0.1291
LR1	0.8	23.02	2.093	8.061	2.839	7.494	7.206	8.968	23.67	0.31664	0.30445	0.37891	0.01431	0.01608	0.00310	0.1220
LR1	1.6	23.06	2.068	7.856	2.803	7.556	7.171	8.749	23.48	0.32185	0.30546	0.37268	0.01429	0.01605	0.00300	0.1138
LR1	3.2	23.13	2.038	7.622	2.761	7.627	7.144	8.505	23.28	0.32767	0.30693	0.36540	0.01449	0.01617	0.00305	0.1086
LR1	6.4	23.25	1.990	7.285	2.699	7.725	7.145	8.162	23.03	0.33543	0.31020	0.35437	0.01461	0.01631	0.00298	0.1092
LR2	0.1	22.99	2.128	8.416	2.901	7.426	7.140	9.352	23.92	0.31047	0.29852	0.39101	0.01398	0.01579	0.00314	0.1444
LR2	0.2	22.99	2.126	8.392	2.897	7.425	7.151	9.327	23.90	0.31062	0.29918	0.39019	0.01378	0.01567	0.00313	0.1430
LR2	0.4	23.00	2.123	8.367	2.893	7.425	7.148	9.301	23.87	0.31100	0.29940	0.38960	0.01387	0.01574	0.00313	0.1426
LR2	0.8	23.02	2.116	8.308	2.882	7.477	7.113	9.241	23.83	0.31374	0.29849	0.38776	-	-	-	0.1414
LR2	1.6	23.06	2.101	8.195	2.863	7.505	7.133	9.123	23.76	0.31585	0.30021	0.38394	0.01395	0.01567	0.00304	0.1400
LR2	3.2	23.13	2.083	8.053	2.838	7.552	7.173	8.982	23.71	0.31857	0.30255	0.37888	0.01383	0.01547	0.00296	0.1438
LR2	6.4	23.27	2.055	7.846	2.801	7.712	7.099	8.786	23.60	0.32683	0.30083	0.37234	0.01429	0.01603	0.00304	0.1550

Table 4.12: Milk production average (MPA), mean of absolute residuals (MAR), mean of squared residuals (MSR), standard deviation of residuals (SDR), variance components (V_a, V_{pe}, V_p and V_e) and their ratios (h^2, c^2 and e^2) and standard errors (SE) and mean log likelihood (MLL) from different regressions for discarding scheme using ASTAR.

Analysis	**Percent**	**MPA**	**MAR**	**MSR**	**SDR**	**V_{pe}**	**V_a**	**V_e**	**V_p**	**c^2**	**h^2**	**e^2**	**SE(c^2)**	**SE(h^2)**	**SE(e^2)**	**MLL**
Full data	0	22.98	2.133	8.462	2.909	7.399	7.154	9.402	23.95	0.30889	0.29864	0.39248	0.01403	0.01582	0.00315	0.1477
SR	0.1	22.99	2.128	8.412	2.900	7.400	7.170	9.348	23.92	0.30937	0.29979	0.39084	0.01400	0.01582	0.00314	0.1442
SR	0.2	22.99	2.124	8.377	2.894	7.406	7.176	9.311	23.89	0.30996	0.30035	0.38969	0.01381	0.01570	0.00314	0.1420
SR	0.4	22.99	2.117	8.299	2.881	7.417	7.168	9.226	23.81	0.31150	0.30102	0.38748	0.01402	0.01578	0.00309	0.1362
SR	0.8	23.00	2.105	8.198	2.863	7.386	7.225	9.119	23.73	0.31125	0.30446	0.38429	0.01389	0.01559	0.00302	0.1298
SR	1.6	23.01	2.087	8.055	2.838	7.360	7.295	8.968	23.62	0.31155	0.30882	0.37963	0.01430	0.01606	0.00312	0.1229
SR	3.2	23.02	2.054	7.787	2.791	7.441	7.274	8.688	23.40	0.31795	0.31081	0.37124	0.01435	0.01616	0.00304	0.1107
SR	6.4	23.04	1.999	7.380	2.717	7.577	7.263	8.267	23.11	0.32790	0.31432	0.35778	0.01474	0.01649	0.00299	0.0933
RR	0.1	22.99	2.127	8.401	2.898	7.407	7.155	9.336	23.90	0.30994	0.29941	0.39065	0.01390	0.01576	0.00313	0.1431
RR	0.2	22.99	2.123	8.361	2.892	7.387	7.185	9.293	23.87	0.30954	0.30106	0.38939	0.01406	0.01587	0.00313	0.1405
RR	0.4	23.00	2.116	8.300	2.881	7.400	7.178	9.228	23.81	0.31085	0.30152	0.38762	0.01345	0.01507	0.00301	0.1363
RR	0.8	23.00	2.105	8.207	2.865	7.436	7.151	9.128	23.72	0.31354	0.30154	0.38492	0.01408	0.01588	0.00308	0.1300
RR	1.6	23.01	2.089	8.073	2.841	7.409	7.187	8.988	23.58	0.31415	0.30474	0.38111	0.01420	0.01595	0.00314	0.1222
RR	3.2	23.03	2.061	7.857	2.803	7.480	7.160	8.765	23.40	0.31959	0.30591	0.37450	0.01425	0.01604	0.00303	0.1104
RR	6.4	23.04	2.032	7.657	2.767	7.544	7.086	8.575	23.21	0.32511	0.30535	0.36954	0.01443	0.01618	0.00301	0.1081
LR1	0.1	22.99	2.131	8.444	2.906	7.409	7.148	9.384	23.94	0.30948	0.29856	0.39196	0.01403	0.01583	0.00314	0.1466
LR1	0.2	22.99	2.130	8.438	2.905	7.406	7.147	9.378	23.93	0.30948	0.29864	0.39188	0.01403	0.01583	0.00315	0.1466
LR1	0.4	22.99	2.129	8.427	2.903	7.402	7.143	9.368	23.91	0.30955	0.29869	0.39175	0.01401	0.01581	0.00314	0.1466
LR1	0.8	23.00	2.126	8.402	2.899	7.405	7.134	9.344	23.88	0.31004	0.29871	0.39125	0.01382	0.01569	0.00314	0.1459
LR1	1.6	23.01	2.120	8.346	2.889	7.415	7.114	9.290	23.82	0.31130	0.29867	0.39004	0.01395	0.01574	0.00309	0.1445
LR1	3.2	23.04	2.109	8.249	2.872	7.429	7.097	9.198	23.72	0.31315	0.29915	0.38771	0.01403	0.01583	0.00311	0.1446
LR1	6.4	23.15	2.082	8.044	2.836	7.458	7.130	9.005	23.59	0.31612	0.30220	0.38168	0.01390	0.01557	0.00299	0.1559
LR2	0.1	22.99	2.132	8.458	2.908	7.400	7.161	9.399	23.96	0.30884	0.29889	0.39227	0.01403	0.01583	0.00315	0.1478
LR2	0.2	22.99	2.131	8.452	2.907	7.408	7.153	9.393	23.95	0.30926	0.29862	0.39212	0.01404	0.01584	0.00315	0.1482
LR2	0.4	23.00	2.131	8.451	2.907	7.409	7.156	9.394	23.96	0.30923	0.29867	0.39210	0.01404	0.01584	0.00315	0.1494
LR2	0.8	23.02	2.129	8.435	2.904	7.428	7.140	9.380	23.95	0.31017	0.29815	0.39168	0.01405	0.01584	0.00315	0.1507
LR2	1.6	23.05	2.127	8.420	2.902	7.444	7.123	9.372	23.94	0.31095	0.29757	0.39149	0.01404	0.01583	0.00315	0.1547
LR2	3.2	23.12	2.124	8.403	2.899	7.449	7.131	9.369	23.95	0.31105	0.29774	0.39121	0.01405	0.01584	0.00315	0.1650
LR2	6.4	23.22	2.117	8.353	2.890	7.489	7.104	9.348	23.94	0.31282	0.29674	0.39044	-	-	-	0.1862

Table 4.13: Milk production average (MPA), mean of absolute residuals (MAR), mean of squared residuals (MSR), standard deviation of residuals (SDR), variance components (V_a, V_{pe}, V_p and V_e) and their ratios (h^2, c^2 and e^2) and standard errors (SE) and mean log likelihood (MLL) from SR and RR for adjusting scheme.

Analysis	**Level**	**Percent**	**MPA**	**MAR**	**MSR**	**SDR**	V_{pe}	V_a	V_e	V_p	c^2	h^2	e^2	**SE(c^2)**	**SE(h^2)**	**SE(e^2)**	**MLL**
Full data	-	-	22.98	2.133	8.462	2.909	7.399	7.154	9.402	23.95	0.30889	0.29864	0.39248	0.01403	0.01582	0.00315	0.1477
SR	1	0.78	22.98	2.132	8.448	2.906	7.400	7.154	9.386	23.94	0.30911	0.29883	0.39207	0.01403	0.01583	0.00314	0.1466
SR	2	1.22	22.98	2.131	8.438	2.905	7.401	7.155	9.376	23.93	0.30925	0.29897	0.39178	0.01404	0.01584	0.00315	0.1459
SR	3	1.71	22.98	2.130	8.426	2.903	7.401	7.156	9.363	23.92	0.30941	0.29915	0.39143	0.01403	0.01583	0.00314	0.1449
SR	4	2.62	22.98	2.129	8.410	2.900	7.402	7.157	9.345	23.90	0.30965	0.29939	0.39096	0.01399	0.01580	0.00314	0.1437
SR	5	3.82	22.99	2.127	8.389	2.896	7.404	7.157	9.322	23.88	0.31000	0.29967	0.39033	0.01380	0.01568	0.00314	0.1420
SR	6	5.10	22.99	2.124	8.362	2.892	7.406	7.157	9.292	23.86	0.31046	0.30003	0.38951	0.01404	0.01585	0.00312	0.1398
SR	7	6.46	22.99	2.121	8.331	2.886	7.410	7.158	9.258	23.83	0.31100	0.30043	0.38857	0.01396	0.01575	0.00308	0.1373
SR	8	7.94	22.99	2.117	8.296	2.880	7.414	7.158	9.219	23.79	0.31163	0.30086	0.38752	0.01406	0.01584	0.00310	0.1345
SR	9	9.46	22.99	2.113	8.256	2.873	7.419	7.157	9.175	23.75	0.31237	0.30134	0.38630	0.01409	0.01590	0.00311	0.1312
SR	10	11.04	22.99	2.108	8.213	2.866	7.424	7.157	9.128	23.71	0.31314	0.30187	0.38499	0.01411	0.01591	0.00309	0.1277
RR	1	0.73	22.99	2.128	8.408	2.900	7.402	7.157	9.343	23.90	0.30968	0.29943	0.39089	0.01399	0.01581	0.00314	0.1435
RR	2	1.13	22.99	2.126	8.385	2.896	7.404	7.157	9.317	23.88	0.31007	0.29973	0.39020	0.01377	0.01566	0.00314	0.1416
RR	3	1.72	22.99	2.123	8.353	2.890	7.409	7.156	9.283	23.85	0.31067	0.30007	0.38925	0.01407	0.01586	0.00314	0.1391
RR	4	2.62	22.99	2.118	8.309	2.883	7.414	7.156	9.234	23.80	0.31145	0.30063	0.38792	0.01407	0.01587	0.00314	0.1355
RR	5	3.78	22.99	2.112	8.249	2.872	7.420	7.158	9.168	23.75	0.31248	0.30143	0.38608	0.01408	0.01589	0.00310	0.1305
RR	6	5.39	22.99	2.103	8.172	2.859	7.430	7.158	9.082	23.67	0.31389	0.30240	0.38370	0.01398	0.01571	0.00301	0.1239
RR	7	6.37	22.99	2.097	8.126	2.851	7.435	7.160	9.032	23.63	0.31468	0.30303	0.38229	0.01413	0.01590	0.00302	0.1200
RR	8	7.50	23.00	2.091	8.076	2.842	7.442	7.159	8.977	23.58	0.31564	0.30364	0.38072	0.01423	0.01602	0.00310	0.1157
RR	9	8.75	23.00	2.084	8.020	2.832	7.450	7.159	8.915	23.52	0.31670	0.30431	0.37899	0.01385	0.01551	0.00297	0.1109
RR	10	10.12	23.00	2.075	7.959	2.821	7.457	7.160	8.848	23.47	0.31779	0.30515	0.37706	0.01429	0.01608	0.00305	0.1056

Table 4.14: Milk production average (MPA), mean of absolute residuals (MAR), mean of squared residuals (MSR), standard deviation of residuals (SDR), variance components (V_a, V_{pe}, V_p and V_e) and their ratios (h^2, c^2 and e^2) and standard errors (SE) and mean log-likelihood (MLL) from LR1 and LR2 for adjusting scheme.

Analysis	**Level**	**Percent**	**MPA**	**MAR**	**MSR**	**SDR**	V_{pe}	V_a	V_e	V_p	c^2	h^2	e^2	**SE(c^2)**	**SE(h^2)**	**SE(e^2)**	**MLL**
Full data	-	-	22.98	2.133	8.462	2.909	7.399	7.154	9.402	23.95	0.30889	0.29864	0.39248	0.01403	0.01582	0.00315	0.1477
LR1	1	0.37	22.98	2.132	8.456	2.908	7.400	7.153	9.395	23.95	0.30900	0.29869	0.39231	0.01403	0.01582	0.00314	0.1472
LR1	2	0.74	22.98	2.131	8.447	2.906	7.400	7.153	9.385	23.94	0.30914	0.29879	0.39206	0.01403	0.01583	0.00314	0.1464
LR1	3	1.23	22.98	2.129	8.433	2.904	7.402	7.151	9.370	23.92	0.30940	0.29892	0.39168	0.01403	0.01583	0.00314	0.1452
LR1	4	2.05	22.98	2.127	8.410	2.900	7.403	7.149	9.345	23.90	0.30980	0.29917	0.39104	0.01397	0.01580	0.00314	0.1432
LR1	5	3.28	22.98	2.122	8.374	2.894	7.406	7.148	9.305	23.86	0.31039	0.29961	0.39000	0.01391	0.01577	0.00313	0.1400
LR1	6	4.05	22.98	2.120	8.349	2.890	7.406	7.148	9.278	23.83	0.31076	0.29994	0.38930	0.01410	0.01581	0.00314	0.1379
LR1	7	5.07	22.97	2.116	8.320	2.884	7.408	7.148	9.245	23.80	0.31124	0.30033	0.38843	-	-	-	0.1353
LR1	8	6.49	22.97	2.111	8.282	2.878	7.412	7.147	9.204	23.76	0.31191	0.30075	0.38734	0.01404	0.01583	0.00309	0.1321
LR1	9	8.61	22.97	2.106	8.234	2.870	7.415	7.148	9.151	23.71	0.31268	0.30143	0.38589	0.01408	0.01588	0.00309	0.1281
LR1	10	11.40	22.97	2.098	8.173	2.859	7.418	7.150	9.083	23.65	0.31365	0.30232	0.38404	0.01400	0.01573	0.00301	0.1229
LR2	1	0.13	22.98	2.133	8.461	2.909	7.399	7.154	9.401	23.95	0.30889	0.29865	0.39247	0.01403	0.01582	0.00315	0.1477
LR2	2	1.64	22.98	2.132	8.457	2.908	7.400	7.153	9.397	23.95	0.30898	0.29868	0.39234	0.01403	0.01582	0.00315	0.1472
LR2	3	3.33	22.98	2.131	8.443	2.906	7.403	7.150	9.382	23.94	0.30931	0.29873	0.39195	0.01404	0.01583	0.00315	0.1456
LR2	4	4.04	22.98	2.129	8.427	2.903	7.405	7.150	9.364	23.92	0.30958	0.29894	0.39148	0.01400	0.01580	0.00313	0.1438
LR2	5	4.50	22.98	2.127	8.412	2.900	7.407	7.149	9.347	23.90	0.30989	0.29907	0.39104	0.01384	0.01571	0.00314	0.1420
LR2	6	5.00	22.97	2.126	8.398	2.898	7.409	7.148	9.332	23.89	0.31015	0.29922	0.39062	0.01379	0.01568	0.00314	0.1402
LR2	7	5.57	22.97	2.124	8.384	2.895	7.412	7.147	9.316	23.87	0.31047	0.29934	0.39019	0.01403	0.01584	0.00315	0.1384
LR2	8	6.51	22.97	2.122	8.369	2.893	7.416	7.144	9.299	23.86	0.31082	0.29944	0.38975	0.01400	0.01581	0.00310	0.1367
LR2	9	8.37	22.97	2.120	8.349	2.889	7.420	7.142	9.277	23.84	0.31126	0.29958	0.38915	-	-	-	0.1347
LR2	10	11.13	22.97	2.117	8.320	2.885	7.424	7.142	9.246	23.81	0.31177	0.29993	0.38830	0.01395	0.01570	0.00308	0.1326

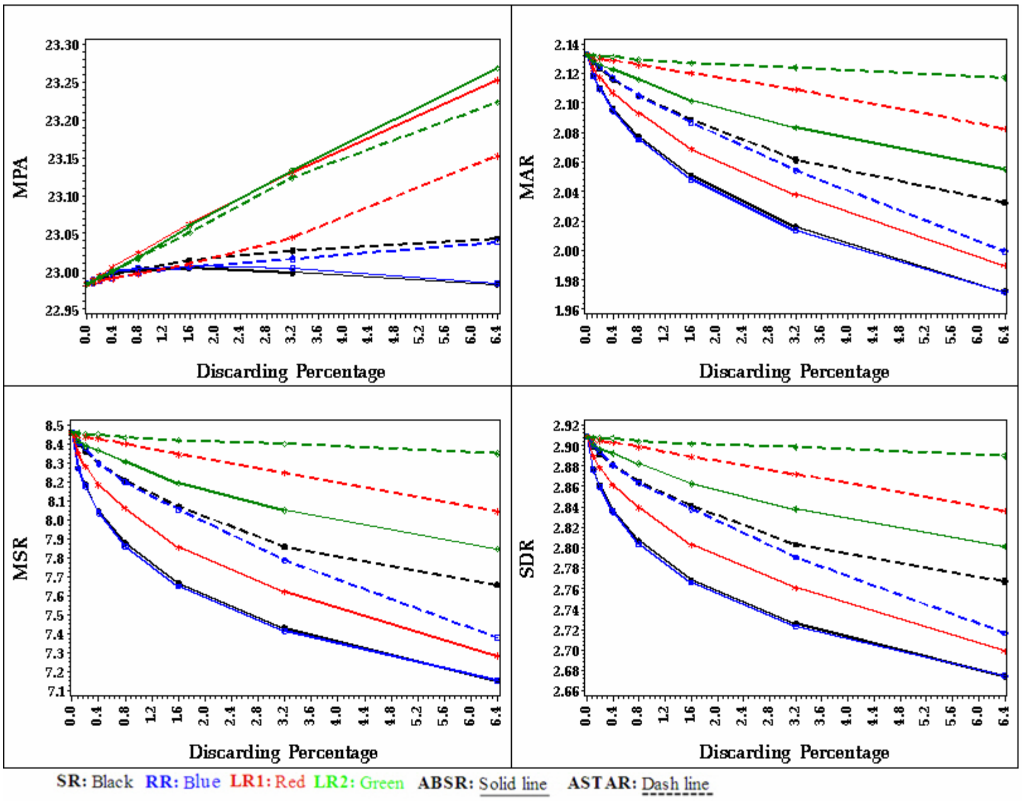

Figure 4.7: Milk production average (MPA), mean absolute residual (MAR), mean squared residual (MSR) and standard deviation of residual (SDR) for both discarding schemes by various percentages.

