DISSERTATION

RANDOM REGRESSION MODELS FOR THE PREDICTION OF DAYS TO FINISH IN BEEF CATTLE

Submitted by

Scott Eugene Speidel

Department of Animal Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2011

Doctoral Committee:

Advisor: R. Mark Enns

Denny Crews Jr.
R. Kraig Peel
Ann Hess

ABSTRACT

RANDOM REGRESSION MODELS FOR THE PREDICTION OF DAYS TO FINISH IN BEEF CATTLE

The idea of reducing the number of days required for livestock to reach their desired endpoint is not new, with its economic importance first discussed in 1957. Given this economic relevance, genetic evaluation research for reducing these required days has received very little attention throughout the pertinent literature with the exception of the swine industry. Many different production systems exist in today's beef industry, and a single prediction for the required number of days to reach a single finish endpoint does not lend itself well to this diversity. Because of this point, random regression models are an attractive alternative to more traditional multiple trait or repeated measures best linear unbiased prediction models in the calculation of days to finish.

Random regression models estimate regression lines for each animal in the pedigree, thereby resulting in the ability to calculate estimated breeding values (EBV) for any age or any number of days on feed. This inherent property allows beef producers to calculate days to finish EBV for finish endpoints that fit individual production scenarios.

The objective of this study was to develop a series of models using random regression techniques for the genetic prediction of the required number of days to reach the finish endpoints of weight, ultrasound back fat and ultrasound rib eye area. This

study performed some basic tasks of describing data and the behavior of random regression models used for the prediction of days to finish.

Genetic predictions for the traits days to weight (DTW), days to ultrasound back fat (DTUBF) and days to ultrasound ribeye area (DTUREA) were prototyped using data obtained from the Agriculture and Agri-Food Canada Research Centre, Lethbridge, Alberta. This data consisted of pedigree, weight, ultrasound back fat and ultrasound ribeye area observations on 1,324 cattle spanning the years 1999 – 2007. Individual animals averaged 5.77 weight observations with weights and ages ranging from 293 kg to 863 kg and 276 to 519 days, respectively. For the ultrasound traits individual animals averaged 5.57 observations. Ultrasound back fat observations ranged from 1.53 mm to 30.47 mm and ultrasound rib eye area observations ranged from 36.77 cm^2 to 129.54 cm^2.

Fixed effects included in the model were determined through a series of regressions to identify those accounting for a significant amount of variation in the age response variable. Results showed for the trait DTW the effects of year, pen and breed type should be included, and due to the confounding of year and breed type, all three were included in the contemporary group definition. Similar results were obtained for both DTUREA and DTUBF. Year of measure, pen and breed were included in the contemporary group definition for both traits. Using these three effects to form contemporary groups resulted in average contemporary group sizes of 21.50 and 21.45 for the days to weight and days to ultrasound traits, respectively. All three models, contained the effects of contemporary group and a fixed regression of age on weight / ultrasound back fat / ultrasound rib eye area to account for the overall mean relationship between age and each of the three finish traits.

Random regression models were built for each of the days to finish traits. Model building exercises for the three traits consisted of conducting likelihood ratio tests to determine the order of the random regression polynomial. For DTW, a linear random regression polynomial was sufficient in describing the genetic variation in days. Depending on how residual variance was modeled, heritability estimates varied. When observations were classified into four distinct residual variance sub-groups, heritability estimates for DTW ranged from 0.56 for the number of days to reach 293 kg all the way to 0.93 for the number of days to reach 863 kg. If residual variance was modeled using a linear random regression, heritability estimates for DTW were more conservative ranging from 0.53 for the number of days to reach 293 kg to 0.76 for the number of days to reach 863 kg.

The significant random regression order for the ultrasound traits was dependent on how the residual variance was modeled. For DTUREA, when residual variance was modeled using four distinct sub-groups, a quartic random polynomial was needed to model the genetic variation in days. When a linear random regression was applied to the residuals, a linear polynomial was all that was needed. The quartic polynomial tended to artificially inflate heritability estimates in the extremes of the data distribution for DTUREA ranging from 0.81 (36.77 cm^2) then dropping to 0.15 around 110 cm^2 and jumping back up to 0.91 at 129.54 cm^2. Heritability estimates obtained from the linear random regression using linear residual random regression were much more sensible, ranging from 0.53 at 36.77 cm^2 to 0.49 at 129.54 cm^2.

For the trait DTUBF, when residual variance was modeled using four distinct sub-groups, a quadratic random polynomial was all that was needed to describe the genetic

variation in days. Similar to DTW and DTUREA, the linear residual random regression model only needed a linear polynomial. Heritability estimates for DTUBF from the linear random regression model using linear residual random regression ranged from 0.54 at 1.53 mm of ultrasound back fat to 0.35 at 30.47 mm of back fat. Heritability estimates from the four residual sub-groups model became much more variable ranging from 0.58 at 1.53 mm of back fat down to 0.08 at 26 mm of back fat then jumping back up to 0.54 at 30.47 mm of back fat.

For all three traits, modeling the residual variance using a linear random regression seemed to be the most ideal, as it required the lowest order polynomial for describing the genetic variation in days. The linear residual random also yielded the most realistic heritability estimates for each of the endpoints. Heritability estimates obtained in this study show the days to finish traits are moderately to highly heritable, depending on endpoint. As such, sufficient genetic variation exists to make fairly rapid progress in reducing the number of days to reach finish endpoints, giving producers tools to increase the profitability of their operations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER I

INTRODUCTION AND OBJECTIVE

Introduction

Traditionally, in feedlot situations, cattle have been fed to a time constant endpoint with price being determined by the live weight of the animal. However, marketing cattle on the basis of their live weight may not be the most desirable or profitable for all cattle types.

With the advent of marketing alliances, carcass premium grids and value-based marketing systems, more emphasis has been placed on the selection of carcass traits as a way to increase the selling price of finished animals. Depending on the system, cattle producers can be paid a higher price if their cattle have less back fat and larger ribeye areas, resulting in a lower yield grade, or more intramuscular fat corresponding to higher quality grades. With these systems, low yield grades and high quality grades result in cattle producers receiving premiums for their animals. Additionally, high yield grades and low quality grades result in discounts for the price of their cattle; however, harvest weight is still an important trait since all prices are paid per unit of weight. Over the past 10 to 20 years, cattle producers wishing to retain ownership on their cattle through the feedlot have benefited from the increase in carcass merit selection tools produced by national cattle evaluation programs.

These marketing systems suggest to producers that cattle be fed to particular finish endpoints. The choice of this endpoint is very dependent on the biological type of cattle being marketed as well as the marketing systems available to the owners (Amer et al., 1994; Williams and Bennett, 1995).

Given a proper diet, all cattle will reach an appropriate finish endpoint at some point. However, a large amount of variation exists in the time it takes individual animals to reach these endpoints. Given the feed and yardage costs cattle owners incur each day an animal is in the feedlot, any shortening of the time it takes cattle to reach their desired economic endpoint would be beneficial.

Reducing the number of days required for livestock to reach a specific weight or finish endpoint has received very little attention throughout literature. With the exception of the swine industry this research has been almost non-existent. Only a handful of studies pertaining to beef cattle having been published going back to 1957. In summary, Lindholm and Stonaker (1957) found an average phenotypic correlation of -0.46 between the number of days to reach a perceived quality grade and net income per 45.4 kg of slaughter weight. Thirty years later, McWhir and Wilton (1987) found the heritability for the number of days to reach a back fat depth of 7 mm to be 0.65, which increased to 0.90 when the trait was adjusted to a constant market weight. Then in 1992, Johnston et al., reported the heritability for the number of days to reach 8.9 mm of back fat to be 0.24. Both of these studies ignored the fact that animals will re-rank depending on the endpoint and marketing system chosen by cattle owners.

Random regression models are a method of analyzing data with repeated observations, and were first used for genetic prediction in the early 1990's (Schaeffer and

Dekkers, 1994). These models are used to rank cattle on the basis of their genetic merit for the amount of time it would take to reach a specific endpoint. This ranking is facilitated by regressing the number of days to a certain endpoint on the endpoint itself. Random regression models assume the regression coefficients for each individual's additive genetic makeup are random, which allows for the genetic prediction of a regression line. Breeding values (or EPD) for a specific endpoint can be calculated from this regression line as follows:

$$\text{Days} = b_0 + b_1(\text{backfat / weight / ribeye area})$$

In this equation, animals would be ranked based on their genetic merit for the number of days it takes to reach the preferred endpoint. The term b_0 in this equation is the animal's breeding value for the intercept and the term b_1 is the animal's linear coefficient breeding value. The producer's desired backfat, weight or ribeye area endpoint could be plugged into the equation and the animals would then be ranked according to the number of days required to reach that endpoint.

Random regression models have been traditionally used to analyze data with several records per animal, which is why their use in the beef industry has been very limited. Typical feedlot data contain very few records of each animal's backfat depth and ribeye area (via ultrasound or harvest measurement) and weight. A field data set with a sufficient number of observations per individual animal would allow the parameterization of these models by obtaining accurate estimates of variance components. Then, once the models were built and variance estimates obtained they could be extended to similar

situations more appropriate to beef cattle production scenarios (such as data sets with fewer observations per individual animal).

Objective

The objectives of this study were the development of an illustrative example of the equivalency between random regression and traditional multivariate models currently used in beef cattle genetic prediction as well as the development of a days to finish genetic prediction using three alternate endpoints. Each of the specific objectives discussed in more detail below.

1) Build an example evaluation that illustrates that the estimated breeding values obtained from a random regression model are equivalent, and in certain instances identical to those obtained from multivariate models.
2) Develop a days to weight genetic prediction. The model building exercise consisted of determining which predictor variables should be included in the fixed portion of the mixed model, and determining the number of parameters that need be included in the random portion of the model. The order of the random regression, whether a random permanent environmental effect should be included as well as different approaches to model the residual variation was determined.
3) Develop a days to ultrasound back fat and a days to ultrasound rib eye area genetic prediction. Model building exercises similar to those presented above for the days to weight genetic prediction were used.

Literature cited

Amer, P. R., R. A. Kemp, J. G. Buchanan-Smith, G. C. Fox, and C. Smith. 1994. A bioeconomic model for comparing beef cattle genotypes at their optimal economic slaughter endpoint. J. Anim. Sci. 72:38-50.

Johnston, D. J., L. L. Benyshek, J. K. Bertrand, M. H. Johnson, and G. M. Weiss. 1992. Estimates of genetic parameters for growth and carcass traits in Charolais cattle. Can. J. Anim. Sci. 72:493-499.

Lindholm, H. B. and H. H. Stonaker. 1957. Economic importance of traits and selection indexes for beef cattle. J. Anim. Sci. 16:998-1006.

McWhir, J. and J. W. Wilton. 1987. Heritabilities of predicted performance measures in beef bull test and their relationships with feedlot efficiency measures. J. Anim. Sci. 64:1323-1331.

Schaeffer, L. R. and J. C. M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. In: Proc. 5th World Congr. Appl. Livest. Prod. XVIII. Pp. 443-446.

Williams, C. B., and G. L. Bennett. 1995. Application of a computer model to optimum slaughter endpoints for different biological types of feeder cattle. J. Anim. Sci. 73:2903-2915.

CHAPTER II

REVIEW OF LITERATURE

Cattle growth and development

Over the past 50 years, studies on the subject of growth and development have been published in both the scientific literature and in textbooks. Growth has been given many different definitions. Amongst researchers, growth is typically defined as the production of new cells, and in the livestock industry it is usually measured as the increase in mass of an individual over a given time frame (Owens et al., 1993). The latter definition of growth seems to be pleasing to researchers and livestock producers alike, whereas definitions of development (the process of immature individuals emerging into adults) leave a little to be desired. Cleveland (2006), differentiated between growth and development nicely by citing Brody (1945). He referred to development as the combination of growth and environmental factors that lead to an adult individual whereas growth alone refers to the addition of new biochemical units. Growth is the area of development concerned with an increase in living substance. Growth has also been defined as "directed biosynthesis" consisting of two phases. The first phase entails the production of new cells commonly referred to as hyperplasia while the second phase covers the increase in cell size referred to as hypertrophy (Brody, 1945, Owens et al., 1993).

One of the first researchers to study the mathematical behavior of growth was Samuel Brody in 1945 through the development of his now famous growth and aging equations. Brody described the growth of individuals by plotting weight versus age, which resulted in a sigmoidal or s-shaped curve. He stated that a large number of differences in the shape of this curve can be seen between breeds of the same specie or between species. For example, certain species such as humans have a very long interval from weaning to puberty whereas this period is almost absent in laboratory and farm animal species. However, when comparisons are made between the growth curves of individuals within a given population or breed, the shape of each individual curve is quite similar to other individuals in the population.

This "growth age curve" can be broken into two segments. The first segment which typically occurs prior to puberty consists of growth increasing as an increasing rate is aptly named the *self-accelerating* phase of growth. The second growth segment it typically a post-pubertal *self-inhibiting* phase of growth.

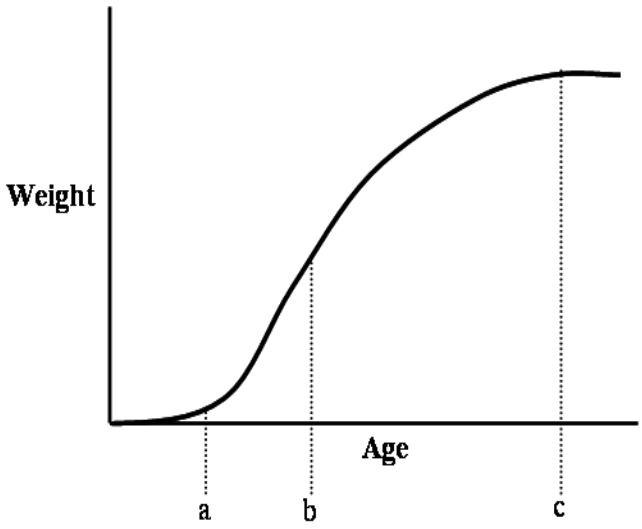


Figure 2.1. Generalized growth curve. Adapted from Owens et al. (1993).

Figure 2.1, which was adapted from Owens et al. (1993) illustrates the important points of this hypothetical curve of weight plotted against age. As is shown, the shape of the curve resulted in the common s-shaped or sigmoidal curve. The points a, b and c represent birth, the curve's inflection point

(puberty) and maturity, respectively. The segment in which the most rapid growth rate, the self-accelerating phase, occurs is between the points a and b, or from birth to puberty. In farm animal species, the point of inflection typically occurs when approximately 30% of the animal's mature weight is reached equating to approximately 6 months of age in cattle (Brody, 1945). Beyond the inflection point is the self inhibiting phase, or the time period in which body mass is still increasing, but at a decreasing rate. The causative factors for this post pubertal deceleration are not well understood despite many hypotheses on the subject (Owens et al., 2003). Point c on the figure represents the animal's mature weight or stage in life where the additional weight is no longer added when food is consumed at will (Marple, 2003).

Mathematical representations of the growth curve

The growth curve shown in Figure 2.1 above can be represented mathematically, as such, many different researchers have presented differing equations on how to do so. Of the many mathematical representations of this growth curve, five equations, ranging from 3 and 4 parameter non-linear models (Brody, Richards, Von Bertalanffy, Gompertz and Logistic) have seen a significant amount of use, with the Brody equation being by far the most used for beef cattle studies (Arango and Van Vleck, 2002). These equations will be presented here, along with their unique features as well as their benefits and drawbacks as summarized by Brown et al. (1976).

Brody (1945) developed the classical empirical equation used to predict body weight. This equation, which is applicable to many different species from mice to farm animals, is represented by the following:

$$W = A\left(1 - Be^{-kt}\right)$$

where body weight (W) at age (t) is a function of mature weight (A) a time scale parameter (B) and the rate at which a logarithmic function of weight changes per unit of time or rate of maturation parameter (k). One of the drawbacks of using this curve is it provides for no inflection point, therefore it fails to denote the break point between the self-accelerating and self-inhibiting phases of growth. As a result, it has been suggested that actual use of this equation be limited to animals who are more than 30% mature (Brody, 1945; DeNise and Brinks, 1985).

Throughout the literature, a number of modifications have been made to this equation presented by Brody. The Richards equation is one of these modified Brody equations that has seen extensive use (Richards, 1959). Richard's equation is represented by the formula:

$$W = A\left(1 - Be^{-kt}\right)^{M}$$

While this equation is quite similar to the Brody equation, the Richard's equation includes an additional shape parameter (M), which allows the modeling of a variable inflection point that represents the age at which puberty occurs. Remaining parameters included in the Richard's equation, are as described during the presentation of the Brody equation above.

The nonlinear growth equations developed by Brody and Richards appear to be the most popular equations used in the livestock industry. Brown et. al, (1976) as well as Fitzhugh Jr. (1976), describe an additional 3 equations that deserve a brief mention here. These equations are referred to as the Von Bertalanffy (Von Bertalanffy, 1957), Gompertz (Winsor, 1932; Laird, 1966) and logistic equations (Nelder 1961, 1962), and

are only mentioned here by name with advantages and disadvantages to using all 5 equations.

Model Comparisons. A number of studies have compared the advantages and disadvantages of using these growth equations in the analysis of beef cattle growth data (Brown et al., 1976; Fitzhugh Jr., 1976; DeNise and Brinks, 1985; Lopez de la Torre et al., 1992; Mezzadra and Miquel, 1994). When comparing the parameters of each of the models, every equation provides for an estimate of mean mature weight (A) and growth rate (K). Given the parameter K, larger values of K, indicate those individuals that are earlier maturing while smaller values are associated with later maturing animals. As mentioned earlier, the Brody equation is the only model which does not provide for some sort of inflection point. The Von Bertalanffy, Gompertz and Logistic models have fixed inflection points relative to mature size, limiting their biological interpretation (Brown et al., 1976). The Richards equation is the only model which allows for a variable inflection point (Arango and Van Vleck, 2002).

Brown et al., (1976) compared the goodness of fit, for the five different models. All five of the models tended to give a poorer description of growth early in live as opposed to later in life. However, the Gompertz and Logistic models overestimated weights taken early in life most severely. The Brody model tended to fit the observed data well after 6 months of age, while the Von Bertalanffy model seemed to overestimate weights at ages prior to 6 months of age. Overall, the Richards model appeared to give an unbiased fit at all ages.

DeNise and Brinks (1985) compared estimated parameters for both Brody and Richards growth curves applied to beef cow growth data. They found that while both

curves modeled similar mature weights (A), Brody's curve was especially dependent on input data. If individuals had missing weight observations, i.e. missing birth or mature weights, Brody's curve fit the data poorly beyond the range of this information. The findings of DeNise and Brinks have been supported in other studies. Both Beltran et al., (1992) and Lopez de la Torre et al., (1992) reported the Richards equation did a better job estimating individual growth curves than did the Brody equation. Advantages and disadvantages of each of the five curves are summarized in Table 2.1 below.

While each of these curves can estimate the growth of an individual animal at any given time, it is often desirable to change the shape of these curves to improve animal populations over time. Fitzhugh Jr. (1976) identified a number of reasons for changing the shape of these curves; each one resulting in an impact on a beef cattle producer's herd

Table 2.1. Advantages and disadvantages of the differing growth curves[1].

Model	Advantage	Disadvantage
Brody	Fit observed data well	Difficulty in estimating weights before six months of age
	Computationally Simple	Difficulty estimating weights outside the range of observed data
Richards	Flexible and accurate due to additional parameters	Computationally Complex
	Generally provides an unbiased fit at all ages	
Von Bertalanffy	Fit reasonably well over all ages	Overestimated weights at early ages prior to 6 months
Gompertz		Overestimated early weights
Logistic		Overestimated early weights Underestimated Mature Weights

[1]Adapted from Cleveland (2006)

dynamics. First, it was necessary to change the shape of the growth curve in order to produce animals that grow quickly to larger sizes while maintaining smaller breeding stock to reduce maintenance costs. The second reason was to improve efficiency through increased maturation rates. Third was reducing incidence of dystocia through decreasing birth weights while maintaining growth. Finally, it was suggested that changing the growth curve was a way to decrease age at puberty thereby increasing fertility and reducing carcass fatness at preferred market weights. Cleveland (2006) cautioned any changes made to the growth curve will result in consequences elsewhere, not just for weight and size. Selection decisions which affect growth need to be weighed against the costs and benefits of altering body composition.

Factors affecting cattle growth and composition

Many different post-weaning measures are economically important to beef cattle producers in the growing and finishing phases of beef cattle production. Given ever increasing feed costs, the main production concerns seem to be maximizing the amount of weight gain, thereby maximizing efficiency while producing a carcass which reaches an optimum harvest endpoint. This section focuses on the growth weight change / carcass composition of post-weaning / finishing cattle. The previous section presented mathematical equations to quantify the change in weight over time throughout the entire life cycle. The following section discusses factors such as breed and biological type, environment, management and sex that can have an influence on the growth of individual animals.

Biological Type. All breeds of cattle are typically classified into two taxonomic categories *Bos Taurus* and *Bos Indicus*. *Bos Taurus* cattle are typical of Europe and North America and are typically adapted to temperate climates. *Bos Indicus* cattle are better suited to hotter climates. Separate biological types exist in both species of beef cattle. The focus of this discussion will be of the differences between "types" and not necessarily between individual breeds.

Swift growth has been a highly sought after trait in all aspects of the beef industry due to its economic importance in determining the number of days an individual animal will be on feed. It has been suggested that the differences observed in growth rate between breed types during the normal growing periods (including the finishing period) will result in animals different in size (Berg and Butterfield, 1976). These differences between breed types have been attributed to the shape of the growth curve which influences how quickly animals will grow. Animals in the pre-inflection growth period will gain more weight as a proportion to an animal's overall body weight than those in the post-inflection growth period (Brody, 1945); therefore animals in the pre-inflective growth period longer will appear to grow at a faster rate than animals that spend more time in the post-inflective growth period.

Average daily gain (**ADG**) or the average weight gain per day on feed has been used as a predictor of an animal's ability to grow swiftly during a given feeding period. Breed and type differences impacting ADG have been reported numerous times. Smith and colleagues (1976) found calves resulting from a cross of Continental (Simmental, Charolais and Limousin) by British (Hereford and Angus) cattle had higher ADG, than British by British crosses during a fixed 180 day feeding period. Calves produced by the

Continental x British cross were generally heavier and had ADG values approximately 10% heavier than the British cross cattle. Similarly, Smith and Rahnefeld (1988) reported similar results for British x Continental crosses fed to a constant number of days. Urick et al. (1991) found differences in ADG between British and Continental sired steers. Steers out of the Continental sires were heavier than their British counterparts with the exception of the calves out of Tarentaise sires. In this study, the Tarentaise sired calves were intermediate to both the Continental and British steers and were not significantly different from either type.

Other studies have come to alternative conclusions concerning ADG than those studies presented above. Anderson et al. (1986) observed no differences between British and Continental sired steers out Hereford dams in post weaning ADG. Conversely, Wyatt et al. (2002) found ADG to be higher in British (Angus) steers than both Continental and Brahman derivatives when fed to a constant fat thickness endpoint. In this study, however, breed did significantly affect final weight. In a separate study, Block and colleagues (2001) found similar results to Wyatt et al. (2002). They determined Continental cross steers to have lower ADG than their British counterparts when fed to a constant back fat. The common theme to the previous two studies was both sets of animals were fed to a constant back fat. When the larger framed Continental animals are fed to an endpoint that takes a longer number of days to reach, they spend a longer period of time in the post-inflective segment of their growth curve where growth, while still increasing, is beginning to slow down (Brody, 1945). Growth occurring during this phase will have the effect of decreasing ADG (Berg and Butterfield, 1976).

When analyzing gain, ADG in this discussion, an important consideration is how efficient the gain was achieved or more specifically how much input was necessary to achieve a certain amount of gain. A breed or type effect does seem to exist for efficiency, but similar to ADG discussed above performance endpoint determines which type of cattle are deemed "more efficient". If performance is measured at a constant age, the larger Continental are more efficient than the smaller British type cattle at converting pounds of feed consumed to pounds of gain (Smith et al., 1976; Urick et al., 1991; Amer et al., 1992). Conversely, when fed to a constant fat endpoint British type cattle are found to be more efficient simply due to the fact they are smaller framed, reach mature size quicker thus begin to deposit fat sooner (Smith et al., 1976; Urick et al., 1991).

Under commercial feeding situations, cattle are penned in groups according to their target endpoint (weight, fat thickness, etc.). ADG and efficiency of the different biological types of cattle are important considerations when marketing various groups on a specific target endpoint because each of the different types of cattle will need a different number of days to reach their target endpoint.

Summarizing the above discussion, given a targeted fat endpoint, British cattle will typically need fewer days to reach a given level of fat thickness than Continental cattle (Block et al., 2001; Wyatt et al, 2002). Similarly, British cattle need fewer days to reach a constant marbling endpoint than do Continental cattle (Wheeler et al., 2004; 2005). However, if the target is a constant age or weight endpoint, Continental cattle will need fewer days to reach this target than will British cattle (Smith et al., 1976; Smith and Rahnefeld, 1988; Urick et al., 1991).

Frame Size. Following closely the discussion of biological type and ADG, beef cattle frame size deserves some discussion given the relationship between the topics. Generally, smaller framed cattle such as Angus and Hereford are considered early maturing breeds meaning they reach puberty and mature size at earlier ages than do larger framed cattle such as Simmental, Charolais or even Brahman. Many of the differences mentioned in the above discussion relative to the effects of breed and biological type on ADG and "efficiency" can be attributed to the differences in frame size of the breeds in the studies. Owens et al. (1993, 1995) alludes to the fact that larger framed cattle will tend to consume greater amounts of feed and reach target endpoints later than smaller framed cattle even though, depending on the endpoint, they can grow more rapidly and efficiently (Smith et al., 1976; Urick et al., 1991; Amer et al., 1992). Many different studies have specifically reported on the influence of cattle frame size on ADG (Cianzio et al., 1982, Tatum et al., 1986a), days on feed (Dolezal et al., 1993), weight (Tatum et al., 1986b; Dolezal et al., 1993), and fat and carcass composition (Cianzio et al., 1982; Tatum et al., 1986c; Dolezal et al., 1993). All studies appear to agree that while larger framed animals grow more quickly, they tend to need a longer time period to reach a constant fat thickness.

Management. Growth performance throughout the post-weaning / finishing phases is greatly influenced by management decisions made by cattle producers and feedlot operators. Management has the ability to make a number of decisions which can either result in changes to the growth curves of individual animals through the use of hormonal implant strategies, different weaning / backgrounding strategies, nutrition, or

by properly matching animals to specific endpoints through the selection of breed and appropriate endpoints for each of the selected breeds (Urick et al., 1991; Wheeler et al., 2004).

Hormonal implants, often considered "metabolic modifiers", are a group of compounds that change an animal's metabolism by altering the manner in which nutrients are absorbed resulting in improved efficiency of production (NRC, 1994). Researchers began studying the effects of hormonal implants in poultry in the 1930s and in cattle in the late 1940s and early 1950s (Hancock et al., 1991). The reason for their use in the cattle industry is because they have been shown to improve growth rates by 10 to 30% and feed efficiency by 5 to 15% (Duckett et al., 1997; Preston, 1999; Montgomery et al., 2001; Nichols et al., 2002). Implants have also been shown to increase ribeye area and improve carcass yield and carcass leanness 5 to 8% (Johnson et al., 1996, Dolezal, 1997; Duckett et al., 1997; Preston 1999; Pritchard, 2000; Roeber et al., 2000; Schneider et al., 2007). The timing of the implant also appears to have an effect on final weight at harvest (Foutz et al., 1997) and on marbling score (Milton et al., 2000; Pritchard, 2000; Bruns et al., 2005), although the effects found by the 3 studies are a bit contradictory. Implants do seem to have an overall negative impact on the ability of an animal to deposit fat. Perhaps this is because they have been shown to increase the mature body weight of steers, thereby causing an increase in the weight for an animal to reach a desired compositional endpoint (Guiroy et al., 2002).

Environment. There are a number of different environmental factors which will influence an animal's production level. Most beef cattle studies have generally looked at

season of the year and the resulting climatic conditions and their effects on ADG, feed intake and feed efficiency. A few studies have also compared these seasonal climatic changes to incidence of sickness or more specifically Bovine Respiratory Disease Complex which will negatively impact any performance traits.

Many studies have concluded the fact that climatic changes have an increase incidence of respiratory tract disease in feedlots. Whether it is daily temperature change during the first 30 days on feed (Alexander et al., 1989), a decrease in mean daily temperature (Ribble et al., 1995), daily ambient temperature range (Cusack et al., 2007; Speidel et al., 2008) or mean daily wind speed (Speidel et al., 2008), as the ambient temperature drops the incidence of respiratory tract disease increases. It has also been shown that a short lag of 2 to 3 days between the climatic event and spike in incidence of sickness could be expected. However, when looking specifically at growth characteristics, Mader (2003) reported increased ADG and decreased dry matter intake and feed efficiency in winter versus summer months. Similarly, Kreikemeier and Mader (2004) reported similar decreases in dry matter intake in heifers during the winter months. Both studies attributed the decrease in intake to daily temperature range, and not just daily minimum temperatures alone.

Live animal and carcass evaluation

Beef cattle carcass evaluation has undergone many changes since the advent of the first voluntary federal grading program in the United States in 1926 (Taylor and Field, 1999) and subsequent establishment of a mandatory federal grading program in the late 1930's and early 1940's (USDA AMS, 1996). Berg and Butterfield (1976) gave a

definition of the ideal carcass as one with "a high proportion of muscle, and acceptable optimum amount of fat and a minimum amount of bone", which is still appropriate in today's market. The only part of this definition which has changed over the years is the "acceptable optimum amount of fat". Recently, the beef industry has seen a shift from the "commodity beef" marketing systems in the past to the advent of value-based systems that place an emphasis on carcass quality and tenderness (Williams, 2002). This change in marketing scheme has caused beef producers to change the way in which they select beef cattle.

The seedstock sector of the beef industry was the first to adopt the use of ultrasound to measure carcass traits in the live animal (Robinson, et al., 1992). Producers which sell bulls for commercial use found it difficult to obtain reliable estimates of their herd's genetic merit for carcass traits given the small number of "cull" animals they have slaughtered. The advent of ultrasound allowed these producers to use non-invasive live animal indicators of their herd's genetic merit for carcass characteristics which enabled them to make more educated selection decisions for their customers.

More recently, given the development of carcass genetic evaluations by breed associations, the importance of the use of ultrasound has shifted slightly. Currently, ultrasound measurements of carcass traits have three uses in the beef industry. First, they are used to add information to the carcass genetic evaluations published by the majority of the beef cattle breed associations. The second and third are described by Lusk et al., (2003) where ultrasound can be used to effectively sort fed cattle into more homogeneous groups to sell at optimum times or to strategically market similar groups of cattle to maximize revenue. The following sections will discuss the definition of the different

ultrasound traits, how they relate to their carcass counterparts and how they can be used to effectively manage groups of cattle.

Ultrasound. Briefly, ultrasound machines consist of a console unit and transducer composed of crystals that emit high frequency sound waves which penetrate the tissues being measured. Once they have entered the tissues, the density of the tissue determines the rate at which the sound waves are bounced back to the transducer. The ultrasound machine then interprets the returned sound waves and converts them into images that can be viewed on the screen of the machine. Considering carcass characteristics of beef cattle, there are currently four traits which are routinely measured in the animal. These traits which include backfat thickness, longissimus muscle area, percentage intramuscular fat and rump fat (Williams, 2002) can be used to estimate the carcass merit of individual animals. More importantly, these estimates of carcass merit can be used to predict characteristics that have more direct economic significance such as yield grade, percent cutability, percent retail product and even the probability of whether or not a carcass will grade choice (Lusk et al., 2003; Walburger and Crews, 2004).

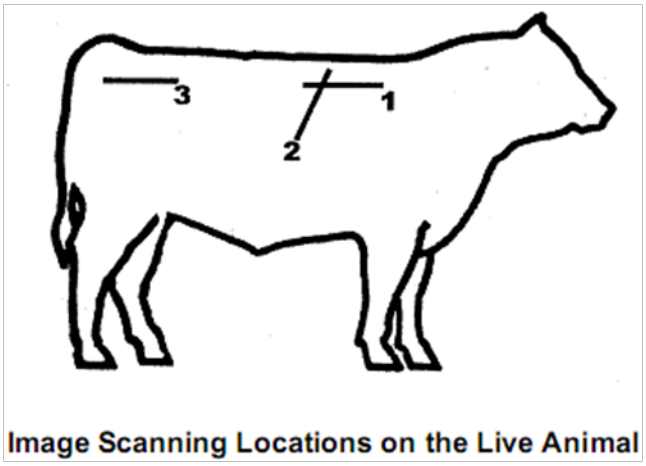


Figure 2.2. Live animal ultrasound image scanning locations. Adapted from Guidelines for Uniform Beef Improvement (2002).

Each of the live animal traits mentioned above are measured differently and each has their own difficulty of measurement. Figure 2.2 illustrates the three different locations in which ultrasound traits are measured in beef cattle which are well described by Williams (2002) and Perkins et al., (2008). Backfat thickness measured at position 2, represents the amount of subcutaneous fat deposited between the 12th and 13th ribs over the longissimus muscle. Of all the traits, this is perhaps the easiest and most accurate to measure due to the fact it is a linear measurement. Longissimus muscle area, another common estimator used in the calculation of the animal's yield grade, is also measured between the 12th and 13th ribs (Position 2). It is a two-dimensional measurement; therefore it is slightly more difficult to measure than the backfat measurement. Rump fat is measured over the rump between the animal's hooks and pins and is used as an indicator of total carcass fat (Position 3). Finally, percent intramuscular fat, an indicator of the animal's marbling score is measured longitudinally over the 11th, 12th and 13th ribs (Position 1). Percent intramuscular fat is probably the most difficult of all the live animal carcass indicators to translate to its carcass counterpart for two reasons. First, it is not a direct measurement like the fat thicknesses and muscle areas of the carcass trait of interest. Percent intramuscular fat is estimated by applying an algorithm to the "backscatter" and "attenuation" of the sound waves (Brethour, 1990, 1991). Second, an increase in live animal percent intramuscular fat is does not represent an equal increase in carcass marbling score (Wilson et al., 1998; Wall et al., 2004) as is evident with reported phenotypic correlations ranging from 0.35 to 0.87 (Wilson, 1992; Perkins et al., 1997).

Relationship to beef cattle carcass traits. The ability of an animal's ultrasound measurement to act as a predictor of its carcass characteristics varies depending on the trait. Numerous studies have looked at the ability of ultrasound measurements to predict their carcass counterparts. Ultrasound backfat seems to do the best job at predicting its carcass component. Simple phenotypic correlation estimates between ultrasound and carcass backfat have been shown to range between 0.76 to 0.93 (Brethour, 1992; Perkins et al., 1992a,b; Perkins et al., 1997; Wall et al., 2004). Even though ultrasound backfat measurements are strongly correlated with their corresponding carcass backfat measurements, ultrasound backfat tends to overestimate carcass backfat in leaner cattle while under estimating backfat in fatter cattle (Brethour, 1992; Robinson et al., 1992; Perkins et al., 1997). Overall, Brethour (1992) found ultrasound backfat measurements to be 8% lower than carcass fat measurements.

As previously mentioned, ultrasound loin muscle area is more difficult to measure than ultrasound backfat due to its two-dimensional nature (Williams, 2002), therefore it is not surprising the correlations between this ultrasonic measurement and carcass loin muscle area are more variable than backfat. Phenotypic correlations between ultrasound and carcass loin muscle areas have been shown to range from 0.43 to 0.95 (Perkins et al., 1992a; Robinson et al., 1992; Smith et al., 1992; Perkins et al., 1997; Wall et al., 2004). Ultrasound loin muscle measurements tend to underestimate carcass loin muscle measurements by an average of 1.7 cm^2 (Perkins et al., 1992b).

Phenotypic correlations between ultrasound percent intramuscular fat and marbling scores are even more variable ranging from 0.35 to 0.87 (Wilson, 1992; Perkins et al., 1997). Perkins et al., (1997) found the correlation between ultrasound percent

intramuscular fat and carcass quality grade to be 0.69. In a comparison of four different systems for measuring intramuscular fat, Herring et al. (1998) found marbling to be overestimated by an average of 1.6675 and 1.075 for percentage ether extractable fat and marbling score converted to percentage ether extractable fat, respectively.

Beef cattle management and ultrasound

Knowledge of live animal characteristics can assist producers in effectively marketing their cattle to increase per head revenues. For example, Schroeder and Graff (2000) reported revenues could be increased by $15.14 / head to $34.74 / head if producers knew before hand the quality and yield grades of their cattle and marketed them appropriately as opposed to simply selling cattle on live-weight, dressed- weight or grid basis. Ultrasound measurements can be used to predict carcass quality and yield grades prior to slaughter, providing a useful tool for sorting, placement, time on feed and marketing strategies (Bergen et al., 1996; Lusk et al., 2003; Walburger and Crews, Jr., 2004).

Two separate studies looked at improving producer marketing decisions (as measured by increased revenue) using ultrasound to predict carcass traits such as the probability of grading choice, yield grade, dressing percent (Lusk et al., 2003) and hot carcass weight, rib eye area, back fat thickness, marbling score and slaughter weight (Walburger and Crews Jr., 2004). Both studies used predicted carcass traits in a simulation study to identify optimum market selection in an effort to increase revenue. Lusk et al. (2003) found ultrasound traits were able to correctly predict choice grade or better 74% of the time and were found to be significant predictors of yield grade and

dressing percent. Using this information, they were able to determine that using ultrasound to strategically market cattle led to increased revenue of $25.07, $4.98 and $32.90 per animal over marketing all animals on live weight, dressed weight or grid basis, respectively. Similarly, Walburger and Crews Jr., (2004) found revenue increases of $11.27 to $27.93 from using both sire and ultrasound data. Lusk et al., (2003) goes on to state "*The most attractive use of ultrasound is to predict when an animal should be slaughtered. By optimally timing cattle, producers can cull low-grading cattle who likely will never grade Choice and can stop feeding higher grade cattle before the marginal benefit of an extra day's feed is greater than the marginal cost*". This is an important statement given that a large proportion of beef cattle are not fed for an appropriate number of days (Brethour, 2000). Brethour (2000) cited a study which found that approximately 25% of beef carcasses had too much back fat for industry standards, an indication they were on feed too long.

Genetic evaluation of longitudinal data

In today's beef industry, many different data types are collected by beef cattle breed associations for the purpose of genetic evaluation. These data points are all biological characteristics of individual animals which can be measured a multitude of times over an animal's lifetime. The number of times a given trait is observed during an animal's life is dictated by the nature of the trait. For example, traits such as carcass characteristics, heifer calving ease, and heifer pregnancy can only be recorded one time on an individual animal. However, traits which monitor the status of an animal as it grows such as weight traits and live animal indicators of carcass merit can be measured a

number of times over the lifespan of an animal. Weight traits such as birth weight and weaning weight describe the same underlying trait, growth as measured by weight gain observed over time. As such, perhaps they can be best described by some type of mathematical function rather than a finite set of data points (Kirkpatrick and Heckman, 1989; Meyer and Kirkpatrick, 2005). As a result, this unique data type has been referred to throughout the literature as "function valued" (Kirkpatrick and Heckman, 1989; Meyer and Kirkpatrick, 2005) or as "infinite-dimensional" or "longitudinal" data by Meyer and Hill (1997).

Longitudinal data. A number of traits currently collected for beef cattle genetic evaluation fall under the umbrella definition of longitudinal data. They can range from commonly collected observations such as weight, height and body condition score measurements to more obscure measures such as feed intake, survival and sperm production and quality (Schaeffer, 2004). Several different methods have been implemented by groups conducting national cattle evaluations to properly model these data types. These methods include more traditional models such as the repeatability and multivariate models, to the more contemporary (and perhaps more appropriate) models such as the suite of random regression models using different base functions (Mrode, 2005).

The analysis of function valued traits is challenging, and each of the different methods has their respective benefits and limitations. Discussion of these benefits and limitations for each of the methods of analysis will be addressed individually beginning with the traditional repeatability model, then move on to the multivariate models and

finally finishing with random regression models that use covariance functions and splines as their base function.

The Repeatability Model. Perhaps the simplest method of analysis of longitudinal data is the "Repeatability Model". The idea behind this model is to treat each observation as a repeated record of the same trait on the same individual. This model has been implemented in the past for traits such as litter size in successive pregnancies in swine and milk yield in successive lactations (Jamrozik et al., 1997b; Interbull, 2000).

The repeatability model is most often described in matrix form by the following (Mrode 2005):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wpe} + \mathbf{e}$$

where **X**, **Z**, and **W** are incidence matrices relating the repeated observations in **y** to fixed (**b**), random additive animal genetic (**u**), and random permanent environmental and non-additive genetic effects (**p**), with **e** defining a vector of random residual errors. The model makes the assumption that the mean of the random effects is zero with variances represented by:

$$\text{var}\begin{bmatrix} \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & 0 & 0 \\ 0 & \mathbf{I}\sigma_p^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

where σ_u^2, σ_p^2, and σ_e^2 are the variances of random additive animal genetic effect, random permanent environmental effect, and random residual error, respectively. In the above **A** is Wright's numerator relationship matrix (Wright, 1922) and **I** is an identity

matrix with an order equal to the number of observations in **y**. The observations in **y** are assumed to have the mean **Xb** and variance equal to $\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma_a^2 + \mathbf{WI}\sigma_p^2\mathbf{W}' + \mathbf{I}\sigma_e^2$.