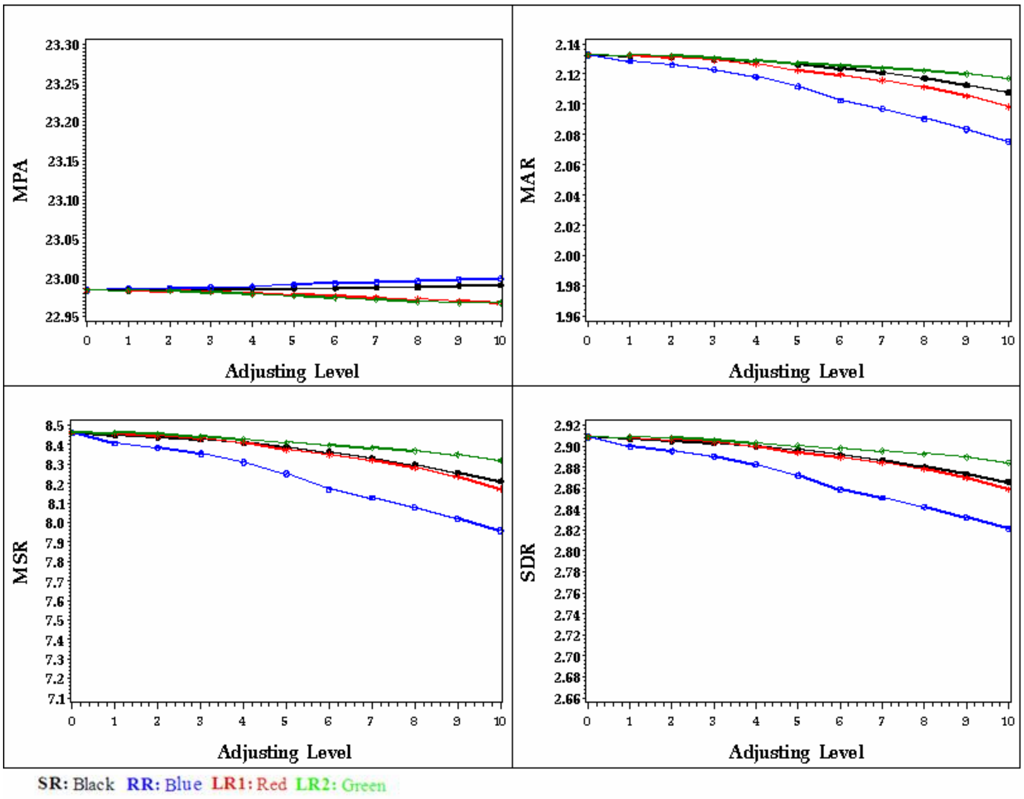

Figure 4.8: Milk production average (MPA), mean absolute residual (MAR), mean squared residual (MSR) and standard deviation of residual (SDR) for adjusting scheme by various levels.

4.6. Impact on variance components and standard deviations

Tables 4.11 to 4.14 also present the influence of removal and correction of outliers on the variance components as well as on the standard errors. Figures 4.9 through 4.13 illustrate the same information as plot. For both phenotypic (V_p) and residual (V_e) variances, a decreasing pattern was observed by omitting 0.1 to 6.4% observations with the largest ABSR, a logical consequence of discarding the extreme individual records which varied widely from their predicted values (Figure 4.9). Using ASTAR instead to ABSR in order to select OTD observations has slowed a decrease in V_p and V_e, likely due to not considering observations with the large predicted residuals which usually have larger TD values, as explained before. Despite small random fluctuations in additive genetic variance (V_a), it remained nearly stable by both ABSR and ASTAR methods. Along with increasing the percentage of discarded data, permanent environment variance (V_{pe}) increased using ABSR and it remained almost constant (without any considerable change) using ASTAR.

When data was treated by adjusting the OTD recordings, both V_p and V_e were moved in opposite direction of V_{pe}, though the variations were very small in scale. Despite negligible increase in V_{pe}, V_a remained constant for all methods through adjusting of outliers (Table 4.13 and Table 4.14 and Figure 4.10). Among all regressions applied to adjusting scheme, RR displayed the highest differences in all variances compared with the original analysis. Overall schemes and regressions, both V_e and V_p declined as a result of removing the observations with large residuals, before implementation of genetic evaluation.

At a particular percentage additive genetic variance ratio or heritability (h^2) and permanent environmental variance ratio (c^2) for all regressions (except LR2) which their extreme observations were identified by ABSR, were slightly higher compared with those selected by ASTAR. In comparison with the complete dataset, error variance ratio (e^2) decreased (similar to V_e) in a nonlinear way in all cases, while the mean of log-likelihood estimates (MLL) increased for LR1 (using ABSR) and LR2 (using both ABSR and ASTAR). These findings were illustrated in the Figure 4.11. Though larger values of log-likelihoods indicate better fitting models, this comparison should not be applied between full and reduced models (Fatehi et. al 2005), consequently in this case MLL is not comparable.

After adjusting of outliers using all regression methods, both h^2 and c^2 increased in contrast to e^2, but the greatest increase in them was found for RR compared to other methods (Table 4.13 and 4.14 and Figure 4.12). Both $h^2=0.2987$ and $c^2=0.3089$ were higher than 0.24 and 0.28, respectively, reported by Swalve (1995a) for German data. Kaya et al. (2003) found a $h^2=0.11$ using milk yield records from 1103 first lactations of Holstein cows. Furthermore, Mrode et al. (2002) using a single-trait fixed regression model, achieved a value 0.25 for first lactation milk yields of 1134042 Holstein cows.

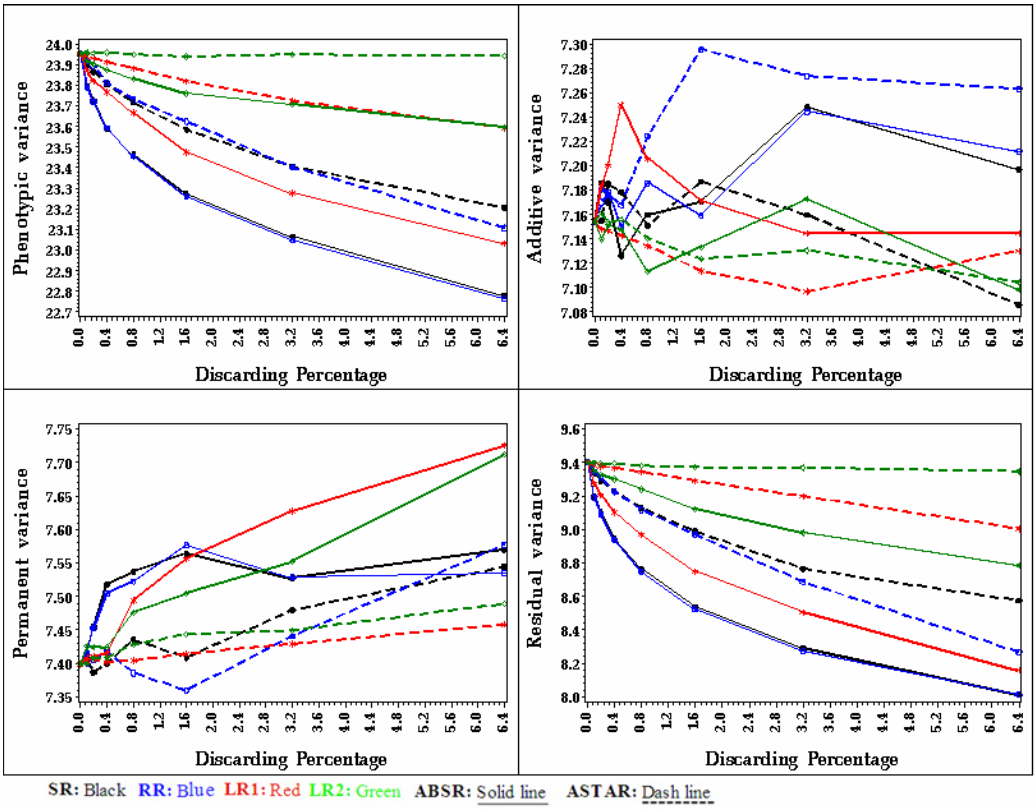

Figure 4.9: Phenotypic, additive genetic, permanent and residual variances for both discarding scheme by various percentages.

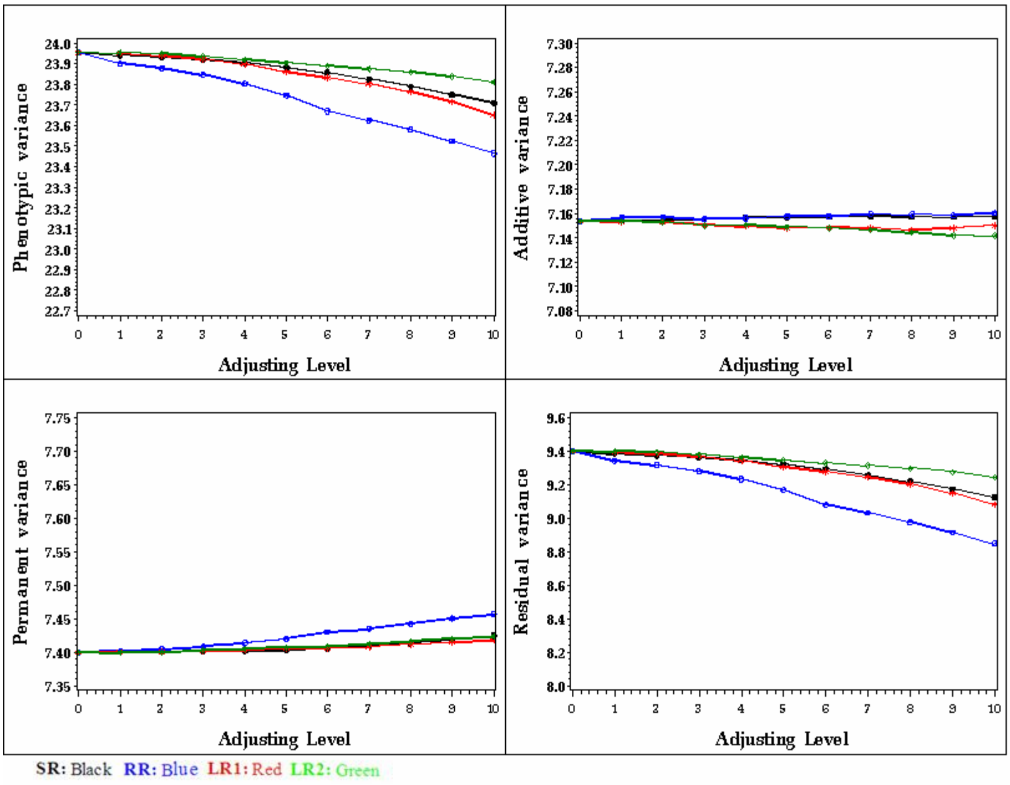

Figure 4.10: Phenotypic, additive genetic, permanent and residual variances for adjusting scheme by various levels.

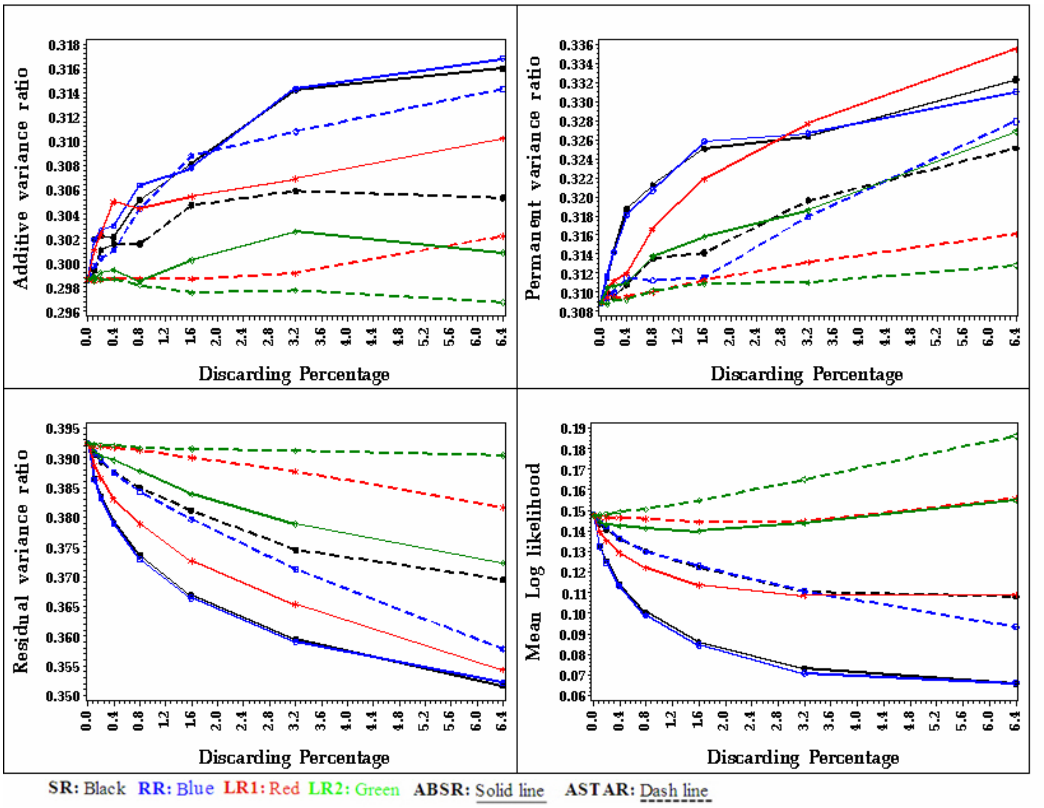

Figure 4.11: Additive genetic, permanent, residual variance ratios and mean log likelihood for both discarding schemes by various percentages.

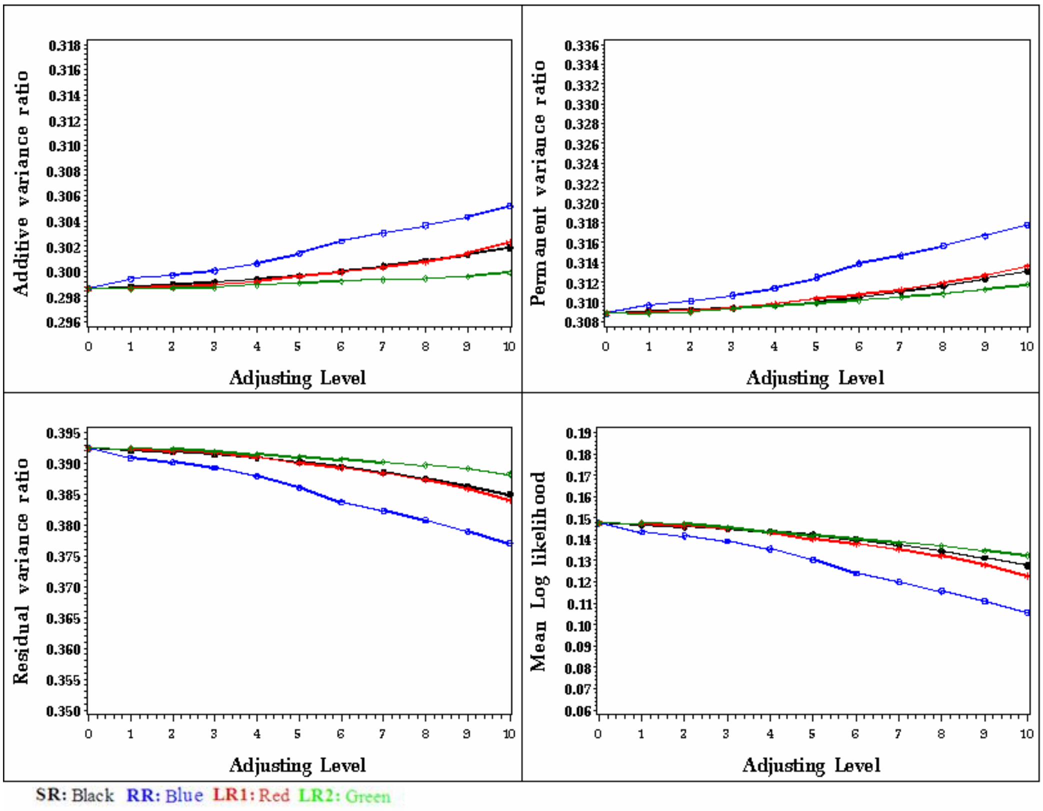

Figure 4.12: Additive genetic, permanent, residual variance ratios and mean log likelihood for adjusting scheme by various levels.