As can be inferred from the model presented above, the repeatability model makes assumptions on the data structure that do not hold under all situations. Under the assumptions of the repeatability model, observations from the same individual measured at different ages are assumed to have a constant variance and a common correlation with each other (Jennrick and Schluchter, 1986). This assumption of constant variance does not hold where individual variance changes according to the amount of time that has passed between measurements (Meyer and Hill 1997). In the situation where the repeated observations typically follow some type of curve (e.g. growth or lactation curves) correlations between observations taken close together in time are higher than those taken farther apart from one another. In this situation, a more complex model that accounts for the differing correlation structure between successive observations is required.

The Multiple Trait Model. Multivariate genetic evaluation, introduced by Henderson and Quaas (1976), predicts genetic values for multiple traits through the incorporation of genetic and residual covariances among the traits (Mrode, 2005). This property can be extended to the analysis of longitudinal data if differing measurements on an individual animal are treated as separate but genetically correlated traits. It is under this assumption that the current national cattle genetic evaluations for growth are performed. For example, birth weight and weaning weight are observations which are analyzed as separate but genetically correlated traits using a multivariate model even though both are observations of the growth of an individual.

This multivariate model as described by Mrode (2005) is shown in matrix form below.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

In the above set of equations, $\mathbf{y}_i$ is a vector of observations for the i^{th} trait, $\mathbf{b}_i$ is a vector of fixed effects for the i^{th} trait, $\mathbf{u}_i$ and $\mathbf{e}_i$ are vectors of random animal genetic and random residual effects for the i^{th} trait, respectively. $\mathbf{X}_i$ and $\mathbf{Z}_i$ are incidence matrices relating the observations in **y** to the fixed effects in **b** and random animal genetic effects in **u**. As with the above repeatability model, the observations in **y** are assumed to have the mean **Xb**. Random effects in the model are assumed to have means of zero and genetic variances equal to:

$$\text{var}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \sigma^2_{g_1} & \sigma_{g_1,g_2} \\ \sigma_{g_2,g_1} & \sigma^2_{g_2} \end{bmatrix} \otimes \mathbf{A}$$

and residual variances equal to:

$$\text{var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma^2_{e_1} & \mathbf{I}\sigma_{e_1,e_2} \\ \mathbf{I}\sigma_{e_2,e_1} & \mathbf{I}\sigma^2_{e_2} \end{bmatrix}.$$

Above, $\sigma^2_{g_1}$, $\sigma^2_{g_2}$, σ_{g_1,g_2}, and σ_{g_2,g_1} are the additive genetic variance for $\mathbf{y}_1$, $\mathbf{y}_2$ and the additive genetic covariances between $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively. Likewise, $\sigma^2_{e_1}$, $\sigma^2_{e_2}$, σ_{e_1,e_2}, and σ_{e_2,e_1} are the residual error variances for $\mathbf{y}_1$ and $\mathbf{y}_2$ as well as the residual covariances between $\mathbf{y}_1$ and $\mathbf{y}_2$. **A** is Wright's numerator relationship matrix and **I** is an n × n identity matrix.

Henderson and Quaas (1976) were the first to implement the multivariate BLUP model illustrated above in the analysis of a three trait beef cattle example (birth weight, weaning weight and post-weaning gain). Following their work, Schaeffer and Jamrozik (1996) first suggested the use of a multivariate model for the analysis of test day records for milk volume, fat, and protein percentages in dairy cattle. In each of these examples, the observations measured on individuals across time were treated as separate and unique traits that are genetically correlated to one another.

The multivariate model is not without its inherent problems when analyzing longitudinal data. Given the fact longitudinal data can be described using some type of function (Meyer and Kirkpatrick, 2005), they tend to have a large number of data points which are of interest to the individuals performing the data collection. In the multivariate model, this can lead to equation systems which have very high dimension and computational requirements. Considering the test day records discussed by Wiggans and Goddard (1996, 1997) three yield traits (milk volume, fat and protein percentages) over two parity groups (first parity versus later parities) and ten stages of lactation (ten different test days per lactation), analyzing this data using a multivariate model would result in an analysis with 60 different traits.

Another issue with the multivariate model is the potential for high correlations between successive measurements. In beef cattle evaluation, weaning weight and yearling weight are two traits of economic importance, with genetic and phenotypic correlations between these two measurements reported to be 0.78 and 0.72, respectively (Koots, 1994). In the analysis of test day records, the correlations are even higher. Pander et al., (1992) reported milk yield correlations ranging from 0.97 (1 test day apart)

to 0.73 (7 test days apart), with correlations between fat yield and protein yield test day records nearly as high. These elevated correlations are undesirable for two main reasons. First, if two variables predict the same information, it doesn't make sense to include both of the variables in the model. Second, the correlation between the two variables has the effect of reducing the power of the tests of significance (Foster et al., 2006).

The high correlations between traits such as weaning and yearling weights as well as between individual test days in dairy cattle evaluation have resulted in studies designed to determine how to specifically handle these elevated correlations. One method, an extension of the multivariate model, allows higher correlations between observations measured close together than those measured farther apart. This technique, referred to as autoregression or autocorrelation, has been documented in the literature numerous times (Harville, 1979; Kachman and Everett, 1993; Carvalheira et al., 1998). Another method to handle this data type is to model the data using a pre-determined function, or data mean. Referred to as fixed regression (Mrode, 2005), these functions can be extended in such a manner where each individual will have its own random function.

Random Regression. Regression models have been used in the analysis of longitudinal data for many years. The use of pre-determined functions as covariates was introduced as random regression or a random coefficients model during the early to mid 1980's (Henderson, 1982; Laird and Ware, 1982; Jennrich and Schluchter, 1986). However, the first study with application to livestock production data was conducted by Ptak and Schaeffer (1993) in the analysis of test day milk production records of dairy

cattle. This first attempt was not a random regression model, but it accounted for the general shape or mean lactation curve for cows within similar herd, year and season. Following this initial trial, Schaeffer and Dekkers (1994) extended the regression coefficients of this fixed regression model to random animal effects. In doing so, they were able to account for the mean shape of the lactation curve within a given herd, year and season, as well as account for the deviation of each individual animal's lactation curve from this mean shape. They were also able to account for the change in correlation structure of repeated records on individuals over time. This ability of the random regression model to properly account for the changing correlation structure has been shown to result in an increase in prediction accuracy of 5.9% when compared to the multivariate model (Meyer, 2004).

The general form of a random regression model as described by Mrode (2005) can be shown in matrix form as:

$$\mathbf{y = Xb + Qu + Zpe + e}$$

where **y** is a vector of repeated test day yields measured on individual animals, **X** is an incidence matrix relating observations in **y** to fixed effects and fixed regression coefficients, **b** is a vector of solutions for fixed effects and fixed regressions, **Q** is an incidence matrix of covariates relating observations in **y** to random additive genetic regression coefficients, **u** is a vector of random additive direct genetic effects, **Z** is an incidence matrix of covariates relating observations in **y** to permanent environmental random regression coefficients, **pe** is a vector of random permanent environmental regression coefficients for each animal, **e** is a vector of random residuals which includes

the temporary environmental effects for each observation. Variances assumed for this model are:

$$\text{var}\begin{bmatrix} \mathbf{u} \\ \mathbf{pe} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \otimes \mathbf{G} & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{P} & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

where **A** is Wright's numerator relationship matrix, **G** is the (co)variance matrix of the additive genetic random regression coefficients, **I** is an identity matrix whose order is equal to the total number of observations, **P** is the (co)variance matrix of the permanent environmental random regression coefficients, and σ_e^2 is the variance of random residuals.

Worth mentioning is that in some studies, the random residual variance has been allowed to vary (between observations taken in multiple years, for example). Jamrozik et al., (1997a) modified the residual variance structure $\mathbf{I}\sigma_e^2$ presented above to the following:

$$\text{var}[\mathbf{e}] = diag\left\{\sigma_{e_k}^2\right\}$$

where k is equal to the total number of differing residual variances. In this example, the authors used k = 29 resulting in **e** having 29 different values depending on the number of days in milk which ranged from 1 to 305. Perhaps, another more appropriate method for modeling heterogeneous residual variance is to allow the variance to follow a continuous function (Rekaya et al., 2000). Both methods account for changing residual variance structures, and López et al. (2004) found the two methods to be equivalent. It is important to note that if the assumption of homogeneous residual variance does not hold across all stages of production, a modification should then be made to the model which

allows the residual variance to change between those stages of production. Olori et al. (1999) determined the assumption of homogeneity of residual variance will bias the residual variance estimates, leading to over- or under-estimation of heritability values. However, the assumption of homogeneous residual variance has no effect on permanent environmental variance (López-Romero et al., 2003).

Covariance Functions. At approximately the same time the techniques for random regression methodology was being introduced and subsequently implemented, covariance functions were introduced in a series of three papers (Kirkpatrick and Heckman, 1989; Kirkpatrick et al., 1990; Kirkpatrick et al., 1994) with the specific goal of how to account for the changes in the covariance structure between successive observations of longitudinal data. Initial groundwork for the development of the covariance function was first reported by Kirkpatrick and Heckman (1989). They defined the covariance function as the infinite-dimensional counterpart to covariance matrices used in standard multivariate analyses and offered three advantages over the conventional methods. Their three advantages are as follows:

1) Covariance functions have the ability to describe the trait at all points, even if measurements were not taken on specific days, rather than at a finite number of data points;
2) Covariance functions help to reduce errors in calculating the response to selection. Conventional methods only select on a specific age window (for example birth weight or weaning weight), however when selection on a part of the curve is performed, the entire trajectory is changed through the genetic correlation

(selecting on increasing birth weight has the correlated effect of increasing weaning weight). Covariance functions help account for the correlated responses observed at other data points as well;

3) Covariance functions estimate parameters more efficiently due simply to the fact that more data points are used in the analysis.

Kirkpatrick et al. (1990, 1994) provided additional insight into the covariance function they introduced in 1989, with examples using a beef cattle growth data set. Calculating the covariance function begins with the standard classical quantitative genetic (co)variance matrix of the traits in question over different time periods, often referred to as **G** (see the multivariate model presented above). Using a beef cattle growth analysis as an example, the genetic (co)variance matrix (**G**) could consist of the additive genetic variance for birth weight and weaning weight. Using this **G**, covariance functions are built by using a smooth curve to interpolate the values of **G** between the measured ages (birth weight and weaning weight). The process starts with the decision as to which smooth curve to use. Kirkpatrick et al. (1990) suggests the use of Legendre polynomials, but states that any orthogonal function could in fact be used. For longitudinal data such as growth, the authors favored polynomials because growth tends to be smooth similar to the curves created using polynomial functions.

A number of sources illustrate the calculation of Legendre polynomial functions. The equations presented here were adapted from Schaeffer (2003). To calculate Legendre polynomials, first we need to define the polynomials:

$$P_0(x) = 1, \text{ and } P_1(x) = x.$$

Then the additional polynomials can be calculated using the recursive formula:

$$P_{n+1}(x) = \frac{1}{n+1}\left((2n+1)xP_n(x) - nP_{n-1}(x)\right).$$

These values are then normalized using:

$$\phi_n(x) = \left(\frac{2n+1}{2}\right)^{0.5} P_n(x).$$

Table 2.2 below illustrates how a fourth order polynomial would be calculated using the above equations for a normalized Legendre Polynomial.

Table 2.2. Normalized Legendre polynomials for up to a fourth order polynomial.

Order	Legendre Polynomial	Normalized Legendre Polynomial
n = 0	$P_1(x) = x$	$\phi_0(x) = 0.7071$
n = 1	$P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}$	$\phi_1(x) = 1.2247x$
n = 2	$P_3(x) = \frac{5}{2}x^3 - \frac{9}{6}x$	$\phi_2(x) = 2.3717x^2 - 0.7906$
n = 3	$P_4(x) = \frac{35}{8}x^4 - \frac{45}{12}x^2 + \frac{3}{8}$	$\phi_3(x) = 4.6771x^3 - 2.8062x$
n = 4	$P_5(x) = \frac{63}{8}x^5 - \frac{35}{4}x^3 + \frac{15}{8}x$	$\phi_4(x) = 9.2808x^4 - 7.9550x^2 + 0.7955$

This series of normalized polynomials $(\phi_n(x))$ shown in Table 2.2 are then put into a matrix Λ such that:

$$\Lambda' \begin{pmatrix} 0.7071 & 0 & 0 & 0 & 0 \\ 0 & 1.2247 & 0 & 0 & 0 \\ -0.7906 & 0 & 2.3717 & 0 & 0 \\ 0 & -2.8062 & 0 & 4.6771 & 0 \\ 0.7955 & 0 & -7.9550 & 0 & 9.2808 \end{pmatrix}.$$

Legendre polynomials are defined over the interval of -1 to 1 (Kirkpatrick et al., 1990), therefore it is necessary to standardize the ages of the observations to the interval of -1 and 1. The formula used to standardize these ages was presented by Schaeffer (2003) and is defined as follows:

$$t_i^* = -1 + 2\left(\frac{t_i - t_{\min}}{t_{\max} - t_{\min}}\right)$$

where t_i^* is the standardized time, t_i is the time point being standardized, and $t_{\min}$ and $t_{\max}$ were the minimum and maximum time points or ages represented in the dataset, respectively. Standardized time values are placed in to a matrix **M** such that an example standardized age vector $t_i^* = \begin{bmatrix} -1 & -0.25 & 0.25 & 1 \end{bmatrix}^T$ would result in:

$$\mathbf{M} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -0.25 & 0.0625 \\ 1 & 0.25 & 0.0625 \\ 1 & 1 & 1 \end{bmatrix}$$

for a quadratic polynomial. The first column of the matrix is a column of ones representing the intercept of the curve; the second column is the standardized age while the third column is the standardized age squared for the quadratic term. Fitting higher order polynomials is done by the addition of columns for the additional parameters needed. The next step is to combine the standardized ages and the polynomials into a matrix $\Phi = \mathrm{M}\Lambda$. Performing this step with the **M** defined above and the first three rows (quadratic) of A' gives the matrix

$$\mathbf{\Phi} = \begin{bmatrix} 0.7071 & -1.2247 & 1.5811 \\ 0.7071 & -0.30618 & -0.64237 \\ 0.7071 & 0.30618 & -0.64237 \\ 0.7071 & 1.2247 & 1.5811 \end{bmatrix}$$

which when combined with the original genetic (co)variance matrix, using the formula

$$\hat{\mathbf{C}}_G = \mathbf{\Phi}^{-1}\hat{\mathbf{G}}\left[\mathbf{\Phi}^T\right]^{-1}$$

results in an estimated coefficient matrix $\hat{\mathbf{C}}_G$ from which the covariance function can be formed (Kirkpatrick et al., 1990).

The estimated **C** matrix can be used in conjunction with the following covariance function to estimate the covariance between any two measurements taken at any two standardized times denoted t_1 and t_2 (Kirkpatrick and Heckman, 1989; Kirkpatrick et al., 1990; Kirkpatrick et al., 1994):

$$f\left(a_1, a_2\right) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left[\mathbf{C}_G\right]_{ij}\phi_i\left(t_1^*\right)\phi_j\left(t_2^*\right)$$

where $\left[\mathbf{C}_G\right]_{ij}$ is the i^{th} and j^{th} element of the estimated matrix $\hat{\mathbf{C}}_G$, and $\phi_{i(j)}$ is the Legendre polynomial coefficient for the i^{th} age and j^{th} order. The use of this equation is somewhat limited though given phenotypic measurements are typically measured at n ages. Therefore, only an n × n truncated version of $\mathbf{C}_G$ can be used (Kirkpatrick et al., 1990).

The preceding discussion details the formation of a covariance functions for a full order fit, meaning the number of orthogonal functions estimated (k) equals the number of ages measured (n) and is equivalent to the multivariate model (Mrode, 2005). Given a situation where a large number of different ages were measured, meaning n becomes large; the problem becomes intractable rather quickly. Kirkpatrick et al. (1990) determined it possible, and in some cases more attractive, to reduce the order of fit ($k < n$) such that the covariance matrix can be fitted with as few parameters as possible. The reduced order covariance function was found using weighted least squares procedures to

identify the simplest orthogonal function in which the reduced (co)variance matrix was not significantly different from the full (co)variance matrix as determined from a χ^2 goodness of fit test. If the reduced (co)variance matrix differed significantly from the full order matrix, the order of the reduced matrix was increased by using higher order Legendre polynomials until the reduced and full matrices did not differ significantly. According to Kirkpatrick et al. (1990), the reduced estimate is the simplest polynomial that is "statistically consistent" with the data. It also smoothes out the fluctuations caused by the sampling error in the initial measurements used to estimate **G**. The authors do caution, however, that this method will exclude higher order terms even if they actually exist if the data is not powerful enough to show their existence.

Random Regression versus Covariance Functions. Meyer and Hill (1997) were the first to show the equivalence of the random regression model to the covariance function, and then Mrode (2005) illustrated this equivalence through the use of an example. He compared the covariance between breeding values calculated from data recorded on an individual animal using both a parametric curve and a set of orthogonal polynomials fitted in a random regression model. The equality of the covariance function to the random regression model allows the estimation of fewer regression coefficients for each source of variation. When used in random regression models, the matrix Φ replaces the standard covariate incidence matrix.

Recently, some issues have surfaced concerning random regression models which employ the use of Legendre polynomials as their basis function. The estimated covariance matrices used to calculate genetic variances over the range of data (over the

range of lactation for instance) tend to result in genetic variances that are much higher at the beginning and end of the data range than in the middle (Schaeffer & Jamrozik, 2008). Perhaps this is due to the fact that polynomials place a large amount of emphasis on observations at the extremes, which compounds the problem with higher orders of fit (Meyer, 2005a). Other reported problems with Legendre polynomials being used in random regressions are the poor modeling capabilities of asymmetrical functions, their lack of information to estimate a large number of parameters, and their sensitivity to each of the many different (co)variance parameters (Misztal, 2006).

Splines. Given the problems with the use of polynomials as a basis function in random regression models discussed by Misztal (2006) and Schaeffer & Jamrozik (2008), several different alternatives such as fractional polynomials (Robert-Granié et al., 2002), cubic smoothing splines (White et al., 1999), and B-splines (Torres & Quaas, 2001; Meyer, 2005b) have been proposed. Spline functions are defined as piecewise polynomials which join together at "knots" and are continuous across the range of data (Wold, 1974). As a result, they do not suffer from the same problems as polynomials where their behavior in one small area determines their behavior across the entire range of data. Since splines are defined as "piecewise polynomials" they represent smooth curves between each knot.

Ruppert et al. (2003) describes simple spline basis functions as an extension of the following standard simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

where y_i is the observed value of the i[th] trial, x_i is the predictor variable of the i[th] trial, β_1 and β_1 are regression coefficients corresponding to the y intercept and slope of the regression line, respectively, and e_i is the random error term with mean equal to 0 and variance equal to σ_e^2. This model can be easily extended to higher order polynomials through the addition of one more regression coefficient and predictor variable squared such that:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i .$$

The quadratic simple linear regression model presented above would result in an **X** incidence matrix for fitting the regression of:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}.$$

Modification of these models for the inclusion of "knots" or points where the piecewise polynomials join together is a rather simple task accomplished by the addition of K columns of $(x_i - \kappa)_+$ where κ is a specific knot point and "+" refers to the positive section of the function, meaning negatives values of $(x_i - \kappa)$ are included as zero. These values are included in the general simple linear regression in such a manner where:

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} (x - \kappa_k)_+ ,$$

and in the quadratic version of this model as:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{k=1}^{K} b_k (x - \kappa_k)_+^2 .$$

The **X** incidence matrix associated with the above quadratic spline equation would then be modified to be:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & (x_1 - \kappa_1)_+^2 & \cdots & (x_1 - \kappa_K)_+^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & (x_n - \kappa_1)_+^2 & \cdots & (x_n - \kappa_K)_+^2 \end{bmatrix}.$$

These "modified" **X** matrices are then included in the Least Squares normal equations as a substitute for the standard simple linear regression **X** incidence matrices. As a result, standard Least Squares regression statistical properties apply and fitted values can be found by solving the normal equations:

$$\hat{\mathbf{y}} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}.$$

The spline basis functions presented above are referred to as truncated power bases of degree p, and useful for understanding the mechanics of spline based regression. They can be used in practice if knots are carefully chosen or a penalized fit (inclusion of a roughness penalty or a value which penalizes fits which are too rough, resulting in a smoother result) is used (Ruppert et al., 2003). Truncated power base functions are at a disadvantage when it comes to orthogonality, meaning numerical instability can result if too many knots are used and the roughness penalty is too small. It has been suggested that the use of equivalent bases such as B-splines or natural cubic smoothing splines with more stable numerical properties would be desirable (Eilers and Marx, 1996; Ruppert et al., 2003).

Given the piecewise nature of spline bases, some of the problems associated with random regression using Legendre polynomials such as instability at the extremes, seem to be avoided. In 1999, Hill and Brotherstone reported that splines can be included rather

easily in the standard mixed model framework and when compared to random regression models, they include more random effects but require fewer (co)variance parameters. Splines also have the advantage of quicker convergence over Legendre polynomials, which may be due to the fact that spline coefficients are more sparse than their polynomial counterparts (Misztal, 2006). One of the most important questions when using spline bases seems to be related to the number of knots needed to accurately describe the data as well as where to place these knots. The use of too many knots will increase the complexity of the model, while the use of too few will reduce accuracy. Misztal (2006) suggests the following guidelines for choosing proper knot placement:

1) Choose knots in such a manner that they encompass the extremes observed in the data.
2) Choose knots in a way that the correlations between knots is in the range of 0.6 to 0.8.

These two suggestions will result in knots being placed close together around areas that have the largest data density (i.e. birth weight, weaning weight, etc.), and will also result in a larger concentration of knots in areas where the data is changing more rapidly.

Until very recently, use of spline based regression techniques by quantitative geneticists in the livestock industry had been almost non-existent. Spline basis functions have been used in the analysis of a number of traits, and as with the random regression and covariance function models, they were first proposed for the analysis of dairy cattle test day records. They have been incorporated into fixed regressions to model the lactation curve in the analysis of dairy cattle test day records (Druet et al., 2003), as well as the modeling of curves for estimated breeding values (White et al., 1999).

The use of splines for the analysis of beef cattle data seems, so far, limited to the analysis of growth traits. Meyer (2005b) used quadratic B-splines to analyze Angus growth data from birth to 820 days of age with knots at 0, 200, 400, 600 and 821 days of age. She found that the B-splines lend themselves well to the modeling of growth data, but they tend to be susceptible to irregularities in the distribution and sparseness of the data. Using simulated beef cattle growth data, Bohmanova et al. (2005) found that despite the fact that splines are simpler with fewer parameters than Legendre polynomials, they are just as accurate (within 0.2%). A series of studies was conducted in 2005 and 2006 investigating the feasibility of using spline basis functions in random regression models with the application to large scale genetic evaluations (Iwaisaki et al., 2005; Robbins et al., 2005; Bertrand et al., 2006). In this set of studies, it was determined that random regression using spline bases is a feasible alternative to random regression with Legendre polynomial bases as well as the more contemporary multivariate model.

Prediction of days to finish

The idea of reducing the number of required days for livestock to reach their desired endpoint (which will differ depending on the species of interest) is not new. In 1957, Lindholm and Stonaker recognized the importance of reducing the number of days cattle are on feed and were able to quantify on a phenotypic level the value of reducing the number of days it takes to "finish" cattle in the feedlot. They reported a phenotypic correlation of -0.46 between the number of days taken to reach a perceived quality grade and net income per 100 lbs of slaughter weight. Jumping ahead in time nearly 50 years, a group beef cattle researchers coined the term "Economically Relevant Traits" (traits

which directly affect profitability by either increasing revenue or by reducing costs), identified three "days to" traits (days to target fat thickness, slaughter weight, and finish or carcass grade) as being economically relevant (Golden et al., 2000).

From 1957 to present time, very little research has been conducted with regards to these "days to" traits for most species of livestock. McWhir and Wilton (1987) analyzed beef cattle days to market finish (7 mm subcutaneous back fat depth, determined ultrasonically) and reported a heritability estimate for this trait of 0.65 ± 0.42. When this trait was adjusted to a constant market weight, the authors found that the heritability increased to 0.90 ± 0.48. Following McWhir and Wilton, Johnston et al., (1992) estimated heritability for the number of days to a constant backfat (8.9 mm) endpoint in Charolais cattle to be 0.24. Surprisingly, they reported negative correlations between the days to finish trait and the growth traits, while the number of days to reach 8.9 mm of backfat was positively correlated with the carcass traits carcass weight, longissimus muscle area and marbling score.

Contrary to this absence of "days to" research in the beef industry, the swine industry has embraced the concept, most notably as a way to reduce the number of days to finish weight ultimately improving economic efficiency (Faust et al., 1992). Within herd evaluations of days to 105 kg were first reported in the swine industry in 1986 (Stewart et al., 1991), and have been conducted since. According to the Swine Testing and Genetic Evaluation System, the swine industry currently uses three genetic predictions for days to finish (STAGES, 2006). The first of these predictions is the days to market weight EPD (currently 113 kg), which over the years has increased somewhat since 1986. This days to weight is the only published genetic prediction. The remaining

two, days to weaning (a linear combination of calculated genetic predictions) and days from weaning to breeding, are included in the economic selection indices sow productivity index and maternal line index (Harris and Newman, 1994; STAGES, 2006).

Given the range of heritability estimates for the number of days required to reach a weight endpoint ranging from 0.26 to 0.69 (Table 2.3), sufficient genetic variation exists which has allowed swine producers to achieve a positive genetic trend in decreasing the number of days required to reach a weight constant endpoint. According to STAGES (2006), who conducts the days to 250 lb (113 kg) genetic evaluation for the Duroc, Hampshire, Yorkshire and Landrace breeds, the average genetic trend for these four breeds from 1986 to 2008 has been -0.34, -0.18, -0.36, and -0.36 days per year for the Duroc, Hampshire, Yorkshire and Landrace breeds, respectively.

Table 2.3. Published heritability estimates for days to finish weight endpoints in swine[1].

Trait	Breed	Estimate[2]	Source
Days to 90 kg	Duroc	0.27	Kennedy et al., 1985
	Hampshire	0.46	
	Yorkshire	0.36	
	Landrace	0.40	
Days to 91 kg	Yorkshire and Duroc	0.18 ± 0.14	Bereskin, 1987
Days to 100 kg	Duroc	0.25 ± 0.01	Keele et al., 1988
	Hampshire	0.11 ± 0.05	
	Yorkshire	0.22 ± 0.04	
	Pooled	0.22 ± 0.01	
Days to 110 kg	Large White	0.26	Kaplon et al., 1991
Days to 113 kg	Duroc	0.69 ± 0.12	Newcom et al., 2005
	All Breeds	0.25	NSIF (2002)

[1]Adapted from Kuehn (2000) and Cleveland (2006)

[2]h^2 ± SE

More recent work on a days to finish endpoint genetic prediction has been conducted at Colorado State University. Kuehn (2000) in a simulation study looked at the feasibility of using random regression models which only included intercept and linear coefficients for days to produce genetic evaluations for days to finish weight and days to finish backfat. He determined an average of 2.5 observations per animal were required to successfully obtain accurate estimates of variance components, but if covariance components are estimated as well, the use of a dataset with at least 5000 records would be advised. Jubileu (2003) compared differing methods of evaluation for a days to weight endpoint using a Simmental field dataset. He looked at the differences between using more traditional approaches such as univariate and multivariate models versus random regression techniques to conduct a days to weight prediction. He was able to conduct a days to finish weight genetic prediction using random regression methodologies.

Calculating days to finish endpoint genetic predictions using random regression seems to have certain advantages over the traditional univariate and multivariate approaches. Besides the statistical advantages mentioned earlier in this section, random regression produces EBV for the regression curves, meaning an EBV can be calculated for any age or any number of days on feed. Kuehn (2000) presented an equation for calculation of any customized EBV as is shown below:

$$EBV(\text{age or weight}) = b_0 + b_1 * (\text{desired endpoint})$$

where b_0 is the EBV for the intercept and b_1 is the linear EBV for each individual sire. EBV resulting from a random regression model are not equivalent to the EBV from the more traditional models, and both Kuehn (2000) and Jubileu (2003) recognize the

possible confusion these predictions can cause, especially if higher order polynomials are used. Therefore, it is necessary to publish these predictions using some sort of decision support system such as is proposed by Cleveland (2006).

Literature cited

Alexander, B. H., D. W. MacVean and M. D. Salman. 1989. Risk factors for lower respiratory tract disease in a cohort of feedlot cattle. J. Amer. Vet. Med. Assn. 195:207-211.

Amer, P. R., R. A. Kemp, and C. Smith. 1992. Genetic differences among the predominant beef cattle breeds in Canada: An analysis of published results. Can. J. Anim. Sci. 72:759-771.

Anderson, D. C., D. D. Kress, D. E. Doornbos and T. J. Lawlor. 1986. Postweaning traits of biological types raised on foundation Hereford cows. Montana Res. 3:3.

Arango, J. A. and L. D. Van Vleck. 2002. Size of beef cows: early ideas, new developments. Genet. Mol. Res. 1:51-63.

Beef Improvement Federation. 2002. Guidelines for Uniform Beef Improvement Programs. 8th ed. Omaha, NE.

Beltran, J. J., W. T. Butts Jr., T. A. Olson and M. Koger. 1992. Growth patterns of two lines of Angus cattle selected using predicted growth parameters. J. Anim. Sci. 70:734-741.

Bereskin, B. 1987. Genetic and phenotypic parameters for pig growth and body composition estimated by intraclass correlation and parent-offspring regression. J. Anim. Sci. 64:1619-1629.

Berg, R. T. and R. M. Butterfield. 1976. New concepts of cattle growth. John Wiley and Sons, New York.

Bergen, R. D., J. J. McKinnon, D. A. Christensen and N. Kohle. 1996. Prediction of lean yield in yearling bulls using real-time ultrasound. Can. J. Anim. Sci. 75:305-311.

Bertrand, J. K., I. Misztal, K. R. Robbins, J. Bohmanova, and S. Tsuruta. 2006. Implementation of random regression models for large scale evaluations for growth in beef cattle. In: Proc. 8th World Congr. Appl. Livest. Prod. Belo Horizonte, Brazil.

Bohmanova, J., I. Misztal, and J. K. Bertrand. 2005. Studies on muyltiple trait and random regression models for genetic evaluation of beef cattle for growth. J. Anim. Sci. 83:62-67.

Brethour, J. R. 1990. Relationship of ultrasound speckle to marbling score in cattle. J. Anim. Sci. 68:2603-2613.

Brethour, J. R. 1991. Estimating quality grade in cattle with ultrasound. Kansas Agric. Exp. Stn., Prog. Rep. No. 627, Manhattan, KS

Brethour, J. R. 1992. The repeatability and accuracy of ultrasound in measuring backfat of cattle. J. Anim. Sci. 70:1039-1044.

Brethour, J. R.. 2000. Using serial ultrasound measures to generate models of marbling and backfat thickness changes in feedlot cattle. J. Anim. Sci. 78:2055-2061.

Brody, S. 1945. Bioenergetics and growth. Reinhold Publishing Corporation, New York.

Brown, J. E., H. A. Fitzhugh Jr., and T. C. Cartwright. 1976. A comparison of nonlinear models for describing weight-age relationships in cattle. J. Anim. Sci. 42:810-818.

Bruns, K. W., R. H. Pritchard and D. L. Boggs. 2005. The effect of stage of growth and implant exposure on performance and carcass composition in steers. J. Anim. Sci. 83:108-116.

Carvalheira, J. G. V., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle in United States Holstein herds. J. Dairy Sci. 81:2738-2751.

Cianzio, D. S., D. G. topel, G. B. Whitehurst, D. C. Beitz and H. L. Self. 1982. Adipose tissue growth in cattle representing two frame sizes: Distribution among depots. J. Anim. Sci. 55:305-312.

Cleveland, M. A. 2006. Sire selection for profit in beef production. Ph D Dissertation. Colorado State University, Fort Collins, CO.

Cusack, P. M. V., N. P. McMeniman and I. J. Lean. 2007. Feedlot entry characteristics and climate: Their relationship with cattle growth rate, bovine respiratory disease and mortality. Aust. Vet. J. 85:311-316.

DeNise, R. S. K. and J. S. Brinks. 1985. Genetic and environmental aspects of the growth curve parameters in beef cows. J. Anim. Sci. 61:1431-1440.

Dolezal, H. G. 1997. Impact of implants on carcass yield grade and cutability. Pages 155-166 in Proc. OSU Symposium: Impact of Implants on Performance and Carcass Value of Beef Cattle, Oklahoma Ag. Exp. Sta. P-957. Stillwater, OK.

Dolezal, H. G., J. D. Tatum and F. L. Williams Jr. 1993. Effects of feeder cattle frame size, muscle thickness and age class on days fed, weight and carcass composition. J. Anim. Sci. 71:2975-2985.

Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test day records of French Holstein cows. J. Dairy. Sci. 86:2480-2490.

Duckett, S. K., F. N. Owens, and J. G. Andrae. 1997. Effects of implants on performance and carcass traits of feedlot steers and heifers. Pages 63-82 in Proc. OSU Symposium: Impact of Implants on Performance and Carcass Value of Beef Cattle, Oklahoma Ag. Exp. Sta. P-957. Stillwater, OK.

Eilers, P. H. C., and B. D. Marx. 1996. Flexible smoothing with B-splines and penalties. Statistical Science. 11:89-121.

Faust, M. A., M. W. Tess, and O. W. Robison. 1992. A bioeconomic simulation model for a hierarchical swine breeding structure. J. Anim. Sci. 70:1760-1774.

Fitzhugh Jr., H. A. 1976. Analysis of growth curves and strategies for altering their shape. J. Anim. Sci. 42:1036-1051.

Foutz, C. P., H. G. Dolezal, T. L. Gardner, D. R. Gill, J. L. Hensley and J. B. Morgan. 1997. Anabolic implant effects on steer performance, carcass traits, subprimal yields and longissimus muscle properties. J. Anim. Sci. 75:1256-1265.

Golden, B. L., D. J. Garrick, S. Newman, and R. M. Enns. 2000. Economically relevant traits, a framework for the next generation of EPDs. Proc. Beef Improvement Federation, Wichita, KS, pp. 2-13.

Guiroy, P. J., L. O. Tedeschi, D. G. Fox, and J. P. Hutcheson. 2002. The effects of implant strategy on finished body weight of beef cattle. J. Anim. Sci. 80:1791-1800.

Hancock, D. L., J. F. Wagner, and D. B. Anderson. 1991. Effects of estrogens and androgens on animal growth. Pages 255-197 in Growth Regulation in Farm Animals: Advances in Meat Research. Volume 7. A. M. Pearson and T. R. Dutson, eds. Elsevier Appl. Sci., New York, NY.

Harris, D. L., and S. Newman. 1994. Breeding for profit: Synergism between genetic improvement and livestock production (A Review). J. Anim. Sci. 72:2178-2200.

Harville, D. A. 1979. Recursive estimation using mixed linear model with autoregressive random effects. In: Variance Components and Animal Breeding. Proc. Conf. In honor of C. R. Henderson. Pp. 157-179. Cornell Univ., Ithaca, NY.

Henderson, C. R. and R. L. Quaas. 1976. Multiple trait evaluation using relatives' records. J. Anim. Sci. 43:1188-1197.

Henderson, C. R. 1982. Analysis of covariance in the mixed model: higher-level nonhomogeneous and random regressions. Biometrics. 38:623-640.

Herring, W. O., L. A. Kriese, J. K. Bertrand and J. Crouch. 1998. Comparison of four real-time ultrasound systems that predict intramuscular fat in beef cattle. J. Anim. Sci. 76:364-370.

Hill, W. G. and S. Brotherstone. 1999. Advances in methodology for utilizing sequential records. British Soc. Anim. Sci. Occaisonal Pub. 24:55-61.

Interbull. 2000. National Genetic Evaluation Programmes for Dairy Production Traits Practised in Interbull Member Countries 1999-2000. Department of Animal Breeding and Genetics, Uppsala, Sweden, Bulletin 24.

Iwaisaki, H., S. Tsuruta, I. Misztal, and J. K. Bertrand. 2005. Genetic parameters estimated with multitrait and linear spline-random regression models using Gelbvieh early growth data. J. Anim. Sci. 83:757-763.

Jamrozik, J., G. J. Kistemaker, J. C. M. Dekkers, and L. R. Schaeffer. 1997. Comparison of possible covariates for use in a random regression model for analyses of test day yields. J. Dairy. Sci. 80:2550-2556.

Jamrozik, J., L. R. Schaeffer, and J. C. M. Dekkers. 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy Sci. 80:1217-1226.

Jennrich, R. I., and M. D. Schluchter. 1986. Unbalanced repeated-measures models with structured covariance matrices. Biometrics. 42:805-820.

Johnson, B. J., M. R. Hathaway, P. T. Anderson, J. C. Meiske, and W. R. Dayton. 1996a. Stimulation of circulating insulin-like growth factor I (IGF-I) and insulin-like growth factor binding proteins (IGFBP) due to administration of a combined trenbolone acetate and estradiol implant in feedlot cattle. J. Anim. Sci. 74:372-379.

Johnston, D. J., L. L. Benyshek, J. K. Bertrand, M. H. Johnson, and G. M. Weiss. 1992. Estimates of genetic parameters for growth and carcass traits in Charolais cattle. Can. J. Anim. Sci. 72:493-499.

Jublieu, J. S. 2003. The use of random regression models to predict days to finish in beef cattle. Master's thesis. Colorado State University, Fort Collins, CO.

Kachman, S. D., and R. W. Everett. 1993. A multiplicative mixed model when the variances are heterogeneous. J. Dairy. Sci. 76:859-867.

Kaplon, M. J., M. F. Rothschild, P. J. Berger, and M. Healey. 1991. Population parameter estimates for performance and reproductive traits in polish large white nucleus herds. J. Anim. Sci. 69:91-98.

Keele, J. W., R. K. Johnson, L. D. Young and T. E. Socha. 1988. Comparison of methods of predicting breeding values of swine. J. Anim. Sci. 66:3040-3048.

Kennedy, B. W., K. Johansson, and G. F. S. Hudson. 1985. Heritabilities and genetic correlations for backfat and age at 90 kg in performance tested pigs. J. Anim. Sci. 61:78-82.

Kirkpatrick, M. and N. Heckman. 1989. A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. J. Math Biol. 27:429-450.

Kirkpatrick. M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124:979-993.

Kirkpatrick, M., W. G. Hill, and R. Thompson. 1994. Estimating the covariance structure during growth and ageing, illustrated with lactation in dairy cattle. Genet. Res. Camb. 64:57-69.

Koots, K. P., J. P. Gibson and J. W. Wilton. 1994. Analyses of published genetic parameter estimates for beef production traits. 2. Phenotypic and genetic correlations. Anim. Breed. Abstr. 62:825-853.

Kreikemeier, W. M. and T. L. Mader. 2004. Effects of growth promoting agents and season on yearling feedlot heifer performance. J. Anim. Sci. 82:2481-2488.

Kuehn, L. A. 2000. Parameterization of random regression models for beef cattle data sets. Master's thesis. Colorado State University, Fort Collins, CO.

Laird, A. K. 1966. Post natal growth of birds and mammals. Growth. 30:349-363.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. Biometrics. 38:963-974.

Lindholm, H. B. and H. H. Stonaker. 1957. Economic importance of traits and selection indexes for beef cattle. J. Anim. Sci. 16:998-1006.

Lopez de la Torre, G., J. J. Candotti, A. Reverter, M. M. Bellido, P. Vasco, L. J. Garcia, and J. S. Brinks. 1992. Effects of growth parameters on cow efficiency. J. Anim. Sci. 70:2668-2672.

López-Romero, P., R. Rekaya, and M. J. Carabaño. 2003. Assessment of homogeneity vs. heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. J. Dairy. Sci. 86:3374-3385.

López-Romero, P., R. Rekaya, and M. J. Carabaño. 2004. Bayesian comparison of test-day models under different assumptions of heterogeneity for the residual variance: the change point technique versus arbitrary intervals. J. Anim. Breed. Genet. 121:14-25.

Lusk, J. L., R. Little, A. Williams, J. Anderson and B. McKinley. 2003. Utilizing ultrasound technology to improve livestock marketing decisions. Rev. Ag. Econ. 25:203-217.

Mader, T. L. 2003. Environmental stress in confined beef cattle. J. Anim. Sci. 81:E110-119.

Marple, D. 2003. Fundamental Concepts of growth. In: C. G. Scanes (Ed.) Biology of growth of domestic animals. Pp. 9-19. Iowa State Press, Ames, Iowa.

McWhir, J. and J. W. Wilton. 1987. Heritabilities of predicted performance measures in beef bull test and their relationships with feedlot efficiency measures. J. Anim. Sci. 64:1323-1331.

Meyer, K. 2004. Scope for a random regression model in genetic evaluation of beef cattle for growth. Livest. Prod. Sci. 89:69-83.

Meyer, K. 2005a. Advances in methodology for random regression analyses. Aust. J. Exp. Agric. 45:847-858.

Meyer, K. 2005b. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel. Evol. 37:473-500.

Meyer, K. and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. Livest. Prod. Sci. 47:185-200.

Meyer, K. and M. Kirkpatrick. 2005. Up hill, down dale: quantitative genetics of curvaceous traits. Phil. Trans. R. Soc. B. 360:1443-1455.

Mezzadra, C. A. and M. C. Miquel. 1994. Heterosis and breed transmitted effects in growth curve parameters in Angus, Criollo and reciprocal crossbred cows. Proc. 5th World Cong. Genet. Appl. Livest. Prod. Guelph, CA. 17:276-279.