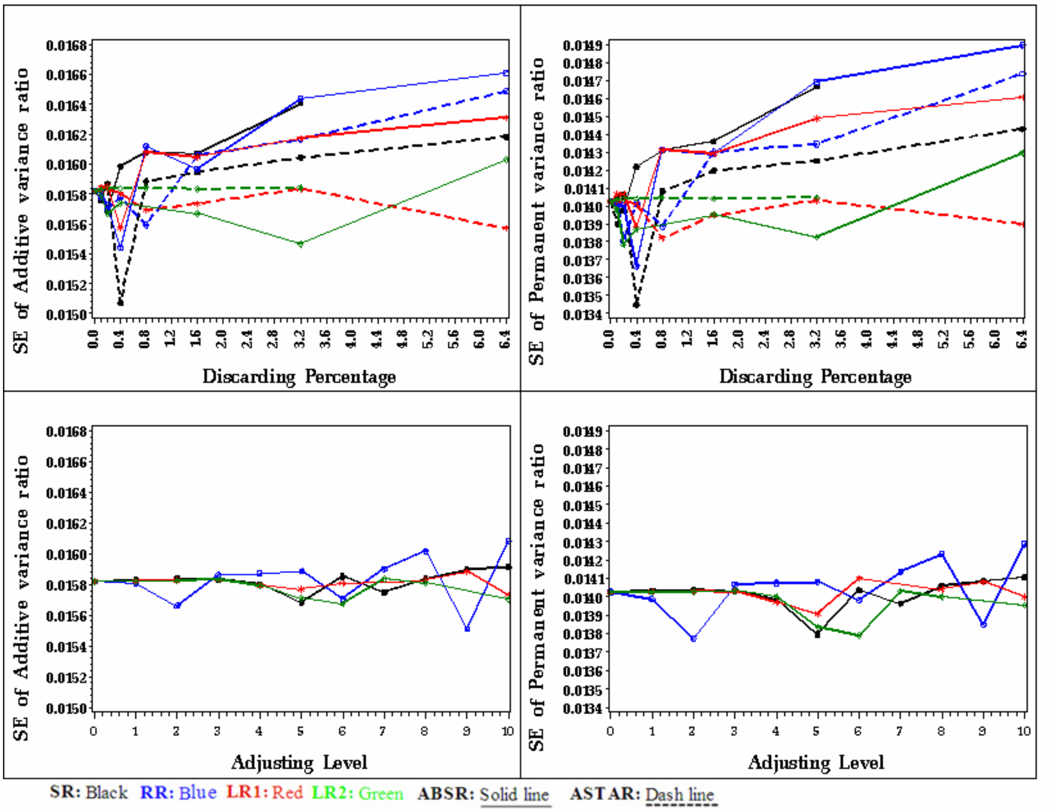

Figure 4.13: Standard error of additive and permanent environmental variance ratios for both discarding and adjusting schemes by various percentages and levels.

For the most discarding strategies, standard deviation of additive genetic variance ratio (SE(h^2)) and permanent environmental variance ratio (SE(c^2)) were minimized at k=0.4. For the adjusting schemes, the lowest level occurred at k=5 for SR with 3.82%, at k=2 for RR with 1.13%, and k=5 for LR1 and LR2 with 3.28% and 4.50% of the data, respectively. Figure 4.13 illustrates these findings for both schemes. This is surprising that standard deviations are more reduced using discarding method compared to adjusting method. In some cases however, the programme was not able to predict the standard deviation of parameters due to the fact, that the optimization algorithm did not reach to the convergence limit.

4.7. Impact on breeding values and ranking of top animals

Tables 4.15 and 4.16 contain the results in respect of the mean (MEBV), standard deviation (SDEBV), minimum (MINEBV) and maximum (MAXEBV) values of estimated breeding values (EBV) of milk yield for all animals in pedigree (N= 58315). Furthermore, they show the correlation (CORR) between EBVs of all animals generated from original data and those from discarding and adjusting schemes, respectively. Figure 4.14 demonstrates the variability of MEBV and CORR over the discarding percentages and adjusting levels. MEBV, SDEBV, MINEBV and MAXEBV practically remained stable for all regressions, although MEBV in most discarding cases followed a quadratic form. For the discarding scheme, however, the peak value of MEBV happened at k=0.8 or k=1.6. Figure 4.14 clearly represents the pattern of variation for MEBV for both discarding and adjusting schemes. Generally, as the proportion of discarded and corrected outliers increased, variations in MEBV became larger. In contrast to discarding, MEBV resulted from adjusting schemes show small differences with the data without treatment, but it follows a certain pattern: constant for SR, increasing for LR1, decreasing for LR2 and a upward and then downward for RR (Table 4.16). Correlations (CORR) were reduced in parallel with an increase in the proportion of discarded outliers, but a slightly greater decrease was seen by ABSR than by ASTAR, as a result of removing records with the largest deviations. The CORR between corrected and full data remained almost stable (Figure 4.14).

Tables 4.15 and 4.16 also give the average of estimated breeding values of top 100 bulls (MEBVS) and cows (MEBVC), number of animals in common in the top 100 bulls (NTSIC) and cows (NTCIC) lists between full data and different analyses, and spearman rank correlation of top 100 bulls (SRCORRS) and cows (SRCORRC) for the discarding and adjusting schemes, respectively. Both MEBVS and MEBVC actually remained stable as compared with those of the full data. The Figures 4.15 and 4.16 show MEBVS, MEBVC, SRCORRS and SRCORRC for discarding and adjusting schemes, respectively. In comparison to the to sires, rankings of top cows (SCORRC) were more highly affected by the estimation procedures, probably for the reason that bulls had a minimum of 50 TD records compared with a minimum of 9 records per cows (because data included only first lactation cows).

The final results from the Tables 4.15 and 4.16 indicate that no more than 6 (9) and 1 (1) bulls (cows) disappeared from the top 100 lists for discarding and adjusting schemes as compared with the full dataset. In comparison to ASTAR, rankings of top bulls and cows severely altered by the use of ABSR. Overall regression procedures, the list of the top 100 bulls which resulted from adjusted datasets, mostly remained without significant changes, in contrast to the discarding approach (Table 4.16). This provides evidence that removal of outlying observations affect the ranking of the top bulls and cows more than adjusting them. The similarity of rankings of bulls and cows between SR and RR methods is also shown in Figures 4.15 and 4.16.

Table 4.15: Mean, SD, Min and Max of EBVs, correlations of EBVs between methods and those for all animals (N= 58315) and ranking top 100 bulls and cows for discarding schemes[1].

Analysis	**scheme**	**k**	**EBV statistics for all animals**					**EBV statistics for top bulls**			**EBV statistics for top cows**		
			MEBV	**SDEBV**	**MINEBV**	**MAXEBV**	**CORR**	**MEBVS**	**NTSIC**	**SCORRS**	**MEBVC**	**NTCIC**	**SCORRC**
Full data	-	0	0.4371	1.6122	-5.5111	9.1054	-	5.1137	-	-	6.9972	-	-
SR	ABSR	0.1	0.4362	1.6165	-5.4798	9.1009	99.979	5.1191	99	99.677	7.0098	98	99.763
SR	ABSR	0.2	0.4386	1.6152	-5.4744	9.0863	99.969	5.1206	98	99.573	7.0047	97	99.645
SR	ABSR	0.4	0.4388	1.6125	-5.4452	9.0465	99.952	5.1245	97	99.488	6.9789	96	99.478
SR	ABSR	0.8	0.4414	1.6176	-5.4615	9.0619	99.932	5.1505	97	99.264	6.9998	95	99.055
SR	ABSR	1.6	0.4451	1.6199	-5.4811	9.0989	99.908	5.1669	97	99.017	7.0032	95	98.166
SR	ABSR	3.2	0.4369	1.6245	-5.5480	9.1719	99.869	5.1614	96	99.058	7.0193	93	96.755
SR	ABSR	6.4	0.4273	1.6137	-5.5806	9.1389	99.812	5.1264	96	98.537	6.9541	92	95.954
SR	ASTAR	0.1	0.4383	1.6143	-5.4716	9.1247	99.995	5.1200	99	99.949	7.0064	100	99.901
SR	ASTAR	0.2	0.4389	1.6136	-5.4689	9.1416	99.990	5.1167	99	99.878	7.0038	100	98.959
SR	ASTAR	0.4	0.4404	1.6133	-5.4569	9.1225	99.978	5.1223	98	99.608	6.9997	99	99.117
SR	ASTAR	0.8	0.4444	1.6219	-5.4904	9.1954	99.963	5.1374	96	99.517	7.0402	98	99.022
SR	ASTAR	1.6	0.4417	1.6268	-5.5445	9.2407	99.937	5.1472	96	99.215	7.0740	97	98.154
SR	ASTAR	3.2	0.4384	1.6264	-5.5444	9.1812	99.888	5.1414	95	98.464	7.0500	96	96.737
SR	ASTAR	6.4	0.4384	1.6231	-5.5244	9.1170	99.796	5.1310	95	97.518	7.0293	91	95.341
RR	ABSR	0.1	0.4364	1.6162	-5.4764	9.0987	99.978	5.1206	99	99.718	7.0090	98	99.756
RR	ABSR	0.2	0.4383	1.6170	-5.4737	9.0972	99.966	5.1269	98	99.546	7.0089	97	99.623
RR	ABSR	0.4	0.4385	1.6154	-5.4558	9.0668	99.950	5.1306	97	99.507	6.9919	96	99.460
RR	ABSR	0.8	0.4417	1.6189	-5.4793	9.0901	99.928	5.1500	97	99.146	7.0150	95	98.838
RR	ABSR	1.6	0.4420	1.6182	-5.4801	9.0806	99.904	5.1583	97	98.944	6.9866	95	97.808
RR	ABSR	3.2	0.4407	1.6254	-5.5447	9.1828	99.869	5.1670	96	98.897	7.0290	94	95.688
RR	ABSR	6.4	0.4297	1.6178	-5.5852	9.1697	99.813	5.1324	94	98.479	6.9730	93	93.672
RR	ASTAR	0.1	0.4399	1.6132	-5.4654	9.1076	99.994	5.1201	98	99.898	7.0002	100	99.840
RR	ASTAR	0.2	0.4405	1.6168	-5.4843	9.1422	99.989	5.1247	98	99.743	7.0159	99	98.855
RR	ASTAR	0.4	0.4414	1.6169	-5.4814	9.1503	99.980	5.1279	98	99.571	7.0130	99	98.657
RR	ASTAR	0.8	0.4390	1.6136	-5.4870	9.1084	99.966	5.1200	97	99.540	6.9896	99	98.043
RR	ASTAR	1.6	0.4340	1.6166	-5.5394	9.1301	99.945	5.1195	98	99.503	7.0009	99	97.506
RR	ASTAR	3.2	0.4328	1.6171	-5.5180	9.0707	99.902	5.1228	96	99.146	6.9829	97	96.821
RR	ASTAR	6.4	0.4287	1.6048	-5.5276	8.9933	99.848	5.0911	95	98.131	6.9264	92	94.839
LR1	ABSR	0.1	0.4394	1.6148	-5.5097	9.1220	99.988	5.1232	100	99.671	7.0126	99	99.717
LR1	ABSR	0.2	0.4393	1.6184	-5.5166	9.1389	99.982	5.1329	100	99.615	7.0204	98	99.596
LR1	ABSR	0.4	0.4413	1.6248	-5.5358	9.1793	99.971	5.1545	99	99.581	7.0447	98	98.727
LR1	ABSR	0.8	0.4463	1.6226	-5.4837	9.1237	99.957	5.1568	98	99.506	7.0148	97	98.380
LR1	ABSR	1.6	0.4393	1.6208	-5.4851	9.0667	99.929	5.1550	99	99.140	6.9917	97	97.319
LR1	ABSR	3.2	0.4303	1.6158	-5.4619	9.0162	99.883	5.1341	98	97.947	6.9504	96	96.223
LR1	ABSR	6.4	0.4250	1.6137	-5.4355	8.9474	99.793	5.1330	97	97.463	6.9209	92	92.986
LR1	ASTAR	0.1	0.4382	1.6119	-5.5105	9.1022	99.997	5.1136	99	99.960	6.9902	100	99.162
LR1	ASTAR	0.2	0.4386	1.6111	-5.5076	9.1002	99.996	5.1127	99	99.861	6.9906	100	99.116
LR1	ASTAR	0.4	0.4403	1.6117	-5.5002	9.1046	99.993	5.1160	100	99.807	6.9930	100	99.016
LR1	ASTAR	0.8	0.4406	1.6109	-5.4927	9.0958	99.986	5.1180	99	99.674	6.9909	98	98.630
LR1	ASTAR	1.6	0.4413	1.6091	-5.4807	9.0632	99.972	5.1150	98	99.338	6.9840	98	97.799
LR1	ASTAR	3.2	0.4356	1.6044	-5.4638	9.0336	99.949	5.1009	98	99.018	6.9683	97	96.581
LR1	ASTAR	6.4	0.4382	1.6123	-5.4738	9.0601	99.893	5.1341	97	98.769	6.9620	94	93.164
LR2	ABSR	0.1	0.4365	1.6101	-5.4981	9.0756	99.994	5.1071	99	99.958	6.9878	100	99.792
LR2	ABSR	0.2	0.4370	1.6114	-5.5047	9.0936	99.991	5.1114	99	99.869	6.9940	100	99.868
LR2	ABSR	0.4	0.4366	1.6107	-5.4987	9.0898	99.988	5.1134	99	99.855	6.9919	100	99.840
LR2	ABSR	0.8	0.4382	1.6074	-5.4708	9.0483	99.978	5.1077	99	99.758	6.9624	99	99.501
LR2	ABSR	1.6	0.4397	1.6151	-5.4573	9.0592	99.952	5.1367	99	99.173	6.9675	97	98.713
LR2	ABSR	3.2	0.4393	1.6221	-5.4707	9.0415	99.908	5.1520	98	98.386	6.9769	96	97.570
LR2	ABSR	6.4	0.4357	1.6111	-5.4224	8.9062	99.813	5.1211	98	96.731	6.8944	91	92.706
LR2	ASTAR	0.1	0.4367	1.6123	-5.5124	9.1088	99.999	5.1138	99	99.921	7.0024	100	99.741
LR2	ASTAR	0.2	0.4350	1.6107	-5.5092	9.0960	99.997	5.1087	99	99.888	6.9930	99	99.337
LR2	ASTAR	0.4	0.4347	1.6114	-5.5157	9.0995	99.995	5.1108	99	99.866	6.9944	99	99.286
LR2	ASTAR	0.8	0.4331	1.6087	-5.5097	9.0488	99.990	5.1021	99	99.734	6.9766	98	99.269
LR2	ASTAR	1.6	0.4302	1.6076	-5.4975	9.0258	99.981	5.0983	99	99.671	6.9622	96	98.785
LR2	ASTAR	3.2	0.4342	1.6107	-5.4766	9.0325	99.964	5.1066	99	99.540	6.9663	95	98.046
LR2	ASTAR	6.4	0.4270	1.6032	-5.4992	8.9963	99.919	5.0767	99	98.899	6.9048	94	96.174

[1] In this table, MEBV stands for mean, SDEBV for standard deviation, MINEBV for minimum and MAXEBV for maximum of estimated breeding values, CORR stands for correlation between cases and regular BLUP for all animals (=58315). Also, MEBVS and MEBVC stand for average of estimated breeding values, NTSIC and NTCIC stand for top 100 lists in common; and SRCORRS and SRCORRC stand for spearman rank correlation for bulls and cows, respectively for both excluding schemes (ABSR and ASTAR).

Table 4.16: Mean, SD, Min and Max of EBVs, correlations of EBVs between methods and those for all animals (N= 58315) and ranking top 100 bulls and cows for adjusting scheme[1].

Analysis	Percent	EBV statistics for all animals					EBV statistics for top bulls			EBV statistics for top cows		
		MEBV	SDEBV	MINEBV	MAXEBV	CORR	MEBVS	NTSIC	SCORRS	MEBVC	NTCIC	SCORRC
Full data	0	0.4371	1.6122	-5.5111	9.1054	-	5.1137	-	-	6.9972	-	-
SR	0.78	0.4373	1.6123	-5.5083	9.1066	100.00	5.1140	100	99.994	6.9975	100	99.990
SR	1.22	0.4374	1.6124	-5.5080	9.1079	100.00	5.1144	100	99.984	6.9980	100	99.988
SR	1.71	0.4375	1.6125	-5.5083	9.1093	100.00	5.1148	99	99.983	6.9986	100	99.977
SR	2.62	0.4374	1.6126	-5.5089	9.1107	100.00	5.1150	99	99.981	6.9989	100	99.977
SR	3.82	0.4374	1.6126	-5.5087	9.1114	100.00	5.1150	99	99.976	6.9988	100	99.957
SR	5.10	0.4373	1.6127	-5.5094	9.1118	100.00	5.1151	99	99.972	6.9990	100	99.930
SR	6.46	0.4373	1.6129	-5.5103	9.1126	99.999	5.1153	99	99.969	6.9992	100	99.914
SR	7.94	0.4372	1.6129	-5.5115	9.1127	99.999	5.1149	99	99.963	6.9986	100	99.884
SR	9.46	0.4372	1.6128	-5.5120	9.1118	99.999	5.1148	99	99.948	6.9980	99	99.860
SR	11.04	0.4371	1.6129	-5.5127	9.1125	99.998	5.1149	99	99.942	6.9985	99	99.810
RR	0.73	0.4378	1.6130	-5.5030	9.1103	99.999	5.1161	99	99.985	6.9999	100	99.953
RR	1.13	0.4379	1.6131	-5.5032	9.1111	99.999	5.1165	99	99.981	7.0000	100	99.934
RR	1.72	0.4380	1.6132	-5.5039	9.1107	99.999	5.1167	99	99.975	6.9991	100	99.886
RR	2.62	0.4379	1.6135	-5.5050	9.1115	99.998	5.1174	99	99.959	6.9993	99	99.827
RR	3.78	0.4377	1.6139	-5.5083	9.1133	99.997	5.1182	99	99.946	7.0003	99	99.751
RR	5.39	0.4377	1.6142	-5.5099	9.1140	99.996	5.1194	99	99.922	7.0009	99	99.618
RR	6.37	0.4377	1.6145	-5.5123	9.1163	99.995	5.1203	99	99.910	7.0019	99	99.578
RR	7.50	0.4375	1.6145	-5.5129	9.1153	99.993	5.1199	99	99.869	7.0007	99	99.503
RR	8.75	0.4374	1.6144	-5.5142	9.1152	99.992	5.1201	99	99.848	7.0003	99	99.361
RR	10.12	0.4372	1.6145	-5.5202	9.1154	99.990	5.1202	99	99.837	7.0003	99	99.262
LR1	0.37	0.4373	1.6121	-5.5107	9.1052	100.00	5.1140	100	99.992	6.9969	100	99.994
LR1	0.74	0.4376	1.6122	-5.5103	9.1069	100.00	5.1144	100	99.981	6.9974	100	99.986
LR1	1.23	0.4379	1.6120	-5.5089	9.1065	100.00	5.1146	99	99.972	6.9969	100	99.975
LR1	2.05	0.4385	1.6121	-5.5077	9.1067	99.999	5.1155	99	99.958	6.9980	100	99.938
LR1	3.28	0.4387	1.6119	-5.5053	9.1068	99.999	5.1160	99	99.934	6.9990	100	99.853
LR1	4.05	0.4389	1.6119	-5.5046	9.1088	99.998	5.1165	99	99.922	7.0005	100	99.819
LR1	5.07	0.4391	1.6119	-5.5044	9.1095	99.998	5.1169	99	99.914	7.0013	100	99.778
LR1	6.49	0.4394	1.6119	-5.5017	9.1096	99.997	5.1172	99	99.897	7.0016	99	99.705
LR1	8.61	0.4396	1.6122	-5.5011	9.1124	99.996	5.1182	99	99.885	7.0034	99	99.677
LR1	11.40	0.4396	1.6124	-5.5062	9.1156	99.996	5.1191	99	99.867	7.0055	99	99.650
LR2	0.13	0.4371	1.6122	-5.5112	9.1054	100.00	5.1138	100	100.00	6.9973	100	99.999
LR2	1.64	0.4370	1.6120	-5.5115	9.1043	100.00	5.1133	100	99.995	6.9967	100	99.986
LR2	3.33	0.4367	1.6116	-5.5108	9.1013	100.00	5.1115	100	99.984	6.9949	99	99.973
LR2	4.04	0.4365	1.6115	-5.5102	9.1016	100.00	5.1108	100	99.968	6.9955	100	99.965
LR2	4.50	0.4365	1.6113	-5.5096	9.1011	99.999	5.1099	100	99.959	6.9955	100	99.939
LR2	5.00	0.4362	1.6109	-5.5105	9.1007	99.999	5.1084	100	99.940	6.9951	100	99.871
LR2	5.57	0.4361	1.6107	-5.5092	9.0991	99.998	5.1071	100	99.935	6.9947	100	99.822
LR2	6.51	0.4358	1.6101	-5.5077	9.0959	99.997	5.1051	100	99.915	6.9924	100	99.791
LR2	8.37	0.4357	1.6097	-5.5069	9.0937	99.997	5.1037	100	99.906	6.9909	99	99.770
LR2	11.13	0.4357	1.6096	-5.5046	9.0932	99.996	5.1033	100	99.898	6.9903	99	99.753

1- In this table, MEBV stands for mean, SDEBV for standard deviation, MINEBV for minimum and MAXEBV for maximum of estimated breeding values, CORR stands for correlation between cases and regular BLUP for all animals (=58315). Also, MEBVS and MEBVC stand for average of estimated breeding values, NTSIC and NTCIC stand for top 100 lists in common; and SRCORRS and SRCORRC stand for spearman rank correlation for bulls and cows, respectively for adjusting scheme.