Milton, T., R. Cooper, D. J. Jordan, and F. Prouty. 2000. Delayed implant strategies using Synovex Plus on performance and carcass characteristics in finishing

yearling steers. Pages 47-50 in Nebraska Beef Rep., University of Nebraska, Lincoln.

Misztal, I. 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet. 123:74-80.

Montgomery, T. H., P. F. Dew, and M. S. Brown. 2001. Optimizing carcass value and the use of anabolic implants in beef cattle. J. Anim. Sci. 79:E296-E306.

Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CABI Publishing Company. Cambridge, MA.

Nelder, J. A. 1961. The fitting of a generalization of the logistic curve. Biometrics. 17:89-110.

Nelder, J. A. 1962. An alternative form of a generalized logistic equation. Biometrics. 18:614-616.

Newcom, D. W., T. J. Baas, C. R. Schwab, and K. J. Stalder. 2005. Genetic and phenotypic relationships between individual subcutaneous backfat layers and percentage of longissimus intramuscular fat in Duroc swine. J. Anim. Sci. 83:316-323.

Nichols, W. T., M. L. Galyean, D. U. Thomson, and J. P. Hutcheson. 2002. Review: Effects of implants in the tenderness of beef. Prof. Anim. Sci. 18:202-210.

NRC. 1994. Metabolic Modifiers: Effects on the nutrient requirements of food-producing animals. National Academy Press, Washington, DC.

NSIF. 2002. Guidelines for uniform swine improvement programs. [serial online]. Available at: http://www.nsif.com/ Accessed Jan. 15. 2009.

Olori, V. E., W. G. Hill, and S. Brotherstone. 1999. The structure of the residual error variance of test day milk yield in random regression models. Workshop on Computational Breeding, Finland. March, 1999. Pp. 18-20.

Owens, F. N., P. Dubeski, and C. F. Hanson. 1993. Factors that alter the growth and development of Ruminants. J. Anim. Sci., 71:3138-3150.

Owens, F. N., D. R. Gill, D. S. Secrist and S. W. Coleman. 1995. Review of some aspects of growth and development of feedlot cattle. J. Anim. Sci. 73:3152-3172.

Pander, B. L., W. G. Hill, and R. Thompson. 1992. Genetic parameters of test day records of British Holstein-Friesian heifers. Anim. Prod. 55:11-21.

Perkins, T. L., R. D. Green and K. E. Hamlin. 1992a. Evaluation of ultrasonic estimates of carcass fat and longissimus muscle area in beef cattle. J. Anim. Sci. 70:1002-1010.

Perkins, T. L., R. D. Green, K. E. Hamlin, H. H. Shepard and M. F. Miller. 1992b. Ultrasonic prediction of carcass merit in beef cattle: Evaluation of technician effects on ultrasonic estimates of carcass fat thickness and longissimus muscle area. J. Anim. Sci. 70:2758-2765.

Perkins, T. L., J. C. Paschal, N. C. Tipton and M. J. De La Zerda. 1997. Ultrasonic prediction of quality grade and percent retail cuts in beef cattle. J. Anim. Sci. 75(Suppl.1):178 (Abstr.).

Perkins, T., A. Meadows and B. Hays. 2008. Study guide for the ultrasonic evaluation of beef cattle for carcass merit. Ultrasound Guidelines Council Study Guide Sub-Committee. http://www.designergenesusa.com/UGCSTUDYGUIDE.pdf . Accessed Nov. 07, 2008.

Preston, R. L. 1999. Hormone containing growth promoting implants in farmed livestock. Adv. Drug Del. Rev. 38:123-138.

Pritchard, R. H. 2000. Comparison of estradiol-trenbolone acetate implant programs for yearling steers of two genotypes. Pages 44-52 in South Dakota Beef Res. Rep. 00-10, Brookings.

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.

Rekaya, R., M. J. Carabaño, and M. A. Toro. 2000. Assessment of heterogeneity of residual variances using change point techniques. Genet. Sel. Evol. 32:383-394.

Ribble, C. S., A. H. Meek, E. D. Janzen, P. T. Guichon and G. K. Jim. 1995. Effect of time of year, weather and the pattern of auction market sales on fatal fibrinous pneumonia (Shipping Fever) in calves in a large feedlot in Alberta (1985-1988). Can. J. Vet. Res. 59:167-172

Richards, F. J. 1959. A flexible growth function for empirical use. J. Exp. Bot. 10:290-300.

Robbins, K. R., I. Misztal, and J. K. Bertrand. 2005. A practical longitudinal model for evaluating growth in Gelbvieh cattle. J. Anim. Sci. 83:29-33.

Robert-Granié C., E. Maza, R. Rupp, and J. L. Foulley. (2002). Use of fractional polynomial for modeling somatic cell scores in dairy cattle. In: Proc. 7th World Congr. Appl. Livest. Prod. CD-ROM, comm.. no 16-05, Montpellier, France.

Robinson, D. L., C. A. McDonald, K. Hammond and J. W. Turner. 1992. Live animal measurement of carcass traits by ultrasound: Assessment and accuracy of sonographers. J. Anim. Sci. 70:1667-1676.

Roeber, D. L., R. C. Cannell, K. E. Belk, R. K. Miller, J. D. Tatum, and G. C. Smith. 2000. Implant strategies during feeding: impact on carcass grades and consumer acceptability. J. Anim. Sci. 78:1867-1874.

Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. Cambridge series in statistical and probabilistic mathematics: Semiparametric Regression. Cambridge University Press. New York, NY.

Schaeffer, L. R. 2003. Random regression models. ANSC637 Course Notes – Quantitative genetics and animal models. Available at: http://www.aps.uoguelph.ca/~lrs/ANSC637/LRS14/LRS14.pdf. Accessed Jan. 8, 2009.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45.

Schaeffer, L. R. and J. C. M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. In: Proc. 5th World Congr. Appl. Livest. Prod. XVIII. Pp. 443-446.

Schaeffer, L. R. and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125:145-146.

Schneider, B. A., J. D. Tatum, T. E. Engle, and T. C. Bryant. 2007. Effects of heifer finishing implants on beef carcass traits and longissimus tenderness. J. Anim. Sci. 85:2019-2030.

Schroeder, T. C. and J. L. Graff. 2000. Estimated value of increased pricing accuracy for fed cattle. Rev. Agr. Econ. 22(Spring/Summer 2000):89-101.

Smith, E. G. and G. W. Rahnefeld. 1988. Economic value of feeder calves between breed crosses. Can. J. Anim. Sci. 68:1041-1050.

Smith, G. M., D. B. Laster, L. V. Cundiff and K. E. Gregory. 1976. Characterization of biological types of cattle II. Post weaning growth and feed efficiency of steers. J. Anim. Sci. 43:37-47.

Smith, M. T., J. W. Oltjen, H. G. Dolezal, D. R. Gill and B. D. Behrens. 1992. Evaluation of ultrasound for prediction of carcass fat thickness and longissimus muscle area in feedlot steers. J. Anim. Sci. 70:1667-1676.

Speidel, S. E., R. M. Enns, G. H. Loneragan, R. L. Weaber, H. Van Campen, J. L. Salak-Johnson, C. C. L. Chase, J. J. Wagner, and E. J. Pollak. 2008. Effect of daily ambient temperature and wind speed on sickness of feedlot cattle. J. Anim. Sci., 86: Suppl 2 (Abstract): 593.

STAGES. 2006. Swine Testing and Genetic Evaluation System. Available at: http://www.ansc.purdue.edu/stages. Accessed Jan. 15, 2009.

Stewart, T. S., D. L. Lofgren, D. L. Harris, M. E. Einstein, and A. P. Schinckel. 1991. Genetic improvement programs in livestodk: swine testing and genetic evaluation system (STAGES). J. Anim. Sci. 69:3882-3890.

Tatum, J. D., F. L. Williams, Jr., and R. A. Bowling. 1986a. Effects of feeder cattle frame size and muscle thickness on subsequent growth and carcass development. I. An objective analysis of frame size and muscle thickness. J. Anim. Sci. 62:109-120.

Tatum, J. D., H. G. Dolezal, F. L. Williams, Jr., R. A. Bowling and R. E. Taylor. 1986b. Effects of feeder cattle frame size and muscle thickness on subsequent growth and carcass development. II. Absolute growth and associated changes in carcass composition. J. Anim. Sci. 62:121-131.

Tatum, J. D., F. L. Williams, Jr., and R. A. Bowling. 1986c. Effects of feeder cattle frame size and muscle thickness on subsequent growth and carcass development. III. Partitioning of separable carcass fat. J. Anim. Sci. 62:132-138.

Taylor, R. E. and T. G. Field. 1999. Beef production and management decisions. Third Edition. Prentice Hall, Inc. Upper Saddle River, New Jersey.

Torres, R. A., and R. L. Quaas. 2001. Determination of covariance functions for lactation traits on dairy cattle using random-coefficient regressions on B-splines. J. Anim. Sci., 79, Suppl. 1, 112, abstract.

Urick, J. J., M. D. MacNeil and W. L. Reynolds. 1991. Biological type effects on postweaning growth, feed efficiency and carcass characteristics of steers. J. Anim. Sci. 69:490-497.

USDA Agricultural Marketing Service. 1996. United States standards for grades of slaughter cattle. USDA-AMS, Washington, DC.

Von Bertalanffy, L., 1957. Quantitative laws in metabolism and growth. Quart. Rev. Biol. 32:217-231.

Walburger, A. M. and D. H. Crews, Jr. 2004. Improving market selection for fed beef cattle: The value of real time ultrasound and relations data. Can. J. Ag. Econ. 52:1-16.

Wall, P. B., G. H. Rouse, D. E. Wilson, R. G. Tait, Jr., and W. D. Busby. 2004. Use of ultrasound to predict body composition changes in steers at 100 and 65 days before slaughter. J. Anim. Sci. 82:1621-1629.

Wheeler, T. L., L. V. Cundiff, S. D. Shackelford and M. Koohmaraie. 2004. Characterization of biological types of cattle (Cycle VI): Carcass, yield, and longissimus palatability traits. J. Anim. Sci. 82:1177-1189.

Wheeler, T. L., L. V. Cundiff, S. D. Shackelford and M. Koohmaraie. 2005. Characterization of biological types of cattle (Cycle VII): Carcass, yield, and longissimus palatability traits. J. Anim. Sci. 83:196-207.

White, I. M., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy. Sci. 82:632-638.

Wiggans, G. R. and M. E. Goddard. 1996. A computationally feasible test day model with separate first and later lactation genetic effects. Proc. New Zealand Soc. Anim. Prod. 56:19-21.

Wiggans, G. R. and M. E. Goddard. 1997. A computationally feasible test day model for genetic evaluation of yield traits in the United States. J. Dairy. Sci. 80:1795-1800.

Williams, A. R. 2002. Ultrasound applications in beef cattle carcass research and management. J. Anim. Sci. 80:E183-E188.

Wilson, D. E. 1992. Application of ultrasound for genetic improvement. J. Anim. Sci. 75:7-13.

Wilson, D. E., G. H. Rouse and S. Greiner. 1998. Relationship between chemical percentage intramuscular fat and USDA marbling score. Iowa State University, 1998 Beef Res. Rep.

Winsor, C. P. 1932. The Gompertz curve as a growth curve. Proc. National Academy of Science. 18:1-8.

Wold, S. 1974. Spline functions in data analysis. Technometrics. 16:1-11.

Wright, S. (1922). Coefficients of inbreeding and relationship. American Naturalist. 56:330-338.

Wyatt, W. E., T. D. Bidner, P. E. Humes, D. E. Franke and D. C. Blouin. 2002. Cow-calf and feedlot performances of Brahman-derivative breeds. J. Anim. Sci. 80:3037-3045.

CHAPTER III

MULTIVARIATE VERSUS RANDOM REGRESSION MODELS: AN EXAMPLE OF EQUIVALENCY

Overview

Methods of evaluating beef cattle data for the purpose of genetic evaluation has been thoroughly reported throughout pertinent scientific literature as reported by Speidel et al., (2010). Given that a majority of information (i.e. body weight observations) recorded in a typical beef cattle genetic evaluation scheme can be measured in many instances over time, this data has often been referred to as "infinite-dimensional" or "longitudinal" data (Meyer and Hill, 1997) due to the fact that one could theoretically record data points on the same trait an infinite number of times. In the context of beef cattle genetic evaluation, this data has been analyzed using primarily three different approaches, which include the simplistic "repeatability" or "repeated measures" model, the multiple trait model (**MTM**) and, perhaps most appropriate, random regression models (**RR**).

Of these three models named above, the repeated measures model is rarely used, except for the genetic evaluation of mature weights, while the latter two have received the most attention in recent genetic evaluation approaches. Upon initial glimpses of both the MTM and RR, the two models seem quite different, where in fact the two models are very similar, even identical under certain situations (Meyer, 1998). The object of this

section is to illustrate the process involved in executing both the RR and MTM models illustrating their equivalency to one another using a small example data set.

Data description

Data used in the example evaluations for this chapter are shown below in Table 3.1. This data set is a selected subset of the Colorado State University's Beef Improvement Center bull test data collected between November 2009 and March 2010. Animals were carefully chosen to minimize the number of necessary fixed effects as well as minimizing the number of animals with missing data. The animals were selected for inclusion based on the criteria that they were born on the same day therefore they are all the same age and no data was missing, thereby minimizing the number of fixed effects to fit.

In order to conduct a genetic evaluation using this data set, a few key components such as the stacked pedigree (Table 3.2), numerator relationship matrix (Figure 3.1) and the genetic covariance matrices (Table 3.3) are needed. Genetic variances for each weight measurement were estimated using the entire year's worth of data in conjunction with five single trait models. Genetic correlations were then estimated using all pairwise combinations between the five weights. This method yields fairly accurate estimates of genetic variance and correlation to be used in this evaluation in a short time frame. For the purposes of this example, residual variance was assumed to be constant across all ages at 1426 lb^2, and was obtained by averaging the residual variance estimates across the five single trait models.

Table 3.1. Sample bull test data for the illustration of the equivalence of calculating Estimated Breeding values from both multivariate and random regression models. Data was obtained from Colorado State University's Beef Improvement Center.

Identification	Sire	Dam	Test Day (Trait)	Weight[1]	Measurement Date
9	1	3	1	627	11/3/2009
10	1	4	1	712	11/3/2009
11	2	5	1	632	11/3/2009
12	2	6	1	605	11/3/2009
13	1	7	1	630	11/3/2009
14	2	8	1	731	11/3/2009
9	1	3	27	732	11/30/2009
10	1	4	27	855	11/30/2009
11	2	5	27	728	11/30/2009
12	2	6	27	731	11/30/2009
13	1	7	27	758	11/30/2009
14	2	8	27	861	11/30/2009
9	1	3	62	828	1/4/2010
10	1	4	62	952	1/4/2010
11	2	5	62	861	1/4/2010
12	2	6	62	869	1/4/2010
13	1	7	62	869	1/4/2010
14	2	8	62	972	1/4/2010
9	1	3	90	927	2/1/2010
10	1	4	90	1039	2/1/2010
11	2	5	90	924	2/1/2010
12	2	6	90	940	2/1/2010
13	1	7	90	957	2/1/2010
14	2	8	90	1058	2/1/2010
9	1	3	119	969	3/2/2010
10	1	4	119	1111	3/2/2010
11	2	5	119	1007	3/2/2010
12	2	6	119	1051	3/2/2010
13	1	7	119	1042	3/2/2010
14	2	8	119	1118	3/2/2010

[1]Weights for the example are measured in pounds.

Table 3.2. Stacked pedigree used in the calculation of estimated breeding values for both the multivariate and random regression models.

ID	Sire	Dam
1	.	.
2	.	.
3	.	.
4	.	.
5	.	.
6	.	.
7	.	.
8	.	.
9	1	3
10	1	4
11	2	5
12	2	6
13	1	7
14	2	8

$$\mathbf{A}^{-1} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \end{matrix} \\ \begin{bmatrix} 2.5 & 0 & 0.5 & 0.5 & 0 & 0 & 0.5 & 0 & -1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 2.5 & 0 & 0 & 0.5 & 0.5 & 0 & 0.5 & 0 & 0 & -1 & -1 & 0 & -1 \\ 0.5 & 0 & 1.5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 1.5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 1.5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 1.5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 1.5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 1.5 & 0 & 0 & 0 & 0 & 0 & -1 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \end{array}$$

Figure 3.1. Inverse of the numerator relationship matrix built from the pedigree used in the calculation of estimated breeding values for both the multivariate and random regression models.

Table 3.3. Genetic variance and covariance matrix used in the multivariate model / random regression model weight (lb^2) equivalency example.

	Weight 1	Weight 2	Weight 3	Weight 4	Weight 5
Weight 1	1709	1467	1939	1894	2185
Weight 2	1467	1395	1809	1811	2045
Weight 3	1939	1809	2599	2580	2855
Weight 4	1894	1811	2580	2838	2883
Weight 5	2185	2045	2855	2883	3960

Example genetic evaluations

To illustrate the application of each technique, the following presents examples of each.

Multiple trait model. The MTM, introduced by Henderson and Quaas (1976), predicts genetic values for multiple traits through the incorporation of genetic and residual covariances among the traits (Mrode, 2005). The use of genetic and residual covariances, allows the modeling of changing relationships between observations, recorded at different times and distinguishes this model from the repeated measures model. This MTM is shown below in matrix form (Equation 3.1) has been described with the notation presented here as used by Mrode (2005).

Equation 3.1. Multiple trait model for two traits presented in matrix form as described by Mrode (2005).

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

In Equation 3.1, $\mathbf{y}_i$ is a vector of observations for the i^{th} trait (weight observation 1, weight observation 2, for example); $\mathbf{u}_i$ and $\mathbf{e}_i$ are vectors containing the random animal genetic and random residual effect for the i^{th} trait, respectively. $\mathbf{X}_i$ and $\mathbf{Z}_i$ are incidence

matrices consisting of zeroes and ones relating the observations in **y** to fixed effects in **b** and random animal genetic effects in **u**. The observations in **y** are assumed to have the mean **Xb**, while the random effects (additive genetic and residual error) are assumed to have means of zero and genetic and residual variances shown in Equation 3.2 and Equation 3.3, respectively.

Equation 3.2. Variance of the random effects for a two trait multiple trait model dispersed by the numerator relationship matrix.

$$\text{var}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \sigma^2_{g_1} & \sigma_{g_1,g_2} \\ \sigma_{g_2,g_1} & \sigma^2_{g_2} \end{bmatrix} \otimes \mathbf{A}$$

Equation 3.3. Variance and covariance matrix for a two trait multiple trait model dispersed by an identity matrix.

$$\text{var}\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma^2_{e_1} & \mathbf{I}\sigma_{e_1,e_2} \\ \mathbf{I}\sigma_{e_2,e_1} & \mathbf{I}\sigma^2_{e_2} \end{bmatrix}$$

.

In the equations above, $\sigma^2_{g_1}$, $\sigma^2_{g_2}$, $\sigma_{g_1g_2}$, and σ_{g_2,g_1} are the additive genetic variance for $\mathbf{y}_1$, $\mathbf{y}_2$ and the additive genetic covariances between $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively. Likewise, $\sigma^2_{e_1}$, $\sigma^2_{e_2}$, σ_{e_1,e_2}, and σ_{e_2,e_1} are the residual error variances for $\mathbf{y}_1$ and $\mathbf{y}_2$ as well as the residual covariances between $\mathbf{y}_1$ and $\mathbf{y}_2$. **A** is Wright's numerator relationship matrix (Wright, 1922) and **I** is an n x n identity matrix.

Solutions $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ from the above model are obtained by solving Henderson's mixed model equations shown in Equation 3.4 (Henderson et al., 1959).

Equation 3.4. Mixed model equations to estimate fixed and random effects for a multiple trait model.

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

In Equation 3.4, **X** and **Z** are incidence matrices that combine the both the $\mathbf{X_i}$ and $\mathbf{Z}_i$ matrices illustrated above in Equation 3.1. Similarly, **y** is a vector containing both $\mathbf{y}_i$ matrices discussed earlier. **G** is the additive genetic (co) variance matrix between random effects and **R** is the residual (co) variance matrix between traits.

With the final dataset, pedigree, numerator relationship matrix, covariance matrices created, the next step is to begin building the fixed and random effect incidence matrices which are used to form the coefficient matrix and right hand side. Since all animals in the example dataset are born on the same day, are of the same sex and managed the same from birth through the end of the test, the only fixed effect needed is the overall mean (Figure 3.2). The purpose of this exercise is to compare the EBV obtained from MTM to those obtained using a random regression model that influences the size of the **X** incidence matrix. A typical five trait MTM would have five incidence matrices for the mean, each corresponding only to those observations in each of the five traits resulting in separate overall means for each of the five weigh days. Since we are comparing the results here to a RR, in which all observations contribute to one overall mean we to will fit one overall mean so all observations are adjusted equally. This has been shown in Figure 3.2**,** where each of the observations contributes to the calculation of the overall mean, even though in this instance they are associated with separate traits.

Formation of the incidence matrices for the random animal genetic effects is more straightforward. For the five trait MTM, each of the five weigh days (or test days) are treated as separate but genetically correlated traits. This assumption results in the formation of five separate **Z** matrices relating the trait specific observations in **y** to the random animal effects in the pedigree. Formation of the matrices in this manner will result in five **Z** matrices, one for each age or test day. In order for the **X** and **Z** matrices to conform to one another when multiplied together such as in $\mathbf{X'R^{-1}Z}$, they must have the same number of rows, meaning each **Z** matrix needs 30 rows. However, unlike the formation of the **X** matrix, the **Z** matrix relates each trait or test day to each of the animal effects contained in the pedigree. Accordingly, each **Z** matrix must have 14 columns. Figure 3.5 shows the form of matrix $\mathbf{Z_1}$. Notice, there are only six ones corresponding to the six observations of each test day for animals 9, 10, 11, 12, 13, and 14 in the first 6 rows corresponding to the first trait or test day. Additional matrices will be formed in the same manner with this block of ones in rows 7-12, etc (see Appendix I).

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 627 \\ 712 \\ 632 \\ 605 \\ 630 \\ 731 \\ 732 \\ 855 \\ 728 \\ 731 \\ 758 \\ 861 \\ 828 \\ 952 \\ 861 \\ 869 \\ 972 \\ 927 \\ 1039 \\ 924 \\ 940 \\ 957 \\ 1058 \\ 969 \\ 1111 \\ 1007 \\ 1051 \\ 1042 \\ 1118 \end{bmatrix}$$

Figure 3.2. Incidence matrix used to calculate the overall mean along with the vector of observations for both the multivariate and random regression model examples.

$\mathbf{Z}_1 =$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	1	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	1	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	1	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	1	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	1	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	1	14	1	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	27	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	62	972
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	90	1058
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	119	1118

Figure 3.3. Incidence matrix relating the traits for weight observations measured on test day 1 to the corresponding animals in the pedigree on which the observations were observed.

Given the inverse of the numerator relationship (Figure 3.1), the **X** incidence matrix corresponding to the overall mean and observation vector **y** (Figure 3.2), each of the five **Z** incidence matrices (Figure 3.3; Appendix I) along with the genetic covariance matrix (Table 3.3), the residual variance of 1426 lb^2 and the mixed model equations (Equation 3.4) expanded in order to accommodate five separate but genetically correlated traits (Equation 3.5), the coefficient matrix (order of 71 rows by 71 columns) and right hand side (order of 71 rows by 1 column) can be built for the five trait example. Figure 3.4 illustrates the first few blocks of the coefficient matrix ($\mathbf{X'R^{-1}X}$, $\mathbf{X'R^{-1}Z_1}$, $\mathbf{Z_1'R^{-1}X}$ and $\mathbf{Z_1'R^{-1}Z_1} + g^{11}\mathbf{A^{-1}}$) and right hand side ($\mathbf{X'R^{-1}y}$ and $\mathbf{Z_1'R^{-1}y}$) from the mixed model equations for the five-trait multivariate model example. Once the entire equation set has been built, the solution vector can be solved for resulting in EBV for every trait and animal (Figure 3.5).

Equation 3.5. Expanded form of the multiple trait mixed model equations to accommodate five separate but genetically correlated traits.

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_2 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_3 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_4 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_5 \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{11}\mathbf{A}^{-1} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_2 + g^{12}\mathbf{A}^{-1} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_3 + g^{13}\mathbf{A}^{-1} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_4 + g^{14}\mathbf{A}^{-1} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_5 + g^{15}\mathbf{A}^{-1} \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{21}\mathbf{A}^{-1} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_2 + g^{22}\mathbf{A}^{-1} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_3 + g^{23}\mathbf{A}^{-1} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_4 + g^{24}\mathbf{A}^{-1} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_5 + g^{25}\mathbf{A}^{-1} \\ \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{31}\mathbf{A}^{-1} & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_2 + g^{32}\mathbf{A}^{-1} & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_3 + g^{33}\mathbf{A}^{-1} & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_4 + g^{34}\mathbf{A}^{-1} & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_5 + g^{35}\mathbf{A}^{-1} \\ \mathbf{Z}_4'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_4'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{41}\mathbf{A}^{-1} & \mathbf{Z}_4'\mathbf{R}^{-1}\mathbf{Z}_2 + g^{42}\mathbf{A}^{-1} & \mathbf{Z}_4'\mathbf{R}^{-1}\mathbf{Z}_3 + g^{43}\mathbf{A}^{-1} & \mathbf{Z}_4'\mathbf{R}^{-1}\mathbf{Z}_4 + g^{44}\mathbf{A}^{-1} & \mathbf{Z}_4'\mathbf{R}^{-1}\mathbf{Z}_5 + g^{45}\mathbf{A}^{-1} \\ \mathbf{Z}_5'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_5'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{51}\mathbf{A}^{-1} & \mathbf{Z}_5'\mathbf{R}^{-1}\mathbf{Z}_2 + g^{52}\mathbf{A}^{-1} & \mathbf{Z}_5'\mathbf{R}^{-1}\mathbf{Z}_3 + g^{53}\mathbf{A}^{-1} & \mathbf{Z}_5'\mathbf{R}^{-1}\mathbf{Z}_4 + g^{54}\mathbf{A}^{-1} & \mathbf{Z}_5'\mathbf{R}^{-1}\mathbf{Z}_5 + g^{55}\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \hat{\mathbf{u}}_3 \\ \hat{\mathbf{u}}_4 \\ \hat{\mathbf{u}}_5 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_4'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_5'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Coefficient Matrix															Right Hand Side
0.0210															18.3
0	0.016														0
0	0	0.016													0
0	0.003	0	0.010												0
0	0.003	0	0	0.010						*Symmetric*					0
0	0	0.003	0	0	0.010										0
0	0	0.003	0	0	0	0.010									0
0	0.003	0	0	0	0	0	0.010								0
0	0	0.003	0	0	0	0	0	0.010							0
0.0007	–0.006	0	–0.006	0	0	0	0	0	0.014						0.440
0.0007	–0.006	0	0	–0.006	0	0	0	0	0	0.014					0.499
0.0007	0	–0.006	0	0	–0.006	0	0	0	0	0	0.014				0.443
0.0007	0	–0.006	0	0	0	–0.006	0	0	0	0	0	0.014			0.424
0.0007	–0.006	0	0	0	0	0	–0.006	0	0	0	0	0	0.014		0.442
0.0007	0	–0.006	0	0	0	0	0	–0.006	0	0	0	0	0	0.014	0.513

Figure 3.4. Example block of equations for the five trait multiple trait model example. Figure consists of the $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X}$ and $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{11}\mathbf{A}^{-1}$ of the coefficient matrix as well as $\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$ and $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y}$ from the right hand side. The bars delineate each sub matrix ($\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{11}\mathbf{A}^{-1}$, $\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$ and $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y}$) from the overall block of the coefficient matrix.

Estimated Breeding Value for each of 14 animals in the pedigree

Weight Trait	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	70.1	73.2	0.1	54.9	5.2	7.3	15.0	60.7	35.2	117.5	44.4	47.5	57.6	127.7
27	87.1	89.9	6.9	58.5	11.5	14.9	21.7	63.5	53.9	131.3	62.3	67.4	76.2	140.2
62	146.9	153.4	18.4	88.6	25.8	31.9	39.9	95.7	101.0	206.4	115.5	124.6	133.3	220.3
90	175.7	180.8	27.7	98.5	33.9	41.1	49.5	105.8	129.4	235.6	141.2	152.1	162.0	249.2
119	229.7	241.3	38.5	121.7	50.3	63.4	69.4	127.5	172.6	297.4	196.1	215.8	219.0	311.9

Figure 3.5. Estimated breeding values obtained from the multiple trait model for all 14 animals in the pedigree for each of the five traits measured at days (1, 27, 62, 90 and 119).

Random regression model. Schaeffer and Dekkers (1994) introduced RR by extending the regression coefficients from the fixed portion of the standard mixed model evaluation to the random animal effects. In doing so, they were not only able to account for the general average shape of the curve but they also were able to account for each animal's individual deviation from that average curve. In addition, since they were estimating regression parameters (intercept, slope, etc) for individual animals they found they were able to account for the changing correlation structure of repeated records measured on individuals over time. So, rather than estimating breeding values for a particular age, RR estimates breeding values for the parameters of the regression of weight on age. Since the resulting estimates are components of a regression line / curve, the prediction of an EBV for any age within the range of the independent variable is possible.

Meyer (1998) spoke of the equivalency of RR and MTM. She stated that traditional multivariate analyses are equivalent to 'full fit' random regression models where the order of the polynomial fit is equal to the number of ages measured. While that might be statistically appropriate, experience showed that as higher order polynomials were fit or observations measured very near one another in time, complications result from extremely high correlations between successive ages. As such, orthogonal base functions such as Legendre polynomials have been suggested to reduce these elevated correlations between successive ages (Kirkpatrick et al., 1990).

Implementation of a RR is accomplished using the same model equations and assumptions of fixed and random effects as for MTM, the only difference between RR and MTM being the formation of the incidence matrix **Z**. In a RR, the incidence matrix

Z is modified by replacing the ones normally present with a covariate consisting of the actual age, or transformed age as would be the case when using Legendre polynomials.

In order to calculate the Legendre polynomials evaluated at each age point in the data set, a matrix **Φ** is created by pre-multiplying the matrix of Legendre polynomial coefficients (**Λ**) by a matrix containing standardized age values (**M**) such that **Φ** = **MΛ**. The first step in this process is to standardize the ages of the observations used in the evaluation. Legendre polynomials are defined over the interval of -1 to 1 (Kirkpatrick et al., 1990, 1994) therefore it is necessary to standardize the ages of the observations to this interval. Equation 3.6 gives the formula presented by Schaeffer (2003) used to transform these ages.

Equation 3.6. Formula for standardizing ages of observations ranging from -1 to 1.

$$t_i^* = -1 + 2\left(\frac{t_i - t_{\min}}{t_{\max} - t_{\min}}\right)$$

In Equation 3.6, t_i^* is the standardized time, t_i represents the time point being standardized, and $t_{\min}$ and $t_{\max}$ were the minimum and maximum time points or ages represented in the dataset, respectively.

Using the example data set above (Table 3.1), the ages of the animals at the time of measurement were standardized using Equation 3.6 and were placed into a matrix **M** (Figure 3.6), shown below. The order of **M** is *k* (the order of fit) by *t* (the number of unique ages). In this example (Figure 3.6), **M** is a five by five matrix with the first column of ones representing the intercept of the polynomial. The second column represents the linear covariate and consists of the calculated standardized ages. The third,

fourth and fifth columns represent the quadratic, cubic and quartic terms, respectively and consist of the standardized age squared, cubed and raised to the fourth power.

	M				
Observed Ages	*Int*	x	x^2	x^3	x^4
1	1.0000	–1.0000	1.0000	–1.0000	1.0000
27	1.0000	–0.5593	0.3128	–0.1750	0.0979
62	1.0000	0.0339	0.0011	0.0000	0.0000
90	1.0000	0.5085	0.2585	0.1315	0.0668
119	1.0000	1.0000	1.0000	1.0000	1.0000

Figure 3.6. Matrix of standardized ages for a fourth order polynomial. Values in M are used to calculate the Legendre polynomials in a random regression model.

Next, the matrix **Λ** containing the Legendre polynomial coefficients are calculated. **Λ** has an order of *k* by *k* or order of fit by order of fit of the Legendre polynomials. To do this, the first two polynomials defined in Equation 3.7 are expanded upon using the recursive formula found in Equation 3.8 used for calculating additional *n*+1 Legendre polynomials.

Equation 3.7. The first two Legendre polynomials defined.

$$P_0(t) = 1, \text{ and } P_1(t) = t$$

Equation 3.8. Recursive formula for calculating additional Legendre polynomials.

$$P_{n+1}(t) = \frac{1}{n+1}\left((2n+1)tP_n(t) - nP_{n-1}(t)\right)$$

The Legendre polynomials calculated in the above two equations are then normalized using Equation 3.9 below. According to Mrode (2005), this is equivalent to the integration of the polynomials from -1 to 1.

Equation 3.9. Equation used to normalize each Legendre polynomial.

$$\phi_n(t) = \left(\frac{2n+1}{2}\right)^{0.5} P_n(t)$$

.

Table 3.4 shows the results from the calculation of a fourth order normalized Legendre polynomial. These normalized polynomials $(\phi_n(t))$ are then put into a matrix **Λ** (see Figure 3.7).

Table 3.4. Normalized Legendre polynomials for up to a fourth order polynomial**.**

Order	Legendre Polynomial	Normalized Legendre Polynomial
n = 0	$P_1(t) = t$	$\phi_0(t) = 0.7071$
n = 1	$P_2(t) = \frac{3}{2}t^2 - \frac{1}{2}$	$\phi_1(t) = 1.2247t$
n = 2	$P_3(t) = \frac{5}{2}t^3 - \frac{9}{6}t$	$\phi_2(t) = 2.3717t^2 - 0.7906$
n = 3	$P_4(t) = \frac{35}{8}t^4 - \frac{45}{12}t^2 + \frac{3}{8}$	$\phi_3(t) = 4.6771t^3 - 2.8062t$
n = 4	$P_5(t) = \frac{63}{8}t^5 - \frac{35}{4}t^3 + \frac{15}{8}t$	$\phi_4(t) = 9.2808t^4 - 7.9550t^2 + 0.7955$

$$\Lambda' = \begin{pmatrix} 0.7071 & 0 & 0 & 0 & 0 \\ 0 & 1.2247 & 0 & 0 & 0 \\ -0.7906 & 0 & 2.3717 & 0 & 0 \\ 0 & -2.8062 & 0 & 4.6771 & 0 \\ 0.7955 & 0 & -7.9550 & 0 & 9.2808 \end{pmatrix}$$

Figure 3.7. Table of normalized values for a fourth order Legendre polynomial.

Figure 3.7 contains the transpose of the matrix **Λ.** Here there are five rows where each row represents one of the normalized polynomials (ϕ_0 to ϕ_4) shown in Table 3.4 above. Columns in Figure 3.7, moving from left to right, represent the intercept, linear, quadratic, cubic and quartic terms for each of the normalized polynomials (ϕ_0 to ϕ_4), respectively. Row 1 corresponds to $\phi_0(t)$ from Table 3.4, which is why there is a 0.7071 in row 1 column 1 and zeroes in the remainder of the row. Row 2 corresponds to $\phi_1(t)$ which is why there is a zero in row 2 column 1 and a 1.2247 in row 2 column 2. The remainder of the matrix is filled in by placing the coefficients for each of the remaining normalized polynomials into the remaining three rows of the matrix.

The matrix **Φ** is then created by multiplying the standardized age matrix **M** and **Λ**, the Legendre polynomial matrix, such that **Φ** = **MΛ**. This matrix is shown below in Figure 3.8 contains the Legendre polynomials evaluated at each of the observed ages present in the example data set (Table 3.1).

	Φ				
Observed Ages	*Int*	x	x^2	x^3	x^4
1	0.7071	–1.2247	1.5811	–1.8704	2.1213
27	0.7071	–0.6850	–0.0486	0.7515	–0.7848
62	0.7071	0.0415	–0.7879	–0.0950	0.7864
90	0.7071	0.6227	–0.1774	–0.8123	–0.6409
119	0.7071	1.2247	1.5811	1.8704	2.1213

Figure 3.8. Matrix of Legendre polynomials for a fourth order polynomial corresponding to each of the observed ages in the data set.

The next step in the construction of a random regression model, is the creation of a random regression genetic covariance matrix from the standard multiple trait genetic covariance matrix in Table 3.3. This conversion is performed using Equation 3.10 below.

Equation 3.10. Formula used to convert a traditional genetic (co-) variance matrix to a genetic (co-) variance matrix describing a polynomial base function used in a random regression model.

$$\hat{\mathbf{C}}_G = \mathbf{\Phi}^{-1}\hat{\mathbf{G}}\left[\mathbf{\Phi}^T\right]^{-1}$$

In the equation above, $\hat{\mathbf{C}}_G$ is the transformed genetic covariance matrix for the terms in a quartic Legendre polynomial matrix. $\mathbf{\Phi}$ is the matrix of Legendre polynomial coefficients evaluated at the five differing ages in the example data set and $\hat{\mathbf{G}}$ is the estimated genetic covariance matrix from the five trait multiple trait model (Table 3.3). Results from this transformation are shown below in Table 3.5. This table contains the genetic variance and covariance estimates needed for performing a quartic random regression genetic evaluation for the example data set.

Table 3.5. Fourth order random regression genetic variance and covariance matrix ($\hat{\mathbf{C}}_{\mathbf{G}}$) used in the multivariate model / random regression model equivalency example.

	Intercept	Linear	Quadratic	Cubic	Quartic
Intercept	4225	638	-140	-89	167
Linear	638	215	-7	-32	9
Quadratic	-140	-7	65	15	-22
Cubic	-89	-32	15	43	18
Quartic	167	9	-22	18	43

Finally, to finish the RR model the **Z** matrices must be modified to accommodate the Legendre polynomials shown in Figure 3.8. These modified incidence matrices, shown below in Figure 3.9 through Figure 3.13 (one figure for each term in the polynomial), contain the modified **Z** matrices relating the specific Legendre polynomial covariate to their respective observation.

$\mathbf{Z}_{Int}$ =

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	0.7071	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	0.7071	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	0.7071	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	0.7071	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	0.7071	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	0.7071	14	1	731
0	0	0	0	0	0	0	0	0.7071	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	0.7071	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	0.7071	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	0.7071	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	0.7071	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	0.7071	14	27	861
0	0	0	0	0	0	0	0	0.7071	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	0.7071	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	0.7071	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	0.7071	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0.7071	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0.7071	14	62	972
0	0	0	0	0	0	0	0	0.7071	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	0.7071	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	0.7071	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	0.7071	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	0.7071	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	0.7071	14	90	1058
0	0	0	0	0	0	0	0	0.7071	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	0.7071	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	0.7071	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	0.7071	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	0.7071	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	0.7071	14	119	1118

Figure 3.9. Random regression incidence matrix relating to the intercept of each random regression equation.

$\mathbf{Z}_{Int}$ =

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	–1.2247	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	–1.2247	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	–1.2247	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	–1.2247	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	–1.2247	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	–1.2247	14	1	731
0	0	0	0	0	0	0	0	–0.6850	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	–0.6850	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	–0.6850	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	–0.6850	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	–0.6850	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	–0.6850	14	27	861
0	0	0	0	0	0	0	0	0.0415	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	0.0415	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	0.0415	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	0.0415	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0.0415	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0.0415	14	62	972
0	0	0	0	0	0	0	0	0.6227	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	0.6227	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	0.6227	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	0.6227	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	0.6227	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	0.6227	14	90	1058
0	0	0	0	0	0	0	0	1.2247	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	1.2247	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	1.2247	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	1.2247	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	1.2247	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	1.2247	14	119	1118

Figure 3.10. Random regression incidence matrix relating to the Legendre polynomial linear covariate to each corresponding observation measurement age per animal.

$\mathbf{Z}_{Quad}$ =

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	1.5811	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	1.5811	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	1.5811	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	1.5811	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	1.5811	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	1.5811	14	1	731
0	0	0	0	0	0	0	0	–0.0486	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	–0.0486	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	–0.0486	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	–0.0486	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	–0.0486	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	–0.0486	14	27	861
0	0	0	0	0	0	0	0	–0.7879	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	–0.7879	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	–0.7879	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	–0.7879	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	–0.7879	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	–0.7879	14	62	972
0	0	0	0	0	0	0	0	–0.1774	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	–0.1774	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	–0.1774	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	–0.1774	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	–0.1774	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	–0.1774	14	90	1058
0	0	0	0	0	0	0	0	1.5811	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	1.5811	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	1.5811	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	1.5811	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	1.5811	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	1.5811	14	119	1118

Figure 3.11. Random regression incidence matrix relating to the Legendre polynomial quadratic covariate to each corresponding observation measurement age per animal.

$\mathbf{Z}_{Cubic}$ =

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	−1.8704	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	−1.8704	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	−1.8704	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	−1.8704	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	−1.8704	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	−1.8704	14	1	731
0	0	0	0	0	0	0	0	0.7515	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	0.7515	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	0.7515	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	0.7515	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	0.7515	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	0.7515	14	27	861
0	0	0	0	0	0	0	0	−0.0950	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	−0.0950	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	−0.0950	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	−0.0950	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	−0.0950	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	−0.0950	14	62	972
0	0	0	0	0	0	0	0	−0.8123	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	−0.8123	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	−0.8123	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	−0.8123	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	−0.8123	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	−0.8123	14	90	1058
0	0	0	0	0	0	0	0	1.8704	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	1.8704	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	1.8704	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	1.8704	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	1.8704	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	1.8704	14	119	1118

Figure 3.12. Random regression incidence matrix relating to the Legendre polynomial cubic covariate term to each corresponding observation measurement age per animal.

$\mathbf{Z}_{Quartic}$ =

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	2.1213	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	2.1213	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	2.1213	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	2.1213	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	2.1213	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	–1.8704	14	1	731
0	0	0	0	0	0	0	0	–0.7848	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	–0.7848	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	–0.7848	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	–0.7848	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	–0.7848	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	–0.7848	14	27	861
0	0	0	0	0	0	0	0	0.7864	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	0.7864	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	0.7864	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	0.7864	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0.7864	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0.7864	14	62	972
0	0	0	0	0	0	0	0	–0.6409	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	–0.6409	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	–0.6409	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	–0.6409	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	–0.6409	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	–0.6409	14	90	1058
0	0	0	0	0	0	0	0	2.1213	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	2.1213	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	2.1213	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	2.1213	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	2.1213	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	2.1213	14	119	1118

Figure 3.13. Random regression incidence matrix relating to the Legendre polynomial quartic covariate term to each corresponding observation measurement age per animal.

Illustrating in the same manner as was performed with the MTM, the first few blocks of the coefficient matrix ($\mathbf{X'R^{-1}X}$, $\mathbf{X'R^{-1}Z_1}$, $\mathbf{Z_1'R^{-1}X}$, and $\mathbf{Z_1'R^{-1}Z_1} + g^{11}\mathbf{A^{-1}}$) and right hand side ($\mathbf{X'R^{-1}y}$ and $\mathbf{Z_1'R^{-1}y}$) are shown in Figure 3.14. Here, the **X** incidence matrices were the same between the RR and MTM which resulted in the blocks of $\mathbf{X'R^{-1}X}$ and $\mathbf{X'R^{-1}y}$ being the same for both models. The differences between the two models are from the differences in the creation of the **Z** matrices. Again, once the entire set of equations were built, the solution vector of EBV for each term in the fourth order polynomial can be calculated for every animal in the pedigree (Figure 3.15). Notice the differences between the EBV in this figure versus the EBV in Figure 3.5. To make comparisons between the two sets of EBV, it is necessary to convert the EBV obtained from the random regression model back to EBV for each age in the data set. Equation 3.11 details the conversion of the EBV for the quartic Legendre polynomial regression back to age specific EBV.

Equation 3.11. Equation for converting transformed EBV obtained from a random regression model using Legendre polynomials as the base function to EBV for observed ages.

$$\mathrm{EBV}_{\mathrm{Observed}} = \Phi * \mathrm{EBV}_{\mathrm{Transformed}}$$

Test day EBV are shown below in Figure 3.16. When compared to the EBV obtained from a five trait MTM, presented earlier in Figure 3.5, they are identical.

Coefficient Matrix														Right Hand Side
0.0210														18.3
0	0.014													0
0	0	0.014												0
0	0.0003	0	0.0008											0
0	0.0003	0	0	0.0008						*Symmetric*				0
0	0	0.0003	0	0	0.0008									0
0	0	0.0003	0	0	0	0.0008								0
0	0.0003	0	0	0	0	0	0.0008							0
0	0	0.0003	0	0	0	0	0	0.0008						0
0.0025	–0.0006	0	–0.0006	0	0	0	0	0	0.003					2.025
0.0025	–0.0006	0	0	–0.0006	0	0	0	0	0	0.003				2.315
0.0025	0	–0.0006	0	0	–0.0006	0	0	0	0	0	0.003			2.059
0.0025	0	–0.0006	0	0	0	–0.0006	0	0	0	0	0	0.003		2.081
0.0025	–0.0006	0	0	0	0	0	–0.0006	0	0	0	0	0	0.003	2.110
0.0025	0	–0.0006	0	0	0	0	0	–0.0006	0	0	0	0	0	2.350

Figure 3.14. Example block of equations for the fourth order random regression model using Legendre polynomials as the base function. Figure consists of the $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X}$ and $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{11}\mathbf{A}^{-1}$ of the coefficient matrix as well as $\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$ and $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y}$ from the right hand side. The bars delineate each sub matrix ($\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + g^{11}\mathbf{A}^{-1}$, $\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$ and $\mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y}$) from the overall block of the coefficient matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Intercept	195.1	202.3	25.5	116.3	34.4	42.7	53.3	125.2	135.7	272.1	152.7	165.1	177.5	289.0
Linear	66.1	68.5	15.8	28.8	17.6	21.2	21.5	29.7	56.8	76.2	60.7	66.1	65.4	78.8
Quadratic	–1.25	–1.69	0.55	–2.07	0.31	0.71	0.27	–2.72	0.21	–3.74	–0.37	0.22	–0.22	–4.92
Cubic	–0.64	0.05	–0.10	–0.98	0.53	1.11	0.44	–1.58	–0.47	–1.79	0.82	1.69	0.35	–2.34
Quartic	6.55	7.96	0.21	4.41	1.39	1.92	1.94	4.65	3.58	9.89	6.07	6.86	6.19	10.96

Figure 3.15. Estimated breeding values obtained from a 4th order Legendre polynomial random regression model for all 14 animals in the pedigree. Rather than obtaining EBV for the test days 1, 27, 62, 90 and 119, EBV are obtained for the components of the line which can be used to calculate a breeding value for any day between 1 and 119.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	70.1	73.2	0.1	54.9	5.2	7.3	15.0	60.7	35.2	117.5	44.4	47.5	57.6	127.7
27	87.1	89.9	6.9	58.5	11.5	14.9	21.7	63.5	53.9	131.3	62.3	67.4	76.2	140.2
62	146.9	153.4	18.4	88.6	25.8	31.9	39.9	95.7	101.0	206.4	115.5	124.6	133.3	220.3
90	175.7	180.8	27.7	98.5	33.9	41.1	49.5	105.8	129.4	235.6	141.2	152.1	162.0	249.2
119	229.7	241.3	38.5	121.7	50.3	63.4	69.4	127.5	172.6	297.4	196.1	215.8	219.0	311.9

Figure 3.16. Estimated breeding values for test day, converted from a 4^{th} order Legendre polynomial random regression model. EBV are calculated by multiplying the matrix of Legendre polynomials for the specific test days by a matrix of transformed EBV.

Summary

The above example has illustrated the equivalence of RR vs MTM as explained by Meyer (1998). In order for the model to be exactly equivalent, not only does the order of the RR need to be the same as the number of traits in the MTM, the fixed effect portion of the two must be equivalent as well.

There are both advantages and disadvantages to fitting both types of models and these typically deal with the parameterization of the two models. For example, if the number of observed ages is large the MTM can become very difficult to solve quickly (especially for very large pedigrees). With a large number of ages measured, an equivalent RR can become large and intractable as well. However, as Kirkpatrick et al. (1990) illustrated, it is possible and perhaps more ideal to fit a reduced order polynomial allowing the covariance matrix to contain a few parameters as possible. Fitting a reduced model that is statistically consistent with the data reduces the fluctuations caused by the sampling error in the initial measurements used to estimate the genetic covariance matrix (Kirkpatrick et al., 1990). This is important given the nature of these polynomials to become unwieldy at the extremes of the independent variable (Meyer, 1998) typically leading to heritability estimates being high at the beginning and ends of the data range and low in the middle.

Literature cited

Henderson, C. R. O. Kempthorne, S. R. Searle, C. M. von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics. 15:192-218.

Henderson, C. R. and R. L. Quaas. 1976. Multiple trait evaluation using relatives' records. J. Anim. Sci. 43:1188-1197.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124:979-993.

Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. Genet. Sel. Evol. 30:221-240.

Meyer, K. and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. Livest. Prod. Sci. 47:185-200.

Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CABI Publishing Company. Cambridge, MA.

Schaeffer, L. R. 2003. Random regression models. ANSC6737 Course Notes – Quantitative genetics and animal models. Available at: http://www.aps.uoguelph.ca~lrs/ANSC637/LRS14/LRS14.pdf. Accessed Jan. 8, 2009.

Schaeffer, L. R. and J. C. M. Dekkers. 2994. Random regressions in animal models for test-day production in dairy cattle. In: Proc. 5th World Congr. Appl. Livest. Prod. XVIII. Pp. 443-446.

Speidel, S. E., R. M. Enns and D. H. Crews Jr. 2010. Genetic analysis of longitudinal data in beef cattle: a review. Genet. Mol. Res. 9:19-33.

Wright, S. 1922. Coefficients of inbreeding and relationship. American Naturalist. 56:330-338.

CHAPTER IV

DATA PREPARATION

Overview

Data used in this project was obtained from a historical database therefore they were not subject to animal care and use committee approval. In the current section a description of the data will be presented, containing an overview of all summary statistics, data distributions, and sifts performed to obtain the final data file used in the evaluations. In subsequent sections, the data will not be discussed other than pertinent preparation techniques specific to the given analyses.

Data description

The Lethbridge Research Centre located in Lethbridge, Alberta, a research station associated with Agriculture and Agri-Food Canada, provided data for this project. The original data set consisted of pedigree, weight and ultrasound observations on 1,375 individual animals spanning the years 1999 – 2007. Each individual animal record contained animal identifier, feedlot pen, breed (Angus, Charolais and Charolais Cross), year of measurement (one feeding period per year), birth date, birth weight, weaning date, weaning weight as well as serial weight and ultrasound measurements taken over the feeding period. Measurement dates also accompanied the serial weight and ultrasound measurements allowing for the calculation of age as measure.

Three separate ultrasound traits were recorded on each sequential measurement date and were recorded in a slightly different manner than is currently performed in the United States. All ultrasound observations were measured between the 12th and 13th ribs of the animal. Ultrasound rib eye area (**UREA**) was measured on each individual animal and was recorded in square centimeters (cm^2). Two ultrasound back fat measurements were recorded. First was an average ultrasound back fat (**UBFa**) where back fat was measured at 3 equally spaced locations across the longissimus muscle and then averaged. The second measurement was an ultrasound back fat observation measured ⅔ of the distance from the medial end of the longissimus dorsi muscle. These two ultrasound back fat measures are slightly different than what is currently measured in the United States where ultrasound back fat is measured as the subcutaneous fat covering at a point of ¾ of the total distance from the medial end of the longissimus dorsi (BIF, 2002; Williams, 2002). Given these two measurements for ultrasound back fat, UBFa is considered to be most consistent with what is currently reported in the United States (D.H. Crews, Jr., Colorado State University, Fort Collins, CO, personal communication) therefore all subsequent uses of UBFa will be referred to singularly as ultrasound back fat (**UBF**).

Eight years of data was compiled. A breakdown of the number of animals per year is shown below in Table 4.1. Animal breed is not represented across all years of data. During the first two years, only Charolais and Charolais cross (calves sired by mating purebred Charolais bulls to un-recorded breeds of dam) individuals were on test. Then in subsequent years, only purebred Angus and Charolais animals were fed. Such a stratification of breed type resulted in a confounding of breed and production year.

Table 4.1. Number of animals reported per year divided into their respective individual breed types.

Year	N	Angus	Charolais	Charolais Cross[1]
1999 – 2000	122	0	22	100
2000 – 2001	161	0	31	130
2001 – 2002	172	79	93	0
2002 – 2003	160	98	62	0
2003 – 2004	178	90	88	0
2004 – 2005	200	110	90	0
2005 – 2006	198	100	98	0
2006 – 2007	184	111	73	0

[1]Sired by Charolais bulls mated to dams of unknown breed type.

Not including the birth and weaning weight observations, the frequency of measurements differed depending on the year in which the measurements were recorded, ranging from approximately every two weeks to four weeks. During the first two years (1999 to 2001), observations were recorded on the individual animal approximately every 14 days. In subsequent years, observations were recorded less frequently approaching 4 weeks between each successive record. The average number of weight observations and ultrasound observations per animal across all years is shown below in Table 4.2 and Table 4.3, respectively.

Table 4.2. Average number of weight observations[1] per animal broken down by breed across the years 1999 to 2007.

	All Animals	Angus	Charolais	Charolais Cross
Mean	9.39	7.66	8.35	16.35
Min	1	1	2	3
Max	19	11	19	19

[1]Not including birth weight or weaning weight records.

Table 4.3. Average number of ultrasound observations[1] per animal broken down by breed across the years 1999 to 2007.

	All Animals	Angus	Charolais	Charolais Cross
Mean	7.46	5.50	6.46	14.9
Min	1	1	1	1
Max	17	7	17	17

[1]Includes both ultrasound rib eye area and ultrasound back fat.

Fewer ultrasound observations exist than weight observations due to the fact that there were phases of the feeding period such as arrival, backgrounding and transition periods where no ultrasound observations were collected. There were greater numbers of observations for Charolais cross animals in years from 1999 to 2001 as the animals were measured every two weeks during these years. The average number of observations for Charolais animals is greater than the average number of observations for Angus due to the fact Charolais were represented in these first two years as well. After the first two years of production, Angus and Charolais animals were treated in the same fashion.

To create a "final" weight file for evaluation, all available weight records (birth and weaning weights included where available) were compiled into a single file. Weight observations were then classified into two separate sub-categories of test-day (**TD**) and test-day on full feed (**OFTD**). Those weight records classified into the TD category were all available observations excluding birth and weaning weight observations. Weight records classified into the OFTD category were those weight observations taken after all backgrounding and transition periods were finished. Individual dates that outlined when the animals were started on full feed were supplied by the data provider and were used to mark the beginning of the OFTD feeding period. An important note is these categories are not mutually exclusive; rather they are inclusive of each other. In other words, TD is a subset of all data and OFTD is a subset of TD.

Ultrasound observations were classified according to feeding periods in a similar manner as the weight records. However, given when the ultrasound observations were recorded, only the first two years of data (1999 to 2001) contained any records outside of the OFTD category. During the years 2001 to 2007, all ultrasound observations were taken when the animals were in the OFTD category. As a result, the decision was made to omit the ultrasound observations exclusively represented in the TD category.

Fixed effects

There are several important factors influencing observations of interest on beef cattle and most of these have been summarized in the Beef Improvement Federation's guidelines for uniform beef improvement programs (2002), a set of guidelines for the standardized reporting of performance information developed by a committee of seedstock breeders and scientists. Age of dam is one of these factors and has been shown to influence weaning weight records by as much as 27.3 kg (BIF, 2002). Age of dam was an effect considered when building the days to finish regression models, however, out of the 1,375 individual animals, 553 (40 percent) were missing dam information. These 553 animals missing dam information represented 6,864 of the 14,325 (48 percent) weight records in the final weight data file. Given the sheer number of animals with unknown dams, age of dam was not included in the analysis of fixed effects, as its inclusion would prohibit the use of 48 % of the data.

Season of birth is another important contributing factor to the observations of an individual (BIF, 2002). This data set is unique in the fact that only one test was conducted per year, therefore all animals must have been born in the same season

(spring). Length of calving season was also considered, however only the animals in one year's group (2000 – 2001) were born in a calving season whose length (113 days) was greater than 90 days. Given the tight calving window across all years as a whole, it was felt unnecessary to include length of calving window in the contemporary group definition to minimize the dissection of contemporary groups, even though it has been documented in the literature that accounting for season of birth in contemporary grouping strategies has the effect of increasing the accuracy of selection (Crump et al., 1997). A noteworthy point to make is that smallest season window described by Crump and his colleagues was 3 months or approximately 90 days not much smaller than the 113 days we see here.

When genetic evaluations are performed, it is common practice to place animals into groups of similarly managed individuals. Comparisons are then made between animals within contemporary group, properly evaluating genetic differences between sires. The dataset used in this project contained weight and ultrasound observations from 3 differing breed types observed across 8 years, fed in 4 separate feedlot pens. Therefore, the identification of proper fixed effect classifications is important to ensure animals are being compared fairly.

In an effort to form the most complete and appropriate contemporary group preliminary analyses of variance was performed using the LMER procedure from the lme4 package in R (R Development Core Team, 2009) to identify significant fixed effects for both the weight and ultrasound traits. A series of three linear mixed models were implemented to determine the significance for each of three predictors; feedlot pen, year of measure, and breed type (Angus, Charolais, and Charolais cross). Each of the models

contained two of the fixed effects listed above, with the third fixed effect added last. The mixed factor regression model used for the preliminary analyses of variance as described by Ruppert et al. (2003) is shown below:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \boldsymbol{\varepsilon}$$

where **y** is a vector of age observations, **X** is an incidence matrix relating age observations in **y** to fixed effects (pen, breed and year) and fixed regression coefficients (weight, UBF or UREA) in **b**, **Z** is an incidence matrix relating age observations in **y** to random effects in **u** (individual animals), and $\boldsymbol{\varepsilon}$ is a vector of random residual error terms. The mixed factor regression model above makes the assumption that the mean of the random effects is zero with variances represented by:

$$\text{Cov}\begin{bmatrix} \mathbf{u} \\ \text{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix},$$

where

$$\mathbf{G} = \sigma_u^2\mathbf{I}, \text{ and } \mathbf{R} = \sigma_\varepsilon^2\mathbf{I}.$$

Given that the statistical package R, has no procedure for estimating denominator degrees of freedom from mixed factor regression models, *F* tests to determine regression relationships were unable to be performed. As a result, likelihood ratio tests (Kutner et al., 2008, p. 580-582) were conducted to determine the significance of the nested fixed effect models using the anova procedure in R (R Development Core Team, 2009) as described by Baayen et al. (2008). The likelihood ratio test statistic was calculated as follows

$$D = 2\left(\log L_F - \log L_R\right),$$

where D is twice the difference between full model REML log likelihood ($logL_F$) and reduced model REML log likelihood ($logL_R$). The values of D are distributed approximately as Chi-square with degrees of freedom equal to the difference between the number of parameters of the two models.

Table 4.4 and Table 4.5 below give the log likelihood estimates, likelihood ratio test statistics and associated p-values for the weight and ultrasound traits, respectively. For the days to weight regression (Table 4.4), all of the effects (year, pen and breed type) account for a significant amount of variation in the age response variable. Therefore including all three effects in the contemporary group definition would be appropriate.

Table 4.5 shows the effects of year and breed explain a significant proportion of the variation for both UREA and UBF ($P < 0.001$). When looking at the effect of pen on the days to UREA and UBF, mixed results were observed. Based on the estimated significance values, pen accounted for a significant proportion of variation of the days to UREA regression, however, when the regression of days to UBF are considered, the

Table 4.4. Log-likelihood estimates, likelihood ratio test statistics and associated significance values for each of the days to weight models.

Effect[1]	Reduced Model[2]	$LogL_R$[3]	$LogL_F$[3]	TS[4]	DF[5]	*P*[6]
Year	Pen, Breed	-63,722	-63,675	94.236	7	< 2.2e-16
Pen	Year, Breed	-63,681	-63,675	12.26	3	0.0065
Breed	Pen, Year	-63,705	-63,675	59.71	2	1.08e-13

[1]Predictor added to the parameters of the reduced model resulting in the full model.
[2]Predictors included in the reduced model.
[3]Log-likelihood estimates for the reduced and full models, respectively.
[4]Likelihood ratio test statistic calculated as two times the difference between the log likelihoods of the full and reduced models.
[5]Difference in model degrees of freedom between the reduced and full models.
[6]Probability of observing a test statistic greater than the one reported the table. Based on a χ^2 distribution with degrees of freedom equal to the difference in model degrees of freedom between the reduced and full models.

Table 4.5. Log-likelihood estimates, likelihood ratio test statistics and associated significance values for each of the days to ultrasound rib eye area and ultrasound back fat models.

Effect[1]	Reduced Model[2]	$LogL_R$[3]	$LogL_F$[3]	TS[4]	DF[5]	*P*[6]
Ultrasound Rib Eye Area						
Year	Pen, Breed	-45,716	-45,532	368.99	7	< 2.0e-16
Pen	Year, Breed	-45,540	-45,532	17.00	3	0.00071
Breed	Pen, Year	-45,640	-45,532	216.03	2	< 2.2e-16
Ultrasound Back Fat						
Year	Pen, Breed	-48.050	-47,950	200.49	7	< 2.0e-16
Pen	Year, Breed	-47,952	-47,950	4.11	3	0.2498
Breed	Pen, Year	-48,548	-47,950	1196.83	2	< 2.2e-16

[1]Predictor added to the parameters of the reduced model resulting in the full model.
[2]Predictors included in the reduced model.
[3]Log-likelihood estimates for the reduced and full models, respectively.
[4]Likelihood ratio test statistic as calculated as two times the difference between the log likelihoods of the full and reduced models.
[5]Difference in model degrees of freedom between the reduced and full models.
[6]Probability of observing a test statistic greater than the one reported the table. Based on a χ^2 distribution with degrees of freedom equal to the difference in model degrees of freedom between the reduced and full models.

amount of variation accounted for by the predictor of pen is not high (P = 0.2498) suggesting the effect of feedlot pen could be excluded from the definition of contemporary group.

If the random animal effect is removed from the model for regression of age on UBF resulting in a simple linear regression of age on UBF, typical Type III tests of significance can be calculated using the Anova (found in the car package) and lm procedures in R (R Development Core Team, 2009). When the effect of pen is added last to the regression of age on UBF, the resulting p-value for the Type III F test of adding pen to the regression containing breed type and year is equal to 0.022. This p-value is an

indication that feedlot pen does influence age that is adjusted for UBF in the absence of a random animal effect.

Given the overall goal of the study is the genetic evaluation of differences between sires for days to finish traits, the inclusion of pen in the contemporary group definition for UBF has the effect of dividing each year x breed groups into each of the four pens. Including pen in the contemporary group definition has the effect of distributing sires more evenly across contemporary groups, with only two sires having progeny in a single contemporary group. From a genetic evaluation perspective, large contemporary group sizes with large numbers of offspring from a given sire is ideal. However, for the sake of simplicity, the same contemporary group definition should be used across all the ultrasound traits.

Results of these preliminary analysis of variance studies suggest contemporary groups for the weight and ultrasound traits should consist of year of feeding period (8 total years), feedlot pen (a total of 4 different pens) and breed type (Angus, Charolais, and Charolais cross). Using these three effects will give the most complete contemporary grouping strategy given the information provided in the data and be uniform across traits. Summary statistics for the number of animals represented per contemporary group are shown below in Table 4.6. Forming contemporary groups in this manner resulted in 62 unique contemporary groups for the final weight and ultrasound outcomes, averaging 21.5 and 21.45 animals per contemporary group for weight and ultrasound, respectively.

Table 4.6. Summary statistics outlining the number of animals represented per contemporary group definition.

	Weight[1]	Ultrasound[2]
N	62	62
Mean	21.5	21.45
SD	9.63	9.62
Min	3	3
Max	42	42

[1]Summary statistics obtained from the weight data.
[2]Summary statistics obtained from the ultrasound data.

Data sifting

Data sifting is typically performed when building genetic evaluations. The goal of sifting is to reduce the size of the final data files by removing incorrect or nonsensical data as well as data that contributes no information to the genetic evaluation. Such data can be obviously incorrect weights, or data coming from contemporary groups where there is no variation in observations within the group (single animal contemporary groups, or all animals have the same observation).

The data set used in the evaluation was extremely complete, quite different than what would be expected from a field dataset. Even so, various data sifting methods were implemented to create useable final data files for both ultrasound and weight traits. Individual animal records missing either the record itself or the date of recording were removed. In order to form the most complete contemporary group possible, year of measurement, breed type and feedlot pen all must have been recorded. Animals identified as being "sick" at the time of measure had that specific observation excluded. Additionally, there were 39 animals from the year 1999 – 2000 missing birth dates, resulting in unknown ages at measure, which were subsequently removed from the analysis.

These data sifts resulted in final weight and ultrasound data files of 14,325 weight observations representing 1,333 unique individual animals and 9,551 ultrasound observations representing 1,330 unique individual animals. The difference between the number of unique animals for weight and ultrasound is due to 3 animals missing ultrasound observations in the raw data files.

Summary Statistics

Summary statistics based on each of the final, post-sifting data sets are shown in the tables below. Table 4.7 gives summary statistics for the final weight data for all animals as well as summary statistics for weight divided by breed type.

The differences between the mean weights across breeds are negligible. Upon initial examination of the standard deviations, Angus and Charolais appear to have more variability in their weight data when compared to the Charolais cross animals. Due to the fact the Charolais cross animals all have missing birth weight records, as evidenced by the minimum weight observation for the Charolais cross animals, explains the difference in the variability of the data.

Table 4.7. Weight observation[1] summary statistics all animals, as well divided into breed type.

	All Animals	Angus	Charolais	Charolais Cross[2]
N	14,325	5,668	5,466	3,191
Mean	404.16	389.22	405.25	428.83
SD	153.00	163.86	162.17	106.54
Min	27.22	27.22	32.66	161.03
Max	863	712	772	863

[1]Includes birth weight and weaning weight observations.
[2]Steers with purebred Charolais sires and dams of unknown breed type.

Summary statistics for the ultrasound traits are shown below in Table 4.8. Looking at the ultrasound summary statistics, there are virtually no differences between the Angus, Charolais, and Charolais cross animals for UREA, averaging 69.59, 72.55, and 65.59 cm^2, respectively. With respect to UBF, Angus animals tended to carry approximately twice the fat cover with an average of 11.58 mm of fat when compared to the Charolais (5.94 mm) and the Charolais cross (5.32 mm) animals.

Table 4.8. Ultrasound Rib Eye Area and Ultrasound Back Fat summary statistics for all animals and divided into individual breed types.

	All Animals		Angus		Charolais		Charolais Cross[1]	
	UREA[2]	UBF[3]	UREA[2]	UBF[3]	UREA[2]	UBF[3]	UREA[2]	UBF[3]
N	9,551		3,217		3,424		2,910	
Mean	69.43	7.65	69.59	11.58	72.55	5.94	65.59	5.32
SD	13.16	4.16	11.91	3.96	12.67	2.47	14.02	2.53
Min	28.89	1.05	40.17	2.54	33.29	1.05	28.89	1.11
Max	129.54	30.47	108.57	30.47	129.54	15.75	111.96	16.39

[1]Steers with purebred Charolais sires and dams of unknown breed type.
[2]Ultrasound Rib Eye Area
[3]Ultrasound Back Fat

Scatter plots of the data points from the final data files are shown in the subsequent figures. Figure 4.1 contains the scatter plot of weight versus age (days) for all animals in the final weight data set. Birth weight observations are clustered on day one, and there appears to be two separate slopes within the data cluster. The breakpoint between the two different slopes appears to be around 250 – 300 days of age, which can be explained by the changing of the rations from the backgrounding / transition ration to the finishing ration.

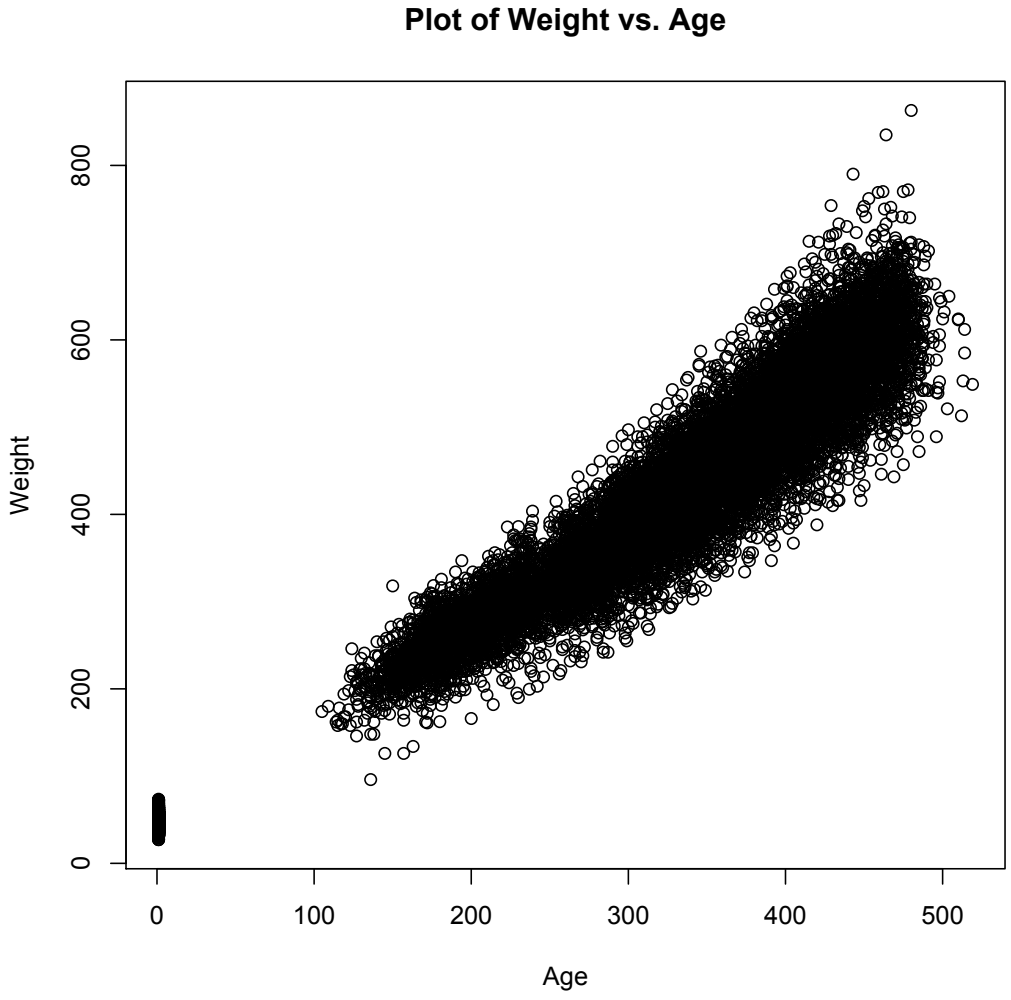


Figure 4.1. Plot of weight versus age (days) for all animals in the final weight data set.

Figure 4.2 and Figure 4.3 are scatter plots of the ultrasound traits versus day of age. In Figure 4.2 UREA is plotted against age of measurement. The figure shows an increasing trend for UREA over time, although it is neither as uniform nor as steep as the trend for increasing weight. This slower growth rate of UREA compared to the weight traits has been previously documented throughout the literature (Butterfield and Berg, 1966; Jones et al., 1980a; Cleveland, 2006).

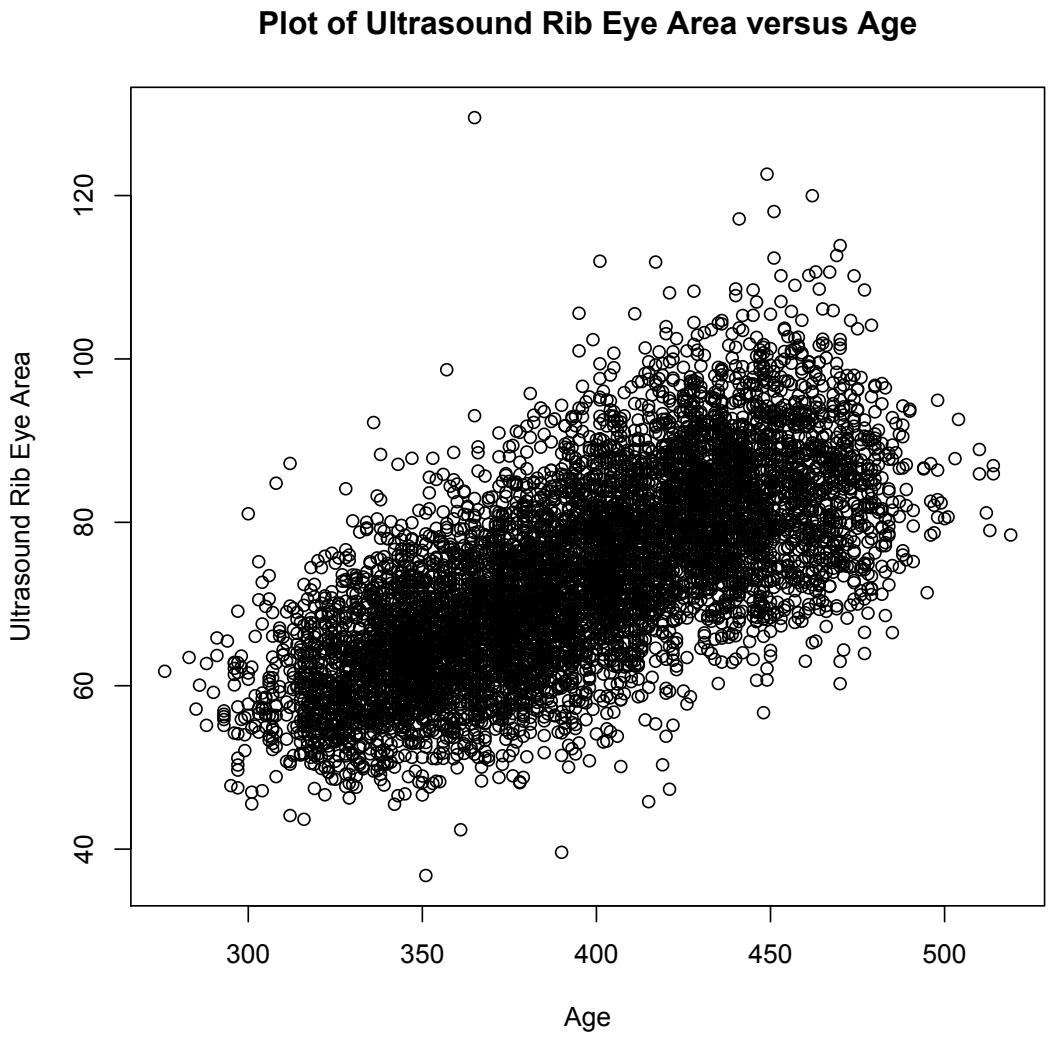


Figure 4.2. Plot of Ultrasound Rib Eye Area versus age (days) for all animals in the final ultrasound data set

Figure 4.3 is the scatter plot of UBF versus day of age. Greater variability appears to exist for UBF as opposed to both weight and UREA, confirming previously reported findings (Jones et al., 1980b). Variability among UBF observations increases as animals increase in age. This increase in variability can be explained by the fact that

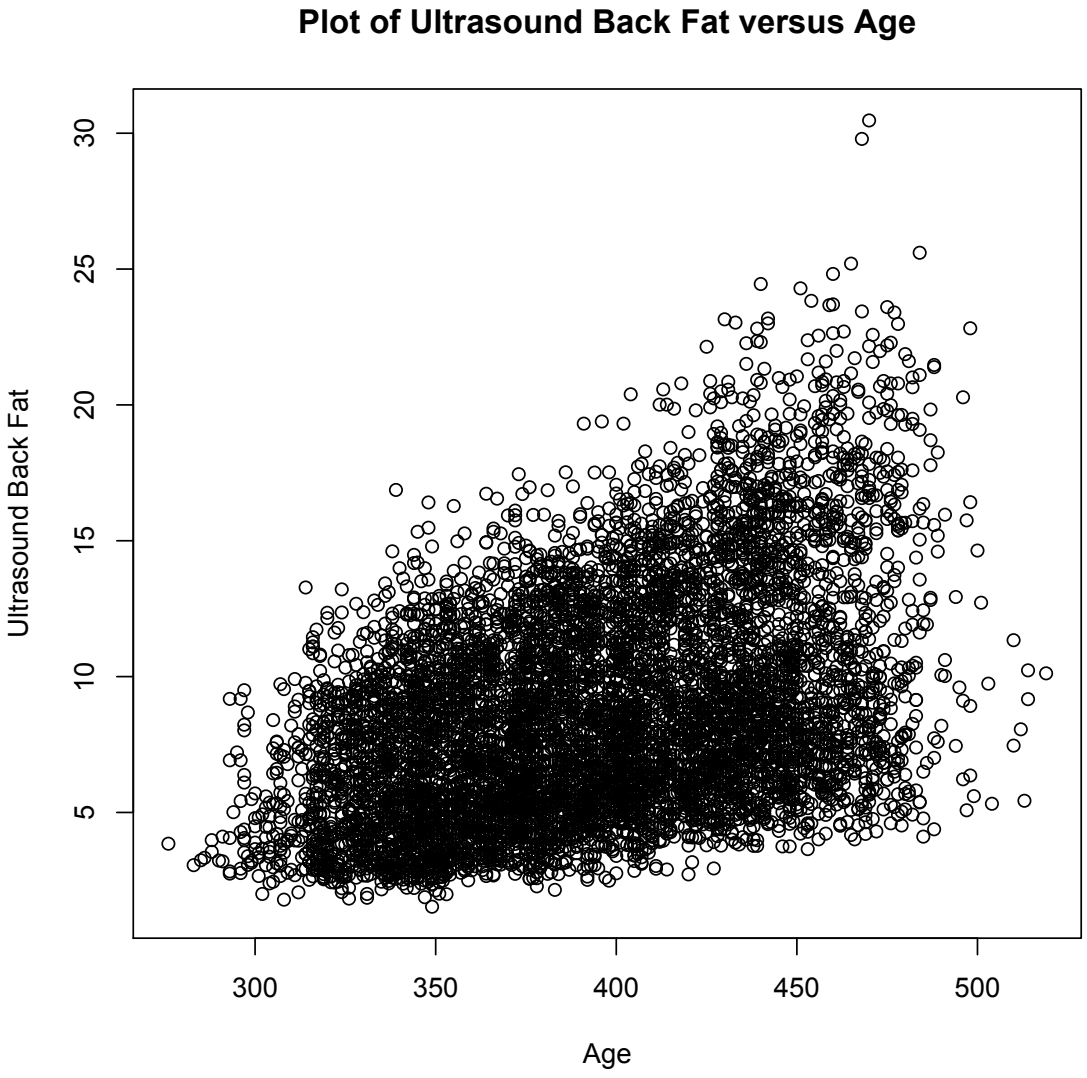


Figure 4.3. Plot of Ultrasound Back Fat versus age (days) for all animals in the final ultrasound data set.

subcutaneous fat is late developing therefore at younger ages less variability would be expected rather than at earlier ages (Berg et al., 1979; Jones et al., 1980a), which can be inferred from this plot.

Age. Given the nature of the "Days to Finish" traits, the distribution of age at measurement is just as important as the typical phenotypic observations (weight and ultrasound traits) given that age is the observation of interest. Age descriptive statistics discussed below are compiled from the final weight data file due to its completeness with respect to the number of individual animals and observations. In Table 4.9 below, summary statistics describing the distribution of individual animal ages are shown. The average age doesn't seem to differ between individual breed types. Again, the Charolais cross individuals appear to have less variability in their age distribution, but this can be explained by the absence of birth date and weight records for this breed category.

Table 4.9. Age[1] summary statistics[2] for all animals and divided into individual breed type.

	All Animals	Angus	Charolais	Charolais Cross[3]
N	14,325	5,668	5,466	3,191
Mean	310.23	298.33	302.32	344.93
SD	121.73	133.30	127.32	75.10
Min	1	1	1	150
Max	519	498	504	519

[1]Measured in day of age.
[2]Includes birth weight and weaning weight observations.
[3]Steers with purebred Charolais sires and dams of unknown breed type.

In Figure 4.4, the frequency and distribution of the ages in the final weight data file are shown. Looking at the histogram, three different distributions appear to exist. The first distribution is day 1, representing the birth weight observations present in the

data file. The second distribution appears to be centered somewhere around 200 days of age, representing the weaning weight observations included in the data file. The last remaining distribution is centered at about 350 days of age, the mean age of the observations with birth and weaning records removed. The distribution of ages is an important consideration to make when analyzing days to finish traits.

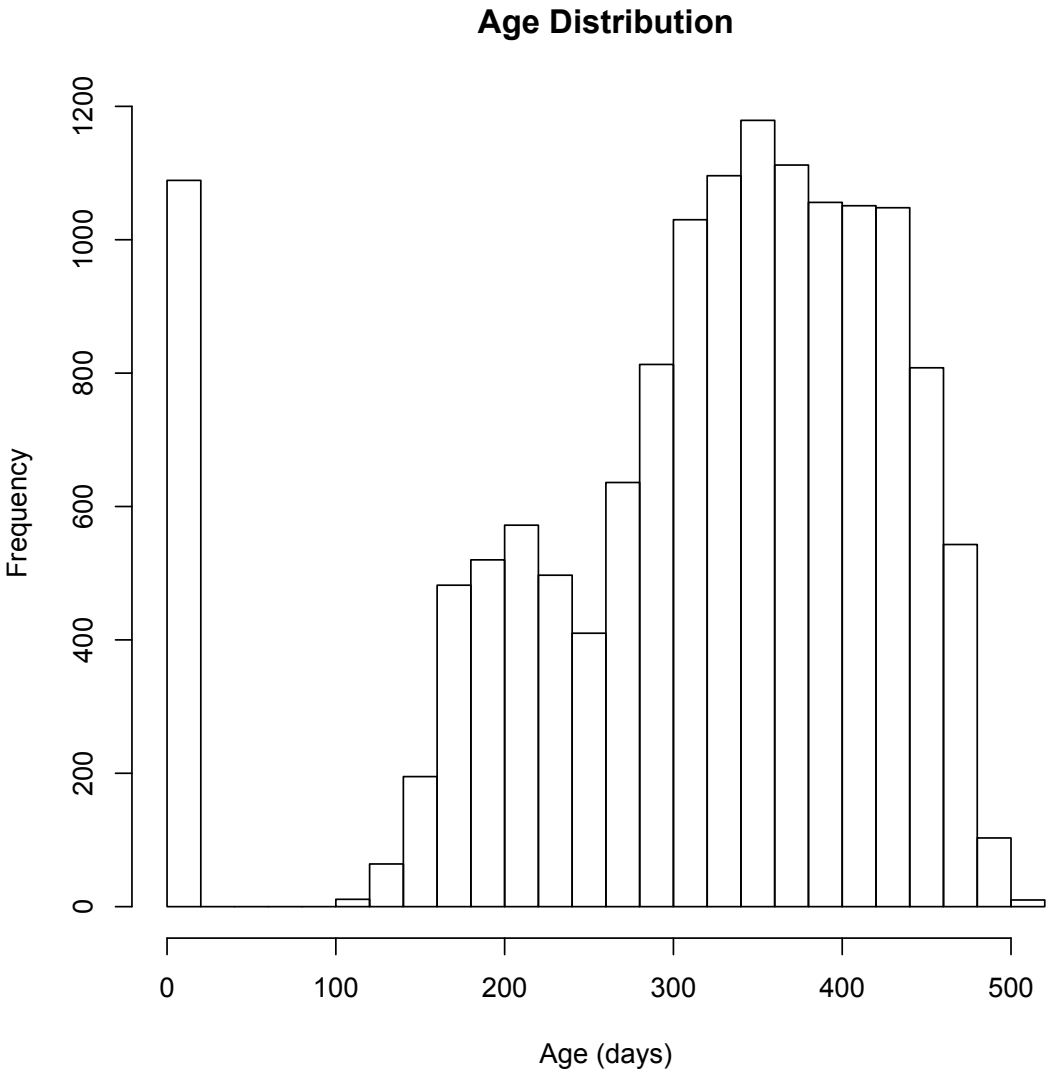


Figure 4.4. Frequency and distribution of day of age from final weight data file.