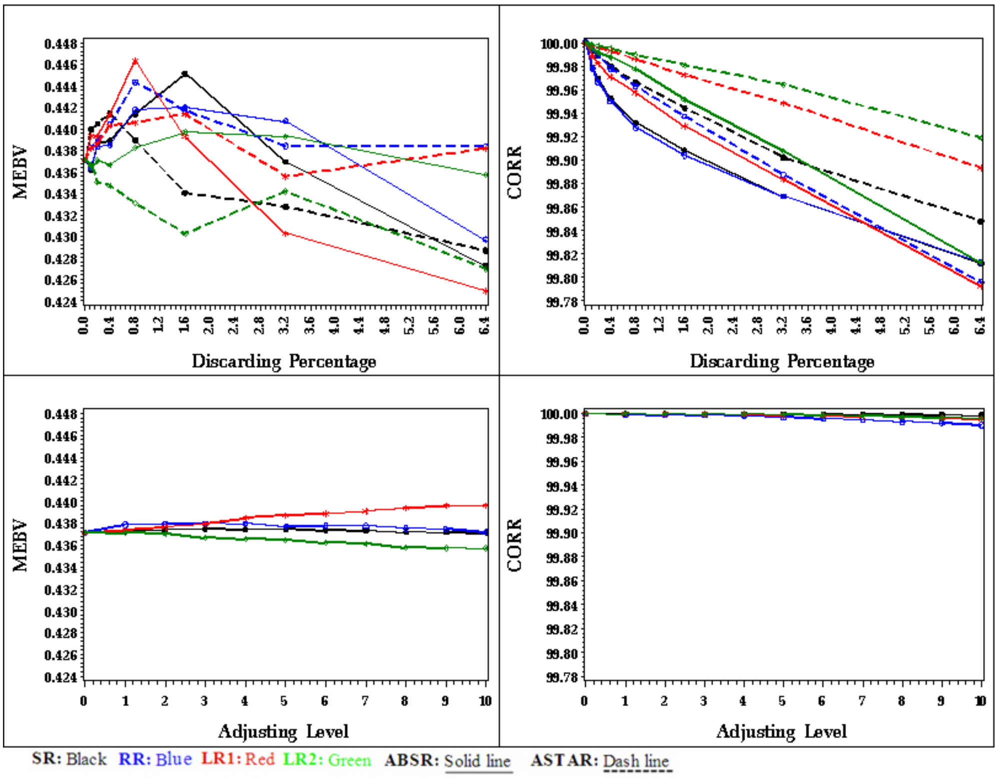

Figure 4.14: Mean of estimated breeding values (MEBV) and correlation of EBVs of animals between full data and those from both discarding and adjusting schemes by various percentages and levels.

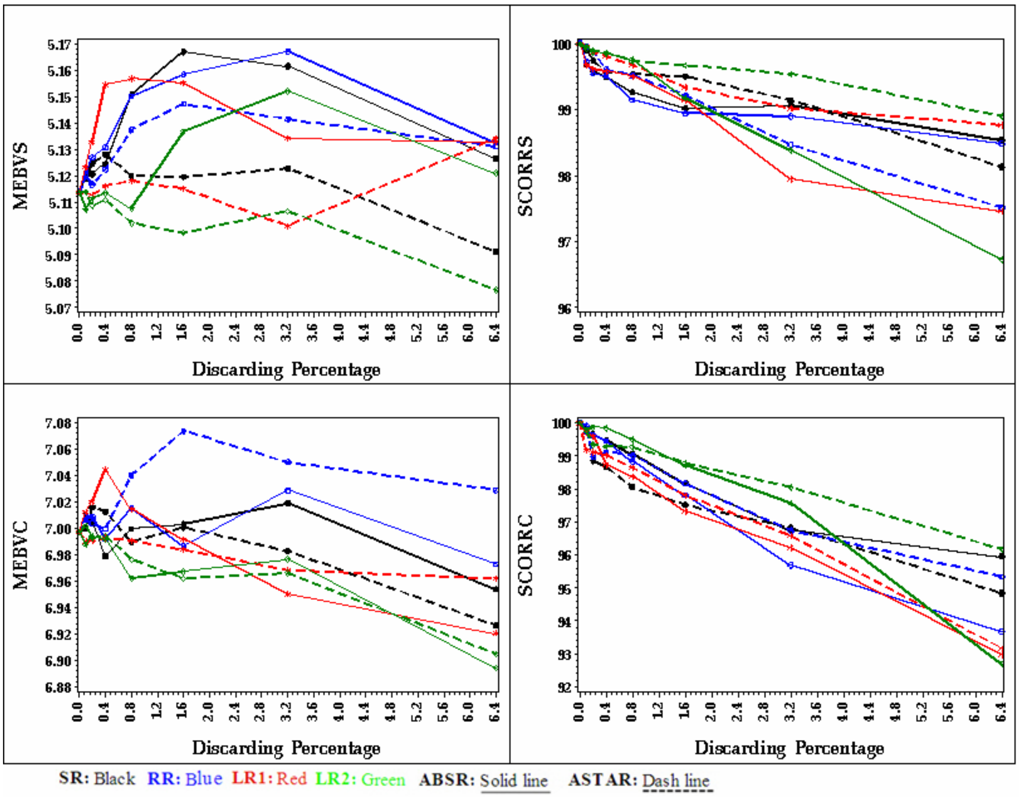

Figure 4.15: Mean of estimated breeding values of the top 100 bulls (MEBVS) and cows (MEBVC) and their Spearman rank correlations (SCORRS and SCORRC) for both discarding schemes by various percentages.

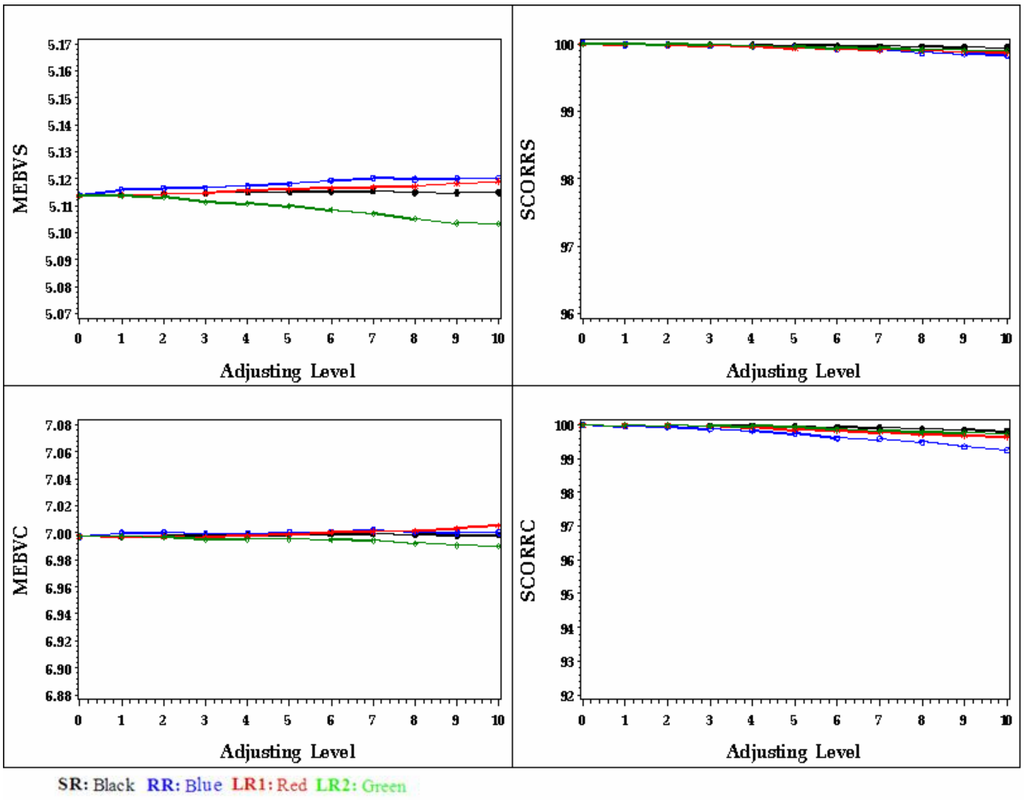

Figure 4.16: Mean of estimated breeding values of the top 100 bulls (MEBVS) and cows (MEBVC) and their Spearman rank correlations (SCORRS and SCORRC) for the adjusting scheme by various levels.

5. *Discussion*

Milk recording data are important management and decision tools for dairy farmers and for genetic evaluation (Bertrand 1996). Consequently, the accuracy of milk recordings will be very important for milk producers and breeding organizations. According to Cardoso et al. (2006), the accuracy of EBVs depends on the quality and validity of the phenotypic and pedigree data. Performance or phenotypic records in the form of TD milk yields are influenced by factors that are not clearly defined in the statistical model, thus leading to large residuals. These large residuals are attributed to outlying observations (or outliers) which can seriously bias genetic parameter estimates and genetic evaluations (Stranden and Gianola 1998). The validity of a TD evaluation relies on the accuracy of (co)variance components that are used (Strable and Mistzal 1999) and the accuracy of a genetic evaluation depends on how well the assumptions of the model match the data (Wiggans and VanRaden 1991).

It is inferred from the above text that there is a need to detect and treat these outlying data for accurate estimation of model parameters and breeding values. Generally, there are two alternative approaches including discarding and adjusting, in order to deal with outliers in TD data. These alternatives can be applied either before or parallel to estimate the variance components and breeding values using mixed linear regression analysis.

Therefore, the general goal of this study was to explore a simple approach to detect abnormal recordings of milk yield. A further purpose was to compare the influence of treating (discarding and adjusting) abnormal records by the use of three regression methods. The final purpose was to explore simple and effective residual analysis and the role that these residuals play in various aspects of genetic evaluation diagnostics related to the used test-day model.

To achieve the above goals, it was required to predict the response variable (TD records) based on the predictor variables (a linear function of days in milk) for individual lactations using a regression procedure. However, different regression methods exhibit different properties e.g. residuals and their standard errors, estimates of the model parameters and their standard errors, and goodness of their fit. In contrast to least-squares regression that here has been called simple regression (SR), robust regressions (RR) provide protection against outliers and high leverage points. Local smoothing regression (LR) combines the simplicity of linear least squares regression with the flexibility of nonlinear regression. An advantage of this method is that no specific function is needed to fit the data to the model (NIST 2003).

In the next step, the limits that specify whether or not a TD value is considered as an outlier must be defined separately for both discarding and adjusting approaches, before the outlying recordings can be identified. In the current study, it was done by defining outliers as values deviating significantly from other records belonging to a given lactation. Finally, all three regression methods SR, RR and LR (with first and second polynomial order) were applied to each of the two treating schemes (discarding and adjusting).

In order to select the required data to be removed, different percentages of the absolute predicted ordinary residual (ABSR) and standardized residual (ASTAR) were examined. Adjusting the different proportions of outlying data through restricting them to the residual limits that are defined by standard deviation (SD) of individual TD records, was another approach to the simple removal of outliers. To adjust an observation, the value of the residual must be large enough to exceed the product of k (constant) with the SD. If k is large, then the largest residuals will be less likely to exceed this product and they will be less likely to be corrected. On the other hand, if k is small the product of k with SD will be smaller and the observation with the greatest residual will be more likely to be corrected.

All three regression methods and two treating schemes were applied to the original data set, prior to a fixed regression test-day model analysis. Finally the residual criterions, genetic and non-genetic variances and parameters, their standard deviations, average of estimated breeding values and ranking of top bulls and cows were inspected.

5.1. Tests for determining the existence of outliers

There are a variety of statistical tests for assessing the existence of outliers in the data depending on the data structure (univariate or multivariate) and the nature of the test which each lead to different set of outliers. A good example in this regard is the study done by Strandén and Gianola (1997). They used three types of linear models including a) a bivariate model, b) a univariate model (repeatability) with homogeneous residual variance, and, c) a heteroscedastic repeatability model. Within each of these models, the residual distributions were either Gaussian (normal) or t-student at the statistical level of less than 1%. Thus, a total of six statistical models was studied. Their results showed that the normal method detects more outliers than the t-probability approach, because of the difficulty for estimating probabilities precisely in the normal procedure. Thus, they found a clear superiority for t-models and considered this method as the more accurate method. For example, within univariate heteroscedastic models, the Gaussian assumption gave 25 outliers whereas the t-distribution produced only 10 (a reduction of 60% in the number of outliers). The bivariate Gaussian models detected 39 outliers (about 2.3% of all records) over 21 (about 1.4%) for t-models.

In the current research, the examination of data for possible outliers using both standardized (STAR) and studentized residuals (STUDR) attested the existence of potential outliers within the data. However, the proportion (percentage) of detected outliers depended on the type of the statistics used and significance level. Using STUDR, a minimum percentage of 0.64% occurred at $p<0.01$ after Bonferroni correction (which identifies fewer outliers than usual case). PRESS statistics discovered approximately half of all TD records to be outliers that can be considered as its drawback. Probably, this statistic can be efficient when data includes weekly instead of monthly milk recordings. In this case, there are many observations per lactation and consequently, there are several TD records prior to peak yield. However, regardless of the amount of the identified outliers by above tests, a high proportion of them are normal data.

5.2. Checking the fit of the models

Although evaluating and comparing goodness of fit among regression methods that has been accomplished by estimating of the correlation between predicted and actual values, was not the primary aim of this study, the outcomes indicate a significant difference in their fit at the end of lactation. These differences influence the efficiency of regression methods in detection and treatment of outlier leading to the different sets of outliers that treating them will affect both the variance components and breeding value estimates differently. The results from the current study confirm that both SR and RR procedures fitted relatively well to the data but local regressions, particularly those with the higher smoothing parameter, suffered from an inaccurate fit at later months of the lactation course. It seems that local regressions yielding a smooth but biased fit in the late lactation, lead to inaccurate outlier detection and treatment conclusions.

Although a common assumption of genetic evaluation models is homogeneity of (co)variances, this assumption is often incorrect across time or herds (Boldman and Freeman 1990). In TD models an additional reason for unequal variances is linked to lactation stage (Gengler and Wiggans 2001) and production level (De Veer and Van Vleck 1987).

The findings from the current research underline the importance of the dependency between the magnitude of TD records and both lactation stages and production levels that should be taken into account before any attempt for dealing with outliers. Factors influencing the variation in milk production in the beginning of the lactation are likely cow factors such as milk let down in the first few days, stress due to parlour introduction, calf removal, change of diet, dietary upsets and calving problems (Quinn et al. 2007). Other contributing cow factors cause the inability of the cow, immediately acalving, to consume sufficient energy to sustain lactation (Buckley et al. 2003, McGuire et al. 2004). An increase in the phenotypic variance at the end of lactation was also observed. Factors such as stage of lactation, production potential, pregnancy status, and changes in nutrition are likely to be the primary factors, contributing to the variation at this time (Quinn et al. 2007).

The current study showed that there is an intermediate correlation between TD values and both ordinary and standardized residuals that can cause problems in the process of correct detection of outliers. Other types of residuals can be also used as diagnostics such as the Cook' distance and studentized residuals.

5.3. The distribution frequency of outliers

The results of genetic evaluations are also influenced not only by how well the model predicts the TD yields for each individual lactation but also by the frequency of outliers detected by each regression method within lactation stages and production levels. Based on Mayeres et al. (2003), the distribution of outliers should be theoretically independent of the production level and lactation stage. Consequently, prediction residuals and standardized residuals are supposed to be homogeneous within lactation period and production level.

From the results of the present research, it is now known that distributions of outliers by DIM (all four regressions combined) were not generally uniform in the interval from 5 to 330 DIM. Distributions become more uniform (smaller differences between DIM 5-30 and the remaining part of the lactation) with increasing values of k (results have not been shown) in agreement with Jamrozik et al. (2007). Moreover, the beginning and the end of lactation were characterized by a lower frequency of detected outliers which is in contrast to what has been reported by Wiggans et al. (2003) and Quinn et al. (2007) who used similar approaches and definitions (<60% or >150% of predicted TD yield) to detect outliers. In the study of Quinn et al. (2007), 51% of the abnormal recordings were in the first lactation week where milk production is affected adversely by several environmental factors. However, this high percentage may be somewhat unrealistic because outliers are extreme observations which occur fairly randomly across lactation. Jamrozik et al. (2007) reported around 39% of outliers on DIM=10 at $k = 1.5$ for first lactation cows.

In the present study, the high phenotypic variance at early and late lactation stages generally leads to a lower probability for a potential outlier to pass the defined range and become an actual outlier. This problem can be solved by providing weights for the residuals or for the product of k and SD.

In contrast to RR and LR procedures, SR showed a nearly constant distribution for OTD records throughout lactation with the highest values at peak lactation. This period of lactation could be associated with erratic or problematic values of milk recording. Mayeres et al. (2003) found a higher number of deviant TD yields at the peak yield and attributed this to the larger phenotypic variances which is in contrast to the current study.