Pedigree. The complete pedigree was built using all animals in the data set; prior to performing any data sifts. Beginning with a list of parents of the animals in the original dataset, the Animal Breeder's Tool Kit (Golden et al., 1992) was used in conjunction with the complete pedigrees obtained from the Canadian Angus Association and Canadian Charolais Association to add ancestral animals generation by generation. The pedigree was deemed complete at the point where with no additional animals added given subsequent generations. This point was reached after 16 and 22 generations for the Charolais and Angus animals, respectively.

Building the pedigrees in this manner resulted in an Angus pedigree of 5,284 individual animals representing 1,685 unique sires and 3,011 unique dams. The Charolais pedigree consisted of 8,175 individual animals resulting in 2,402 unique sires and 4,986 unique dams. The average inbreeding coefficients for both the Angus and Charolais pedigrees were 0.024 and 0.020, respectively. On a side note, even though the Angus animals were slightly more inbred than the Charolais, the Charolais did have the highest inbred animal with an inbreeding coefficient of 0.344 (the maximum inbreeding coefficient for the Angus animals was 0.281). After the individual breed pedigrees were built to completion, they were combined to form one large pedigree to be used in the genetic evaluation of days to finish. Combining the two pedigrees resulted in 13,459 individual animals representing 4,087 unique sires and 7,997 unique dams. The average inbreeding coefficient for all animals in this combined pedigree was 0.022.

The full pedigree was then used to form the individual pedigrees to be used in the genetic evaluation of days to finish.

Literature cited

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. and Lang. 59:390-412.

Beef Improvement Federation. 2002. Guidelines for Uniform Beef Improvement Programs. 8th ed. Omaha, NE.

Berg, R. T., S. D. M. Jones, M. A. Price, F. Fukuhara, R. M. Butterfield, and T. Hardin. 1979. Patterns of carcass fat deposition in heifers, steers and bulls. Can. J. Anim. Sci. 59:359-366.

Butterfield, R. M. and R. T. Berg. 1966. Relative growth patterns of commercially important muscle groups of cattle. Res. Vet. Sci. 7:326-332.

Cleveland, M. A. 2006. Sire selection for profit in beef production. Ph D Dissertation. Colorado State University, Fort Collins, CO.

Crump, R. E., N. R. Wray, R. Thompson and G. Simm. 1997. Assigning pedigree beef performance records to contemporary groups taking account of within-herd calving patterns. Anim. Sci. 65:193-198.

Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's toolkit user's guide and reference manual. Agric. Exp. Stn. Tech. Bull. LTB92-2. Colorado State University, Fort Collins, CO.

Jones, S. D. M., M. A. Price, and R. T. Berg. 1980a. The growth and distribution of muscle in bulls and heifers of two breeds. Can. J. Anim. Sci. 60:669-675.

Jones, S. D. M., M. A. Price, and R. T. Berg. 1980b. Fattening patterns in cattle 1. Fat partition among the depots. Can. J. Anim. Sci. 60:843-850.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. Applied Linear Statistical Models. 5th ed. McGraw-Hill, New York, NY.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Ruppert, D, M. P. Wand, and R. J. Carroll. 2003. Cambridge series in statistical and probabilistic mathematics: Semiparametic Regression. Cambridge University Press. Cambridge, NY.

Williams, A. R. 2002. Ultrasound applications in beef cattle carcass research and management. J. Anim. Sci. 80:E183-E188.

CHAPTER V

DAYS TO WEIGHT ENDPOINT

Introduction

Reducing the number of required days for livestock to reach a specific weight endpoint has received very little attention throughout literature. With the exception of the swine industry this research has been almost non-existent, with only a handful of studies pertaining to beef cattle having been published going back to 1957. In summary, this research has shown a phenotypic correlation of -0.46 between the number of days to reach a perceived quality grade and net income per 45.4 kg of slaughter weight (Lindholm and Stonaker, 1957). More recently, Kuehn (2000) determined it feasible to obtain accurate variance component estimates for a linear random regression of days to finish weight using simulated data while Jubileu (2003) looked at differences between more traditional approaches such as multivariate models versus random regression techniques using Simmental weight data. Both Kuehn and Jublieu stressed the advantages of using random regression methodologies in the calculation of days to finish EPD.

Random regression allows for the calculation of EPD along any given point of the polynomial which is attractive for days to finish because each individual producer's "finish" endpoint can be different. RR has been implemented in many instances in other livestock industries for the genetic evaluation of test day records in dairy cattle (Ptak and

Schaeffer, 1993; Guo and Swalve, 1997; Brotherstone et al., 2000) growth data in pigs (Andersen and Pedersen, 1996) and beef cattle (Meyer, 1999; Legarra et al. 2004) just to name a few.

The lack of published research in this area of beef cattle genetic improvement is puzzling given the nature of "days to finish" as one of the economically relevant traits described by Golden et al. (2000). The objective of this study was to explore the feasibility of creating a days to weight (**DTW**) genetic prediction from a field data set using random regression methodologies.

Methodology

A genetic prediction for DTW was built using the previously described data set (Chapter IV), which consisted of pedigree and multiple weight observations on 1,375 animals. Two separate models were used to evaluate DTW. First, an evaluation was built using random regression methodology. Second, a more traditional repeated measures model was used to make comparisons to the RR for the purposes of model validation.

The Random Regression Model. In constructing the RR for the genetic prediction of DTW, a model building exercise somewhat similar to that described by Brommer et al. (2008) was implemented beginning with the most basic model. In this process, random effects were sequentially entered with their significance as a predictor of days to reach a weight endpoint being tested using a likelihood ratio test. All models were implemented using the statistical package ASReml (Gilmour et al., 2009).

First, the general form of the RR used in the genetic evaluation of DTW is shown in matrix form (Equation 5.1) as described by Mrode (2005).

Equation 5.1. General form for a random regression model presented in matrix notation.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Qu} + \mathbf{Zpe} + \mathbf{e}$$

In Equation 5.1, **y** represents a vector of age observations recorded on individual animals, **X** is an incidence matrix relating age observations in **y** to contemporary group and fixed regression coefficients containing weight observations for the regression of age on weight to their solutions in **b**, **Q** is an incidence matrix consisting of weight covariates (representing the random regression effects of the age on weight regression) relating the age observations in **y** to the random additive genetic regression coefficients in **u**, **Z** is a matrix of weight covariates relating the age observations in **y** to the permanent environmental random regression coefficients for each animal in **pe**, and **e** is a vector of random residuals that includes the temporary environmental effects for each observation. As the order of the random regression increases, the columns of the incidence matrix **Q** increase by one. Variances for the additive genetic, permanent environmental effects and random residuals in the model are:

$$\text{var}\begin{bmatrix} \mathbf{u} \\ \mathbf{pe} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \otimes \mathbf{G} & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{P} & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

where **A** is Wright's numerator relationship matrix (Wright, 1922), **G** is the (co)-variance matrix of the random additive genetic regression coefficients whose order is equal to the order of the polynomial in the random regression, **I** is an identity matrix whose order is equal to the total number of observations in **y**, **P** is the (co)-variance matrix of the random permanent environmental regression coefficients, and σ_e^2 is the variance of random residuals.

Following suggestions by Jamrozik et al. (1997) and Gilmour (2009) the structure of this residual variance was allowed to vary. First, models were fit using a heterogeneous residual variance structure as mentioned by Jamrozik et al., (1997). In that paper, the authors presented an equation, which allows for changing residual variance across time by enabling subsequent observations to be classified into alternate categories using Equation 5.2. An important note about this equation is it ignores the error covariance between classification levels similar to a multiple trait analysis mentioned by Arnold et al. (1992) with zero covariance between error variances.

Equation 5.2. Equation used to classify observations in random regression models to predefined residual variance classifications.

$$\text{var}[\mathbf{e}] = diag\left\{\sigma_{e_k}^2\right\}$$

Above, k is equal to the number of differing residual variances. In a standard evaluation, such as the regression of weight on age, the values of k could have some biological basis such as the designations of important weight measurement times like birth weight, weaning weight, yearling weight, etc. Here, the trait DTW is being analyzed using a regression of age on weight. No longer can these observations be divided in such a

manner because individual animal weights could be classified into separate levels of k. For example, one animal's start weight could be heavier than the next animal's second weight measured two weeks later. Therefore, somewhat arbitrary cutoff points were used to determine its value. For the regression of age on weight in the DTW evaluation, the cutoff points for the different values of k were determined by calculating the quartiles of the data set using the boxplot.stats package in R (R Development Core Team, 2009).

Second, error covariance was added to the model through the inclusion of a random regression for residuals. To implement this model in ASReml (Gilmour, 2009), a standard single residual variance is included in the evaluation; similar to a residual variance included in a repeated measures analysis. Accounting for covariance at the error level is accomplished by including a linear random regression on permanent environmental effects. Permanent environmental variance is included in genetic evaluations to account for the random non-genetic factors present between animals (Henderson, Jr., C. R., 1982) affected by multiple observations on an animal. Allowing permanent environment effects to vary with increasing weight captures the non-genetic covariance between subsequent observations on individual animals, which is essentially residual covariance.

While building up to a complete DTW genetic prediction model, the order of both the fixed and random regressions were taken into consideration. Fixed regression coefficients contained in the matrix **Xb** are important to the overall random regression model. As suggested by Schaeffer (2003), RR are intended to model deviations around the phenotypic trajectory and Gilmour (2009) suggests their order not be reduced below the order specified in the random terms. The question in this study remained what the

proper order of the fixed regression should be, and how much would that actually influence the resulting heritability estimates.

The proper order of the fixed regression of age on weight was determined by using methods similar to the forward stepwise regression procedures described by Kutner et al. (2005). They describe this procedure as the incremental inclusion of several fixed effects, with the ultimate selection of the effect with the largest t^* statistic. Here, the order of the regression was chosen by a series of sequential models that fitted incrementally higher order polynomials for the fixed regression within a given random regression order. Partial *F*-tests of significance, as described by Kutner et al. (2005) and Gilmour (2009), were used to determine the highest significant term or order of the fixed regression coefficients within a given order for the random regression polynomial. Beginning with a polynomial whose order is equal to the order of the random polynomial and working toward a more complex model, the highest coefficient was tested as to whether or not it was significantly different from zero. *F*-statistics used in the partial *F*-test were constructed by squaring the *t*-statistic calculated by taking the ratio of the estimate to its standard error, using Equation 5.3 (Gilmour, 2009).

Equation 5.3. Equation for constructing a t-distribution test statistic to test whether or not the regression coefficient b_k is equal to zero.

$$t^* = \frac{b_k}{s\{b_k\}}$$

Above, t^* is the test statistic for testing whether or not β_k is equal to 0, b_k is the solution for the highest order or kth term and $s\{b_k\}$ is the standard error for the highest order term obtained from the ASReml solution output files. This t^* was then converted to an *F*

statistic by squaring $\left(F^* = \left(t^*\right)^2\right)$. F^* is the F-value for adding the last effect to the whole model and is distributed as $F^* \sim F(1, n-p)$. Allowing ASReml to estimate the denominator degrees of freedom $(n-p)$ using the Kenward & Roger approximation method (Kenward & Roger, 1997), this highest order polynomial term can be tested as to its difference from zero. Using a pre-determined significance value of 0.05, if the highest order term was dropped, the reduced model was then refit testing the next lowest term.

Once the proper fixed regression polynomial (within a given random regression order) was determined, incrementally higher orders of random regressions were fit until additional higher order polynomials no longer statistically accounted for additional variation in trait "days". Likelihood ratio tests (**LRT**) were used to conduct these tests of significance for each nested random effect or random polynomial term. The LRT test statistic as described by Beckman et al. (2007) and Brommer et al. (2008) is shown below in Equation 5.4.

Equation 5.4. Likelihood ratio test statistic for testing differences between equivalent fixed effect models where there are differences in the number of parameters between the two models.

$$D = 2\left|\log L_F - \log L_R\right|$$

In the LRT test statistic equation above, D is the absolute difference between the full model REML log-likelihood $(\log L_F)$ and the reduced model REML log-likelihood $(\log L_R)$. Here, the null hypothesis stated the full model did not significantly fit better than the reduced or simpler model. This test statistic is distributed approximately as χ^2

with degrees of freedom equal to the difference in the number of parameters fit between the full and reduced models.

Likelihood ratio tests are only valid if, as mentioned above, the parameters of the full model fully encompass the parameters of the reduced model and if the fixed effects of both the full and reduced model are the same. Therefore, comparisons pertaining to the order of random polynomials were only made within equivalent fixed effect models (within the same order of the fixed regression polynomials). For example, if the quadratic term of the random regression polynomial was tested for significance, this was done within each significant fixed regression polynomial order.

Orthogonal Legendre polynomials were used as the base random regression function. Legendre polynomials were chosen because they help to reduce the correlation between successive observations (Kirkpatrick et al., 1990) with the realization that these polynomials tend to place a large emphasis on observations at the extreme ends of the data range (Meyer, 2005). This detractor to the use of these polynomials, as discussed by Meyer (2005), is compounded with higher orders of fit. Also, these polynomials tend to return estimates of genetic variance that are much higher at the beginning and ends of the data range than in the middle (Schaeffer & Jamrozik, 2008).

Weight observations used in the regression of age on weight were standardized using Equation 3.6. Standardization is a necessary restriction with the use of Legendre polynomials because these types of polynomials, specifically, are only defined over the interval of -1 to 1 (Kirkpatrick et al., 1990). Normalized Legendre polynomials for each of the regression orders (linear through quartic) were calculated using the standardized weights in conjunction with Equations 3.7 though 3.9. These polynomials are shown in

Table 3.4. All calculations were performed internally within ASReml using the leg(*v*, *n*) model function. This function forms *n* + 1 Legendre polynomials of order 0 (intercept), 1 (linear), … to *n* from the values in *v*. For instance, to model a linear regression using the predictor variable weight, the ASReml model function would be leg(weight, 1). Example ASReml command files for calculating both linear and quadratic DTW are included in Appendix II.

Estimates of variance obtained from a RR are not interpreted in the same manner as similar estimates obtained from a conventional multiple trait model. Estimating (co)variances for genetic evaluations using RR result in genetic and phenotypic variances for the shape of the polynomial. This means, that for a linear random regression, the resulting variance estimates will be estimates for the intercept and slope of the random polynomial. These estimates can be used to calculate heritabilities for the curve parameters and correlations between the shape parameters. Through a simple conversion, observed variance estimates can be calculated for the range of data. This conversion is done using the formula

$$\mathbf{G_0} = \mathbf{\Phi G_{rr} \Phi}^T$$

where G_0 is the observed genetic (co)variance matrix between the orthogonalized weights in F. G_{rr} is the random regression genetic (co)variance matrix as described by Schaeffer (2003).

The Repeated Measures Model. In order to compare the results obtained from the RR evaluation of DTW, a repeated measures model was used to estimate heritability

and predict breeding values for the same weight data. This model is presented in matrix form in Equation 5.5.

Equation 5.5. General form for a repeated measures model presented in matrix notation.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{e}$$

In the equation for the repeated measures model above, **X**, **Z**, and **W** are incidence matrices relating the repeated age observations in **y** to fixed contemporary group effects (**b**), random additive genetic effects (**u**) and random permanent environmental and non-additive genetic effects (**p**), with **e** defining a vector of random residual errors. Age observations in y are assumed to have the mean **Xb** while random effects in u and p have means of zero with variances represented by:

$$\text{var}\begin{bmatrix} \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & 0 & 0 \\ 0 & \mathbf{I}\sigma_p^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

where σ_u^2, σ_p^2, and σ_e^2 are the variances of random additive animal genetic effect, random permanent environmental effect, and random residual error, respectively. **A** is Wright's numerator relationship matrix, and **I** is an identity matrix with an order equal to the number of observations in **y**.

In order to obtain heritability estimates and resulting EBV for a DTW evaluation from the repeated measures model, and to make proper comparisons to a corresponding endpoint from the DTW RR, data used in the repeated measures model were adjusted to a

constant weight endpoint of 500 kg prior to evaluation. This adjustment allows the direct comparison of the DTW from the RR and the repeated measures model.

Results and Discussion

For the evaluation of DTW, the final weight data set described in Chapter 4 was subset to allow the proper modeling of the random regression polynomials. Depending on the order of random polynomial (linear through quadratic), individual animals were removed from the data set if they lacked the sufficient number of observations to fully fit the current line. Animals were removed from the final data set if they had fewer than five, four, three or two observations for the quartic, cubic, quadratic and linear polynomials, respectively. In order to make LRT comparisons, using forward stepwise regression procedures, the following algorithm was implemented.

1) Animals with fewer than two observations were removed from the data set.
    a. A linear age on weight random regression model was implemented.
2) Animals with fewer than three observations were removed from the data set.
3) Random regressions were modeled using the data set containing those animals from step two.
    a. First, the quadratic random regression model was fit
        i. Fixed regression orders were increased incrementally until the addition of higher order terms was found to be not significant.
    b. Second, the linear random regression model was fit using the fixed regression polynomial order chosen in step 3a.

c. LRT between the linear and quadratic random regression models were performed to determine whether the quadratic terms should be dropped from the regression. Significance levels were set at $P = 0.05$.

4) If the results from the LRT suggested the quadratic term should not be included in the model, the process was ended.

5) If the results from the LRT suggested the quadratic random regression term should be included in the model, the algorithm was repeated back at step two removing animals with fewer than the four observations needed to fit a cubic polynomial and its significance was tested.

   a. A quartic random regression is the highest order polynomial considered for inclusion due to restrictions in the required number of observations needed by individual animals.

Sub-setting the data in this manner resulted in four individual weight data sets whose summary statistics are shown below in Table 5.1.

Table 5.1. Summary statistics for the average number of observations for each of the individual regression data sets, ranging from all data to requiring individual animals to have five or more observations corresponding to that needed for a quartic regression.

	All Data	5 or more	4 or more	3 or more	2 or more
N animals	1,324	1,150	1,311	1,317	1,323
N	7,633	6,958	7,602	7,620	7,632
Mean	5.77	6.05	5.80	5.79	5.77
Minimum	1	5	4	3	2
Maximum	9	9	9	9	9

As a benchmark for illustrating relative sizes in the subsets of data, summary statistics for all available data was included in the table as well. A quartic random regression requires individual animals to possess five or more observations. This requirement results in 675 fewer observations representing 174 animals (13% of the total number of animals) as compared to all data. In fact, an entire year's (2002 – 2003) worth of data is removed from the evaluation. Reducing the random regression order to a cubic polynomial, the observation requirement can be lowered to four or more observations. Here, only 31 observations from 13 different animals are removed. As random regression polynomial orders are reduced even further, the number of useable observations increases, and once the linear random regression is modeled, only one individual does not possess the minimum number of observations (two observations are required for a linear random regression).

Grouping observations for the purpose of specifying differing residual variance structures by biological definition didn't make sense for this evaluation. Given the nature of the regression of age on weight, where age is the variable of interest, no obvious delineation points exist in order to classify individual animal records into residual variance sub-groups. An initial attempt at classifying observations into residual variance categories was conducted by visual inspection. If there were naturally occurring break points in the distribution of weight observations, these break points could be used to classify weight into their respective residual variance sub-groups. A histogram of weight observations (Figure 5.1) was created to view the distribution of these weight observations. As illustrated in Figure 5.1, the distribution of weight observations appeared to be continuous with no obvious breaks.

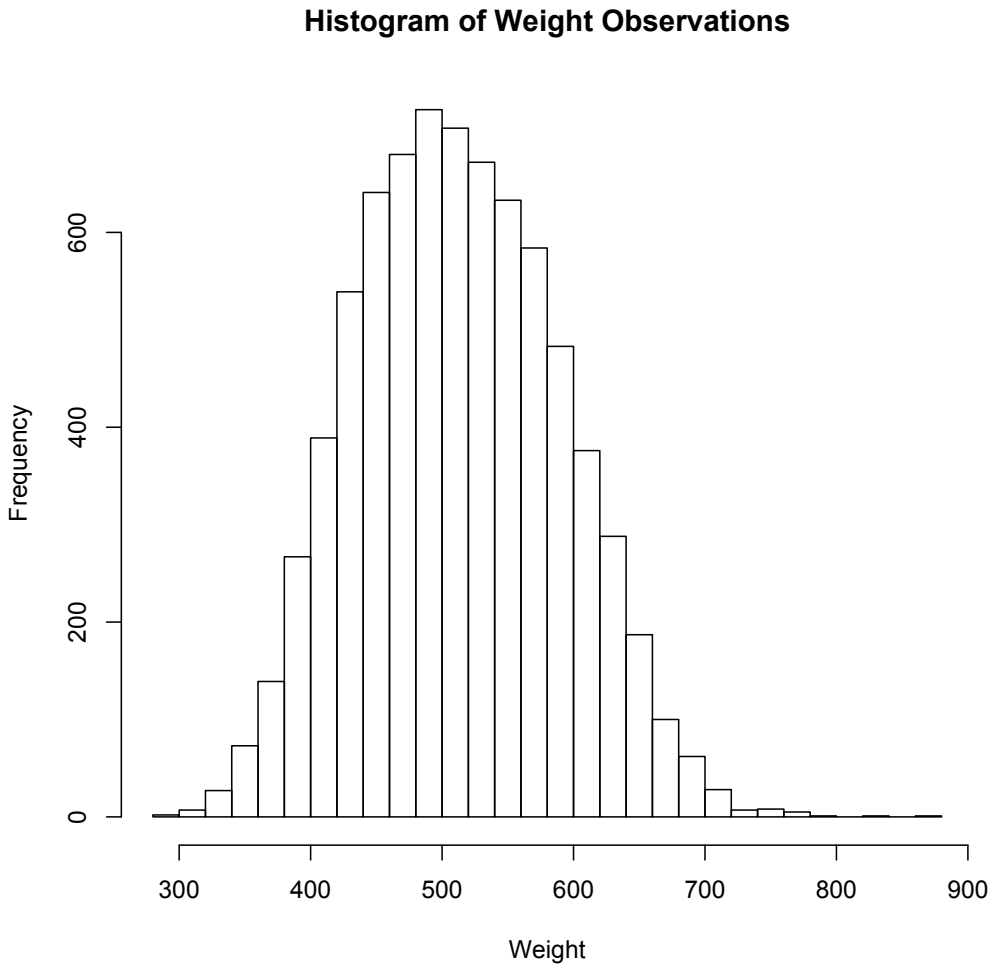


Figure 5.1. Histogram of weight observations for the days to weight genetic prediction.

Given the distribution of these weight observations, the decision was made to define the residual variance categories according to the quartiles of the weight distribution. Quartiles are used to calculate the variation statistic "inner quartile range", an interval that contains roughly 50% of the data. An attractive property of using this inner quartile range is that outliers in the data have no influence. The center of the inner quartile range is identified by the median observation. The first and third quartiles, along with the median, divide the data set into four sub-groups, each containing approximately 25% of the data. Calculation of the quartiles resulted in break points of 456, 509 and 568

pounds. These points resulted in 1,822; 1,662; 1,717 and 1,757 weight observations to be placed in the first, second, third and fourth residual categories, respectively.

Linear Random Regression. A linear RR was implemented using the weight data set sifted to contain only those individuals with two or more observations, described above in Table 5.1. This data set contained 7,632 age and weight observations on 1,323 individual animals that resulted in an average of 5.77 observations per animal. Age and weight summary statistics for this data set are shown below in Table 5.2.

Table 5.2. Age and weight summary statistics for the data used in the linear random regression of age on weight for the calculation of the days to weight genetic prediction.

	Age[1]	Weight
N	7,632	7,632
Average	395.9	513.0
Variance	2,036.5	5,978.9
Minimum	276	293
Maximum	519	863

[1]Age is reported in days.

As mentioned in Chapter IV, contemporary groups were formed on the basis of feedlot pen, year of test and breed composition. Formation of contemporary groups in this manner resulted in 62 unique groups containing an average of 21.3 animals per group. For the purpose of estimating variance components, a 4-generation pedigree was built from this final data file. Formation of the pedigree in this manner resulted in a stacked pedigree that contained 5,414 individual animals, with 1,386 unique sires and 2,705 unique dams. The average inbreeding level for the animals in this pedigree was 1.5% with minimum and maximum inbreeding levels of 0% and 25%, respectively.

Direct genetic (co)variances were estimated for both the intercept and linear terms for the linear DTW random regression for two separate models. First, DTW was modeled using a heterogeneous residual variance (**HRV**) structure as presented above in Equation 5.2. Here, permanent environment was included in the model as a constant or intercept variance to account for the environmental effects that permanently influence the repeated observations on individual animals. Second, DTW was modeled by including error covariance in the model through the implementation of a linear random regression on residuals (**LRRR**). Here, a single error variance was included for the entire trait "days" and permanent environmental variance was estimated by including in the model a linear random regression on permanent environmental effects.

Beginning with a linear fixed regression model, variance components from both the HRV and LRRR models were calculated for models containing increased fixed regression orders. Test statistics and associated p-values corresponding to the fixed regression tests of significance are shown below in Table 5.3. For both residual variance models, the linear fixed regression polynomial was sufficient and models containing higher order fixed regression polynomials did not significantly account for any additional variation in the trait as evidenced by the non-significant quadratic terms. The linear fixed regression coefficients were very similar between the two models, only in the model containing the random regression on residuals was this value estimated more precisely as evidenced by the smaller standard error.

Table 5.3. Best linear unbiased estimates and associated significance values for the different fixed regression polynomial orders obtained from the linear random regression of age on weight using both heterogeneous residual variance and linear residual random regression.

Polynomial Order[1]	Last Coefficient[2]	SE[2]	F[3]	*P*-value
	Heterogeneous Residual Variance			
1	157.1	1.54	10433.7	0.00
2	-0.414	0.60	0.481	0.49
	Linear Residual Random Regression			
1	157.0	1.41	12415.8	0.00
2	0.004	0.61	0.004	0.95

[1]Order of the fixed regression polynomial
[2]Best linear unbiased estimate and standard error of the highest order term from the fixed regression.
[3]F distribution test statistic and associated *P*-value.

Variance estimates obtained from both linear random regression models are shown below in Table 5.4 and Table 5.5. There is very little difference between estimates obtained from the linear and quadratic fixed regressions for both the HRV model (Table 5.4) and LRRR model (Table 5.5), as was suggested by the significance tests shown above in Table 5.3.

Comparing both residual variance models, more differences are apparent. First, estimates of residual variance (43.67 days^2) obtained from the LRRR is very close to the average of the four residual variances (43.5 days^2) obtained from the HRV model. After the residual variances, the similarities between the two models end. Estimates of genetic variance obtained from the model containing the random regression on residuals is smaller than those obtained from the HRV model by a magnitude of 200 days^2, 140 days^2 and 100 days^2 for the intercept, intercept and linear covariance and linear variance, respectively.

Table 5.4. Variance estimates (SE) obtained from a linear random regression model for days to weight using heterogeneous residual variances.

	1[1]	2[1]
LogL[2]	-1520.23	-1520.56
Intercept[3]	1323 (107.4)	1311 (108.1)
Int, Lin[4]	605.1 (34.2)	595.9 (35.3)
Linear[3]	406.6 (22.5)	400.9 (23.0)
PE[5]	278.7 (65.8)	280.8 (66.2)
R11[6]	33.42 (1.46)	33.40 (1.46)
R22[6]	43.62 (1.78)	43.63 (1.78)
R33[6]	54.14 (2.16)	54.26 (2.17)
R44[6]	42.82 (1.83)	42.96 (1.84)

[1]Order of the polynomial used as the mean regression of age on weight.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic random regression variance estimates for the intercept and linear random regression terms. (SE)
[4]Genetic covariance between the intercept and linear random regression terms. (SE)
[5]Permanent environmental variance (SE)
[6]Residual variance estimates corresponding to each of the four weight quartiles. R11, R22, R33, and R44 are the residual variances for the first, second, third and fourth quartiles, respectively. (SE)

In the model containing the LRRR, permanent environmental intercept variance was increased by approximately 151 days^2, which suggests that much of the variation attributed to the additive genetic effects in the HRV model has been re-partitioned to permanent environment effects and ultimately to the error covariance between observations. The model containing the HRV has a higher REML log-likelihood estimate (-1520.23 versus -1551.41) suggesting it is a more appropriate model given the data. However, this HRV model may not be appropriately classifying observations to residual variance subsets, and it is definitely not accounting for the error covariance between observations.

Table 5.5. Variance component estimates obtained from a linear random regression model for days to weight using linear residual random regression.

	1[1]	2[1]
LogL[2]	-1551.41	-1551.9
Intercept[3]	1123 (172.7)	1124 (173.0)
Int, Lin[4]	462.6 (80.1)	463.3 (80.5)
Linear[3]	307.9 (49.3)	308.4 (49.5)
PE Intercept[5]	429.8 (128.0)	429.9 (128.1)
PE Int, Lin[5]	110.7 (60.0)	111.0 (60.1)
PE Lin[5]	76.55 (38.4)	76.79 (38.45)
Residual[6]	43.67 (0.88)	43.67 (0.88)

[1]Order of the polynomial used as the mean regression of age on weight.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic variance for the intercept and linear random regression terms (SE).
[4]Genetic covariance between the intercept and linear random regression terms (SE).
[5]Permanent environmental intercept variance, intercept and linear covariance, and linear variance (SE).
[6]Residual variance (SE).

In an effort to further verify the linear fixed regression is the appropriate order to be included in the evaluation, the observed DTW genetic variance was calculated for each 20 kg increment within the range of weight observations (293 kg to 863 kg). These estimates are then plotted versus their corresponding weight endpoint. Figure 5.2 contains the estimates of genetic variance from the HRV model while Figure 5.3 contains those estimates from the LRRR model. Both plots show no difference between the observed estimates of resulting from each of the differing fixed regression models further indicating the linear fixed regression seems to be the most appropriate order for the random regression of days to weight endpoint.

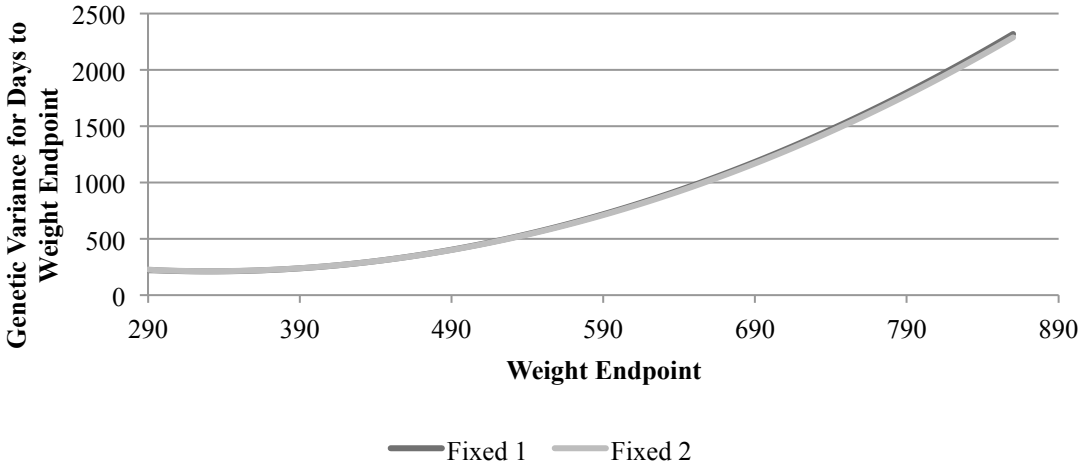


Figure 5.2. Plot of genetic variance obtained from the heterogeneous residual variance linear random regression of age on weight for the number of days to reach a specific weight endpoint for both linear and quadratic fixed regression orders.

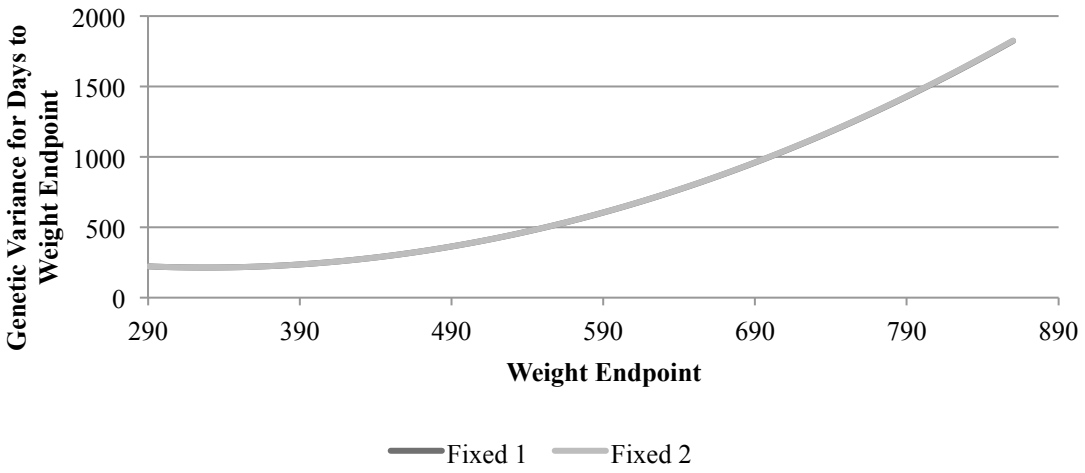


Figure 5.3. Plot of genetic variance obtained from the linear residual random regression using a linear random regression of age on weight for the number of days to reach a specific weight endpoint for both linear and quadratic fixed regression orders.

In order to make comparisons between the HRV model and LRRR model observed genetic variance, permanent environmental variance, phenotypic variance and heritability were calculated, from the same weight observations as mentioned in the previous paragraph, and shown below in Figures 5.4, 5.5, 5.6 and 5.7, respectively.

Figure 5.4 contains the plot of observed genetic variance for days with increasing weight. Both models predicted similar genetic variances for the lighter weight endpoints, but as weight increased, particularly above 590 kg, the model containing the HRV estimated higher genetic variance than the model containing the random regression on residuals. This is most likely due to the HRV model not properly accounting for the error covariance structure and attributing those differences to genetic variability.

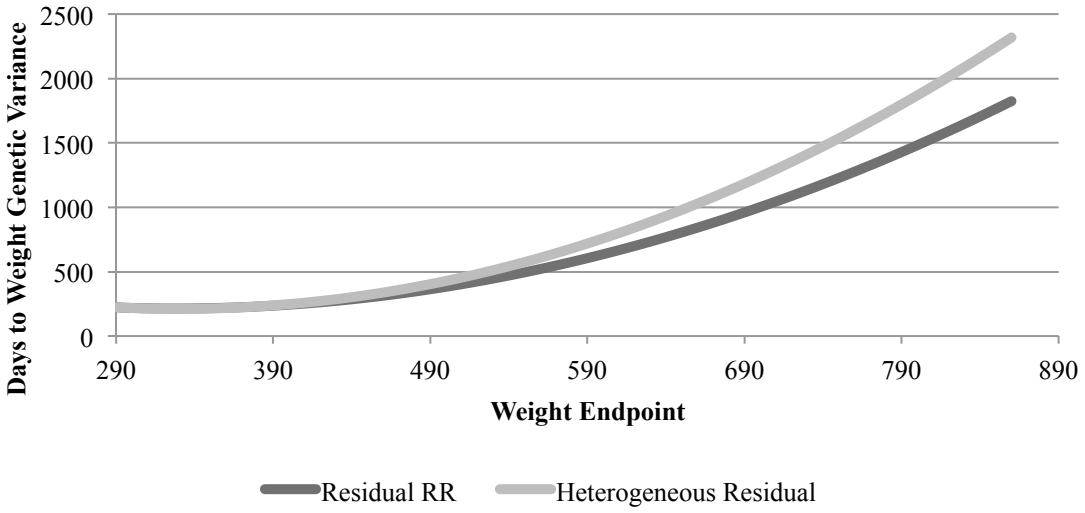


Figure 5.4. Plot of observed days to weight genetic variance obtained from the linear random regression of age on weight using both heterogeneous residual variance and random regression on residuals.

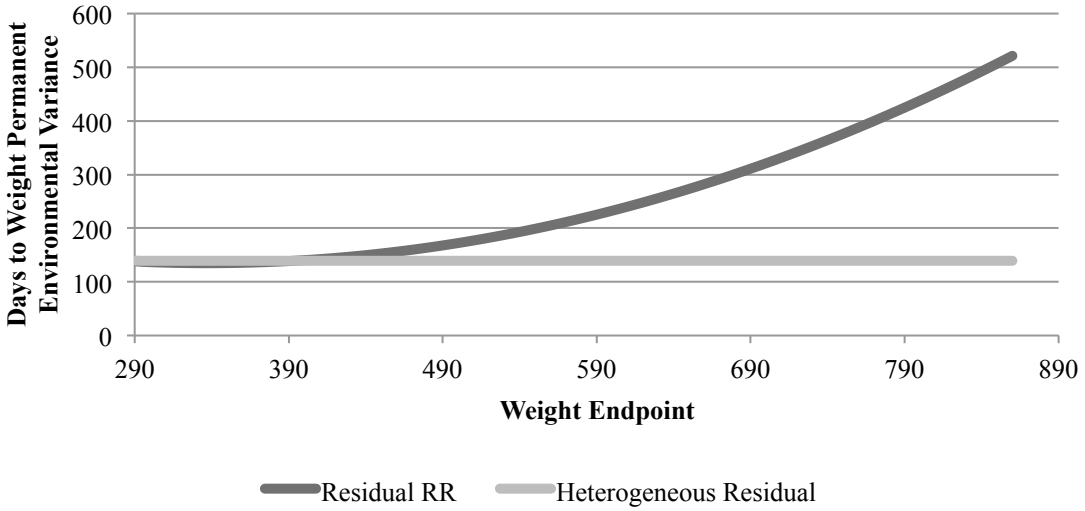


Figure 5.5. Plot of observed days to weight permanent environmental variance obtained from the linear random regression of age on weight using both heterogeneous residual variance and random regression on residuals.

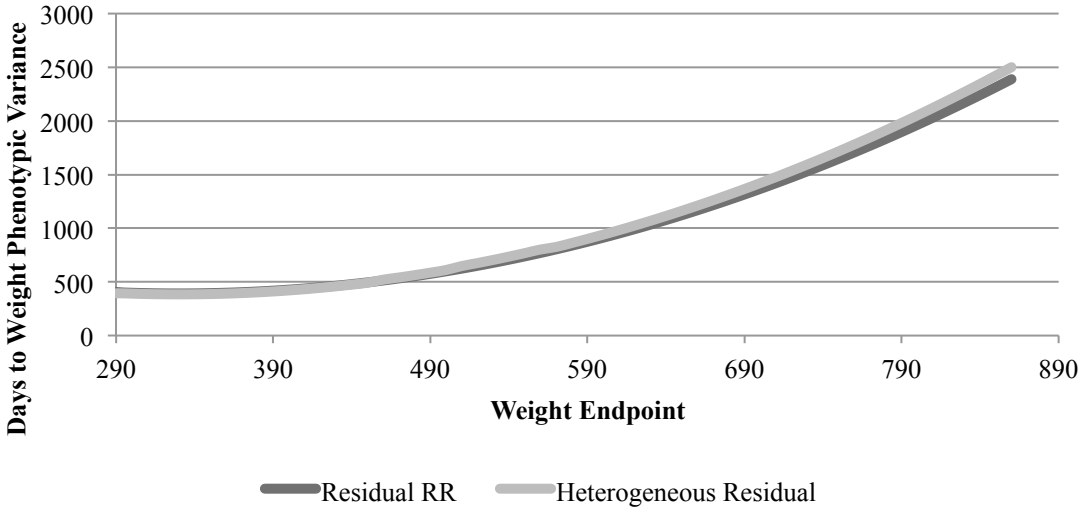


Figure 5.6. Plot of observed days to weight phenotypic variance obtained from the linear random regression of age on weight using both heterogeneous residual variance and random regression on residuals.