Mayeres et al. (2003) also found very few outliers for low milk production, contrary to high levels. They attributed this to the weakness of their method in the detection of outliers for the lower productions of the beginning and the end of the lactation period (Mayeres et al. 2003). They used a fixed SD (=1.83 kg) for product SD by k, in order to define the limits of the prediction error. Thus, an animal having a TD production of 15 kg has less chance to be out of the 15±5×1.83 interval than another one with the 35±5×1.83 interval. Since the phenotypic variance varies across lactation, a constant SD may cause inconvenience in detecting outlying records, consequently some data will incorrectly be defined as outliers or some outliers may not be detected. Although in the current study the percentage of low and high outliers is still a little greater than what is expected (especially for SR) but it is acceptable.

In addition to the need for determining the non-existence of the dependency between production level or lactation stage and the portion of the identified outliers, particular tests also need to be conducted if there are any other given patterns of heterogeneous variances for various subclasses of the data (Cardoso et al. 2006). In this regard, animal breeders are encouraged to employ extra precautions so that no accidental selection of data or the occurrence of bias becomes possible (Interbull 2001).

The statistical tests (Tukey's Studentized Range Test) show that the variances were heterogeneous for different production levels and lactation stages. This means that with no correction for variance heterogeneity in the model, the homogeneity of the variance will be absent. It can also cause some problems in identification of the outlying data like removing animals with extreme high or low production level (Mayeres et al. 2003) and consequently leads to bias in the selection.

For this purpose, robust heteroskedastic error models proposed by Cardoso et al. (2005) for genetic evaluations, are a desirable alternative where there is heterogeneity of the variance among data subclasses. These models use features of structural variance models (Foulley et al. 1992, Kizilkaya and Tempelman 2005) and heavy-tailed density models such as t-student (Lange and Sinsheimer 1993, Rosa et al. 2003), representing an alternative to the deletion of records (Cardoso et al. 2006).

5.4. Counting outliers per lactations

In order to check and control the impact of each regression procedure and treatment method on genetic evaluation results, inspection of the number of detected outliers corresponding to the different sample subclasses is of interest. It is very critical that there are no systematic patterns in the number of outliers in the different subclasses of the data. In addition to the lactation stage and production level, this study has analyzed the distribution of outliers within lactations. Another important issue that needs to be examined is the frequency of low (and high) outliers in each lactation phase and during lactation that influences not only the average of milk production (MPA) but also the results of breeding value estimates (EBV). Because of a large number of applied analyses, it was not possible to determine the frequency of the low and high outliers for each lactation subclasses (week or month) separately.

The results of the incidence frequency of the detected outliers for the current study show that they were less at the beginning and end of the lactation, and were more in the middle of the lactation period with an exception for the local regressions. According to Wiggans et al. (2003), a cow is less likely to have a large increase in yield than a large decrease. Their preliminary investigation showed that a lower limit of 60% of predicted TD yield as specified by ICAR (1995) and an upper limit of 150% are reasonable lower and higher limits. Wiggans et al. (2003) reported about 95% of abnormal milk records (outliers) to be low (1.8% out of total data). Furthermore, in the study done by Quinn et al. (2007), 96% of the abnormal recordings from first lactation cows were declared abnormally low. Both studies which used a similar definition for detecting outliers, found a higher proportion for lower outliers in very early lactation and an intermediate proportion in late lactation.

A probable cause for the observed difference in the percentage of low outliers between the current study and the two above mentioned studies is that they utilized different functions to estimate the slope of the lactation curve for first and subsequent milk yield using adjacent TD yields. In more precise terms, a separate set of regression coefficients was estimated based on subsequent rather than preceding TD yield to predict the first TD yield. This is achieved by comparison between the results of the current study and those of two above studies figuring out that the occurrence of low and high outliers is associated with the model fitted to the data, definition of restrictions and parameters used (predicted values, residuals and standard errors) in these calculations. Therefore, the high frequency of lower outliers in the last two studies may be a consequence of a problem in one or both of the indicated cases.

The relatively higher incidence of low recordings which were obtained by the current survey and the two previous studies, may be due to difference of the adopted procedures in the definition of outliers. In the current study, as the percentage of outliers (discarded or adjusted) data increased, the quotient of abnormally low and high records approached to 50:50. However, for both SR and RR methods the percentage of low outliers was more than 50 and it was lower than 50 for both local regressions.

The percentage of the lactations containing a single outlier decreased along with the increase in the proportion of outliers, as a logical consequence. The higher values for this percentage reflects the fact that the used methods accurately identified OTD records from a given lactation with less regard to the measure of the other TD yields from that given lactation. In addition, the relatively uniform distribution of outliers per lactation shows that there was no evident pattern for the detection of outliers in the level of lactation. In other words, there is a less chance to detect several outliers per a lactation, if only a small amount of data is selected as outliers. Furthermore, the occurrence of low outliers changed along with the increase in the frequency of detected outliers. For example, for SR and RR, it varied from 70% to 50%. From the application point of view, this variation may suggest the need for changing the limitations <60% (for lower yields) and >150% (for upper yields) of predicted values to detect abnormal TD records.

5.5. Impact on residual criteria

One of the most important and primary goals of treating outliers is reducing the measures of error terms. There are many criteria, such as residuals and squares of residuals that can be used to assess the relative size of predicted residuals from the model. The mean error and standard deviation of errors are mathematical measures of the magnitude and dispersion of the errors in estimating actual yields, respectively. In this study, the average of absolute residuals (MAR) and squared (MSR) were used because data sets were different in number of observations after discarding the detected outliers. Establishing such a standardization makes these criteria appropriate for comparing results from both discarding and adjusting schemes.

Yang et al. (2004) applied two types of robust estimation procedures to a single trait random regression TDM to reduce the effects of OTD. The first robust procedure (k method) estimated the residuals from the genetic evaluation model, computed the standard deviation of residuals across all records, and restricted the outlier residuals to be within k standard deviations. Consequently, a new observation was created (similar to the current study) for use in the genetic evaluation model. Four different values of k were examined. The second robust procedure (W method) utilized different weights with each TD record which were estimated from the residuals for all observations. Outliers tended to receive smaller weights, and thus their influence tended to be reduced. Results were compared to usual best linear unbiased prediction (BLUP) ignoring the outlier problem. Their findings showed that the BLUP method resulted in the smallest sum of squared residuals (SSR) and the smallest standard deviation of residuals (SDR) and consequently showed a better performance. For the k robust methods, a k=2.0 (4% of the data) gave the smallest sum of absolute residuals (SAR) that was only 0.85% smaller than BLUP method, but at same time it yields a SDR that was about 4.5% greater than SDR from the usual BLUP. The W robust methods where each observation is weighted according to its residual, performed poorly with much higher values for SSR, SAR, and SDR than BLUP method.

Jamrozik et al. (2007) applied a robust method in the format of a multiple trait random regression model for the correction of outliers. Their method was similar to the k robust method described by Yang et al. (2004) except that the residuals were calculated for all observations within each DIM and the standard deviation of residuals was computed for each DIM rather than only four DIM groups, as used by Yang et al. (2004). They also reported that higher values of k reduced SSR for robust methods but none of them reached the level of BLUP while values of k = 2.75 and 3.0 (for protein yield) gave the smallest SAR compared with BLUP (only about 0.41% and 0.38% smaller than usual BLUP).

In contrast to the above studies, in the current study along with an increase in the percentage of treated outliers, the corresponding MAR, MSR and SDR became smaller in comparison with those from the full data. The discarding scheme with k=0.4 or =0.8 gave small enough values for MAR, MSR and SDR and removed a large portion of the inconsistency in the predicted residuals from the model. For the adjusting case, the best k was dependent on the regression method. The RR method presented the largest reduction in residual terms among all regression methods. For example, at k=2.0 which includes only 1.13% of the data corrected by RR, MAR was 0.33% smaller than that for data with no treatment.

Milk production is known as the most important source of income for many farms and for successful dairy production management. Therefore, the impact of any treatment of outliers should also include a milk production assessment which here was evaluated by average of milk production (MPA). The results clearly indicated that the local regressions with both linear and quadratic orders were different in the pattern of variations for MPA from usual parametric regressions (i.e. SR and RR). The variation in MPA is more likely attributable to the combination of factors including ratio of low to high outliers, accuracy of the fit of the various regressions to the data over the lactation period, the regression technique and the used model.

5.6. Impact on variance components and standard deviations

Estimation of genetic parameters is necessary for prediction of breeding values, prediction of response to selection and determination of selection procedure and also is essential before the development of breeding strategies. Estimating additive genetic and non-additive genetic variances contributes to a better understanding of the genetic structure of a population.

The results from the current research indicated that the error variance (V_e), phenotypic variance (V_p) and error variance ratio (e^2) considerably decreased through treatment of the abnormal TD records by both discarding and adjusting schemes. The general pattern of variations observed for additive genetic (V_a) and permanent environmental (V_{pe}) variances remained relatively constant using adjusting scheme along with the increasing outlier percentage but they showed a random pattern using discarding scheme.

It is proposed that smaller standard errors show the superiority of a model or method because fewer unexpected and more reliable predictions are produced. Naturally, the usefulness of a model would depend mainly on prediction accuracy and also, the accuracy of estimating heritability is dependent on its standard error. Therefore, a high standard error of heritability also shows the low accuracy of the estimates (Khanitta et al. 2005, Basualdo et al. 2007). In the current study it was found that despite a small increase in h^2 (heritability) and c^2 (permanent environmental variance ratio), their standard errors slightly decreased. These results indicate the usefulness and practicality of using the defined methods in detection and treatment the outlying datas in a proper way. Although, there are no other studies to support these results, many studies (Brotherstone and Hill 1986, Boldman and Freeman 1990, Dong and Mao 1990) have found that genetic variance is not constant and increases with herd yield and herd phenotypic variance. Meinert et al. (1988) found that herd mean was only moderately correlated (≤ 0.5) with phenotypic variance.

Brügemann (2008), in a detailed study, estimated the heritability of one hundred of the largest dairy herds in the eastern Germany which also included the data records from those six dairy herds that were used for the present study. The average of the estimated heritability using a fixed regression test-day model for the six herds was considerably lower than that for the current study (0.206 versus 0.2987) which included all six herds as a whole dataset. The corresponding average for V_{pe}, V_e and V_a were 15.060, 9.443 and 6.332, respectively which indicate higher values of V_{pe} and V_e and smaller value of V_a compared to those for the present study (7.399, 9.402 and 7.154, respectively). In the current research, however, the data were subject to considerably more editing than previous study which included various milking frequencies (2x, 3x and robotic milking) and without the limitation of nine TD records per lactation. However, the existence of these sources of the heterogeneous variances can lead to decrease in the heritability coefficient. On the other hand, the differences may just be by chance.

5.7. Impact on breeding values and ranking of top animals

The genetic evaluation has a key role in selection decisions and breeding improvement programs. Nowadays, the estimation of breeding values and the estimation of the variance components are implemented simultaneously with linear mixed models (Krejčová et al. 2007). Therefore, the analysis of impacts of the outlier treatment on them seems logical and necessary because these impacts affect the results of the variance component estimation and finally the results of genetic evaluations.

Jamrozik et al. (2007) estimated the average, standard deviation, minimum and maximum values of EBV for protein yield for BLUP and those data which were corrected for the outlying yields. In their study, the average estimated breeding values (EBV) were practically constant for all methods with a small increase from -0.026 to -0.023 and small changes in standard deviations and ranges of EBV did not show any apparent pattern. Similar findings were observed in the present study where an increase in the percentage of outliers did not significantly affected the EBV of concerned animals (fluctuated in a narrow range), especially for the adjusting scheme. It is expected from the results that upward and downward changes in the MEBV are dependent on the frequency of low and high outliers that are likely belong to animals with smaller and larger EBVs, respectively.

An important component of each breeding program is the identification of animals with the highest genetic merit with a high accuracy, because these animals should be used as parents for the next generation. Treating a higher number of data was always associated with the lower correlation between EBV of animals from entire data and those produced by methods from the current research. Generally, correlations (with minimum of 99.9) in the present study were higher than that (range of 96.1 to 99.7) reported by Jamrozik et al. (2007) with nearly the same proportion of corrected milk yield TD records.

For the present study, EBV average of a top 100 bulls (MEBVS) and cows (MEBVC) remained relatively constant, though they slightly increased in some cases. Furthermore, the variation of EBV for cows was less than for bulls. In the study of Yang et al. (2004), the median EBVs of the top 100 bulls and cows using the method k robust estimation was greater than usual analysis with BLUP.

From the results of the current study it was concluded that the Spearman rank correlations of the top 100 bulls and cows (SCORRS and SCORRC) decreased and some bulls and cows changed their position on the list of superior animals that could be attributed to the number of outlier records for these animals in the current research. However, the variation of SCORRC was remarkably larger than that of SCORRS, most likely due to fewer records per cows.

Using k robust method, Yang et al. (2004) reported a range of 99.2 to 99.3 and 91.4 to 94.3 for spearman rank correlation of bulls and cows, respectively. In the current study, the number of top bulls and cows in common in the top 100 lists between full data and data sets from different analyses was higher than those from robust k method used by the study of Jamrozik et al. (2007) who found a range of 92 to 97 for bulls and a range of 84 to 98 for cows. Using k method, Yang et al. (2004) reported a range of 99 to 100 for bulls and 90 to 91 for cows.

In general, the results of the current research are in a good agreement with the results of Yang et al. (2004) for the single trait model and of Jamrozik et al. (2007) for the multiple trait model. However, their methods require somewhat more computational cost and/or programming effort to their implementation. For these reasons, the strategy described by the current study may be superior in the treatment of outlying data to methods presented by above studies with regard to the ranking of top animals. Therefore, due to simplicity of the procedure used here, it may be comparable with the method used by previous studies.

5.8. Comparison between treating schemes

Outlier data can be excluded or otherwise replaced (substituted) with corrected values in genetic evaluations. In traditional statistical analysis, mean and standard deviation are estimated without taking outliers into account through removing them (Chernobai and Rachev 2006). Removing outliers is drastic, consequently without detailed analysis of the data it is difficult to decide whether they should be excluded or not. The current research describes the methods which initially were used to identify the outliers with residual diagnostic techniques, and then to determine the impact of removing and adjusting outliers on various aspects of genetic evaluations.

As explained in the previous sections, two treatment schemes discarding and adjusting were applied to the data. For the discarding scheme two types of residual definitions, i.e. ordinary (ABSR) and standardized (ASTAR) residuals in absolute value were used in order to select the outlying observations. These two kinds of residuals represented significant difference in their performance not only on outlier detection but also on genetic evaluations. For example, the incidence pattern of the detected outliers using ASTAR during lactation was more uniform than ABSR. The distribution of outliers corresponding to different production levels and lactation stages were more consistent using ASTAR. Moreover, ASTAR identified a higher proportion of lactations with only one TD as outlier. Meanwhile, rankings of top bulls and cows severely altered by the use of ABSR in comparison to ASTAR. Therefore, using ASTAR of records instead of ABSR leads to adjustment of most sources of variance that are not specified in the model and directly or indirectly affect a given TD such as herd, DIM, cow's age and production level. The only superiority of ABSR over ASTAR was its smaller values for three residual criteria.

Thus, analysis of residuals is useful because inadequate models will yield structured residuals (Box and Jenkins 1976). Standardized residuals represent the distance of an individual observation from the sub-sample (herein "lactation") and thus they are effective in assisting analysts in examining data for multivariate outliers (Osborne and Overbay 2004).

However, it can be concluded from the results that the standardized residuals are a good substitute for fitted residuals because they are independent of the units of measurement of the variables and provide a unitless statistical score for judging the size of the residual. They are also more robust and conservative than ordinary residuals for outlier detection applications. Perhaps because of these properties, ICAR agrees with modification the assessment of the bias related to the meter due to measuring errors using standardized residuals to find out if outlier data exist. The bias of the meter is tested both with all results and after removing outlier data and if one of the tests is failed, then the meter is rejected (ICAR 2007).

Based on the results of the current research, correction of the outliers generally less affected the residual criteria, genetic and non-genetic variances and parameters and their standard deviations as well as breeding values in comparison with the discarding scheme. For instance, the number of detected outliers was less dependent on DIM and production level. By adjustment of outliers, a maximum of four OTD per lactation was corrected by the adjusting scheme. It also led to smaller residual terms of MAR, MSR and SDR as well as both V_p and V_e. The MEBV resulted from adjusting schemes showed small differences with the data with no treatment. The CORR between corrected data sets and full data remained almost stable. Rankings of top bulls and cows were not altered by the adjustment of outliers. Though the correction of outliers less influenced the different criteria that have been examined, it may be a much safer and secure solution than discarding them, because it does not remove the potential information associated with the TD records which may be informative for other aspects of the survey. Based on Mayeres et al. (2003), the factors causing an outlier may be different consequently the safest treating is to reduce them towards their expected value predicted by a convenient model.

Taken together, the above points imply that it is not very easy to deal with outliers and making decision about them. Despite these difficulties, some researches attempted to handle outliers in milk production data analysis. Some of them only focused on detection of outliers (Strandén and Gianola 1997, Jamrozik et al. 2004) and the others substituted these abnormalities with values predicted by model prior to (Wiggans et al. 2003) or parallel to genetic evaluation (Mayeres et al. 2003, Yang et al. 2004). It seems that few researches have been undertaken on how to deal with outliers in practice and especially analyzing their impacts on genetic evaluation. In addition, there was no study to compare the impact of discarding and adjusting outliers on residual statistics, variances and genetic evaluation results. Therefore, there is still a need for more research investigating the use of new methods to assess the effect of treatment of outlying data on genetic evaluation. According to Yang et al. (2004) the better method will be the one that gives the higher correlation of EBVs with the genetic values from the data without treating the extreme points, or the greater probability of correctly ranking animals.

According to the obtained results it can be concluded that both discarding and adjusting approaches have been shown to be effective in alleviating obstructive effects of outliers in TD data, irrespective of their advantages and disadvantages. The results are based on their impacts on residual terms, genetic and non-genetic variances and both h^2 and c^2 estimates as well as their standard deviations. However, using both discarding and adjusting schemes together to treat the TD data may give the better performance than each of them individually. This would potentially make it a more attractive data cleaning procedure and may encourage a greater uptake of editing TD data without losing or manipulating a high number of records.

5.9. Comparison between regression methods

Various regression methods have different analytical and statistical characteristics and each of them predicts different values for response variable, residuals and standard deviations. Results of the current study have shown somewhat these differences. The obtained outcomes indicated the superiority of the robust regression technique over simple and local regressions in detection and treatment of outlying TD milk yields. In the following, some differences between these regression methods will be pointed out.