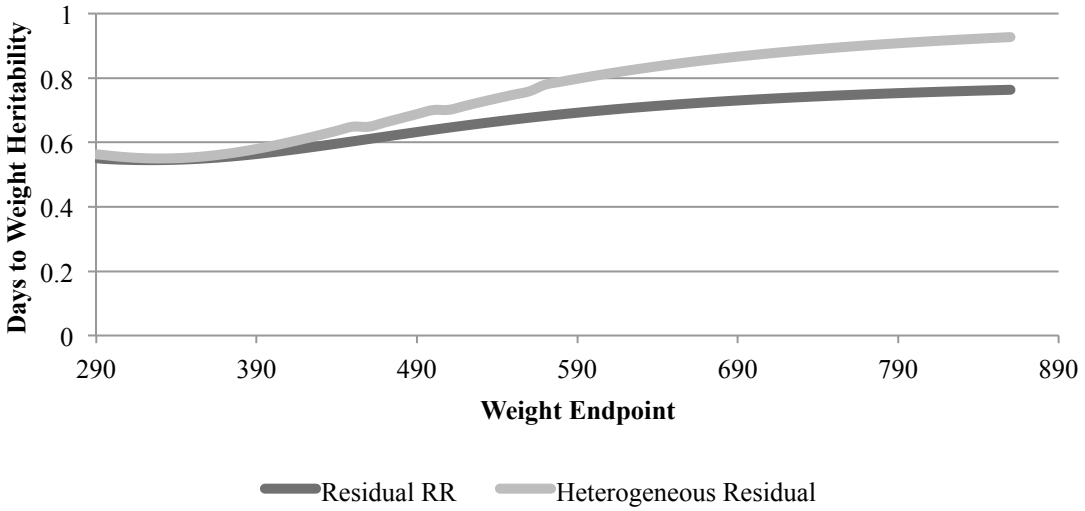


Figure 5.7. Plot of observed days to weight heritability obtained from the linear random regression of age on weight using both heterogeneous residual variance and random regression on residuals.

Figure 5.5 shows the plots of permanent environmental variance between the two models. Permanent environment was included in the HRV model as an intercept for all weight measures and was therefore forced to remain constant. In the LRRR model it was included as a linear random regression term to account for the non-genetic covariation between observations. So as weight increases, the permanent environmental variance also increases.

Perhaps the most interesting plots are of the phenotypic variance and heritability for DTW contained in Figure 5.6 and Figure 5.7. For both methods of modeling residual variance, the DTW phenotypic variance shown in Figure 5.6 is the same. This shows that both models are capturing the same phenotypic variation in days. The only difference being, the manner in which the genetic variation is being partitioned. The LRRR model is attributing more variation to non-genetic components than to genetic effects as the

HRV model is doing. Lastly, Figure 5.7 contains the plots of heritability estimates from the two models. As can be seen here, heritability estimates are the same for the lighter weights, but as weights increase above approximately 500 kg, heritability estimates from the LRRR model are giving more sensible estimates as opposed to the HRV model. At the extreme end of the range of weight observations, the LRRR model is estimating heritability to be 0.76 versus the 0.93 from the HRV model, a rather large difference.

Quadratic Random Regression. Following the linear RR, a quadratic RR was implemented using the data set sifted to require 3 or more observations on individual animals whose summary statistics are shown in Table 5.1. This data set used in the quadratic random regression contained 7,620 age and weight observations on 1,317 individual animals averaging 5.79 observations per individual. Age and weight summary statistics for this data set are shown below in Table 5.6. Comparing this more restrictive data set to that used in the linear random regression model, it contained 6 fewer animals resulting in 12 fewer observations. Looking at contemporary groups (see Chapter IV for contemporary group formation), this data set contained the same number of unique contemporary groups (62 groups) with just a slight difference in the number of animals per group, 21.2 versus the 21.3 from the linear data set.

For the purpose of estimating variance components, a 4-generation pedigree was built from this sifted, final data file. The resulting pedigree contained 5,408 individual animals with 1,386 unique sires and 2,705 unique dams. The average inbreeding coefficient in this pedigree was 1.5% with minimum and maximum inbreeding levels of 0% and 25%, respectively.

Table 5.6. Age and weight summary statistics for the data used in the quadratic random regression of age on weight for the calculation of the days to weight genetic prediction.

	Age[1]	Weight
N	7,620	7,620
Average	396.0	513.1
Variance	2,035.0	5,976.2
Minimum	276	293
Maximum	519	863

[1]Age is reported in days.

Similar to the linear model, direct genetic (co)variances were estimated for the quadratic random regression model, a model that included intercept, linear and quadratic terms as well as all covariances, for both the HRV and the LRRR models. Beginning with the linear fixed regression model, incrementally higher orders of the fixed regression were included in the evaluation until the test statistics and associated p-values showed that increasing this regression an order higher did not significantly account for any additional variation. These estimates are shown below in Table 5.7. Again for the trait DTW, the linear fixed regression polynomial is sufficient in describing the mean relationship between age and weight. The quadratic fixed regression accounts for slightly more variation in DTW for the model containing the HRV even though it is still non-significant ($P > 0.24$). As an additional comparison, genetic variance estimates on the observed scale showed no change as the order of the regression increases for both residual variance models meaning the linear fixed regression was sufficient. However, following the suggestion of Gilmour (2009) that the order of the fixed regression not be reduced below the order of the random terms, further discussions of variance will be in regards to the model containing the quadratic fixed regression.

Table 5.7. Best linear unbiased estimates and associated significance values for the different fixed regression polynomial orders obtained from the quadratic random regression of age on weight using both heterogeneous residual variance and linear residual random regression.

Polynomial Order[1]	Last Coefficient[2]	SE[2]	F[3]	*P*-value
	Heterogeneous Residual Variance			
1	156.6	1.52	10558.8	0.00
2	-0.831	0.71	1.385	0.24
	Linear Residual Random Regression			
1	156.7	1.40	12599.9	0.00
2	-0.004	0.68	0.003	0.96

[1]Order of the fixed regression polynomial
[2]Best linear unbiased estimate and standard error of the highest order term from the fixed regression.
[3]F distribution test statistic and associated *P*-value.

Estimates of variance from both these residual variance models for the quadratic random regression of DTW are shown below in Table 5.8 and Table 5.9. Again, similar to the results seen with the linear random regression model, there were few differences between the estimates obtained from the alternate fixed regression orders within a given residual variance model. Estimates of variance changed more for the HRV model when the fixed regression order was increased from a linear to a quadratic as opposed to the LRRR model where there were virtually no differences between the variance estimates as is indicated by the significance tests shown in Table 5.7.

Table 5.8 Variance estimates obtained from a quadratic random regression model for days to weight using heterogeneous residual variances.

	1[1]	2[1]
LogL[2]	-1477.53	-1477.39
Intercept[3]	1217 (114.5)	1185 (115.7)
Int, Lin[4]	539.7 (43.8)	516.1 (45.3)
Linear[3]	380.6 (28.8)	365.8 (29.6)
Int, Quad[4]	-35.22 (22.7)	-41.00 (23.0)
Lin, Quad[4]	-2.012 (12.9)	-5.74 (13.2)
Quadratic[3]	16.77 (11.0)	16.93 (11.2)
PE[5]	268.3 (66.6)	271.2 (67.3)
R11[6]	32.77 (1.46)	32.68 (1.46)
R22[6]	43.55 (1.78)	43.58 (1.78)
R33[6]	54.05 (2.17)	54.24 (2.18)
R44[6]	42.36 (1.83)	42.56 (1.85)

[1]Order of the polynomial used as the mean regression of age on weight.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic random regression variance estimates for the intercept, linear and quadratic terms (SE).
[4]Direct genetic covariance between the intercept / linear, intercept / quadratic and linear / quadratic terms (SE)
[5]Permanent environmental variance (SE)
[6]Residual variance estimates corresponding to each of the four weight quartiles. R11, R22, R33, and R44 are the residual variances for the first, second, third and fourth quartiles, respectively (SE)

Comparisons between the two residual variance models (HRV versus the LRRR) show more differences. The estimate of residual variance obtained from the LRRR model (43.46 days^2) is very similar to the average of the estimates (43.18 days^2) obtained from the HRV model. Estimates of genetic variance for the quadratic random regression terms obtained from the HRV model are higher than those obtained from the LRRR model by 159 days^2, 77.1 days^2, and 7.33 days^2 for the intercept, linear and quadratic variances, respectively. In the model containing the LRRR variance structure,

Table 5.9. Variance estimates obtained from a quadratic random regression model for days to weight using linear residual random regression.

	1[1]	2[1]
LogL[2]	-1510.55	-1510.94
Intercept[3]	1058 (174.7)	1056 (175.4)
Int, Lin[4]	426.4 (82.8)	425.6 (83.8)
Linear[3]	303.5 (51.8)	303.2 (52.5)
Int, Quad[4]	-27.01 (22.5)	-26.94 (23.2)
Lin, Quad[4]	3.60 (12.6)	3.98 (13.2)
Quadratic[3]	9.44 (10.6)	10.47 (10.9)
PE Intercept[5]	413.5 (127.9)	413.0 (127.8)
PE Int, Lin[5]	108.5 (59.7)	108.1 (59.7)
PE Lin[5]	75.47 (38.3)	75.04 (38.3)
Residual[6]	43.46 (0.88)	43.46 (0.88)

[1]Order of the polynomial used as the mean regression of age on weight.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic variance for the intercept, linear and quadratic (SE).
[4]Direct genetic covariance between the intercept / linear, intercept / quadratic and linear / quadratic terms (SE)
[5]Permanent environmental intercept variance, intercept / linear covariance, and linear variance (SE)
[6]Residual variance (SE)

the intercept variance for the permanent environmental effect was 413.5 days^2 compared to the 268.3 days^2 obtained from the HRV model. The variance attributed to the additive genetic effect appears to again be re-partitioned to the error covariance between observations. Much like the linear random regression model, the REML log-likelihood is higher for the HRV model (-1477.53) than it is for the LRRR model (-1510.55).

Observed DTW estimates of genetic variance, phenotypic variance, permanent environmental variance and heritability were calculated for each 20 kg increment within the range of weight observations, and then plotted versus their corresponding weight endpoint. These plots are shown in Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11.

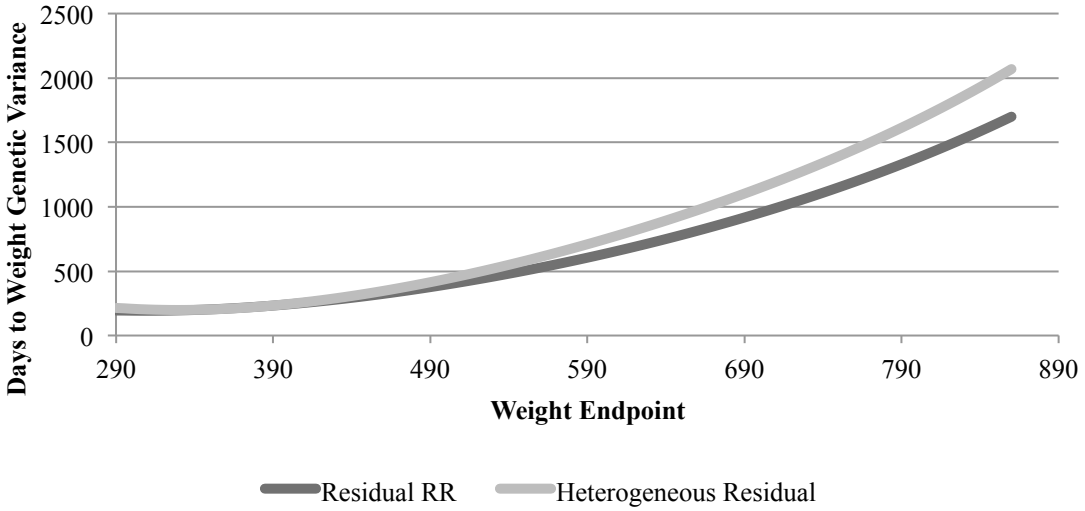


Figure 5.8. Plot of observed days to weight genetic variance obtained from the quadratic random regression of age on weight using both heterogeneous residual variance and linear residual random regression.

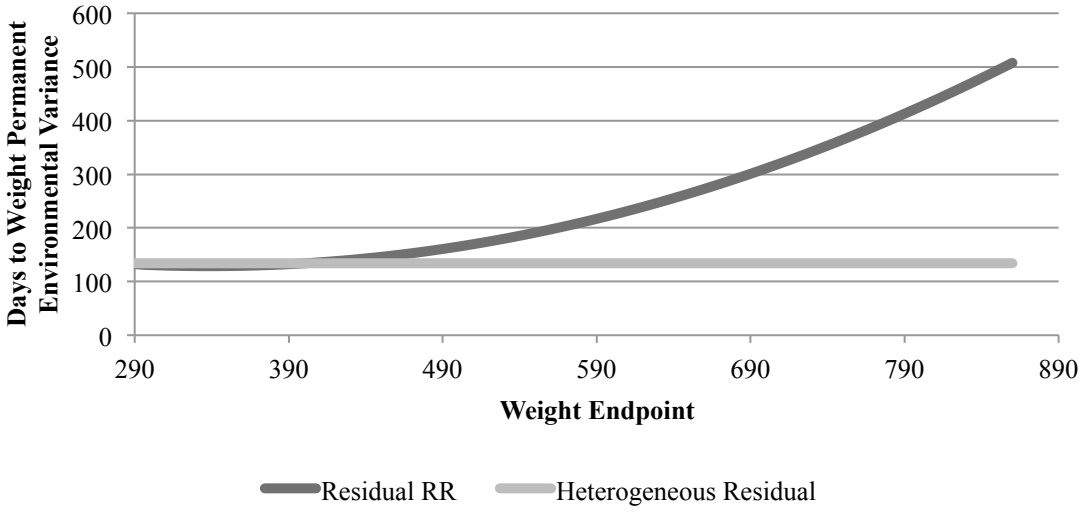


Figure 5.9. Plot of observed days to weight permanent environmental variance obtained from the quadratic random regression of age on weight using both heterogeneous residual variance and linear residual random regression.

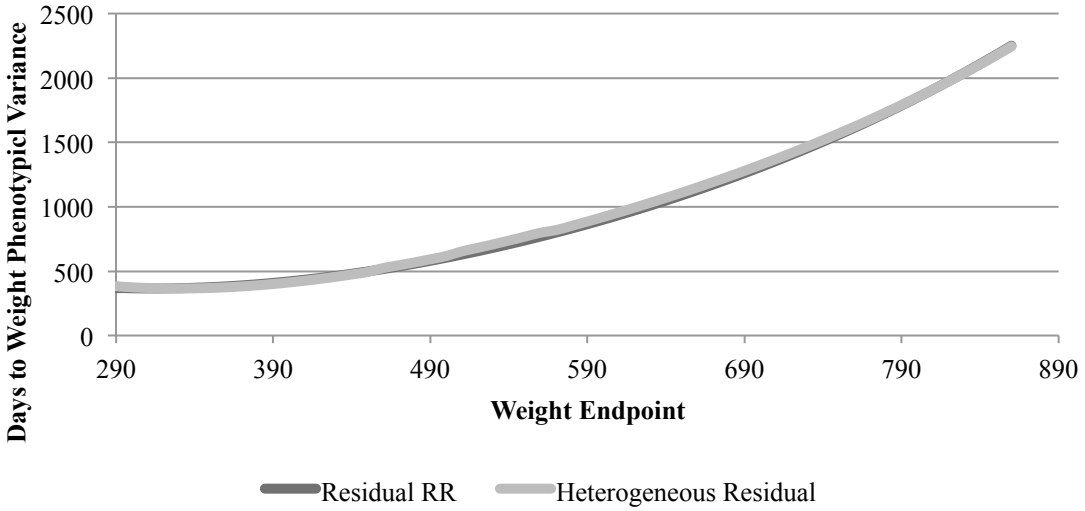


Figure 5.10. Plot of observed days to weight phenotypic variance obtained from the quadratic random regression of age on weight using both heterogeneous residual variance and linear residual random regression.

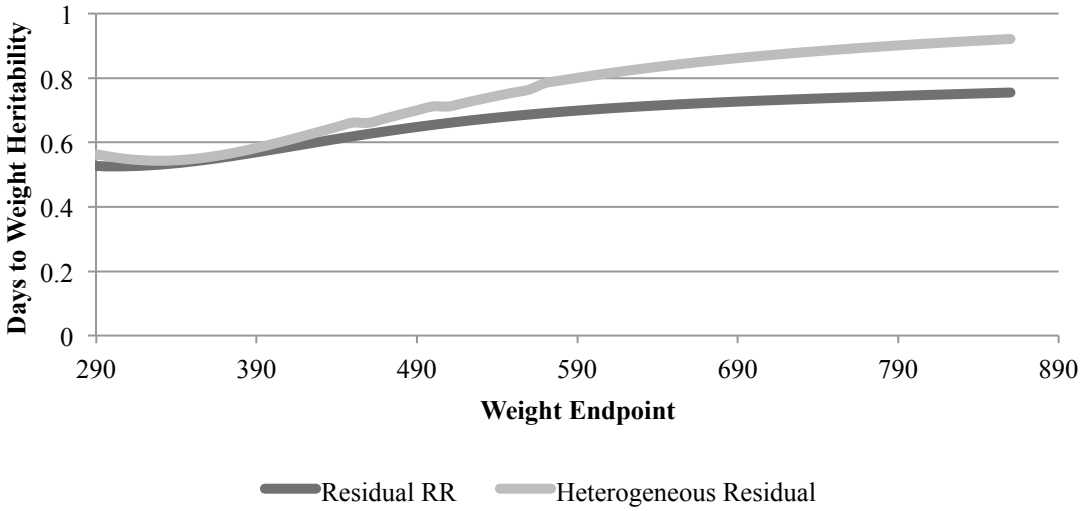


Figure 5.11. Plot of observed days to weight heritability obtained from the quadratic random regression of age on weight using both heterogeneous residual variance and linear residual random regression.

Plots of genetic and permanent environmental variance show the differences in genetic variance between the two residual variance models are due to a re-partitioning of variation from the genetic variance in the HRV model to permanent environmental variance in the LRRR model. The differences between the plots in Figure 5.8 are the same magnitude as the differences seen in the permanent environmental variance in Figure 5.9. Further evidence of this trend is shown in Figure 5.10 where the phenotypic variance estimates obtained from both residual variance models are the same. This re-partitioning has the effect of reducing the magnitude of the heritability estimates (Figure 5.11) observed with increasing weight, where at the maximum weight observation of 863 kg, heritability has been reduced from 0.92 in the HRV model to 0.76 in the LRRR model.

Random Regression Model Selection. The above sections illustrate the point that a linear fixed regression was sufficient in accounting for the mean relationship between age and weight for both the linear and quadratic random regression models where residual variance is modeled as a random regression and as four distinct residual variance sub-classes.

LRT were conducted to determine the statistically significant random regression order with results from these tests presented in Table 5.10. Requirements of LRT state that models being tested have equivalent fixed effect specifications as well as no differences in data. Given this requirement and the results presented earlier in this chapter, comparisons were made between the quadratic and linear random regression models using the data set requiring three or more observations (the necessary number of data points to fit a quadratic polynomial) per animal. Since the linear fixed regression

Table 5.10. REML log likelihood (logL) estimates and associated significance values used to determine the random regression order for the days to weight evaluation for both the heterogeneous residual variance and linear residual random regression models.

	Heterogeneous Residual Variance[1]	Linear Residual Random Regression[1]
Full Model logL[2]	-1477.53	-1510.55
Reduced Model logL[2]	-1480.86	-1512.62
DF[3]	3	3
LRT Test Statistic[3]	6.66	4.14
P-value[3]	0.0836	0.2467

[1]Comparisons reported here are from the comparison between the linear and quadratic random regression models.
[2]Full and reduced model correspond to the more complex versus simpler models, respectively.
[3]Likelihood ratio test statistic and associated *P*-values obtained from a Chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the two models.

polynomial was the highest significant order for both the linear and quadratic random regressions, LRT comparisons between the linear and quadratic random regression models were made using this polynomial.

Comparing the two random regression orders for the model containing HRV in Table 5.10, the addition of the quadratic term approached significance (*p*-value = 0.0836) at the 0.05 level. The addition of the quadratic random regression term to the LRRR model was less significant with the *p*-value of 0.2467. These results suggest that the addition of the linear random regression for permanent environment to the LRRR model is accounting for more of the variability in days than the HRV model is accounting for. As such, the additional variation that has been captured with this model is non-genetic resulting in the lack of ability of higher order random regression terms to significantly account for additional genetic variation in days. Even examination of the scatter plot of

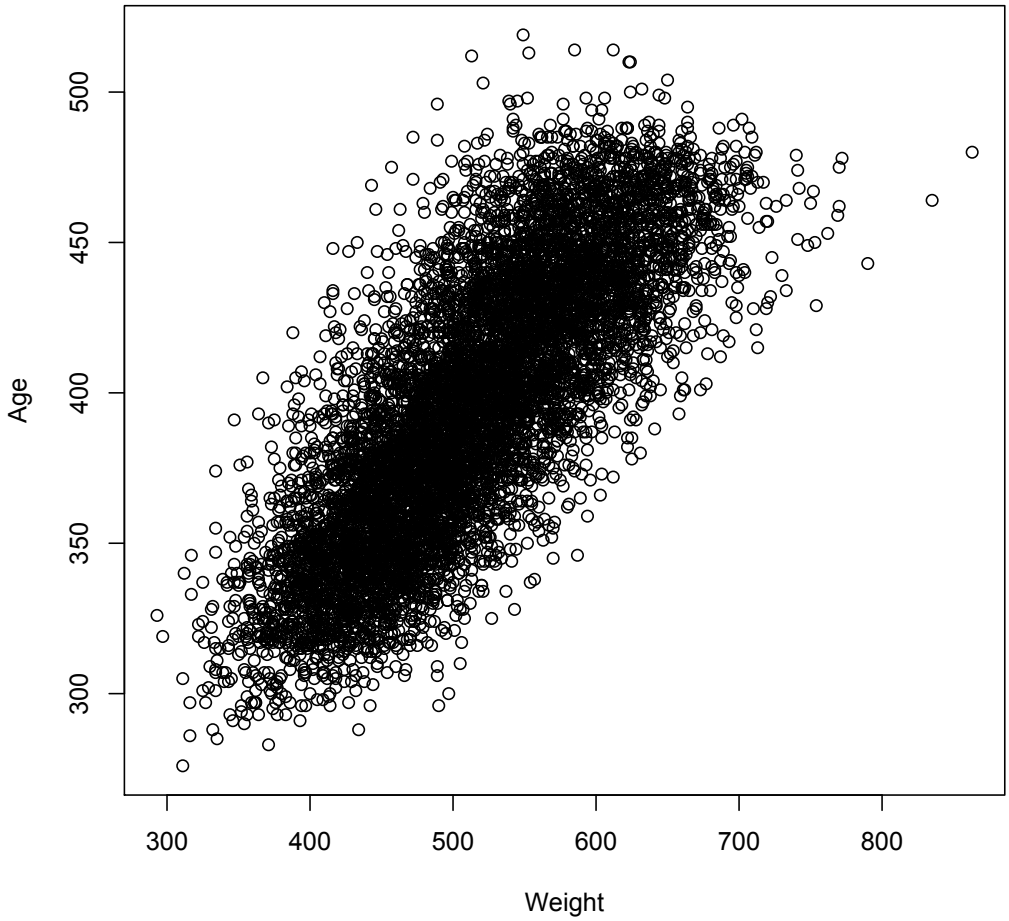


Figure 5.12. Plot of age and weight observations that were analyzed in the days to weight genetic evaluation using a cubic random regression model.

the age and weight observations (Figure 5.12) suggest the lack of need for higher orders than linear terms for both the fixed and random regressions of age on weight.

Model significance tests were only one technique used to determine which random regression model should be used for the genetic prediction of DTW. Next, observed genetic variance, permanent environmental variance, phenotypic variance and heritability for both the HRV (Figure 5.13, Figure 5.14, Figure 5.15, Figure 5.16) and LRRR models (Figure 5.17, Figure 5.18, Figure 5.19, Figure 5.20) were plotted versus weight. For the range of weight endpoints where the data is most dense (Figure 5.1), at

approximately 360 kg to 660 kg, differences in genetic and phenotypic variance obtained from the two different residual variance models are small. Outside of this range, particularly on the upper end of the weight range, weight observations become increasingly sparse resulting in the genetic and phenotypic variance estimates from the linear and quadratic random regressions becoming more variable. This variability may be due to the nature of the higher order polynomials becoming unwieldy near the limits of the data as they have a tendency to place a large emphasis on observations at the tails of the polynomial (Meyer, 1997; 2005).

Permanent environmental variance estimates for the HRV model (Figure 5.14) and the LRRR model (Figure 5.18) are nearly identical between the linear and quadratic random regressions. Moving on to the plots of heritability (Figure 5.16 for the HRV model and Figure 5.20 for the LRRR model), virtually no differences are observed between the linear and quadratic random regressions.

Given the lack of difference in heritability estimates and the results from the LRT presented above in Table 5.10, the linear random regression model is sufficient for the prediction of DTW for both the HRV model and the LRRR model.

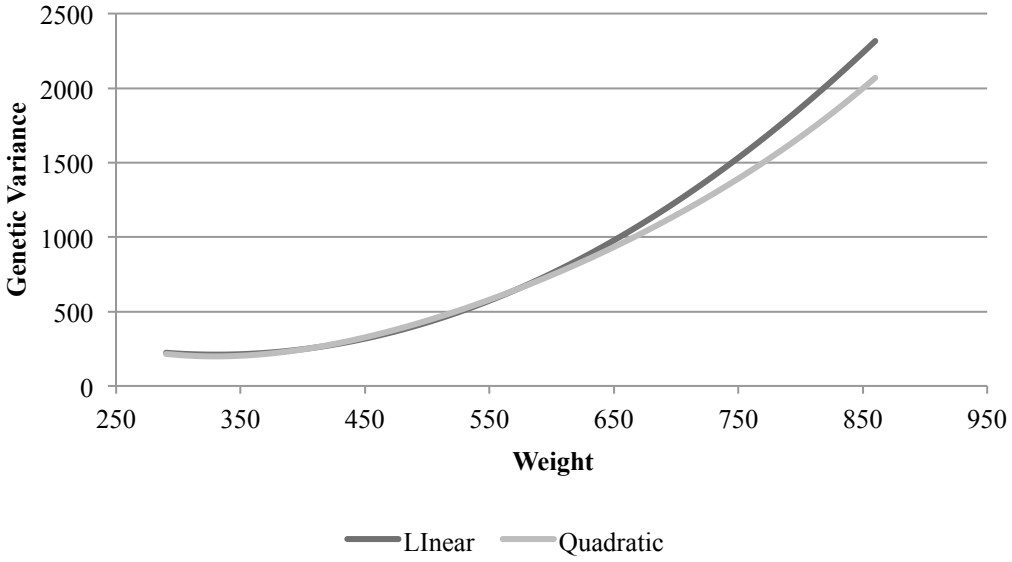


Figure 5.13. Plot of genetic variance (d^2) versus weight obtained from the linear and quadratic random regressions using heterogeneous residual variance for the trait days to weight.

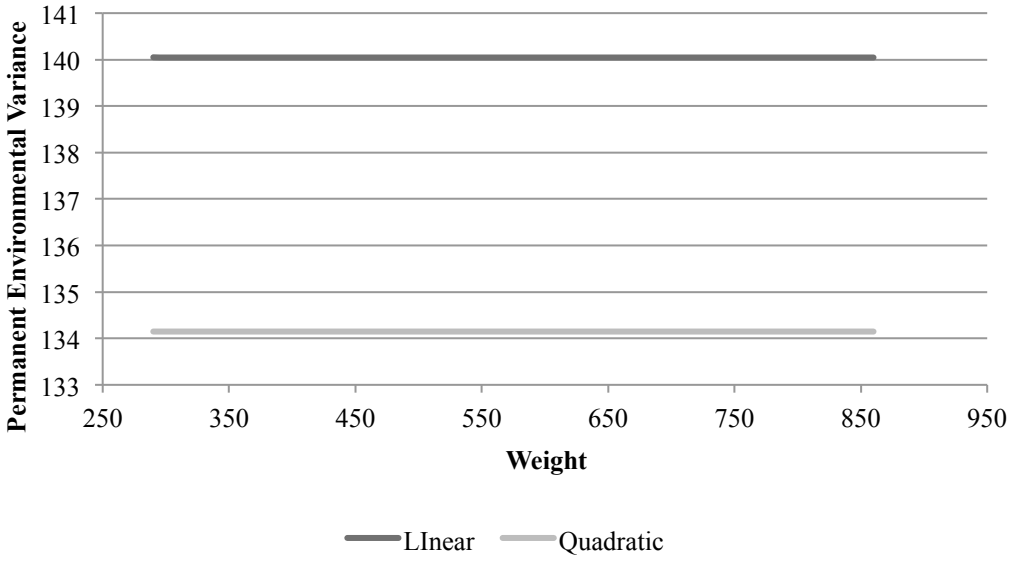


Figure 5.14. Plot of permanent environmental variance (d^2) versus weight obtained from the linear and quadratic random regressions using heterogeneous residual variance for the trait days to weight.

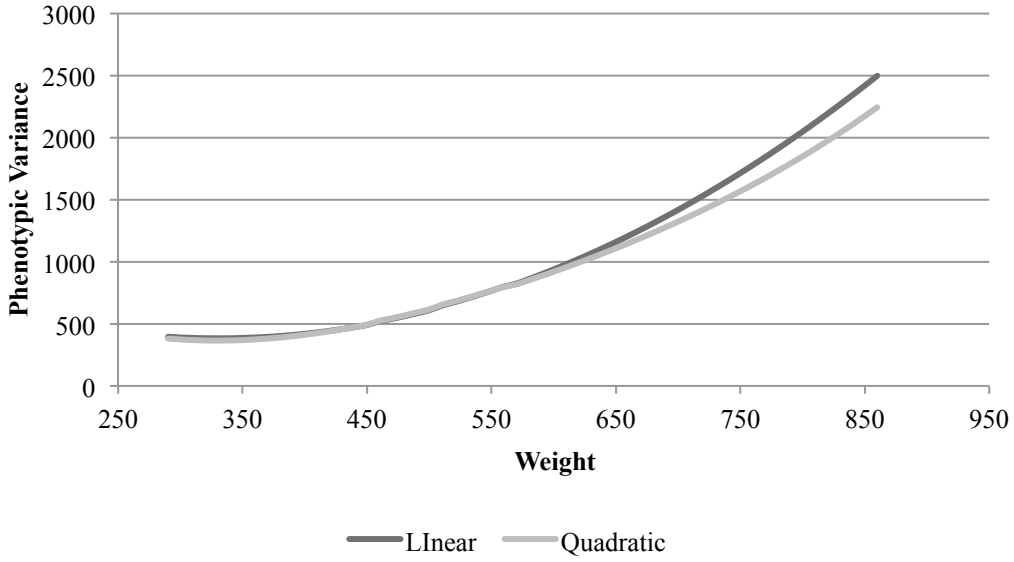


Figure 5.15. Plot of phenotypic variance (d^2) versus weight obtained from the linear and quadratic random regressions using heterogeneous residual variance for the trait days to weight.

.

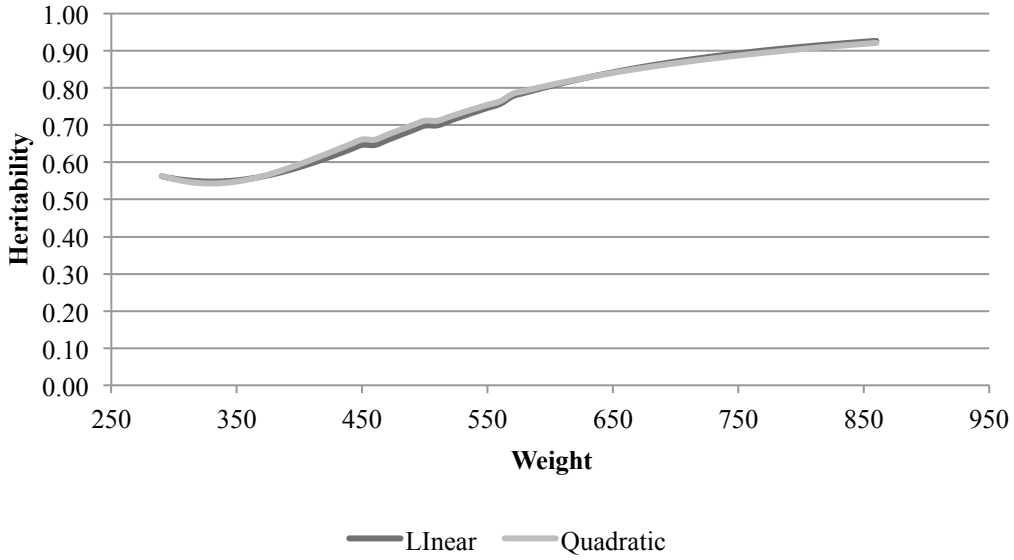
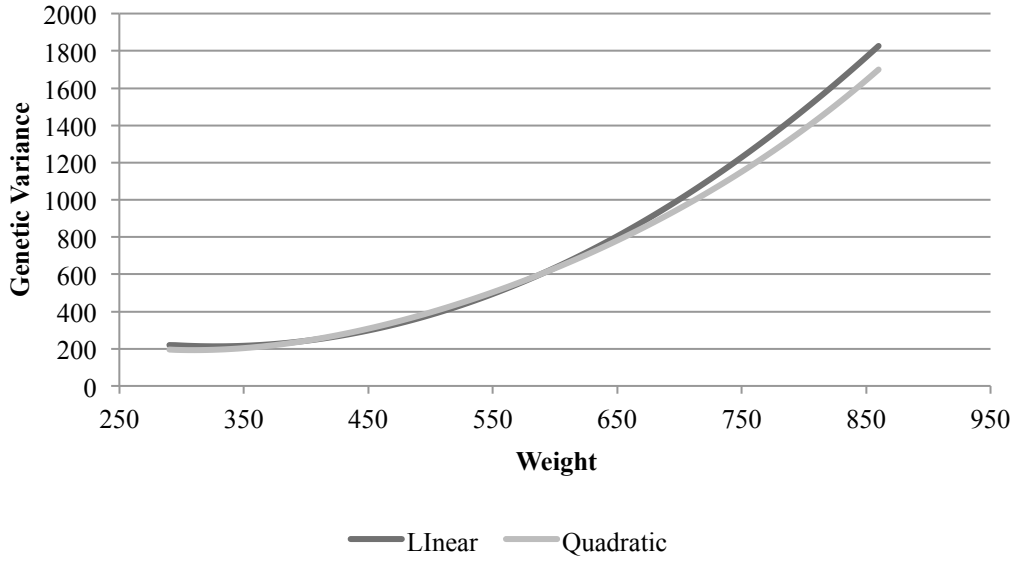


Figure 5.16. Plot of heritability versus weight obtained from the linear and quadratic random regressions using heterogeneous residual variance for the trait days to weight.



.

Figure 5.17. Plot of genetic variance (d^2) versus weight obtained from the linear and quadratic random regressions using linear residual random regression for the trait days to weight.

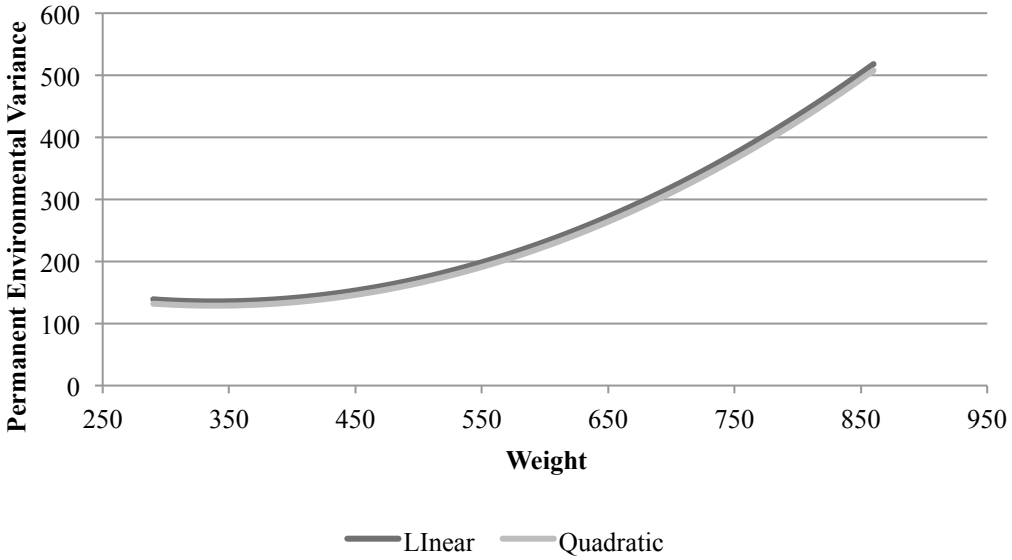


Figure 5.18. Plot of permanent environmental variance (d^2) versus weight obtained from the linear and quadratic random regressions using linear residual random regression for the trait days to weight.

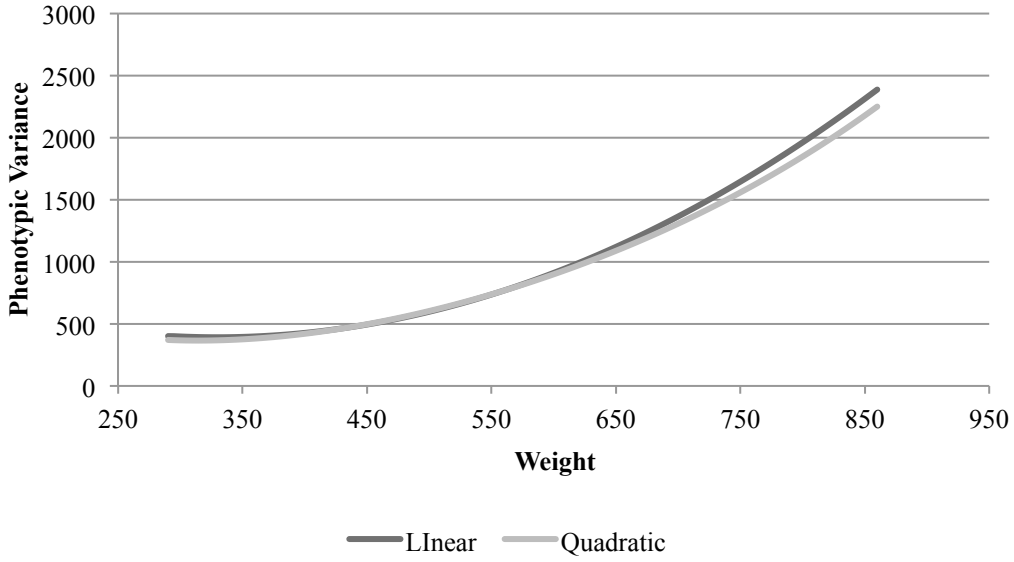


Figure 5.19. Plot of phenotypic variance (d^2) versus weight obtained from the linear and quadratic random regressions using linear residual random regression for the trait days to weight.

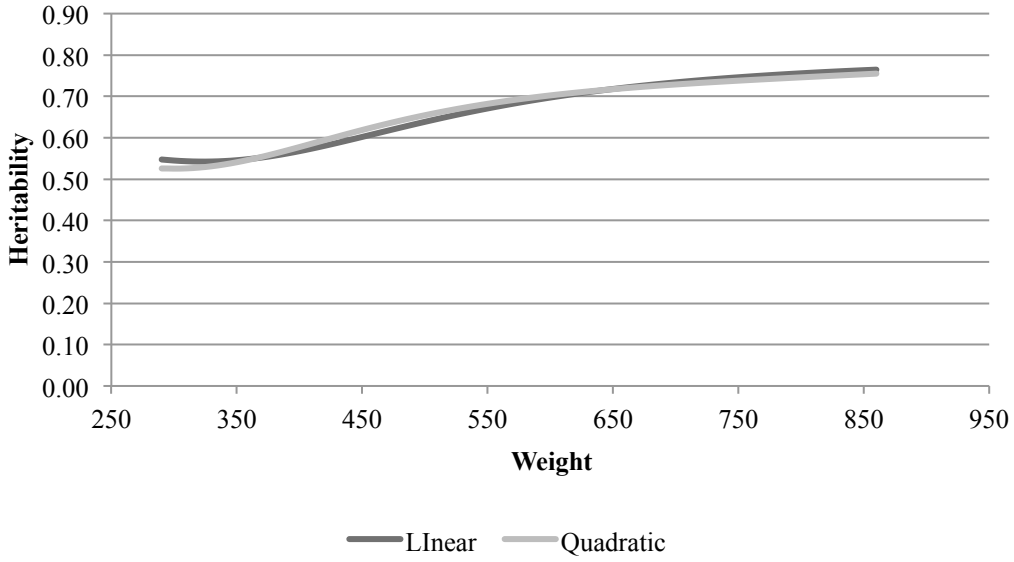


Figure 5.20. Plot of heritability versus weight obtained from the linear and quadratic random regressions using linear residual random regression for the trait days to weight.

The question that remains to be answered is which residual variance model is appropriate most appropriate for the evaluation of DTW. Sire EBV for DTW were calculated from both residual variance models for each 20 kg increment in the range of weight observations (293 kg to 863 kg). EBV corresponding to weight endpoints of 293, 573 and 863 kg were correlated to one another with the results presented in Table 5.11 below. Correlation coefficients between both residual variance models were very high, ranging from 0.993 for the weight endpoint 863 kg to 0.996 for the 293 kg weight endpoint. Spearman rank correlation coefficients are high as well, and looking at the coefficients for the 573 kg and 863 kg weight endpoints, no additional re-ranking is occurring between animal EBV.

Table 5.11. Correlation coefficients along with the EBV regression coefficient from the regression of EBV obtained from the heterogeneous residual variance model on those EBV obtained from the linear residual random regression model for each of three weight endpoints representing minimum, median and maximum observations.

	293 kg	573 kg	863 kg
Pearson Correlation	0.996	0.994	0.993
Spearman Rank	0.992	0.990	0.990
Regression	0.999	1.150	1.110

Regression coefficients obtained from these EBV comparisons are interesting. At the weight endpoint of 293 kg, the two sets of EBV are predicting one another nearly perfectly with a 1 unit increase in EBV obtained from the LRRR model corresponding to a 0.999 unit increase in EBV obtained from the HRV model. As we increase weight endpoint, we can see that the EBV obtained from the LRRR model are under-predicting the EBV obtained from the HRV model. This trend is most likely a function of the higher heritability estimates for the HRV model. Increasing heritability results in an

increase in the spread of the resulting breeding values. Referring back to Figure 5.7, we can see that beginning with the weight endpoint of approximately 490 kg, the HRV model predicts a higher heritability estimate than the LRRR model for corresponding endpoints. This difference in heritability increases as weight increases, and at the weight endpoint of 863 kg, the maximum weight observation in this data set, the difference in heritability between the two models is 0.17.

As another point of comparison between each of the residual variance models, DTW EBV were calculated from each model for 20 kg increments across the entire feeding period. Figure 5.21 contains the plots for each of the 5 most used sires in the pedigree while Figure 5.22 contains the plots of DTW EBV for each of the 5 least used sires in the pedigree. Looking at these plots, we can see that for the most heavily used sires in the data set (averaging 38.4 progeny per sire) the residual variance model has virtually no effect on the prediction of DTW EBV. For animal CA974986, there looks to be a rather large difference, but this is a result of the scale of the graph with the largest difference being only 2 days. For the least used sires in Figure 5.22 (averaging 2 progeny per sire), EBV obtained from the LRRR model are regressed more toward zero meaning that estimates of breeding value for sires with very few progeny are more conservative when they are obtained from the LRRR model than they are from the HRV model.

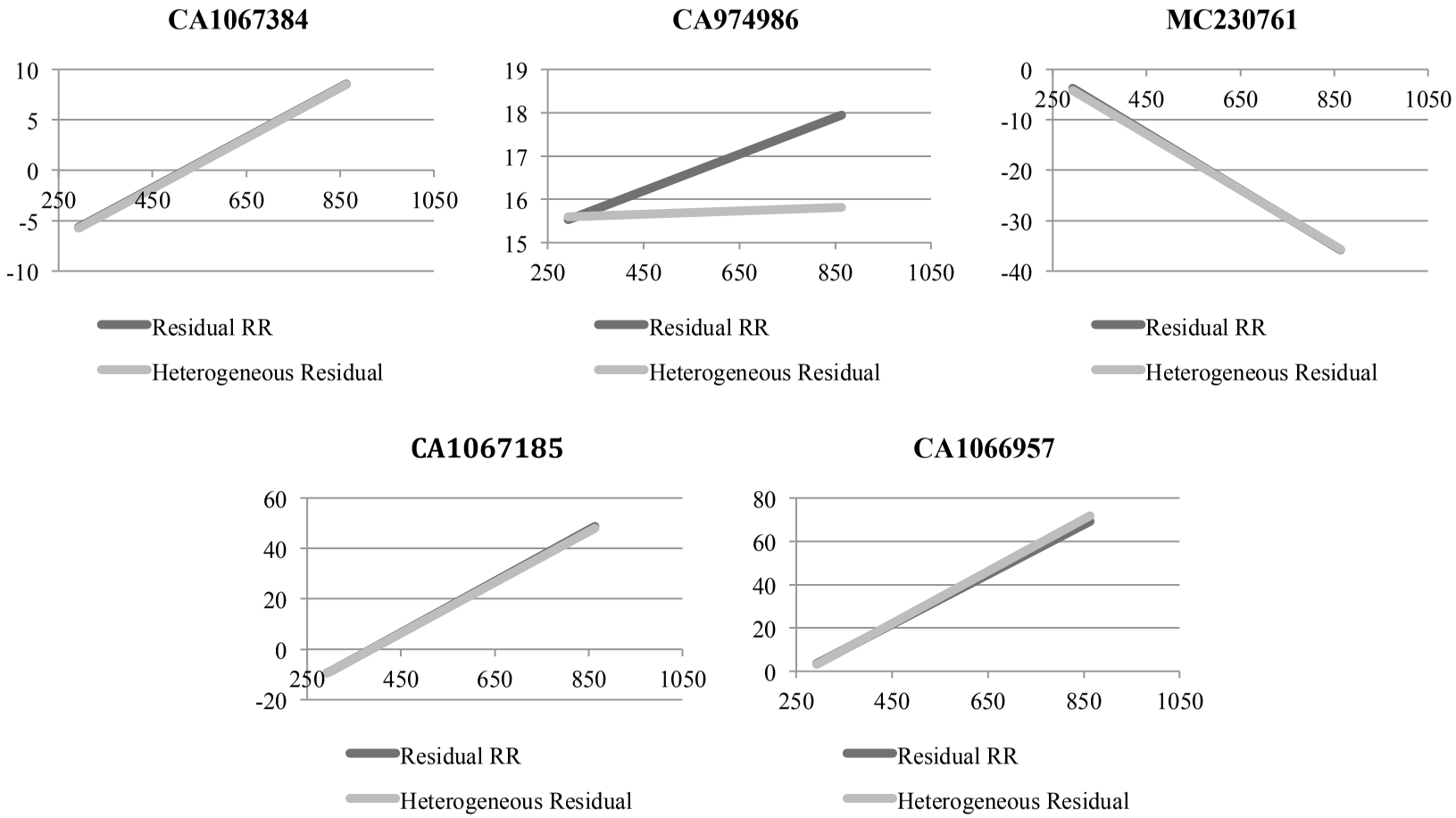


Figure 5.21. Days to weight estimated breeding values obtained from the linear random regression model using both linear residual random regression and heterogeneous residual variance for the five most used sires represented in the weight data set. For the figures above, the vertical axis represents the estimated breeding value for days to weight while the horizontal axis is the weight endpoint.

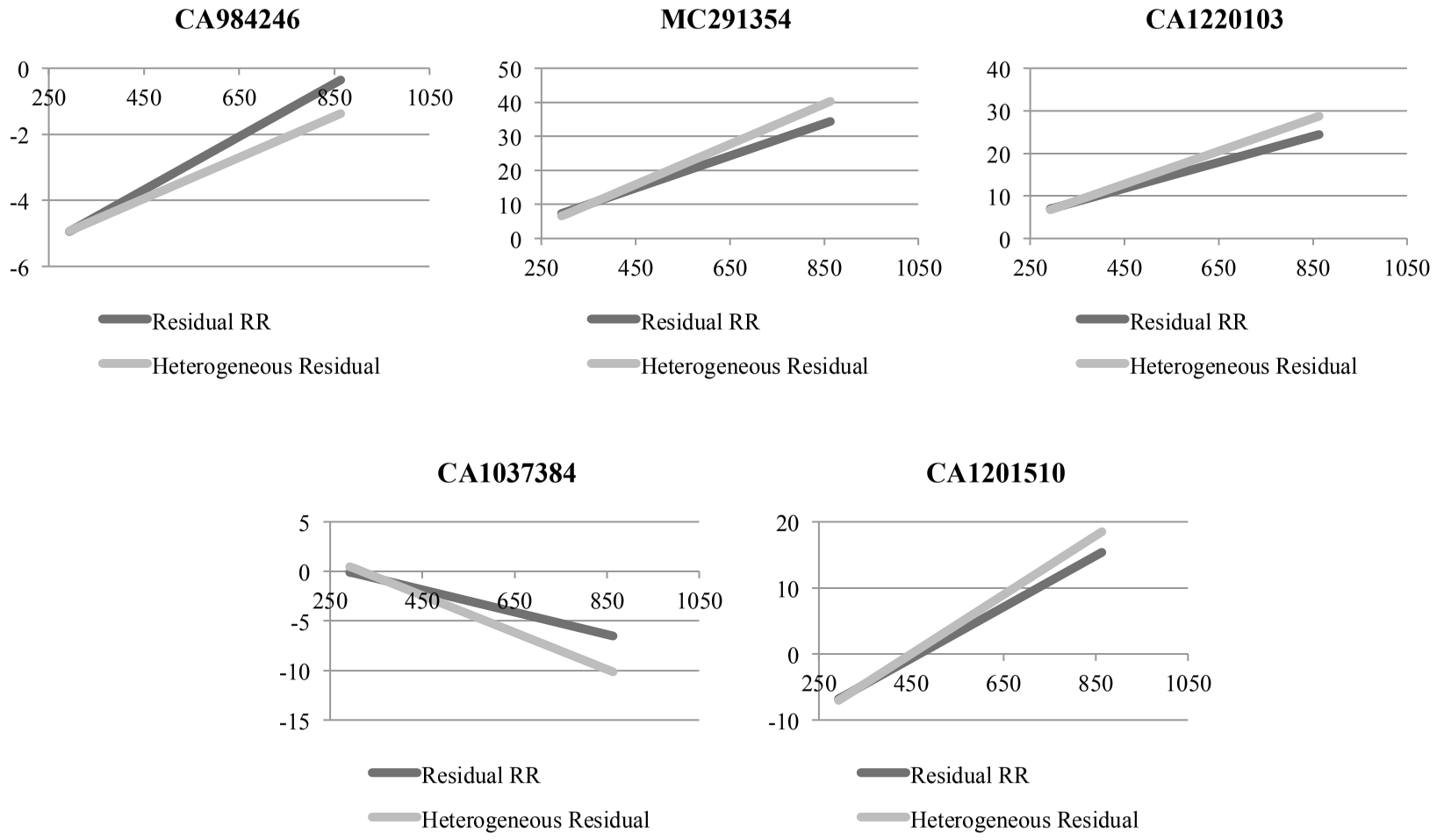


Figure 5.22. Days to weight estimated breeding values obtained from the linear random regression model using both linear residual random regression and heterogeneous residual variance for the five least used sires represented in the weight data set. For the figures above, the vertical axis represents the estimated breeding value for days to weight while the horizontal axis is the weight endpoint.

Comparison to Repeated Measures Model. In an effort to compare the results for the genetic evaluation of DTW using random regression to a more traditional model, a repeated measures analysis was performed using the same weight data set that was used for the linear random regression model (see Table 5.1 for a description).

Heritability estimates obtained from a repeated measures model are endpoint indifferent; meaning no matter the endpoint heritability remains constant. However, depending on the endpoint of interest, variance estimates are scaled according to the magnitude of the observations. These properties of the repeated measures model, resulted in the age observations being adjusted to a constant weight endpoint of 500 kg. This endpoint was chosen because it is in the middle of the distribution of weight observations (Figure 5.1), a location where these observations are most dense, resulting in the most stable predictions obtained from the random regression model.

Repeated measures model estimates of genetic variance and heritability for the number of days to reach 500 kg were 460 ± 73.4 $days^2$ and 0.66 ± 0.09, respectively. Genetic variance and heritability estimates from the HRV linear random regression model for the number of days to reach 500 kg were 420 ± 46.4 $days^2$ and 0.68 ± 0.06, respectively. Genetic variance and heritability estimates from the LRRR model for the number of days to reach 500 kg were 383 ± 61.37 $days^2$ and 0.64 ± 0.08, respectively.

Comparing all three sets of values, the estimates of genetic variance and heritability are well within the standard errors of one another which is evidence the random regression models are estimating the same trait as the repeated measures model. The HRV model is over-estimating the heritability for the number of days to reach 500 kg in comparison to the repeated measures model. Heritability estimated by the LRRR

model is lower than the repeated measures model. When comparing all three values, there is virtually no difference between the estimates when the standard errors are considered.

Summary

This chapter has presented the results from the development of a DTW genetic evaluation using RR with Legendre polynomials as the base random regression polynomial function. Two differing procedures were used to model the residual variation for DTW for both linear and quadratic random regression models. The first residual variance sub-model, the HRV grouped observations into four groups based on their quartile tended to inflate estimates of heritability, particularly for the upper end of the weight range. The LRRR, which modeled the changing residual covariance with increasing weight, resulted in more realistic heritability estimates. The model containing the LRRR appears to be most appropriate for this data set given the resultant heritability estimates. However, which is best?

Several studies have illustrated the necessity of accounting for changing residual variance structure (Olori et al., 1999; Jamrozik et al., 1997; Rekaya et al., 2000). Olori et al., 1999 determined the assumption of the homogeneity of residual variance would bias the resulting heritability estimates upward or downward. Perhaps this is why the heritability estimates obtained from the HRV model are inflated when compared to the LRRR model. The LRRR model accounts for the residual covariance within the estimate of permanent environmental variance. Figure 5.5 illustrates the magnitude of these effects with the permanent environmental variance increasing approximately 400 $days^2$

over the permanent environmental variance from the HRV model. This shows that the assumption of homogeneous residual variance within each residual variance sub-group in the HRV model may not hold, and is likely causing the elevated heritability estimates. Therefore, I am recommending the linear random regression model containing the LRRR for use in a national cattle evaluation scheme because of the more realistic heritability estimates and more conservative breeding value estimates for low accuracy sires.

Literature Cited

Andersen, S. and B. Pedersen. 1996. Growth and food intake curves for group-housed gilts and castrated male pigs. Anim. Sci. 63:457-464.

Arnold, J. W., J. K. Bertrand and L. L. Benyshek. 1992. Animal model for genetic evaluation of multibreed data. J. Anim. Sci. 70:3322:3332.

Beckman, D. W., R. M. Enns, S. E. Speidel, B. W. Brigham and D. J. Garrick. 2007. Maternal effects on docility in Limousin cattle. J. Anim. Sci. 85:650-657.

Brommer, J. E., K. Rattiste, and A. J. Wilson. 2008. Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull Larus canus. Proc. R. Soc. B. 275:687-693.

Brotherstone, S., I. M. S. White, and K. Meyer. 2000. Genetic modeling of daily milk yield using orthogonal polynomials and parametric curves. Anim. Sci. 70:407-415.

Gilmour, A. 2009. Random regression – maternal effects. Available at: http://www.vsni.co.uk/forum/viewtopic.php?t=205&sid=dd107da8ac86898f99dae5f0c7580f8d Accessed Feb. 7, 2010.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Relaease 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk.

Golden, B. L., D. J. Garrick, S. Newman, and R. M. Enns. 2000. Economically relevant traits, a framework for the next generation of EPDs. Proc. Beef Improvement Federation, Wichita, KS, pp. 2-13.

Guo, Z. and H. H. Swalve. 1997. Comparison of different lactation curve sub-models in test day models. Proc. Ann. Interbull Mtg. Vienna. Interbull Bulletin. 16:75-79.

Henderson Jr., C. R. 1982. Analysis of covariance in the mixed model: Higher-Level, Nonhomogeneous, and random regressions. Biometrics. 38:623-640.

Jamrozik, J., G. J. Kistemaker, J. C. M. Dekkers and L. R. Schaeffer. 1997. Comparison of possible covariates for use in a random regression model for analyses of test day yields. J. Dairy Sci. 80:2550-2556.

Jamrozik, J., L. R. Schaeffer, and J. C. M. Dekkers. 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy. Sci. 80:1217-1226.

Jublieu, J. S. 2003. The use of random regression models to predict days to finish in beef cattle. Master's thesis. Colorado State University, Fort Collins, CO.

Kenward, M. G. and J. H. Roger. 1997. The precision of fixed effects estimates from restricted maximum likelihood. Biometrics. 53:983-997.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124:979-993.

Kuehn, L. A. 2000. Parameterization of random regression models for beef cattle data sets. Master's thesis. Colorado State University, Fort Collins, CO.

Kutner, M. H., C. J. Nachtsheim, J. Neter and W. Li. 2005. Applied linear statistical models. 5th ed. McGraw-Hill/Irwin Publishing Company. New York, NY.

Legarra, A., I. Misztal and J. K. Bertrand. 2004. Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle. J. Anim. Sci. 82:1564-1571.

Lindholm, H. B. and H. H. Stonaker. 1957. Economic importance of traits and selection indexes for beef cattle. J. Anim. Sci. 16:998-1006.

Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. Genet. Sel. Evol. 30:221-240.

Meyer, K. 1999. Estimates of genetic and phenotypic covariance functions for post-weaning growth and mature weight of beef cows. J. Anim. Breed. Genet. 116:181-205.

Meyer, K. 2005. Advances in methodology for random regression analyses. Aust. J. Exp. Agric. 45:847-858.

Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CABI Publishing Company. Cambridge, MA.

Olori, V. E., W. G. Hill, and S. Brotherstone. 1999. The structure of the residual error variance of test day milk yield in random regression models. Workshop on Computational Breeding, Finland. March, 1999. Pp. 18-20.

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.

Rekaya, R., M. J. Carabaño, and M. A. Toro. 2000. Assessment of heterogeneity of residual variances using change point techniques. Genet. Sel. Evol. 32:383-394.

R Development Core Team. 2009. R. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Schaeffer, L. R. 2003. Random regression models. ANSC637 Course Notes – Quantitative genetics and animal models. Available at: http://www.aps.uoguelph.ca/%7Elrs/ABModels/NOTES/RRM14a.pdf. Accessed Jan. 8, 2009.

Schaeffer, L. R. and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125:145-146.

Wright, S. 1922. Coefficients of inbreeding and relationship. American Naturalist. 56:330-338.

CHAPTER VI

DAYS TO FINISH ENDPOINT

Introduction

Very little research has been published related to reducing the number of days to reach a specific finish endpoint in the feedlot other than that related to increasing ADG. The term "finish endpoint" is a catch-all phrase and can refer to any point in the life cycle of livestock in which the farmer / rancher / producer has determined an animal to be ready for harvest, thereby hopefully maximizing profits for their operation and marketing program. These endpoints can range from weight, back fat, marbling, rib eye area, to yield grade, etc. The swine industry has been the leader in days to finish research with genetic evaluations for the number of days to reach harvest weight, weaning and breeding endpoints (Stewart et al., 1991; Harris and Newman, 1994; STAGES, 2006). All three evaluations are for various weight endpoints.

Contrary to days to finish research in the swine industry, research in the beef industry has been severely limited. Even so, two studies were uncovered which looked at the number of days to reach a constant back fat. McWhir and Wilton (1987) found the heritability for the number of days to reach a back fat depth of 7 mm to be 0.65, which increased to 0.90 when the trait was adjusted to a constant market weight. Johnston et al. (1992) found the heritability for the number of days to reach 8.9 mm of back fat to be 0.24.

This lack of previously published research for days to finish in beef cattle is puzzling given its economic relevance (Golden et al., 2000). The objective of this study was to use random regression techniques in the creation of genetic predictions for both days to back fat and days to rib eye area using ultrasound measurements from a field data set.

Methodology

Genetic predictions for days to ultrasound rib eye area (**DTUREA**) and days to ultrasound back fat (**DTUBF**) were built using the ultrasound data set previously described in Chapter IV. The initial raw data set consisted of pedigree and multiple ultrasound observations on 1,375 individual animals.

Following the days to weight genetic prediction described in Chapter V two differing evaluations were implemented for the evaluation of DTUREA and DTUBF. First, random regression models were used to predict the genetic merit of individual animals for both traits. Here, two model subsets were implemented using alternate approaches to account for residual variation. Residual variance was modeled both heterogeneously by assigning observations to four different groups (**HRV**) and by using a linear random regression on residuals (**LRRR**) which allows for changing residual variance. The advantage of LRRR models is that they allow for the residual variance to change with increases in the predictor variables, and they account for the changing error covariances as observations become farther apart from one another. Second, in an effort to make comparisons to the random regression models, a more traditional repeated measures model was used to evaluate DTUREA and DTUBF. Here, age observations

were used to adjust ultrasound rib eye area and ultrasound back fat observations to to age-constant measurements.

Model building exercises used in the days to weight evaluation (Chapter V) were also implemented here for the ultrasound traits. Descriptions of both random regression models as well as the repeated measures model along with associated significance tests are described in Chapter V and are omitted here to avoid duplication.

Results and Discussion

For the evaluation of both DTUREA and DTUBF, the final weight data set described in Chapter IV was subset, removing individual animals who did not possess the proper number of observations to fit the random regression line. Depending on the order of the random regression, individual animals were removed from the data set if they had fewer than five, four, three or two observations for the quartic, cubic, quadratic and linear random polynomials, respectively. In order to use likelihood ratio tests to make comparisons between the models, a forward stepwise regression algorithm was implemented. This algorithm is the same as was performed for the days to weight chapter and is shown below.

1) Animals with fewer than two observations were removed from the data set.
    a. A linear random regression of age on ultrasound observation was implemented.
2) In subsequent analyses, animals with fewer than three observations were removed from the data set.

3) Random regressions were modeled using the data set containing animals from step two.
    a. First, a quadratic random regression model was fit identifying the proper order of the fixed regression.
    b. Second, a linear random regression model was fit using the fixed regression polynomial order chosen in 3a.
    c. LRT between the linear and quadratic random regression models were performed to determine whether the quadratic terms should be dropped from the regression. Significance levels were set at $P = 0.05$.
4) If the results from the LRT suggested the quadratic random regression term should be included in the model, the algorithm was repeated beginning with step two removing animals with fewer than the four observations needed to fit a cubic polynomial.
    a. A quartic random regression is the highest order polynomial considered for inclusion due to restrictions on the required number of observations needed by individual animals.
5) If the results from the LRT suggested the quadratic (or higher) term should not be included in the model, the process was ended.

Sub-setting the data in this manner resulted in four individual ultrasound data sets whose summary statistics are shown in Table 6.1. As a benchmark for making comparisons, summary statistics for the all ultrasound data are included in the table as well. Realizing that higher order regressions require more observations per animal, with

Table 6.1. Summary statistics for the average number of observations for each of the individual regression data sets, ranging from all data to requiring individual animals to possess five or more observations, the number needed for a quartic regression.

	All Data	5+	4+	3+	2+
N[1]	7,374	6,600	7,340	7,361	7,373
N Animals[2]	1,324	1,125	1,310	1,317	1,323
Mean[3]	5.57	5.87	5.6	5.59	5.57
Minimum[3]	1	5	4	3	2
Maximum[3]	9	9	9	9	9

[1]Total number of records in the data set.
[2]Number of unique animals in the data set.
[3]Mean, minimum and maximum number of observations per animal.

the quartic random regression requiring individual animals to have five or more observations. The data point requirement for a quartic random regression resulted in the removal of 199 individual animals representing a total of 774 observations when compared to the full data set. With that restriction an entire year's worth of data (2002 – 2003) is removed from the evaluation. Requiring four or more observations for the cubic random regression results in the removal of 34 observations from 14 animals. A total of 13 observations are removed in order to fit a quadratic random regression while only one animal is removed while fitting a linear random regression. Similar trends were seen in the weight data set from Chapter V, although not as severe, meaning animals possessed more useable weight observations than they do useable ultrasound observations.

One of the random regression sub-models, the HRV, used four residual variance classes of grouped observations. Initial attempts at forming these classes were performed by visual inspection of the data distribution of both UREA and UBF observations. Histograms provided below show the distribution for UREA (Figure 6.1) and UBF (Figure 6.2). These figures indicate UREA observations to be fairly normally distributed

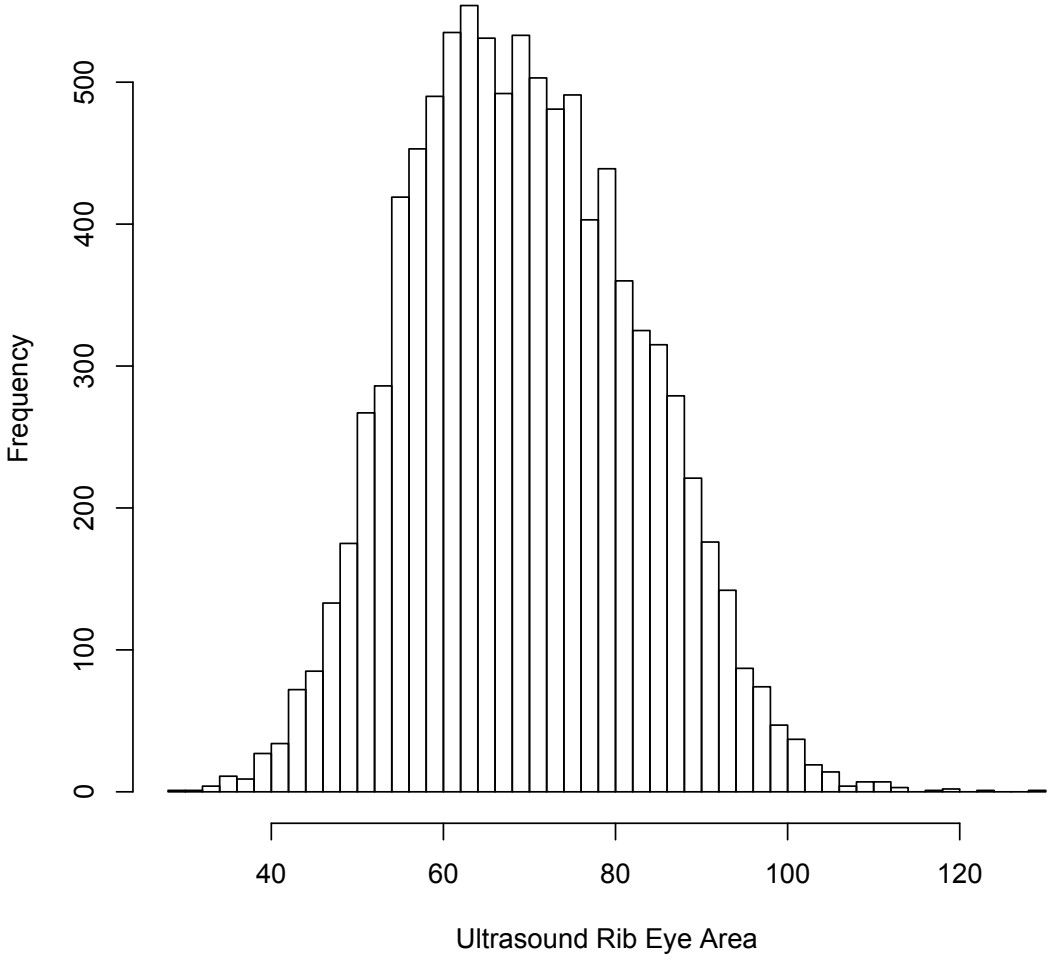


Figure 6.1. Histogram of ultrasound rib eye area observations used in the days to rib eye area genetic evaluation.

while the UBF observations seem to be clustered around lower levels of back fat with a heavy tail extending all the way out to approximately 30 mm of back fat thickness. Neither figure indicates any sort of naturally occurring break point for the specification of residual variance groups; therefore, as was performed with the days to weight genetic prediction, data points were divided according to quartiles. Quartiles for the UREA data were divided at the breakpoints of 64.55, 72.71 and 81.48 cm^2. These points resulted in

1,844, 1,844, 1,844 and 1,841 UREA observations to be placed in the first, second, third and fourth quartiles, respectively.

UBF data was divided into each of the quartiles according to the breakpoints of 5.61, 7.97, and 11.04 mm resulting in quartile sizes of 1,852, 1,839, 1,839 and 1,843 observations in the first, second, third and fourth quartiles, respectively.

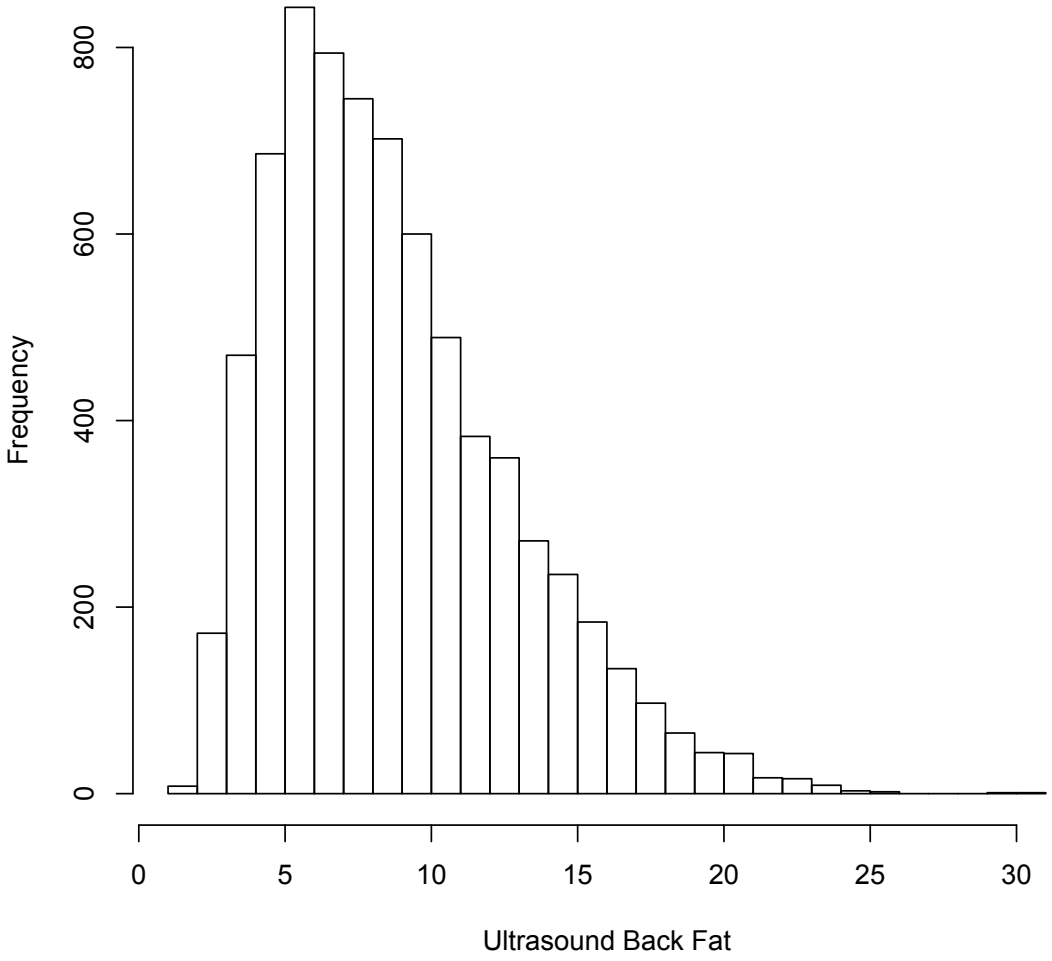


Figure 6.1. Histogram of ultrasound rib eye area observations used in the days to rib eye area genetic evaluation.

Linear Random Regression. A linear random regression was implemented using the ultrasound data set that contained only those individuals possessing two or more observations described above in Table 6.1**.** This data set included 7,373 age and ultrasound observations on 1,323 individual animals resulting in an average of 5.57 observations per animal. Summary statistics for this data set are shown below in Table 6.2. Similar to the days to weight genetic prediction, the range of ages is 276 days to 519 days.

Table 6.2. Summary statistics for age, ultrasound rib eye area and ultrasound back fat used in the linear random regression for the days to "finish" genetic prediction

	Age (d)	UREA[1]	UBF[2]
N	7,373	7,373	7,373
Average	394.1	73.39	8.71
Variance	1,960.5	135.64	16.40
Minimum	276	36.77	1.53
Maximum	519	129.54	30.47

[1]Ultrasound rib eye area.
[2]Ultrasound back fat.

Ultrasound contemporary groups were formed on the basis of feedlot pen, year of test and breed composition (see Chapter IV for more detailed information on the formation of contemporary groups). Formation of contemporary groups in this manner resulted in 62 unique groups averaging 21.3 animals per group. For the purpose of estimating days to finish variance components, a 4-generation pedigree was built from the final ultrasound data set. This final pedigree consisted of 5,414 individual animals, 1,386 unique sires and 2,705 unique dams. The average inbreeding for the animals in this

pedigree was 1.5% with minimum and maximum inbreeding levels of 0% and 25%, respectively.

Beginning with the linear fixed regression model; genetic, permanent environmental and residual variance estimates were obtained from both the HRV and the LRRR models for each of the increasing fixed regression polynomial orders.

Table 6.3. Best linear unbiased estimates and associated significance values for the different fixed regression polynomial orders obtained from the linear random regression of age on ultrasound rib eye area using both heterogeneous residual variance and linear residual random regression.

Polynomial Order[1]	Last Coefficient[2]	SE[2]	F[3]	*P*-value
	Heterogeneous Residual Variance			
1	130.00	1.73	5672.90	0.00
2	-25.13	1.50	290.32	0.00
3	-17.40	1.60	118.11	0.00
4	12.57	1.50	70.24	0.00
5	-3.28	1.39	5.64	0.02
6	-4.42	1.30	11.65	0.00
7	-1.74	1.23	2.02	0.16
	Linear Residual Random Regression			
1	131.50	1.71	5899.89	0.00
2	-25.83	1.55	278.78	0.00
3	-18.33	1.69	117.36	0.00
4	13.19	1.59	69.08	0.00
5	-3.09	1.47	4.44	0.04
6	-4.57	1.37	11.07	0.00
7	1.71	1.31	1.72	0.19

[1]Order of the fixed regression polynomial
[2]Best linear unbiased estimates and standard errors corresponding to the highest order term in the fixed regression.
[3]Fdistribution test statistic and associated *P*-value

Test statistics and associated p-values corresponding to each of these fixed regression polynomials are shown in Table 6.3 and Table 6.4 for DTUREA and DTUBF,

respectively. Considering the results shown in Table 6.3 for DTUREA, the highest order fixed regression polynomial accounting for a significant amount of variation in days is a 6th order polynomial. This order is the same for both the heterogeneous residual variance model as well as the residual random regression model. Between the two models, estimates were similar.

Table 6.4. Best linear unbiased estimates and associated significance values for the different fixed regression polynomial orders obtained from the linear random regression of age on ultrasound back fat using both heterogeneous residual variance and linear residual random regression.

Polynomial Order[1]	Last Coefficient[2]	SE[2]	F[3]	*P*-value
Heterogeneous Residual Variance				
1	183.10	2.91	3969.95	0.00
2	-37.25	1.32	792.74	0.00
3	6.26	1.23	25.82	0.00
4	-7.09	1.04	46.12	0.00
5	8.04	0.97	68.79	0.00
6	-1.07	1.09	0.96	0.33
Linear Residual Random Regression				
1	187.60	3.04	3803.18	0.00
2	-44.12	1.60	765.16	0.00
3	8.05	1.31	37.67	0.00
4	-7.71	1.14	45.57	0.00
5	7.99	1.06	56.94	0.00
6	-0.48	1.19	0.16	0.69

[1]Order of the fixed regression polynomial
[2]Best linear unbiased estimates and standard errors corresponding to the highest order term in the fixed regression.
[3]Fdistribution test statistic and associated *P*-value

For the trait DTUBF (Table 6.4), a 5th order fixed regression polynomial was sufficient in accounting for variation of the mean relationship between age and back fat. This is lower than the order needed for ultrasound rib eye area, but not surprising given

that ultrasound back fat has less variability than ultrasound rib eye area as evidenced by their histograms shown above in Figure 6.1 and Figure 6.2. Comparing the DTUBF models containing the heterogeneous residual variance structure and residual random regressions, the 5th order fixed regression order was sufficient for both models and estimates between the two models were very similar.

Variance estimates obtained for the trait DTUREA from both the heterogeneous residual variance model and residual random regression model are shown below in Table 6.5 and Table 6.6, respectively. In both tables, estimates are presented for each of the seven polynomials discussed above (Table 6.3). Estimates from the lower fixed polynomials (first and second orders) fluctuated wildly, but with further increasing order the variance estimates stabilized with increasing fixed regression order, with the genetic covariance between the intercept and linear random regression terms being the last to stabilize. This was true for both residual variance models. Once the fixed regression order was increased to a 6th order polynomial, very little changes were observed by increasing the order higher as was suggested by the significance testing results (Table 6.3).

Focusing on the estimates from the 6th order fixed regression model, additional differences can be observed between the heterogeneous residual variance model and the residual random regression model. First, the average estimate of residual variance from the heterogeneous residual variance model (442.2 days^2) is very similar to the estimate obtained from the residual random regression model (439.7 days^2). The residual random regression model contained a higher intercept genetic variance estimate than the heterogeneous residual variance model (643.3 days^2 versus 578.9 days^2) and a lower

estimate of linear genetic variance as well as the genetic covariance between the intercept and linear random regression terms. An interesting observation is the estimate standard errors obtained from the residual random regression model are much smaller than those obtained from the heterogeneous residual variance model suggesting perhaps that the residual random regression model is doing a better job at estimating these variances.

Table 6.5. Variance estimates (SE) obtained from a linear random regression model for days to ultrasound rib eye area using heterogeneous residual variance.

	1[1]	2[1]	3[1]	4[1]	5[1]	6[1]	7[1]
LogL[2]	-7307.76	-7221.13	-7164.78	-7129.81	-7126.72	-7120.75	-7119.55
Intercept[3]	899.3 (118.5)	589.7 (105.6)	579.1 (103.4)	583.2 (102.9)	582.3 (102.8)	578.9 (102.1)	578.1 (102.0)
Int, Lin[3]	346.0 (45.8)	-35.64 (37.8)	-67.41 (32.3)	-63.05 (35.1)	-70.38 (35.2)	-78.29 (34.8)	-81.74 (34.9)
Linear[3]	315.1 (47.5)	259.6 (42.6)	189.3 (36.4)	165.7 (34.4)	167.2 (35.5)	161.9 (33.9)	163.6 (34.1)
PE Intercept[4]	121.7 (79.3)	135.9 (81.2)	123.4 (79.8)	114.5 (79.2)	113.0 (79.3)	108.3 (78.8)	107.6 (78.8)
R11[5]	392.5 (15.6)	398.8 (16.0)	389.5 (15.5)	386.6 (15.4)	386.7 (15.4)	388.7 (15.4)	388.0 (15.4)
R22[5]	525.6 (19.8)	512.3 (19.5)	509.2 (19.3)	506.1 (19.2)	505.8 (19.2)	505.2 (19.1)	505.3 (19.1)
R33[5]	528.8 (20.2)	524.3 (19.8)	522.2 (19.7)	520.8 (19.6)	521.0 (19.6)	521.4 (19.6)	521.3 (19.6)
R44[5]	376.3 (15.7)	351.9 (14.2)	357.8 (14.3)	356.3 (14.2)	354.6 (14.1)	353.6 (14.1)	353.6 (14.1)

[1]Order of the polynomial used as the mean regression of age on ultrasound rib eye area.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept and linear random regression terms.
[4]Permanent environmental variance.
[5]Residual variance estimates corresponding to each of the four ultrasound rib eye area quartiles.

Table 6.6 Variance estimates (SE) obtained from a linear random regression model for days to ultrasound rib eye area using linear residual random regression.

	1[1]	2[1]	3[1]	4[1]	5[1]	6[1]	7[1]
LogL[2]	-7340.86	-7257.79	-7200.84	-7166.19	-7163.62	-7157.71	-7156.6
Intercept[3]	1063 (64.3)	667.3 (48.4)	655.5 (46.8)	649.7 (46.2)	645.7 (46.0)	643.3 (45.6)	641 (45.5)
Int, Lin[3]	380.9 (48.0)	-5.06 (40.0)	-51.11 (38.4)	-48.39 (37.12)	-54.95 (37.1)	-65.24 (36.8)	-67.79 (36.8)
Linear[3]	293.9 (47.7)	226.7 (42.8)	166.0 (36.7)	143.9 (34.7)	144.6 (34.8)	140.7 (34.2)	141.9 (34.4)
PE Intercept[4]	36.83 (0.71)	87.93 (1.7)	75.72 (1.4)	75.79 (1.44)	76.61 (1.46)	72.57 (1.4)	73.1 (1.4)
PE Int, Lin[4]	-16.90 (0.3)	-47.01 (0.9)	-38.73 (0.7)	-38.72 (0.74)	-39.24 (0.8)	-36.58 (0.7)	-36.92 (0.7)
PE Lineaar[4]	10.09 (0.2)	27.69 (0.5)	22.20 (0.4)	22.14 (0.42)	22.47 (0.4)	20.75 (0.4)	20.97 (0.4)
Residual[5]	455.1 (8.7)	445.3 (8.5)	442.6 (8.4)	440.0 (8.4)	439.6 (8.4)	439.7 (8.4)	439.6 (8.3)

[1]Order of the polynomial used as the mean regression of age on ultrasound rib eye area.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept and linear random regression terms.
[4]Permanent environmental (co) variance for the intercept and linear random regression terms.
[5]Residual variance.

Variance estimates obtained for the trait DTUBF are shown below in Table 6.7 and Table 6.8. Here, the estimates are presented in the same manner as for DTUREA for each of the fixed regression polynomials and for both the HRV and LRRR models. For both residual variance sub-models, genetic variance estimates fluctuated greatly between the low fixed regression orders. Then as the order of the fixed regression increased above a quadratic, this fluctuation subsided, and by a 5^{th} order polynomial, the order suggested by the significance tests (Table 6.4), both variance and covariance estimates stabilized to the point where the addition of higher order terms changed very little.

According to the significance testing shown above in Table 6.4, the highest order fixed regression term accounting for a significant amount of variation in DTUBF was the 5^{th} order polynomial. Estimates for both the HRV model and the LRRR model show similar trends to those obtained in the DTUREA evaluation. The average estimate from the four residual variance classes from the HRV model was 505.2 $days^2$. This is very similar to the residual variance estimate obtained from the LRRR model of 513.3 $days^2$. The intercept for the permanent environmental variance from the LRRR model was much smaller than the estimate obtained from the HRV model (112.3 $days^2$ versus 482.0 $days^2$). Also, as was mentioned earlier, the estimate standard errors are much smaller from the LRRR model than from the HRV model, suggesting that the estimates from the residual random regression model were estimated more precisely

Table 6.7. Variance estimates (SE) obtained from a linear random regression model for days to ultrasound back fat area using heterogeneous residual variance.