Very extreme TD records were less detected by RR method in comparison to SR and LR that is a consequence of its independency on the level of production. Contrary to RR and LR, SR represented the lowest percentage of single outlier per lactation because of its dependency on the production levels. This reflects the problem of attributing more than one outlying TD to a given lactation of a high- or low-producing dairy cow. However, it is essential to point out that superiority of local regressions over SR is not necessarily true, because most outlier points that have been detected by these regressions, corresponded to the end of lactation. The placing of the most outliers in the end lactation, incorrectly causes a decrease in the probability of detection of two or more outlying records in a given lactation. Among all tested methods, SR method showed little effect on the ranking of EBVs for the top animals. While LR2 represented the highest correlation between observed and predicted breeding values, it showed a high inequality in distribution of outliers during lactation. Thus, the usefulness and efficiency of a regression procedure for outlier detection is a function of its fit to the data and the incidence frequency of outliers at different lactation stages.

Among all methods, RR was the regression procedure that provided the largest differences between the usual analysis and other procedures in terms of residual criteria, variances, parameters and their standard deviations while it affected only a small portion of data. These performance advantages make it a more accurate and convenient regression procedure than all other methods. Therefore, RR methods can attenuate the adverse effects of outliers on residual terms and model parameters. Despite these good features, however, robust regressions are still not broadly used. Several reasons have been explained for their unpopularity (see Hampel et al. 1986). The M-estimation which has been already briefly described, uses weight functions of the residuals that increase less rapidly than the square of the residuals used by OLS. The weighting function attenuates the influence of observations with large residuals (O'Kelly 2006). It became obvious from results (see Tables 4.11 to 4.14) that local regression with lower polynomial degree (i.e. LR1) led to a smaller value for the corresponding residual criteria. Thus, nonparametric local regression as a new tool can be applicable in dealing techniques with outliers because it does not need a predefined model to be fitted to the data. In this context, Quist et al. (2007) estimated the missing milking records from the daily milk data set by using local regression techniques via PROC LOESS in SAS (Statistical Analysis Software).

Generally, the local regression estimators produced the residual terms with less difference from those for the original analysis. There are two main reasons for this phenomenon. The first one is their unsuitable fitting at the end of lactation, and the second one is employing different regression techniques and functions for the detection of outliers and for genetic evaluations. However, the earlier problem seems to be more problematic. The examinations clearly show that the performance of this kind of regressions is greatly influenced by order of polynomial regression and value of smoothing parameter. Choosing a higher smoothing parameter with this procedure will provide a smoother curve which may hide the main features of the data and the fit will have large bias. If it is too small, insufficient data fall within the smoothing window and consequently a large variance will result (Loader 1999).

Like the smoothing parameter, the degree of the local polynomial affects the bias and variance. A local polynomial with higher degree provides a better prediction than a low polynomial degree, and consequently leads to an estimate with less bias but a lower degree usually produces better fits with less variance, especially at the edges (Cleveland and Loader 1995, Loader 1999). The fitted value at a data point for LOESS procedure is obtained by blending them at the initial cell through partitioning the initial cell into sub-cells by median (SAS 1999). However, like other least squares methods, the local smoothing local regression which was applied to the LOESS procedure of SAS, is also sensitive to the effects of outliers in the data set (NIST 2003). The two of the biggest advantages of local regressions over parametric methods are: a) the fact that it does not need to specify a function prior to fitting the model to the data and b) its flexibility. These two advantages make it one of the modern regression methods for applications that use least squares regressions (NIST 2003).

The known drawbacks of LOESS are: a) it requires fairly large and dense data sets in order to produce good models, b) it does not produce a regression function that is easily represented by a mathematical formula, and c) it is a computational intensive method. However, the last drawback is not usually a real problem in current computing environment, unless the data sets being used are very large (NIST 2003).

When the length of lactation is shorter, probably local regressions will show different performance and results compared with the longer lactations. However, owing to a limited number of TD milk records per lactation on each individual lactation, LOESS procedure was unable to meet iterative reweighting least squares method which improves the robustness of the fit in the presence of outliers in the data (Cohen 1999). In order to face such situations where there are possible outliers in the data, robust smoothing methods have been developed to restrict the influence of outliers. LOWESS (local weighted scatterplot smoother) which is sometimes called a robust version of LOESS, could be used (Takezawa 2006). Finally, using thin-plate smoothing splines by PROC TSPLINE is another alternative (Freund and Littell 2000).

Class or group of data that should be searched for outliers, also could be important. Some researches have focused on the detection of outliers within larger DIM levels (Yang et al. 2004) or within smaller DIM levels (Jamrozik et al. 2007) rather than within lactation (Wiggans et al. 2003). Difference in the selected level for searching for outliers, results in considerable differences in conclusions. For example, Yang et al. (2004) and Jamrozik et al. (2007) employed a similar robust estimation (k method) in order to correction of outliers using a random regression TDM. Jamrozik et al. (2007) calculated residuals and standard deviation of residuals for all observations within each of the DIM classes from 5 to 305 (301 classes), separately rather than within only four DIM classes used for the residual variances in the study of Yang et al. (2004). Jamrozik et al. (2007) reported lower values for SAR at k = 2.75 and 3.0 (for protein yield) than BLUP compared with Yang et al. (2004). Generally, larger groups and levels provide higher variance and standard deviation for residuals that decreased the probability of any data point correctly lying outside the particular deviation boundaries and becoming an actual outlier due to heterogeneity among observation within those groups.

What seems certain is that dealing with outliers needs a model to catch them and results depend on the ingredients of the model fitted on the data, regression procedure and limits defined. Similarity of the model applied to both detection and treatment of outlying observations will be critical to obtain accurate results.

In the current research, the model to estimate the variance components and breeding values was a linear mixed model which included both fixed and random effects as well as co-variables. But the model to detect outliers only included the independent co-variables of Ali and Schaeffer's function (1987). However, this dissimilarity may lead to unnecessary removal of data or identifying some records as outliers which may not really be. In this context, Cox and Snell (1968) presented a general definition of residuals for the models with a single source of variability (i.e. fixed effects). Authors like Hilden-Minton (1995) extended such ideas to define three types of residuals (ordinary, standardized and studentized residuals) where the other source of variability presents in linear mixed models (Nobre and Singer 2007). While Haslett and Hayes (1998), and Haslett (1999) considered residuals for the fixed effect linear model with correlated covariance structure, their approaches cannot be directly applied to a mixed model.

The definition of residuals in the fixed effects model is $y - X\beta$, whereas in the mixed model is $y - X\beta - Zu$ (Zewotir and Galpin 2007). Since mixed models have two sources of variability (within and between-subjects), two different types of residuals (marginal and conditional) may be defined and thus, the corresponding analysis would be more complex (Nobre and Singer 2007). Fei and Pan (2003) showed when the covariance structure is misspecified, incorrect identification of outlying observations may occur. Consequently, it is important to know whether genetic and residual covariance structures are correctly specified before performing the outlier test through the residual analysis in linear mixed models.

Apart from how to define the high and low limits which dictate the allowable range for a TD to be declared as outlier, the model features such as the type of distribution for the residuals and modelling of milk production data (305-day or test-day models) are of importance as well.

Most genetic evaluation programs for livestock are based on generalized linear mixed model analyses (Tempelman 1998) with breeding values or genetic effects modeled as random with a covariance structure (Kizilkaya and Tempelman 2005). As it is well known, the Gaussian (normal) linear mixed models are often used by animal breeders when estimating genetic parameters and breeding values for a quantitative trait (Strandén and Gianola 1997). These models are sensitive to departures from assumptions causing by outliers. One such method that uses the t-student distribution for the residuals was proposed by Strandén and Gianola (1997, 1998, and 1999). Since the t-distribution has thicker tails than the Gaussian, it allows more variation and pushes the critical *t* into the tails of the distribution, giving it less chance to obtain a significant residual. Consequently, outliers should influence inferences from t-distribution less than in a Gaussian model (Strandén and Gianola 1997).

In a similar research context, Jamrozik et al. (2004) adopted such method to test the performance of the robust statistical models that use heavy-tailed distributions for the residual effect. The proportion of outlier observations in the research of Jamrozik et al. (2004) which used random regression TD model with normality assumptions, was larger than reported by Strandén and Gianola (1997) for the 305-d model (3.7 vs. 1.4%). In other words, in the study of Jamrozik et al. (2004) the number of outliers using the t-distribution was slightly higher compared with the study of Strandén and Gianola (1997). Probably, utilizing a different model (TD model vs. 305-d model) could be led to a higher percentage of outliers.

Lange et al. (1989) has suggested that the t-distribution model is appropriate when outliers are neither extreme nor common, which according to Strandén and Gianola (1997) is reasonable to expect in TD milk yield records of dairy cattle. In contrast, the findings of the current investigation showed that the outlying records are both extreme and relatively common because a high proportion lactations which contained outliers, included only a single outlier.

5.10. Reword

Outlier detection is a critical part of the data analysis, and the use of ordinary and standardized residuals from various regression models is a very common approach for the identifying discordant observations in the linear regression problems. In general, there are two common alternatives in the face of such a situation including discarding and adjusting of outliers.

While there was no strong evidence to propose a certain coefficient k for both discarding and adjusting schemes, however, it seems from the results reported here that removing 0.4% to 0.8% of data in the editing process could be done. Without losing a large number of observations, the discarding method yielded results more accurate than those obtained with the regular analysis using full data, because of having smaller standard deviation for both h^2 and c^2. Likewise, when data was corrected by various regression methods, for RR a k=2 (1.13%), SR a k=5 (3.82%), and a k=5 for both LR1 (3.28%) and LR2 (4.50%) seem to be better values than other coefficients according to the results of the current survey. However, selection of a smaller k will influence a higher number of data but these did not introduce great changes in residual terms and other criteria. These limits also seem to be a good compromise between correction of abnormal TD values and percentage of detected outliers that has been reported by others, for example, 3.7% (Strandén and Gianola 1997), 1.9% (Wiggans et al. 2003), 0.113% (Mayeres et al. 2003), 1.4% (Jamrozik et al. 2004), < 4% (Jamrozik et al. 2007) and 3.3% (Quinn et al. 2007).

Furthermore, other techniques such as data mining or knowledge discovery which is the process of extracting hidden patterns and predictive information from large datasets, can be used for identifying trends, correlations, discrepancies, irregularities and disruptions (Banos et al. 2003). An attractive feature of this method is that no assumptions on data structure are required to validate consistent and replicable pattern hypotheses (Banos et al. 2003).

In this study, outliers were defined arbitrarily and based on self and rational understanding using residuals calculated for each lactation, coefficient k, regression method and treating scheme. The methods and coefficients k were chosen on the basis of their suitability, methodological considerations and previous studies as well as with the aim of determining whether reasonable adjustments can be made to decrease the obstructive influences of the outliers. Definition of outliers was, therefore, method dependent (i.e. regression, scheme, k level) as done by some authors (Wiggans et al. 2003, Yang et al. 2004, Jamrozik et al. 2007, Quinn et al. 2007). It means that the different regression methods with the various values of k defined outliers in a different way. Definition of outliers, however, is an important and a difficult matter underlying all the studies discussed in literatures which attempted to detect and to treat the outlying data. Observation that was an outlier for a higher percentage could be considered as a probable outlier candidate for a lower percentage. According to the results, although adjusting methods with the smaller k values detected relatively large proportions of outlying data, most of them could likely be considered as normal data. Thus, definition of outliers in the current study was somewhat different from the common definition of outliers.

Besides the known factors altering the overall pattern of the lactation curve, there is also high local variation due to the environmental effects as well as due to the biological nature of the lactation itself (Olori et al. 1999). According to Swalve (2000), while in heat, a cow's performance commonly drops to about 40 to 50% of the yield that was produced in the previous day. Therefore, a large proportion of outliers detected by a statistical model may not be an actual outlier due to factors which affect performance but were not included in the model.

Although, determining the causes of abnormal records was not the goal of this study, but there are several sources for their origins according to Quinn et al. (2007) : 1) manual or human errors in sampling, 2) mechanical or equipment sampling errors, 3) cow factors associated with the general wellbeing or stage of lactation of the cow, 4) post sampling treatment, and 5) the ability of the method outlined to identify true abnormalities.

Among all causes of outlying recordings, preferential treatment of cows has been taken into more consideration. Preferential treatment can be described as any management practice that increases production and is applied to one or several cows, but not to their contemporaries (Kuhn et al. 1994). Some studies have reported that the bias of predicted breeding values in genetic evaluation of dairy cattle may be due to outliers that cannot be accommodated by the statistical model which are caused, for instance, by preferential treatment of cows (Kuhn et al. 1994, Kuhn and Freeman 1995). However, using a suitable robust distribution (e.g. t-student) may alleviate the impact of outliers on genetic evaluations (Strandén and Gianola 1997).

In recent years, a number of countries have been forced to provide flexible service and pricing for farmers due to lack of funding required for their activities. However, this increased milk recording membership of dairy farmers (Miglior et al. 2002). According to Miglior et al. (2000), almost 20% of the countries (60% of milk recorded cows worldwide) were offering flexible options to farmers in terms of milk recording. Consequently, these activities decrease the ability of the model to make a good prediction and can cause some difficulties in accurately identifying the true outliers in the face of many different types of heterogeneous resources.

In this connection, INTERBULL that is a sub-committee of the International Committee for Animal Recording (ICAR), strongly recommends the milk recording organisations to keep the quality of records in genetic evaluations through data quality assessment methods (Interbull 2001). However, in each country a number of criteria are used in the data editing process in order to exclude the logical inconsistencies and those records that make little biological meaning. National genetic evaluation centres, are also recommended by ICAR to devise simple methods of checking for detection of outliers and exclusion of logical inconsistencies in the input data. While it is recommended that sick cows not to be recorded, some evidence shows that their test values may be predicted by the farmer or by the milk recorder, or TD values are declared missing (Quinn et al. 2007). Thus, data editing for genetic evaluation systems generally needs deleting records that are considered extremely deviant from the phenotypic mean of its subclass (e.g. three standard deviations) or if a record falls outside the range of 60% to 140% or 60% to 150% (ICAR 1995, Bertrand and Wiggans 1998, Wiggans et al. 2003). ICAR (1995, 2002) in the revised recording guidelines proposed that true daily-test values labelled abnormal (sick, injured, under treatment or in heat) must be used in the computation of the lactation record.

However, if the test value is less than 50% of the previous test value or less than 60% of the predicted test value, values may be considered as missing (ICAR 1995, 2007). Although ICAR suggested suitable strategies for the very obvious data recording errors, these methods could be inefficient where the residual does not follow a normal distribution and/or variances are heterogeneous across various data subclasses (Cardoso et al. 2006).

The most important conclusion to be drawn from the above paragraphs is that ICAR allows the pre-evaluation filters that exclude a number of data during the editing process. While editing instructions should follow the recommendations of ICAR (Interbull 2001), some breeding organizations conduct different high and low limits such as 3.0 and 99.9 kg for milk (Mayere et al. 2003) other than one proposed by ICAR. Discarding method using ASTAR, as described herein, may be an acceptable substitute in such cases where there is a necessity to exclude very extreme TD records before genetic evaluation in the phase of preparation and editing of data.

6. *Conclusion*

Dealing with outlying or abnormal TD records includes two main parts. The first one is detection of outliers using a suitable statistical model that fits the data closely. In this context, there are several models which the best of them should be selected after a careful examination. If data is not correctly matched or model does not include all data features, predictions will be erroneous and some data will incorrectly be defined as outliers as well as some outliers will not be detected.

The second and more important one is treating the identified outliers through either discarding or adjusting schemes by defining the proper limits of very low and very high records. There is an expectation that the model should have relatively uniform distribution of incidence of outliers across lactation period. The frequency of outliers over the various phases of the lactation severely depends on defining the allowable limits of error.

This implies that dealing with outlier needs a model to detect the outliers and results critically depend on the model fitted on the data and defined limits. The results indicate that after k value, the regression model or procedure is a major contributing factor to the number of abnormal recordings.

The attained results show that all procedures were clearly superior in residual terms (MAR, MSR and SDR) over the data without treatment, but these results do not determine which coefficient k, is more convenient to use for each treating scheme. The best coefficient k highly depends on the data set, model function, regression strategy and definition of limits for the high and low TD yields whether discarding or adjusting schemes are applied. In accordance with the other criteria such as measure of additive genetic variance and heritability as well as their standard deviation, discarding 0.4% to 0.8% and adjusting 1% to 5% of data are recommended depending on the regression methodology.

This study has not provided evidence to determine which method (deleting or adjusting) is a better alternative to usual analysis with no treatment. The better method will be one that give a smaller residual measures over data with no manipulation, higher correlation of EBVs with true genetic values, higher probability for correctly ranking animals and smaller standard deviation for both h^2 and c^2.

Although elimination of outliers is a convenient way to protect the results from their obtrusive effects but removing outlying observations is not ideal because it discards any other information associated with that observation. This will be particularly extreme for the lactations with a small number of TD records. Thus, due to the severe impact of removing the TD records with the large deviation on residual criteria which may bias the obtained model parameters and genetic evaluations, researcher should avoid discarding a large amount of them.

Therefore, in order to protect the data from biases that could be caused by discarding outliers, application of an appropriate adjusting procedure will be a justifiable alternative to keeping them. Through adjusting the outlying records to a minimum or maximum, outliers are retained and at the same time, their impacts on calculation of residual criteria is reduced. Adjusting of outliers also led to a relative reduction in all residual terms, although reductions were less than discarding scheme. The results of this study indeed confirm that either of the methods which have been outlined, could be implemented prior to usual genetic evaluation systems without greatly increasing costs. However, a simultaneous use of both approaches (discarding and adjusting) may be a better choice than doing each of them individually.

Since the robust analysis described in this study limited the influence of outliers due to having the greatest outcomes and maximum benefits, it is recommended to apply these methods into prospective animal breeding. The current study also suggests to simultaneously use the residual terms (MAR, MSR and SDR) and standard deviation of heritability (or additive genetic variance) to evaluate and compare the capability of the lactation models in detecting and treating the outliers. Comparing the treatment of the outlying TD observations in the lactation level to the more general case of multiple lactation (contemporary groups) and using both discarding and adjusting together are the future direction of this research.

While residuals or transformations of them are highly informative to assess the fit of the specified model and reveal the relative inconsistencies between model and data, outliers resulting from unclear causes may be of interest as well. Therefore, detection of outliers using different models may lead to a direct assessment of the models' accuracy and consequently can give an evaluation and comparison criteria of the of fit obtained from the different models.

7. *Zusammenfassung*

Durch die große Anzahl an Daten, die für die Zuchtwertschätzung erhoben werden, ist es vorhersehbar, dass Fehler auftreten. Deshalb ist es normal, dass Ausreißer und abnormale Testtagsleistungen (TD) in den Leistungsdaten des Milchrindes vorkommen. Diese Ausreißer können die Zuchtwertschätzung nachteilig beeinflussen, sofern nicht darauf Einfluss genommen wird. Deshalb müssen geeignete quantitative Methoden entwickelt werden, um Ausreißer zu finden und zu behandeln. Die Überprüfung der Resteffekte ist eine der ersten Methoden, um Ausreißer zu finden. Die Skalierung der Resteffekte mittels der Divison durch die Standardfehler ist eine Alternative zur Beschreibung der unterschiedlichen Varianz.

Diese Arbeit beschäftigt sich mit zwei unterschiedlichen Verfahren, um Ausreißer zu behandeln. Die Varianzkomponentenschätzung mit einem fixed regression TD Model wurde zum Einen ohne Ausreißer und zum Anderen mit angepassten Ausreißern gerechnet. Beide Methoden wurden mit drei unterschiedlichen Regressionen gerechnet. Die Einfache (SR), die Robuste (RR) und die lokale Regression mit linearem (LR1) und quadratischem (LR2) Polynom. Ziel war es, die als Ausreißer gekennzeichneten TD zu entfernen oder zu reduzieren. Zum Einen wurde der prozentuale Anteil an Ausreißern der TD (0.1, 0.2, 0.4, 0.8, 1.6, 3.2 und 6.4%), welche die höchsten Werte der tatsächlichen Resteffekte (ABSR) oder der tatsächlichen standardisierten Resteffekte (ASTAR) darstellen, ausgeschlossen.

Die Methode, welche mit angepassten Ausreißern rechnet, verlangt für jede der drei Regressionen die Schätzung der Resteffekte und die Berechnung der Standardabweichung der Resteffekte für jede Kuh sowie die Reduzierung der Ausreißerresteffekte innerhalb k Standardabweichungen (SD). Anschließend wurde eine neue Beobachtung generiert (=k×SD), um sie in der Zuchtwertschätzung zu nutzen.

Weiterhin sollten drei Regressionen, angewandt auf 111,599 TD für Milch von 11,620 deutschen Holstein-Kühen bezüglich des Einflusses der Behandlung der Ausreißer in den Milchleistungsdaten in Abhängigkeit von der Anzahl und Verteilung der Ausreißer, der mittleren Abweichung (MAR), bezüglich der mittleren Abweichungsquadrate (MSR), der Standardfehler der Resteffekte (SDR) und der Varianzkomponenten verglichen werden. Es erfolgte außerdem eine Schätzung der genetischen und der nicht genetischen Parameter, der mittleren Zuchtwerten (MEBV) und der Rangierung der Tiere.

Die Ergebnisse werden mit den Originaldaten verglichen, wobei jede Art der Veränderung einbezogen wird. Die fixen Effekte im Model zur Auswertung der Milchleistung in der ersten Laktation waren Herde – Jahr – Kalbesaison, Herdentesttageffekt, fixe Regeression mit Ali und Schaeffer (1987) für die Tage in Milch, das Tier als zufälliger Effekt und die permanente Umwelt.

Vorläufige Auswertungen haben ergeben, dass Ausreißer in den Daten als auch Ungleichheit zwischen der Verteilung der Ausreißer in der Regression und den Modellen auftreten. Die relativen Häufigekeiten der Ausreißer lassen vermuten, dass die Abweichungen in einem spezifischen Abschnitt der Laktation in Abhängigkeit der jeweiligen Regressionstechnik öfter auftraten.

In allen Fällen zeigten SR und RR ein unterschiedliches Verhalten zur lokalen Regression (z.B. LR1 und LR2), in der die Ausreißer, die sich mit niedrigeren Werten vom normalen Mittel unterschieden, weniger als 50% der gesamt Ausreißer einnahmen.

Die Auswertung mit ASTAR anstelle von ABSR zeigte, dass MAR, MSR und SDR einen deutlich schwächeren Abfall hatten. Allerdings zeigte ASTAR dabei eine geringere Abhängigkeit von den Tagen in Milch und der Leistung.

In allen aufgezeigten Variationen reduzierte sich die phenotypische Varianz und die Varianz der Resteffekte mit zunehmender Anzahl Ausreißer in Abhängigkeit von der Methode und dem Prozentsatz an betrachteten Ausreißern. Die Korrektur auf Ausreißer hatte wenig Einfluss auf die MEBV der Tiere. Kühe zeigten eine größere Abweichung in den MEBV als Bullen.

Die Rangierung der besten einhundert Bullen und Kühe unterschied sich nicht von den Gesamtdaten. Die Ergebnisse dieser Untersuchung zeigten, dass die RR Methode den Ergebnissen der Methoden SR, LR1 und LR2 zu bevorzugen ist, weil sie eine kleinere Anzahl an behandelten Daten beeinflusst.

Die Ergebnisse der vorliegenden Studie lassen ebenfalls schlussfolgern, dass die Effektivität der einzelnen Regressionsbestimmungen für die Ausreißer sowohl eine gut funktionierende Methode bezüglich der Auswertung des Datenmaterials ist als auch eine angepasste Bewertung des Vorkommens an Ausreißern in der Laktation zulässt.

Obwohl die Ergebnisse der Resteffekte nicht hinreichend genug waren, um den angemessenen Koeffizienten k in Abhängigkeit des Verfahrens zu bestimmen, konnte generell eine verbessert Anpassung, sowohl für die Ausschluss- als auch Anpassungsmethode der Ausreißer erzielt werden. Jedoch sind die Ergebnisse, vor allem hinsichtlich der Standardabweichung für den additiv-genetischen und permanenten Umwelteffekt, von 0,4% bis 0,8% für die Ausschluss- und 1% bis 5% für die Anpassungsmethode der Ausreißer, je nach Methode der Regression zu empfehlen.

Die Auswahl des besten Koefficienten (k) ist stark von den Daten, dem Model, der Regressionsmethode und der Abgrenzung der hohen und niedrigen TD abhängig.

Die Arbeit mit Ausreißerdaten setzt sich aus zwei Hauptbestandteilen zusammen: erstens das Auffinden eines passenden Models für die Ausreißerproblematik und zweitens die Bewertung der Ausreißer. Außerdem muss Entweder das Weglassen der Ausreißer bei der Bearbeitung der Daten oder das Anpassen der Ausreißer unter der Festlegung von oberen und unteren Grenzen erfolgen. Sollte das Model nicht korrekt an die Daten angepasst sein oder die Limitierung der Daten nicht korrekt definiert sein, werden Daten inkorrekterweise als Ausreißer definiert oder nicht als Ausreißer erkannt.

8. References

Ali T. E., and L. R. Schaeffer (1987) Accounting for (co)variances among test day milk yields in dairy cows. Can. J. Anim. Sci. 67: 637-644.

Anderson C. R., and R. E. Schumacker (2003) A comparison of five robust regression methods with ordinary least squares: Realative efficiency, bias, and test of the null hypothesis. Understanding Statistics 2(2): 79-103.

Anscombe F. J. (1960) Rejection of outliers. Technometrics 2: 123-147.

Anscombe, F. J., and J . W. Tukey (1963) The examination and analysis of residuals. Technometrics 5:141.

Bacon D. (1995) A maximum likelihood approach to correlational outlier identification. Multivariate Behavioral Research 30: 125-148.

Banos G., P.A. Mitkas, Z. Abas, A. L. Symeonidis, G. Milis, and U. Emanuelson (2003) Quality control of national genetic evaluation results using data mining techniques; a progress report. Interbull Bull. 31: 8-15.

Barnett V., and T. Lewis (1984) Outliers in statistical data. Chichester, England: John Wiley & Sons.

Barnett V., and T. Lewis (1994) Outliers in statistical data. New York: John Wiley & Sons.

Basualdo M., E. M. Rodríguez, E. Bedascarrasbure, and D. De Jong. (2007) Selection and estimation of the heritability of sunflower (Helianthus annuus) pollen collection behavior in Apis mellifera colonies, Online Journal of Genetics and Molecular Research Volum 2-6.

Belsley D.A., E. Kuh, and R. E. Welsch (1980) Regression diagnostics. Identifying influential data and sources of collinearity. New York: John Wiley & Sons.

Ben-Gal I. (2005) Outlier detection, In: Maimon O. and Rockach L. (Editors.) Data mining and knowledge discovery handbook: A complete guide for practitioners and researchers, Kluwer Academic Publishers.

Ben-Gal I., G. Morag, and A. Shmilovici (2003) CSPC: A Monitoring procedure for state dependent processes. Technometrics 45(4): 293-311.

Bertrand J. A. (1996) Influence of shipping container, preservative, and breed on analysis of milk components of shipped samples. J. Dairy Sci. 79: 145-148.

Bertrand J. K., and G. R. Wiggans (1998) Validation of data and review of results from genetic evaluation systems for us beef and dairy cattle. Proc. 6th World Congr. Gen. Appl. Livest. Prod. Armidale, Australia. 27: 425-432

Birkes D., and Y. Dodge (1993) Alternative Methods of Regression, John Wiley & Sons, New York.

Boldman K. G., and A. E. Freeman (1990) Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. J. Dairy Sci. 73: 503-512.

Bormann J., G. R. Wiggans, T. Druet, and N. Gengler (2003) Within-herd effects of age at test day and lactation stage on test-day yields. J. Dairy Sci. 86: 3765-3774.

Box G. E. P., and G. M. Jenkins (1976) Time series analysis: Forecasting and control. San Francisco: Holden-Day (Rev. ed.).

Brothestone S., and W. G. Hill (1986) Heterogeneity of variance amongst herds for milk production. Anim. Prod. 42: 297-303.

Brügemann K. (2008) Untersuchungen zur Heterogenität der Varianzen von Merkmalen der milchleistung unter differenzierter Stratifikation des Gesamtmaterials beim Holsteinrind in Deutschland. PhD dissertation, Martin-Luther university Halle-Wittenberg.

Buckley F., K. O'Sullivan, J. Mee, R. Evans, and P. Dillon (2003) Relationships among milk yield, body condition, cow weight and reproduction in spring-calved Holstein-Friesians. J. Dairy Sci. 86: 2308-2319.

Cameron N. D. (1997) Selection indices and prediction of genetic merit in animal breeding. CAB International,. Wallingford, UK. 203 pp.

Cardoso F. F., G. J. M. Rosa, and R. J. Tempelman (2005) Multiple breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. J. Anim. Sci. 83: 1766-1779.

Cardoso F. F., G. J. M. Rosa, and R. J. Tempelman (2006) Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of post-weaning gain of Hereford-Nelore Cattle. J. Anim. Sci. (Online Published) Access on March 5, 2008.

Chen C. (2002) Robust regression and outlier detection with the ROBUSTREG procedure. SUGI Paper 265-27. SAS Institute: Cary, NC.

Chernobai A., and S. T. Rachev (2006) Applying Robust Methods to Operational Risk Modeling. Tech. rep., University of California Santa Barbara.

Cleveland W. S., and S. J. Devlin (1988) Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Statist. Assoc. 83: 596-610.

Cleveland W. S., S. J. Devlin, and E. Grosse (1988) Regression by local fitting. J. Econometrics 37: 87-114.

Cleveland W. S., and E. Grosse (1991) Computational methods for local regression. Statistics and Computing 1: 47-62.

Cleveland W. S., and C. Loader (1995) Smoothing by local regression: principles and methods, AT&T Bell Laboratories, Murray Hill, New York, NY, Technical report.

Coenders G., and M. Saez (2000) Collinearity, heteroscedasticity and outlier diagnostics in regression. Do they Always Offer what they Claim? in Ferligoj, A. & Mrvar, A. (Eds.) New Approaches in Applied Statistics, Metodološki Zvezki. 16. FDV, Ljubljana: 79-94.

Cohen R. A. (1999) An introduction to PROC LOESS for Local Regression, Proceedings of the 24th SAS Users Group International Conference (SUGI), Paper no. 273, 1584 -1592.

Cook R. D., and S. Weisberg (1982) Residuals and influence in regression. Chapman & Hall, London.

Cox D. R., and E. J. Snell (1968). A general definition of residuals (with discussion). J. Roy. Statistical Society. Seri B 30: 248-275.

De Veer J. C., and L. D. Van Vleck (1987) Genetic parameters for fust lactation milk yields at three levels of herd production. J. Dairy Sci. 70:1434.

Dong M. C., and I. L. Mao (1990) Heterogeneity of (co)variance and heritability in different levels of intraherd milk production variance and of herd average. J. Dairy Sci. 73: 843-851.

Douzenis C., and E. A. Rakow (1987) Outliers: a potential data problem. Paper presented at the annual meeting of the Mid-South Educational Research Association, Mobile, AL. (ERIC Document Reproduction Service no. ED 291 789)

Evans V. P. (1999) Strategies for detecting outliers in regression analysis: An introductory primer. In B. Thompson (Ed.), Advances in social science methodology 5: 213-233.

Fallon A., and C. Spada (1997) Environmental sampling and monitoring primer: detection and accommodation of outliers in normally distributed data sets. Online available http://ewr.cee.vt.edu /environmental/teach/smprimer/outlier/ outlier.html.

Fatehi J., L.R. Schaeffer, and J. Jamrozik (2005) Models with herd by year interaction and different definitions of season of birth for two heifer fertility traits. Technical Report to the Canadian Genetic Evaluation Board, Centre for Genetic Improvement of Livestock, University of Guelph.

Fei Y., and J. Pan (2003) Influence assessments for longitudinal data in linear mixed models. In 18th international workshop on Statistical Modelling. G. Verbeke, G. Molenberghs, M. Aerts and S. Fieuws (eds.). Leuven: Belgium, 143-148.

Foulley J. L., M. S. Cristobal, D. Gianola, and S. Im (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. Comput. Stat. Data. Anal. 13: 291-305.

Franklin S., and M. Brodeur (1997) A Practical application of a robust multivariate outlier detection method. Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 186.

Freund T. J., and R. C. Littell (2000) SAS system for regression, 3rd Edition. Cary: SAS Institute.

Gengler N., and G. R. Wiggans (2001) Heterogeneity in (co)variance structures of test-day yields. Interbull Bull. 27:179–183.

Gianola D. (2000) Statistics in animal breeding. Special 2000 issue on Statistics in the Life and Medical Sciences. J. Amer. Statistical Assoc. 95: 296-299.

Gill J. L. (1986) Outliers, residuals, and influence in multiple regression. J. Anim. Breed. Genet. 103: 161-175.

Gray J. B., and H. W. William (1994) The maximum size of standardized and internally studentized residuals in regression analysis. The American Statistician 48(2): 11-13.

Groeneveld E. (2003) PEST Users' Manual. Institute of Animal Husbandry and Animal Behaviour Federal Research Centre, Neustadt, Germany.

Groeneveld E., M. Kovac, and T. Wang (1990) Pest, a general purpose BLUP package for multivariate prediction and estimation, in: Hill W.G., Thompson R., Woolliams J.A. (Eds.), Proceedings of the 4th World Congress on Genetics Applied to Livestock Production, Edinburgh, July 23-27, Paper no. 13: 488-491.

Grubbs F. E. (1969) Procedures for detecting outlying observation in samples. Technometrics 11: 1- 21.

Guo Z., and H. H. Swalve (1995) Modelling of the lactation curve as a sub-model in the evaluation of test day records. Proc. Interbull Mtg. Prague, Czechoslovakia. International Bull Evaluation Service, Uppsala, Sweden. Interbull Bull. 11: 15-21.

Hadi A. S. (1992) Identifying multiple outliers in multivariate data. J. Roy. Statistical Society. Series B 54: 761-771.

Hadi A. S., and J. S. Simonoff (1993) Procedures for the identification of multiple outliers in linear models. J. Amer. Statistical Assoc. 88(424) 1264-1272.

Hamilton L.C. (1992) Regressions with graphics: A second course in applied statistics. Monterey, CA.: Brooks/Cole.

Hampel F. R. (1971) A general qualitative definition of robustness. Annals of Mathematic Statistics 42: 1887-1896.

Hampel F. R. (1974) The influence curve and its role in robust estimation. J. Amer. Statistical Assoc. 69: 382-393.

Hampel F. R., E. M. Ronchetti, P. J. Rousseeuwand, and W. A. Stahel (1986) Robust statistics, The Approach Based on Influence Functions. John Wiley & Sons, New York.

Haslett J. (1999) A simple derivation of deletion diagnostics for the general linear model with correlated errors. J. Roy. Statistical Society. Series B 61:603-609

Haslett J., and K. Hayes (1998) Residuals for the linear model with general covariance structure. J. Roy. Statistical Society. Series B 60:201-215

Hawkins D. M. (1980) Identification of outliers. London: Chapman and Hall.

Hecht J. B. (1991) Least squares regression and Schrodingers cat: Perspectives on the analysis of regression residuals. Paper presented at the annual meeting of the American Educational Research Association, Chicago, USA, paper no. ED 333 020.

Henderson C. R. 1984. Applications of Linear Models in Animal Breeding. Guelph, Ontario, University of Guelph.

Hoaglin D. C., and R. E. Welsch (1978) The Hat matrix in regression and ANOVA. The American Statistician 32: 17-22.

Huber P. J. (1973) Robust regression: Asymptotics, conjectures and Monte Carlo. Ann. Stat. 1: 799-821.

Huber P. J. (1990) Between robustness and diagnostics. In: StahelWA,Weisberg S (eds) Directions in robust statistics and diagnostics, Part I. Springer, New York, pp 121-130.

Iglewicz B., and D. C. Hoaglin (1993) How to detect and handle outliers. Milwaukee, WI: ASQC Quality Press.

International Bull Evaluation Service (2001) Interbull guidelines for national and international genetic evaluation systems in diary cattle with focus on production traits. INTERBULL, Uppsala, Sweden, Available online. Accessed July 11, 2005.

International Committee for Animal Recording (1995) Recording Guidelines: Appendices to the International Agreement of Recording Practices. ICAR, Rome, Italy. Available online. Accessed September 1, 2005.

International Committee for Animal Recording (2002) Recording Guidelines: International Agreement of Recording Practices. Approved by the General Assembly held in Interlaken, ICAR, Switzerland. Available online. Accessed June 10, 2005.

International Committee for Animal Recording (2007) International Agreement of Recording Practices, ICAR, Approved by the General Assembly held in Kuopio, Finland, Available online. Accessed August 20, 2007.

Jacobs R. (2001) Outliers in statistical analysis: Basic methods of detection and accommodation. Paper presented at the Annual Meeting of the Southwest Educational Research Association, February 1-3, New Orleans, USA.

Jamrozik J., J. Fatehi, and L.R. Schaeffer (2007) Application of robust procedures for estimation of breeding values in multiple-trait random regression test-day model. J. Anim. Breed. Genet. 124: 3-11

Jamrozik J., I. Stranden, and L. R. Schaeffer (2004) Random regression test-day models with residuals following a Student's-t distribution. J. Dairy Sci. 87: 699-705.

Jensen J. (2001) Genetic evaluation of dairy cattle using test-day models. J. Dairy Sci. 84: 2803-2812.

Johnson R. (1992) Applied multivariate statistical analysis. Prentice Hall.

Kaya I., Y. Akbas, and C. Uzmay (2003) Estimation of breeding values for dairy cattle using test-day milk yields. Turkey J. Veterinary Anim. Sci. 27: 459-464.

Khanitta C., D. Monchai, P. Virote, and R. Chancharat.2005. Model comparison for genetic evaluation of milk yield in crossbred Holsteins in the tropics. J. Appl. Genet 46(4):387-393.

Kizilkaya K., and R.J. Tempelman (2005) A general approach to mixed effects modeling of residual variances in generalized linear mixed models. Genet. Sel. Evol. 37 31-56.

Kovac M., and E. Groeneveld (2002) VCE, User’s guide and reference manual, Version 5.1. Institute of Animal Science Federal Agricultural Research Center (FAL), Neustadt, Germany.

Krejčová H., N. Mielenz, J. Přibyl, and L. Schüler (2007) Comparison of breeding values for daily gains of bulls estimated with multi- trait and random regression models. Arch. Tierz. Dummerstorf 50 (2): 147-154.

Kuhn M. T., P. J. Boettcher, and A. E. Freeman (1994) Potential biases in predicted transmitting abilities of females from preferential treatment. J. Dairy Sci. 77: 2428-2437.

Kuhn M. T., and A. E. Freeman (1995) Biases in predicted transmitting abilities of sires when daughters receive preferential treatment. J. Dairy Sci. 78: 2067-2072.

Lane K. (2002) What is robust regression and how do you do it? Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX.

Lange K. L., and J. S. Sinsheimer (1993) Normal/independent distributions and their applications in robust regression. J. Comput. Graph. Stat. 2: 175-198.

Larsen W. A., and S. J. McCleary (1972) The use of partial residual plots in regression analysis. Technometrics 14: 781-790.

Lazenby D., E. Bohlsen, K. J. Hand, D. F. Kelton, F. Miglior, and K. D. Lissemore (2002) Methods to estimate 24-hour yields for milk, fat and protein in robotic milking herds. Presented at the 33th biennial Session of ICAR, May 26-31, Interlaken, Switzerland.

Lidauer, M., and E. Mäntysaari (1996) Genetic constitution of the Finnish Black and White cattle population and the influence of Holsteinization on protein yield, days open and somatic cell count. Acta Agric. Scand. Section A, Anim. Sci. 46: 193-200.

Liu H., S. Shah, and W. Jiang (2004) On-line outlier detection and data cleaning. Computer and Chemical Engineering 28: 1635-1647.

Liu Z., J. Jamrozik, and G. Jansen (1998) A comparison of fixed and random regression models applied to dairy test day production data. Proceedings of the 1998 Interbull meeting, Rotorua, New Zealand, 17: 60-63.

Liu Z., R. Reents, F. Reinhardt, and K. Kuwan (2000) Approaches to estimating daily yield from single milk testing schemes and use of a.m.-p.m. records in test-day model Genetic evaluation in dairy cattle. J. Dairy Sci. 83:2672-2682.

Loader C. R. (1999) Bandwidth selection: classical or plug-in, The Annals of Statistics 27(2): 415-438.

Macciotta N. P. P., D. Vicario, and A. Cappio-Borlino (2005) Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. J. Dairy Sci. 88: 1178-1191.

Mao I. L. (1995) Comparison of methods for calculating lactation total yield from test-day records. Report to Lactation Working Group of ICAR. (mimeo).

Martens H., and T. Naes (1989) Multivariate calibration. John Wiley & Sons, Chichester, UK.

Martin R. J. (1992) Leverage, influence and residuals in regression models when observations are correlated. Comm. Statist. Theory Methods 21: 1183-1212

Mayeres P., A. Gillon, and N. Gengler (2003) Integrated detection and correction of outliers in a random regression test-day model. Interbull Bull. 31: 3-7.

McGuire M., M. Theurer, J. Vicini, and B. Crooker (2004) Controlling energy balance in early lactation. Advances in Dairy Technology 16: 241–252.

Meinert T. R., and H. D. Norman (1998) Merit of outliers for milk yield as indicators of accuracy of genetic evaluations of sires. J. Dairy Sci. 81(11): 2951-2955.

Meinert T. R., R. E. Pearson, W. E. Vinson, and B. G. Cassen (1988) Prediction of daughter's performance from dam's cow index adjusted for within-herd variance. J. Dairy Sci. 71: 2220-2231.

Miglior F., P. Galesloot, Z. Liu, M. Mathevon, A. Rosati, L.R. Schaeffer, and P. VanRaden (2000) Report of the ICAR working group on lactation calculation methods: A daily yield lactation survey in dairy cattle. In Proceedings of the 32th biennial Session of ICAR, EAAP publication no. 98.

Miglior F., S. de Roos, Z. Liu, M. Mathevon, A. Rosati, L.R. Schaeffer, and P. VanRaden (2002) Report of the ICAR working group on lactation calculation methods: Review and update of guidelines for milk recording. Presented at the 33th biennial Session of ICAR, Interlaken, Switzerland, May 26-31.

Militino A. F., and M. D. Ugarte (1997) Bounded influence estimation in a spatial linear mixed model. In: Gregorie TG, Brillinger DR, Diggle PJ, Russek-Cohen E, Wolfinger WWG, Wolfinger RD (eds) Modelling longitudinal and spatial correlated data. Springer, New York, pp 211-220.

Montanari A. (1995) Descriptive tests for the identification of outliers in the errors in variables linear model. J. Italian Statistical Society 4(1):89-100.

Mrode R. A. (1996) Linear models for the prediction of animal breeding values, CAB International, Wallingford, Oxon, UK.

Mrode R. A., G. J. T. Swanson, and C. M. Lindberg (2002) Efficiency of part lactation test day records for genetic evaluations using fixed and random regression models. Anim. Sci. 74: 189-197.

Naes T., T. Isaksson, T. Fearn, and T. Davies (2002) A user friendly guide to multivariate calibration and classification. Chichester, UK: NIR Publications.

National Dairy Herd Improvement Association (2002) National Dairy Herd Improvement Program uniform operating procedures (effective June 2002). Online. Available: http://www.dhia.org/uop.pdf. Available online. Accessed June 23, 2005.

Nevitt J., and H. P. Tam (1997) A comparison of robust and nonparametric estimators under the simple linear regression model. Multiple Linear Regression Viewpoints 25:54 - 69.

NIST/SEMATECH (2003) Engineering Statistics Internet Handbook. Tobias P. and C. Croarkin, eds., National Institute of Standards and Technology, U.S. Dept. of Commerce, Washington, DC, Online access http://www.itl.nist.gov/div898/handbook/.

Nobre J. S., and J. M. Singer (2007) Residual analysis for linear mixed models. Biometrical J. 49(6): 863-875.

Norman H. D., J. R. Wright, and J.S. Clay (1998) Comparison of the test interval method with best prediction for estimating lactation yield. Proc. 6th World Congr. Genet. Appl. Livest. Prod. 23:343-346. Armidale, Australia, Jan. 11-16.

O'Kelly M. (2006) A tour around PROC ROBUSTREG. Online Access: On June 27, 2008 http://www.lexjansen.com/phuse/2006/st/st01.pdf

Olive D. (2005) A Course in Statistical Theory. Southern Illinois University Department of Mathematics.

Olori V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk (1999) Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. Livest. Prod. Sci. 58: 55-63.

Orr J. M., P. R. Sackett, and C. L. Z DuBois (1991) Outlier detection and treatment in I/O Psychology: A survey of researcher beliefs and an empirical illustration. Personnel Psychology 44: 473-486.

Osborne J. W. (2002) Notes on the use of data transformations. Practical Assessment, Research, and Evaluation., 8, Available online http://ericae.net/pare/getvn.aspv=8&n=6.

Osborne J. W., and A. Overbay (2004) The power of outliers (and why researchers should always check for them). Practical Assessment, Research & Evaluation, 9(6). Retrieved May 2, 2006 from http://PAREonline.net/getvn.asp?v=9&n=6 .

Pedhazur E. J. (1982) Multiple regression in behavioral research (2nd ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Peeters R., and P. J. B. Galesloot (2002) Estimating daily fat yield from a single milking on test day for herds with a robotic milking system. J. Dairy Sci. 85: 682-688.

Perochon L., J. B. Coulon, and F. Lescourret (1996) Modelling lactation curves of dairy cows with emphasis on individual variability. Anim. Sci. 63: 189-200.

Pinheiro J. C., and D. M. Bates (2000) Mixed-effects in S and S-PLUS. Springer, New York.

Ptak E., and L. R. Schaeffer (1993) Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34: 23-34.

Quinn N., L. Killen, T. Guinee, and F. Buckley (2007) Detection of abnormal recordings in Irish milk recorded data. Ir. J. Agric . Food Res. 46: 117-127.

Quist M. A., S. J. LeBlanc, K. J. Hand, D. Lazenby, F. Miglior, and D. F. Kelton (2007) Agreement of predicted 305-day milk yields relative to actual 305-day milk weight yields. J. Dairy Sci. 90: 4684

Rawlings J. O., S. G. Pentula, and D. A. Dickey (1998) Applied regression analysis: A research tool, Second Edition. Springer-Verlag New York, Inc.

Reents R., J. C. M. Dekkers, and L. R. Schaeffer (1995) Genetic evaluation for somatic cell score with a test day model for multiple lactations. J. Dairy. Sci. 77: 2858- 2870.

Rekik B., and A. B. Gara (2004) Factors affecting the occurrence of atypical lactations for Holstein-Friesian cows. Livest. Prod. Sci. 87:245-250.

Rencher A. C., and G. B. Schaalje (2008) Linear models in statistics. 2. Edition, Wiley Series in Probability and Statistics.

Rosa, G. J. M., C. R. Padovani, and D. Gianola (2003) Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. Biometrical J. 45:573-590.

Rousseeuw P., and A. Leory (1987) Robust regression and outlier detection. Wiley Series in Probability and Statistics.

Rousseeuw P. J., and K. Van Driessen (1998) Computing LTS regression for large data sets. Technical Report, University of Antwerp.

Rousseeuw P. J., and V. Yohai (1984) Robust regression by means of S estimators in robust and nonlinear time series analysis. edited by J. Franke, W. Härdle, and R.D. Martin, Lecture Notes in Statistics 26, Springer Verlag, New York, 256-274.

Rutter, C. M., and R. M. Elashoff (1994) Analysis of longitudinal data: random coefficient regression modelling. Statistics in Medicine 13: 1211-1231.

Sachs L. (1982) Applied statistics: A handbook of techniques (2nd ed.). New York: Springer-Verlag.

SAS/STAT (1999) User's Guide, Version 8, Cary, NC: SAS Institute Inc.

SAS/STAT (2003) SAS/STAT 9.1 User's Guide, Cary, NC: SAS Institute Inc.

Schaeffer L.R., and J. Jamrozik (1996) Multiple-trait prediction of lactation yields for dairy cows. J. Dairy Sci. 79: 2044-2055.

Schiffler R. E. (1988) Maximum Z Score and outliers. The American Statistician 42(1): 79-80.

Scott M. L (2003) SOC 504 Course: Social Statistics, www.princeton.edu/~slynch. Online access March 28, 2008.

Serdahl E. (1996) An introduction to graphical analysis of residual scores and outlier detection in bivariate least squares regression analysis. Paper presented at the annual meeting of the Southwest Educational research Association, New Orleans, USA, paper no. ED 395 949.

Shoemaker J. (1999) Robust outlier identification using SAS. 24th SAS User's Group International proceedings.

Silvestre A. M. D., F. Petim-Batista, and J. Colaço (2006) The accuracy of seven mathematical functions in modelling dairy cattle lactation curves based on test day records from varying sample schemes. J. Dairy Sci. 89: 1813-1821

Stevens J. P. (1984) Outliers and influential data points in regression analysis. Psychological Bulletin, 95: 334-344.

Strabel T., and I. Misztal (1999) Genetic parameters for first and second lactation milk yields of Polish Black and White cattle with random regression test-day models. J. Dairy Sci. 82: 2805-2810.

Strabel T., J. Szyda, E. Ptak, and J. Jamrozik (2005) Comparison of random regression test-day models for Polish Black-and-White cattle. J. Dairy Sci. 88: 3688-3699.

Stranden I., and D. Gianola (1997) Gaussian versus Student's-t mixed effects linear models for milk yield in Ayrshire Cattle. 48th Annual Meeting of the EAAP, Vienna, Austria.

Stranden I., and D. Gianola (1998) Attenuating effects of preferential treatment with Student-t mixed linear models: A simulation study. Genet. Sel. Evol. 30: 565-583.

Stranden I., and D. Gianola (1999) Mixed effects linear models with t-distributions for quantitative genetic analysis: A Bayesian approach. Genet. Sel. Evol. 31:25-34.

Stromberg A. J. (2004) Why write statistical software? The case of robust statistical methods. J. Stat. Softw. 10(5): April 2004.

Suzuki M., J. A. Pereira, K. Kuchida, Y. Saito and Y. Ikeuchi (1998) Interaction due to genetic and environmental effects on milk production in Japanese Holstein. 6th World Congress on Genetics Applied to Livestock Production. 23:455.

Swalve H. H. (1995a) The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. J. Dairy Sci. 78:929-938.

Swalve H. H. (1995b) Test day models in the analysis of dairy production data: A review. Arch. Tierz. 38:591-612.

Swalve H.H. (1998) Use of test day records in genetic evaluation. Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, New South Wales, Australia 23:295-302.

Swalve H. H. (2000) Theoretical basis and computational methods for different test-day genetic evaluation methods. J. Dairy Sci. 83: 1115-1124.

Szyda J. and Z. Liu (1999) Modelling test day data from dairy cattle, J. Applied Genetics 40: 103-116.

Takezawa K. (2006) Introduction to nonparametric regression. John Wiley & Sons, Inc.

Tekerli M., Z. Akinci, I. Dogan, and A. Akcan (2000) Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. J. Dairy Sci. 83:1381-1386.

Tempelman R. J. (1998) Generalized linear mixed models in dairy cattle breeding. J. Dairy Sci. 81: 1428-1444.

Tietjen G. L., and R. H. Moore (1972) An extension of some Grubbs-type statistics for the detection of several outliers. Technometrics 14: 583-598.

Tukey J. W. (1977) Exploratory data analysis. Reading, MA: Addison-Wesley.

VanRaden P. M. (1997) Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. J. Dairy Sci. 80: 3015-3022.

VanRaden P. M. (2006) Normality and skewness of genetic evaluations. Interbull Annual Meeting Proceedings. Interbull Bull. 35: 164-167.

Venables W. N., and B. D. Ripley (1999) Modern applied statistics with S-PLUS. 4th ed. Springer: New York.

Verbeke G., and E. Lesaffre (1996) A linear mixed-effects model with heterogeneity in the random-effects population. J. Amer. Statistical Assoc. 91: 217-221.

Wainer H. (1976) Robust statistics: A survey and some prescriptions. J. Educ. Stat. 1(4): 285-312.

Weisberg S. (2005) Applied Linear Regression. 3rd edition. New York: Wiley.

Wiggans G. R., and P. M. VanRaden (1991) Method and effect of adjustment for heterogeneous variance. J. Dairy Sci. 74: 4350-4357.

Wiggans G. R., P. M. VanRaden, and J. C. Philpot (2003) Detection and adjustment of abnormal test-day yields. J. Dairy Sci. 86: 2721-2724.

Wiggins B. C. (2000) Detecting and dealing with outliers in univariate and multivariate contexts. Paper presented at the annual meeting of the Mid-South Educational Research Association, Bowling Green, KY, USA, paper no. ED 448 189.

Williams G. J., R. A. Baxter, H. X. He, S. Hawkins, and L. Gu (2002) A Comparative study of RNN for outlier detection in data mining, IEEE International Conference on Data-mining (ICDM'02), Maebashi City, Japan, CSIRO Technical Report CMIS-02/102.

Williams V. S. L., L. V. Jones, and J. W. Tukey (1999) Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. J. Educ. Behav. Stat 24, 42-69.

Wilmink J. B. M. (1987) Adjustment of test-day milk, fat and protein yields for age, season and stage of lactation. Livest. Prod. Sci. 16: 335-348.

Wolfinger R. (1993) Covariance structure selection in general mixed models. Communications in Statistics-Simulation 22: 1079-1106

Wood P. D. P. (1967) Algebraic model of the lactation curve in cattle. Nature 216: 164-165.

Yaffee R. A. (2002) Robust Regression Analysis: Some Popular Statistical Package Options Available online: http://www.nyu.edu/its/socsci/Docs/RobustReg2.pdf.

Yang R., L. R. Schaeffer, and J. Jamrozik (2004) Robust estimation of breeding values in a random regression test-day model. J. Anim. Breed. Genet. 121:221-228.

Yohai V. J. (1987) High breakdown point and high efficiency robust estimates for regression. Annals of Statistics 15: 642-656.

Zewotir T., and J. S. Galpin (2007) A unified approach on residuals, leverages and outliers in the linear mixed model. Test 16: 58-75.

Zimmerman D. W. (1994) A note on the influence of outliers on parametric and nonparametric tests. J. General Psychology 121(4): 391-401.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation:

“Impact of dealing with outliers on variance estimations and genetic evaluations with milk yield test-day models in German Holstein cows”

selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner Prüfungsbehörde vorgelegt.

Halle/Saale, den 08.06.2009 *SeyedFaridoddin Rezazadeh Baei*

CURRICULUM VITAE

Personal data:

- **Surname:** Rezazadeh Baei
- **Name:** Seyed Faridoddin
- **Birth date and place:** 1972, Iran
- **Sex:** Male
- **Marital status:** Married
- **Address:** Institute of Animal Breeding and Animal Husbandry
  Martin-Luther-University Halle-Wittenberg
  Adam-Kuckhoff-Str. 35, D-06108 Halle / Saale, Germany
- **Phone number**: +49-345-5522345, +49-345-7792800
- **Fax number**: +49-345-5527105
- **E-mail**: seyad.rezazadeh@landw.uni-halle.de

Educational background and employment history:

Martin-Luther University Halle-Wittenberg, Germany
Ph.D. in Animal Breeding and Genetics **since Oct. 2004**
Title of thesis: Impact of dealing with outliers on variance estimations and genetic evaluations with milk yield test-day models in German Holstein cows

Department of Animal Science, Azad University, Ghaemshahr, Iran
Lecturer, Teaching undergraduate courses in animal science and breeding **2000-2004**
Title of courses taught: Animal husbandry, General genetics, Animal breeding (Elementary and Advanced)

Chamran University, Ahvaz, Iran
M.Sc. in Animal Science **1997-1999**
Title of thesis: Estimating genetic parameters for some of economical traits in Najdi Cattle

Tehran University, Tehran, Iran
B.Sc. in Agricultural Engineering **1992-1996**