	1[1]	2[1]	3[1]	4[1]	5[1]	6[1]
LogL[2]	-8082.53	-7910.95	-7898.69	-7876.33	-7843.42	-7842.86
Intercept[3]	3294 (302.7)	1046 (215.9)	1079 (219.5)	1048 (216.4)	942.4 (204.6)	939.2 (204.2)
Int, Lin[3]	1616 (151.3)	-4.594 (63.2)	36.52 (66.7)	43.02 (67.6)	-45.07 (61.2)	-45.92 (61.3)
Linear[3]	961.2 (110.8)	56.18 (47.2)	93.85 (54.9)	129.8 (58.2)	84.96 (53.3)	87.85 (53.5)
PE Intercept[4]	337.1 (127.4)	480.1 (154.6)	500.8 (155.4)	504.9 (154.2)	482.0 (151.6)	481.2 (151.4)
R11[5]	605.6 (23.0)	574.0 (21.4)	555.9 (20.9)	537.0 (20.2)	528.3 (19.8)	528.6 (19.8)
R22[5]	655.3 (24.8)	644.6 (24.1)	633.3 (23.7)	628.1 (23.5)	636.9 (23.8)	637.9 (23.9)
R33[5]	519.8 (19.9)	516.7 (19.5)	520.8 (19.7)	528.3 (20.0)	532.4 (20.1)	531.6 (20.1)
R44[5]	327.1 (13.3)	330.7 (13.4)	332.6 (13.5)	329.9 (13.4)	323.3 (13.1)	322.9 (13.1)

[1]Order of the polynomial used as the mean regression of age on ultrasound back fat.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept and linear random regression terms.
[4]Permanent environmental variance.
[5]Residual variance estimates corresponding to each of the four ultrasound back fat quartiles.

Table 6.8. Variance estimates (SE) obtained from a linear random regression model for days to ultrasound back fat using linear residual random regression.

	1[1]	2[1]	3[1]	4[1]	5[1]	6[1]
LogL[2]	-8171.26	-7998.27	-7971.46	-7948.88	-7922.83	-7922.6
Intercept[3]	4183 (289.1)	1053 (17.4)	1323 (120.9)	1289 (119.7)	1008 (16.7)	1007 (16.7)
Int, Lin[3]	1973 (174.0)	-56.22 (0.9)	-48.49 (48.44)	-42.32 (66.2)	-164.2 (2.7)	-164.3 (2.7)
Linear[3]	1082 (121.9)	195.4 (3.2)	15.11 (0.3)	58.90 (56.2)	42.82 (0.7)	44.81 (0.7)
PE Intercept[4]	88.93 (1.7)	145.5 (2.4)	107.9 (2.0)	107.1 (2.0)	112.3 (1.9)	112.3 (1.9)
PE Int, Lin[4]	-46.53 (0.9)	-71.10 (1.2)	-57.76 (1.1)	57.33 (1.1)	-60.15 (1.0)	-60.14 (1.0)
PE Lineaar[4]	27.30 (0.5)	37.37 (0.6)	33.89 (0.6)	33.6 (0.6)	35.15 (0.6)	35.15 (0.6)
Residual[5]	522.7 (10.0)	516.5 (8.5)	515.8 (9.4)	511.2 (9.6)	513.3 (8.5)	513.3 (8.5)

[1]Order of the polynomial used as the mean regression of age on ultrasound back fat.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept and linear random regression terms.
[4]Permanent environmental (co) variance for the intercept and linear random regression terms.
[5]Residual variance.

In order to make comparisons between the residual random regression model and heterogeneous residual variance models, observed genetic variance, phenotypic variance, permanent environmental variance and heritability were calculated. These estimates were plotted versus their respective endpoint for DTUREA and DTUBF. For the trait DTUREA, both sub-models produced variance and heritability estimates very similar to one another. Genetic variance (Figure 6.3) was nearly identical between the two models as was heritability (Figure 6.6). Permanent environmental variance (Figure 6.4) was obviously different given the inherent differences between the two models, even though the scale of the graph tends to exaggerate these differences making them appear larger. Estimates of phenotypic variance (Figure 6.5) obtained from the LRRR model tended to be higher than those obtained from the heterogeneous residual variance model, by a magnitude of approximately 100 days^2 across the entire range of observations.

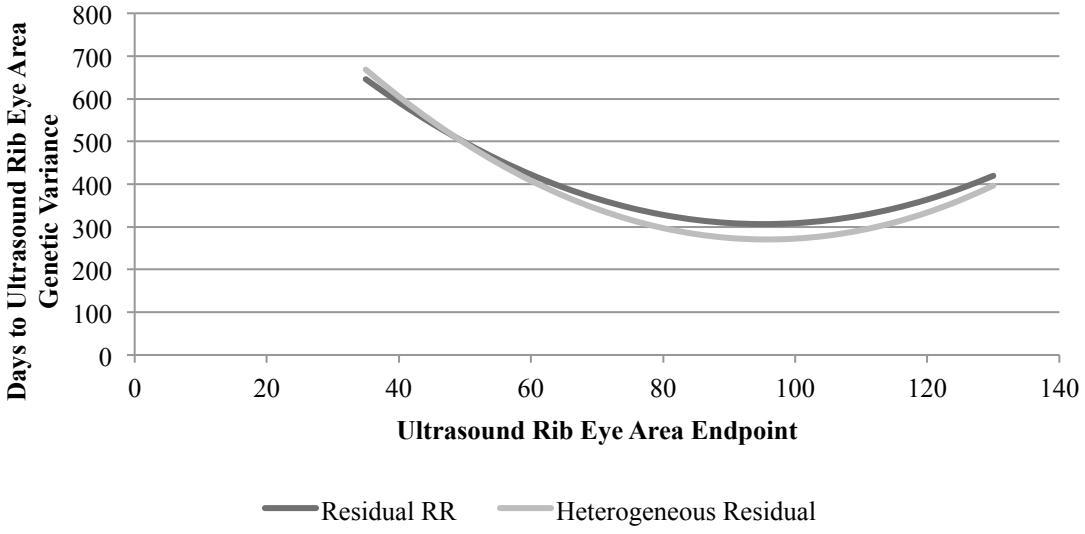


Figure 6.3. Plot of observed days to ultrasound rib eye area genetic variance obtained from the linear random regression of age on ultrasound rib eye area using models containing both heterogeneous residual variances as well as linear residual random regression.

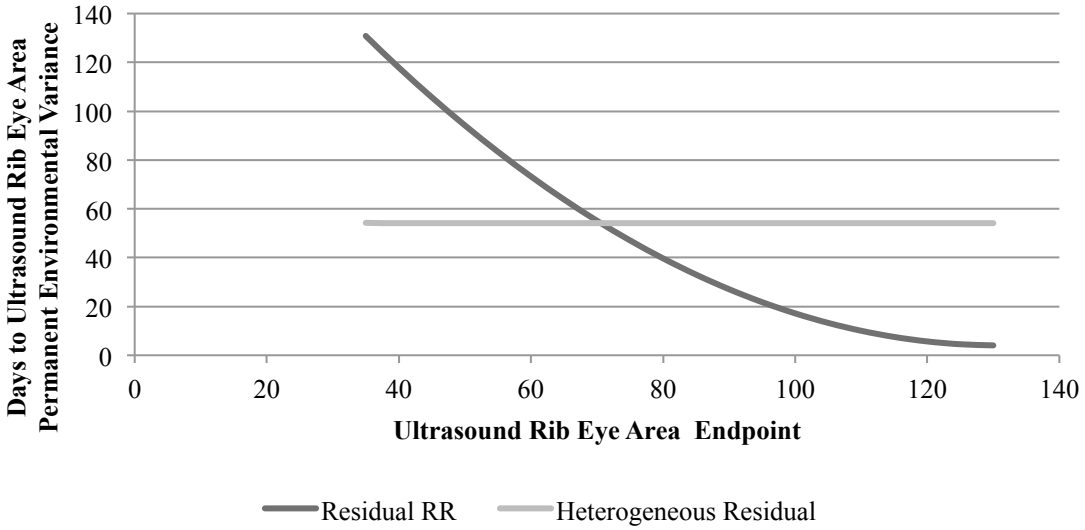


Figure 6.4. Plot of observed days to ultrasound rib eye area permanent environmental variance obtained from the linear random regression of age on ultrasound rib eye area using both heterogeneous residual variance and linear residual random regression.

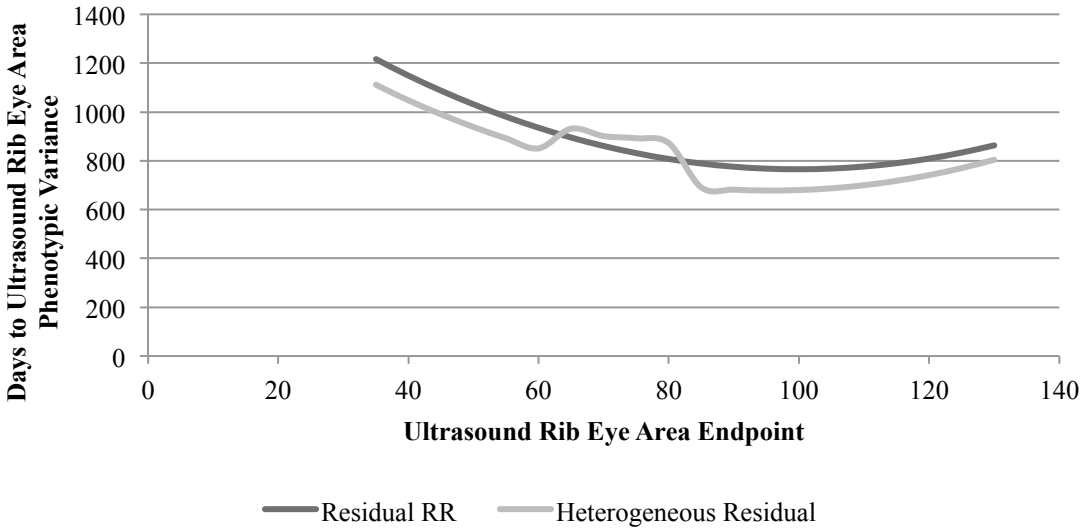


Figure 6.5. Plot of observed days to ultrasound rib eye area phenotypic variance obtained from the linear random regression of age on ultrasound rib eye area using both heterogeneous residual variance and linear residual random regression.

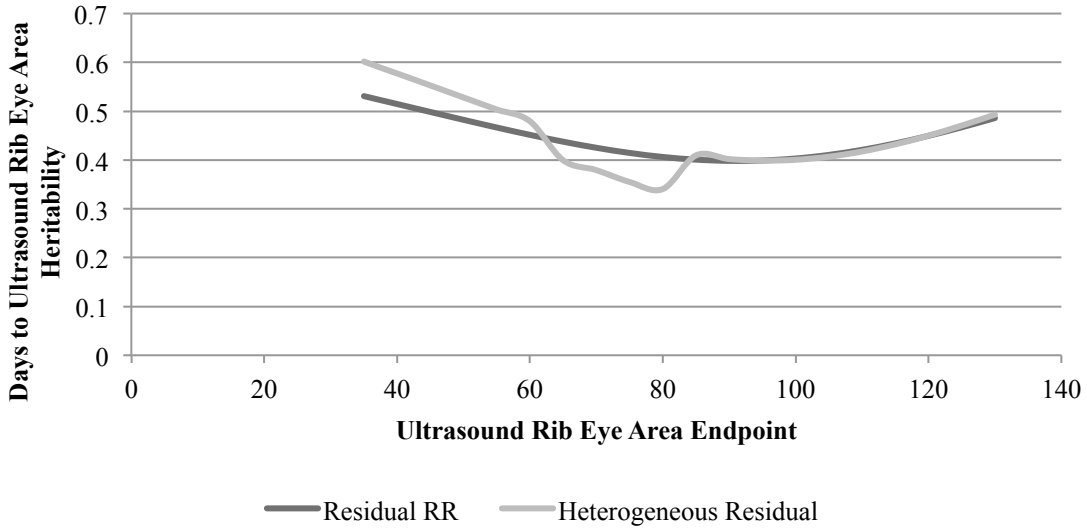


Figure 6.6. Plot of observed days to ultrasound rib eye area heritability obtained from the linear random regression of age on ultrasound rib eye area using both heterogeneous residual variance and linear residual random regression.

For DTUBF, estimates of observed genetic variance (Figure 6.7) differ more between the two sub-models. For lower amounts of deposited fat, the LRRR model predicted higher genetic variance in days by about 200 days^2 (at the endpoint of 1.53 mm). As the target amount of back fat increases, the two models re-rank resulting in the HRV model predicting higher genetic variation in days by 232 days^2 for the upper end of fat deposition. The models re-ranked at the endpoint 18 mm of back fat. Estimates of permanent environmental variance (Figure 6.8) from the LRRR model were lower than those from the HRV model. At the upper end of fat deposition in this data set, the permanent environmental decreases to nearly zero meaning there are no longer any permanent environmental influences in fat deposition. Perhaps this is due to the nature in which the animals were deemed ready for harvest, mostly by visual appraisal.

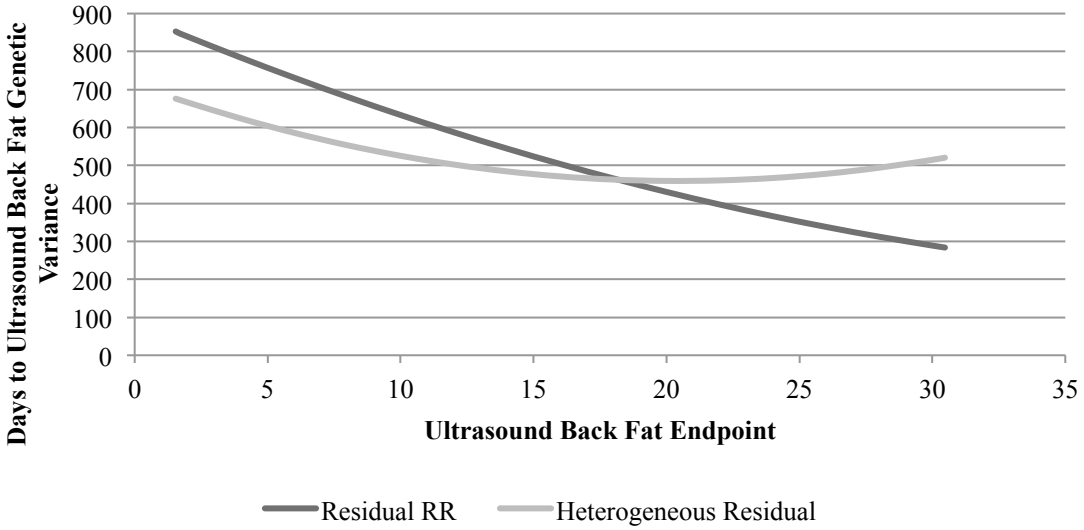


Figure 6.7. Plot of observed days to ultrasound back fat genetic variance obtained from the linear random regression of age on ultrasound back fat using both heterogeneous residual variances and linear residual random regression.

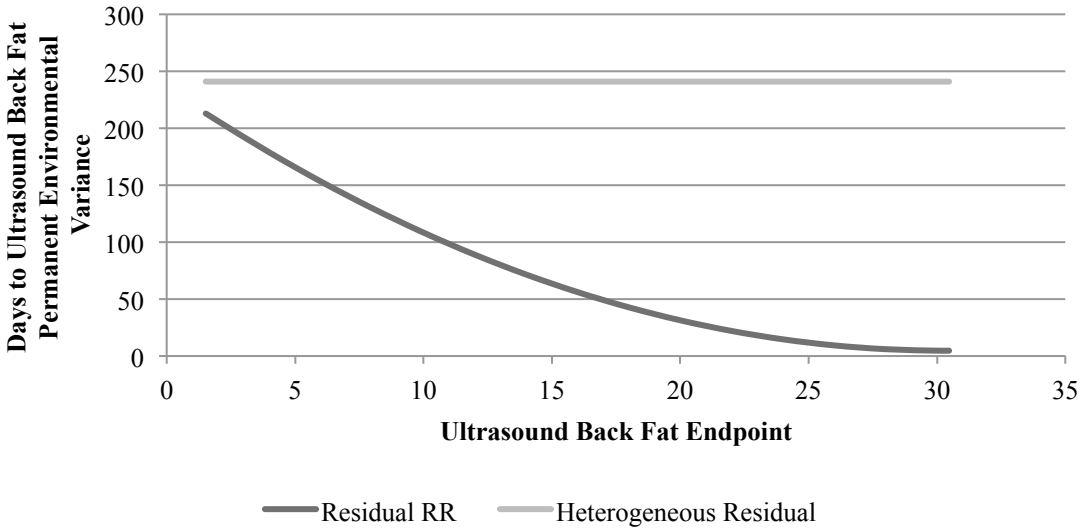


Figure 6.8. Plot of observed days to ultrasound back fat permanent environmental variance obtained from the linear random regression of age on ultrasound back fat using both heterogeneous residual variance and linear residual random regression.

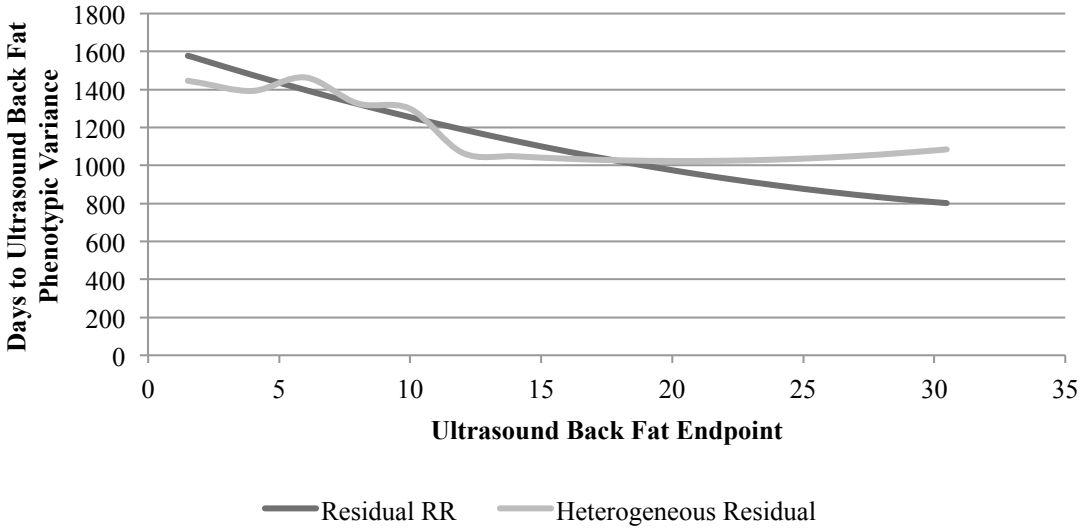


Figure 6.9. Plot of observed days to ultrasound back fat phenotypic variance obtained from the linear random regression of age on ultrasound back fat using both heterogeneous residual variance and linear residual random regression.

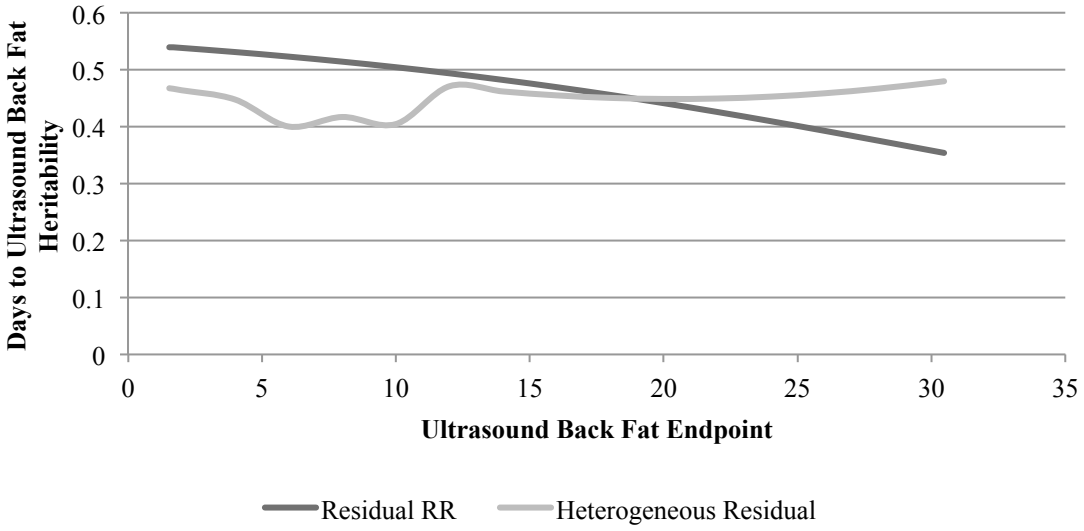


Figure 6.10. Plot of observed days to ultrasound back fat heritability obtained from the linear random regression of age on ultrasound back fat using both heterogeneous residual variance and linear residual random.

Given the differences in genetic and permanent environmental variance, estimates of phenotypic variance (Figure 6.9) obtained from the two models tend to agree with one another for the leaner cattle. As the target UBF endpoint increases, the HRV model begins to result in higher estimates of phenotypic variation for cattle with more than 20 mm of fat thickness. For the UBF endpoint range of 12 to 22 mm, both residual variance sub-models resulted in approximately the same phenotypic variance estimate. In the range of UBF observations where the data was most dense, the LRRR model gave higher heritability estimates. As the UBF endpoint increases data density decreases, and it is in this range where the LRRR model resulted in lower heritability estimates than the HRV model.

Quadratic Random Regression. Following the linear random regression, a quadratic random regression was implemented using the data set sifted to require three or more observations on individual animals for both DTUREA and DTUBF (see Table 6.1). This data set contained 7,361 age and ultrasound observations on 1,317 individual animals averaging 5.59 observations per individual. Age and ultrasound measurement summary statistics are shown below in Table 6.9. Compared to the data set used in the linear random regression model, the requirement of three or more observations on individual animals reduced the total number of animals in the final data set by six resulting in 12 fewer observations. Contemporary group numbers from this restricted data set are the same as observed with the data set used in the linear random regression containing 62 unique contemporary groups. Here, the average number of animal per group is slightly

Table 6.9. Summary statistics for age, ultrasound rib eye area and ultrasound back fat used in the quadratic random regression for the days to "finish" genetic prediction

	Age (d)	UREA[1]	UBF[2]
N	7,361	7,361	7,361
Average	394.2	73.41	8.71
Variance	1,959.1	135.57	16.42
Minimum	276	36.77	1.53
Maximum	519	129.54	30.47

[1]Ultrasound rib eye area.
[2]Ultrasound back fat.

smaller 21.2 animals per group, compared to the 21.3 animals per group in the linear random regression data set.

For the purpose of estimating variance components, a 4-generation ancestral pedigree was built from this final data file. The resulting pedigree contained 5,408 individual animals with 1,386 unique sires and 2,706 unique dams. The average inbreeding coefficient in this pedigree was 1.5% with minimum and maximum inbreeding levels of 0% and 25%, respectively.

Similar to the linear model, direct genetic (co) variance estimates were obtained for the quadratic random regression model. These estimates contained intercept, linear and quadratic genetic variances estimates as well as all associated covariances for both the HRV model as well as the LRRR model. Here again, appropriate fixed regression orders were obtained for both DTUBF and DTUREA. Since random regression models are intended to model deviations around the phenotypic trajectory (Schaeffer, 2003), Gilmour (2009) suggests their order not be less than the order specified in the random terms. Starting with the quadratic fixed regression model, incrementally higher fixed regression orders were included in the evaluation until the test statistics and associated p-

Table 6.10. Best linear unbiased estimates and associated significance values for the different fixed regression polynomial orders obtained from the quadratic random regression of age on ultrasound rib eye area using both heterogeneous residual variance and linear residual random regression.

Polynomial Order[1]	Last Coefficient[2]	SE[2]	F[3]	*P*-value
	Heterogeneous Residual Variance			
1[4]				
2	-27.37	1.62	293.85	0.00
3	-40.06	1.96	419.45	0.00
4	12.99	1.93	45.11	0.00
5	-2.33	1.74	1.66	0.19
	Linear Residual Random Regression			
1[4]				
2	-29.38	1.19	608.43	0.00
3	-41.52	2.02	422.49	0.00
4	13.45	1.99	45.50	0.00
5	-2.08	1.80	1.34	0.25

[1]Order of the fixed regression polynomial
[2]Best liner unbiased estimated and standard error corresponding to the highest order term in the fixed regression
[3]Fdistribution test statistic and associated *P*-value
[4]Fixed regression order omitted because it is lower than the order of the regression specified in the random terms.

values show the increased order did not account for any additional variation. The estimates for DTUREA are shown below in Table 6.10.

For the trait DTUREA using the quadratic random regression model, the 4th order fixed regression polynomial was sufficient in describing the mean relationship between age and UREA for both the HRV model and the LRRR model. This order is lower than that from the linear random regression model where a 6th order polynomial was needed to sufficiently describe the relationship between age and UREA.

The estimates for DTUBF are shown below in Table 6.11. Here, some strange behavior is observed as the fixed polynomial orders are increased.

Table 6.11. Best linear unbiased estimates and associated significance values for the different fixed regression polynomial orders obtained from the quadratic random regression of age on ultrasound back fat using both heterogeneous residual variance and linear residual random regression.

Polynomial Order[1]	Last Coefficient[2]	SE[2]	F[3]	*P*-value
	Heterogeneous Residual Variance			
1[4]				
2	-44.57	1.67	715.71	0.00
3	-1.26	1.45	0.76	0.38
4	-7.62	1.14	44.49	0.00
5	8.04	0.99	66.08	0.00
6	-0.83	1.08	0.59	0.44
	Linear Residual Random Regression			
1[4]				
2	-49.22	2.11	543.12	0.00
3	0.34	1.77	0.03	0.85
4	-9.46	1.32	51.47	0.00
5	8.66	1.17	54.90	0.00
6	-0.70	1.25	0.32	0.57

[1]Order of the fixed regression polynomial
[2]Best liner unbiased estimated and standard error corresponding to the highest order term in the fixed regression
[3]Fdistribution test statistic and associated *P*-value
[4]Fixed regression order omitted because it is lower than the order of the regression specified in the random terms.

As the order of the polynomial is increased from a quadratic to a cubic, according to the significance tests, the cubic term does not significantly account for any additional variation in age over the previously fitted quadratic term. A quick look shows Legendre polynomial genetic variance estimates are still changing. Increasing the polynomial further shows the 4th and 5th order polynomials are accounting for variation in days, and once we reach the 6th order polynomial, higher order terms are no longer significant and the estimates of genetic variance have stabilized (see Table 6.14 and Table 6.15).

Variance estimates for DTUREA corresponding to the HRV model (Table 6.12) and LRRR model (Table 6.13) are shown below. Both tables contain the estimates obtained from the fixed regression orders where significance values were reported above in Table 6.10. Genetic variance estimates obtained from the HRV model appear to stabilize as the order of the fixed regression model increases. Also, fixed regression orders above a quadratic polynomial sent the estimate of permanent environmental variance to zero suggesting the absence of any non-genetic permanent environmental effects influencing DTUREA.

The effect of increasing fixed regression orders on DTUREA variance estimates using LRRR (Table 6.13) indicated more variable changes are observed with increasing orders of the fixed regression when compared to the linear DTUREA model. Here, estimates of variance seem to fluctuate wildly, especially where the covariances are concerned. Once the fixed regression order is increased to a 5^{th} order polynomial, the first non-significant order, the genetic variance estimates for the linear, quadratic and linear-quadratic covariance, seem to be largely inflated. For the LRRR model this erratic behavior is perhaps due to the fact that the quadratic random regression is not significantly accounting for any additional variation in days, a point that will be addressed in more detail later on in the chapter.

Observed DTUREA genetic variance, permanent environmental variance, phenotypic variance and heritability were calculated for the range of UREA (35 cm^2 to 130 cm^2) and are shown below in Figure 6.11, Figure 6.12, Figure 6.13, and Figure 6.14, respectively. Here we see similar trends as were present in the linear random regression for DTUREA presented earlier.

Table 6.12. Variance estimates (SE) obtained from a quadratic random regression model for days to ultrasound rib eye area using heterogeneous residual variance.

	1[1]	2[1]	3[1]	4[1]	5[1]
LogL[2]		-7165.51	-7053.09	-7031.08	-7029.73
Intercept[3]		637.1 (126.8)	165.8 (63.9)	184.4 (65.6)	174.0 (64.5)
Int, Lin[3]		-128.9 (41.5)	-46.9 (52.7)	-14.3 (56.2)	-27.3 (55.3)
Linear[3]		98.2 (46.5)	461.1 (76.6)	479.5 (80.2)	467.3 (79.2)
Int, Quad[3]		60.6 (46.1)	-6.6 (68.2)	-11.2 (68.4)	-28.0 (67.5)
Lin, Quad[3]		-84.0 (31.1)	414.7 (76.7)	423.4 (79.1)	406.7 (78.4)
Quadratic[3]		87.6 (44.1)	591.8 (99.5)	552.7 (98.4)	532.1 (97.6)
PE Intercept[4]		135.4 (81.8)	0.0 (0.0)[6]	0.0 (0.0)[6]	0.0 (0.0)[6]
R11[5]		385.4 (15.9)	371.4 (15.2)	372.6 (15.1)	373.0 (15.1)
R22[5]		512.8 (19.5)	494.5 (19.1)	490.7 (18.9)	490.7 (18.9)
R33[5]		521.3 (19.9)	492.9 (19.1)	493.1 (19.1)	493.6 (19.1)
R44[5]		354.8 (14.3)	347.6 (14.0)	346.9 (14.0)	346.6 (14.0)

[1]Order of the polynomial used as the mean regression of age on ultrasound rib eye area.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept, linear and quadratic random regression terms.
[4]Permanent environmental variance.
[5]Residual variance estimates corresponding to each of the four ultrasound rib eye area quartiles.
[6]Estimates of permanent environmental variance were fixed to keep the variance estimates positive definite by ASReml.

Table 6.13. Variance estimates (SE) obtained from a quadratic random regression model for days to ultrasound rib eye area using linear residual random regression.

	1[1]	2[1]	3[1]	4[1]	5[1]
LogL[2]		-7143.8	-7080.9	-7043.3	-7038.6
Intercept[3]		766.7 (12.7)	115.6 (63.2)	145.5 (48.7)	504.8 (8.4)
Int, Lin[3]		-108.8 (1.8)	-40.4 (54.7)	9.7 (0.2)	563.3 (9.3)
Linear[3]		164.1 (2.7)	473.3 (80.6)	510.0 (59.3)	1287.0 (21.3)
Int, Quad[3]		108.0 (1.8)	-14.1 (70.6)	4.3 (0.1)	572.0 (9.5)
Lin, Quad[3]		-184.8 (3.1)	441.6 (80.6)	473.5 (45.9)	990.3 (16.4)
Quadratic[3]		208.5 (3.5)	610.8 (103.4)	602.8 (55.6)	1291.0 (21.4)
PE Intercept[4]		51.8 (0.9)	15.6 (0.3)	16.1 (0.3)	40.2 (0.7)
PE Int, Lin[4]		-25.1 (0.4)	-10.7 (0.2)	-10.7 (0.2)	-14.9 (0.2)
PE Lineaar[4]		12.4 (0.2)	7.4 (0.1)	7.4 (0.1)	5.5 (0.1)
Residual[5]		432.4 (7.2)	425.3 (8.2)	424.1 (8.1)	406.4 (6.7)

[1]Order of the polynomial used as the mean regression of age on ultrasound rib eye area.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept, linear and quadratic random regression terms.
[4]Permanent environmental (co) variance for the intercept and linear random regression terms.
[5]Residual variance.

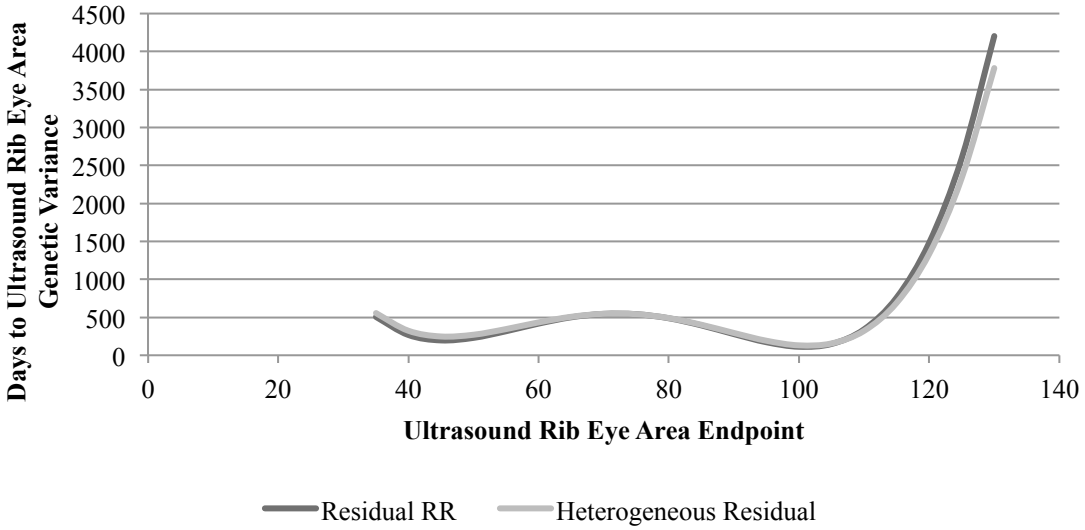


Figure 6.11. Plot of observed days to ultrasound rib eye area genetic variance obtained from the quadratic random regression of age on ultrasound rib eye area using models containing both heterogeneous residual variances as well as linear residual random regression.

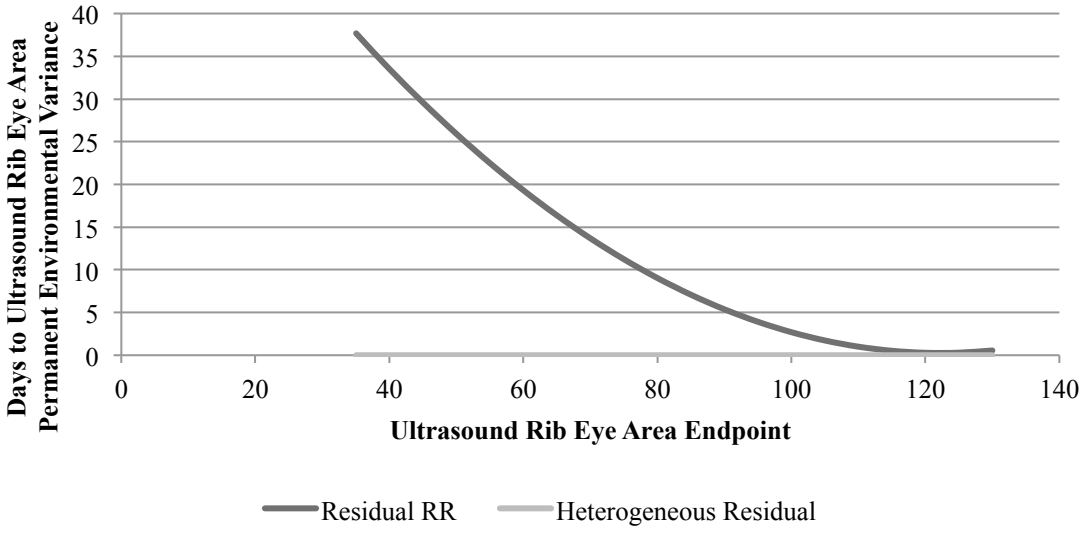


Figure 6.12. Plot of observed days to ultrasound rib eye area permanent environmental variance obtained from the quadratic random regression of age on ultrasound rib eye area using models containing both heterogeneous residual variances as well as linear residual random regression.

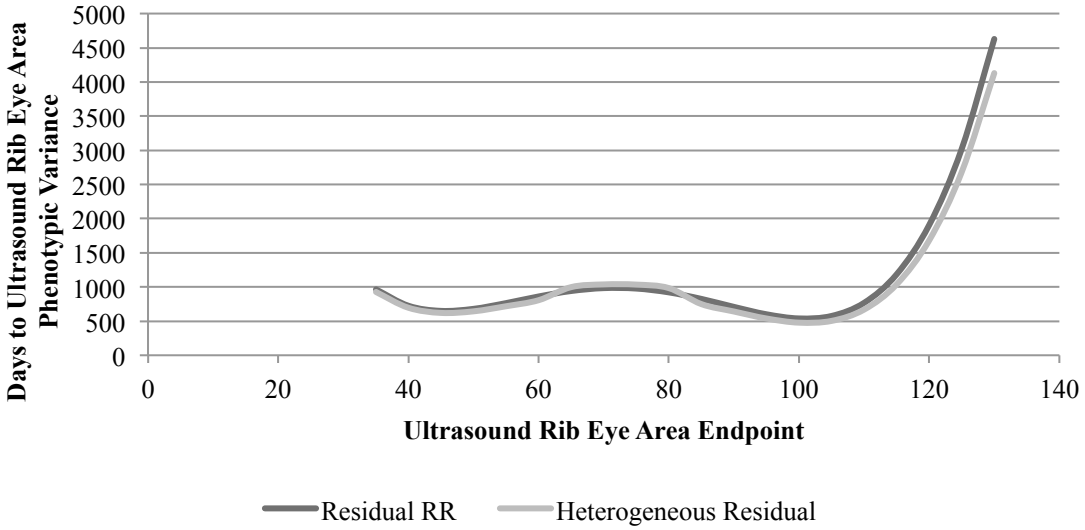


Figure 6.13. Plot of observed days to ultrasound rib eye area phenotypic variance obtained from the quadratic random regression of age on ultrasound rib eye area using models containing both heterogeneous residual variances as well as linear residual random regression.

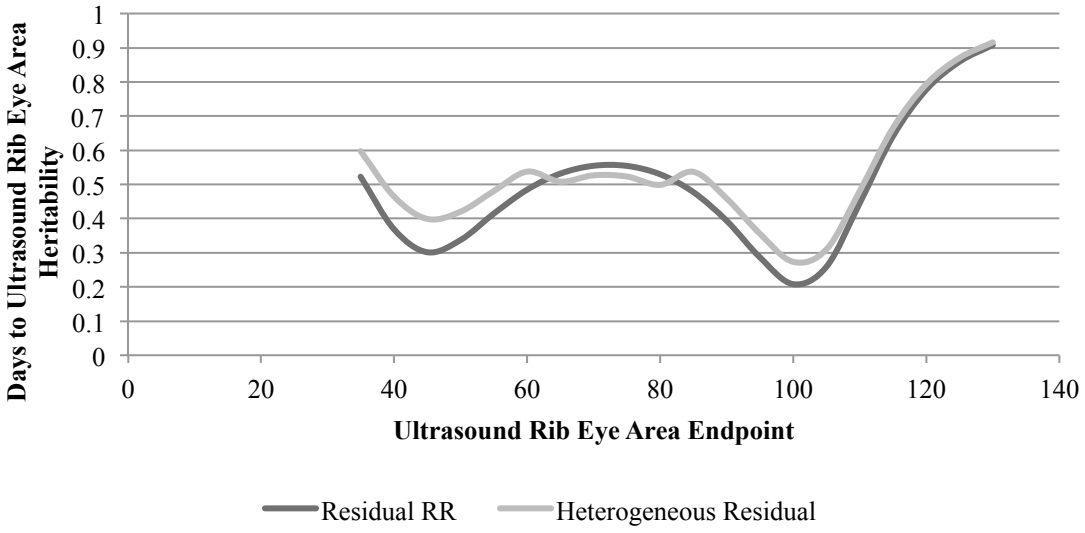


Figure 6.14. Plot of observed days to ultrasound rib eye area heritability obtained from the quadratic random regression of age on ultrasound rib eye area using models containing both heterogeneous residual variances as well as linear residual random regression.

Variance estimates obtained from both the HRV and LRRR models are nearly identical. The figure containing the observed estimates of permanent environmental variance, Figure 6.12, is the only figure that presents any sort of differences. Permanent environmental variance estimated from the LRRR model begins at 37.7 $days^2$ and approaches zero as UREA increases. The magnitude of this variance is rather low, and accounts for very little of the overall phenotypic variation of DTUREA. This chart suggests that while there is some additional non-genetic variation being accounted for in DTUREA by the LRRR model, in the overall scheme of things, the amount of variation is actually quite low. The benefit of the residual random regression model is that it properly accounts for the changing error covariance structure as UREA increases. Compared to the estimates of variance obtained from the linear random regression model on DTUREA, genetic variance obtained from this quadratic random regression typically tended to be higher, with a very large spike in variance as UREA increases.

Variance estimates for DTUBF obtained from the HRV model and LRRR model are shown below in Table 6.14 and Table 6.15, respectively. For both residual variance sub-models, variance estimates for the random polynomials tended to stabilize with increasing fixed regression orders. Looking specifically at the variance estimates obtained from the HRV model in Table 6.14, an absence of standard errors for the genetic variance estimates was observed. These standard errors were not estimated by the software package ASReml. The estimates were flagged as being on the verge of changing from an estimate resulting in a positive definite genetic (co)variance matrix to a boundary estimate. A boundary estimate in ASReml means the software will fix the estimate at a small value within the parameter space in an effort to keep the estimated

variance matrix positive definite. It is hypothesized that as sufficient variation in the data becomes limiting, proper standard errors are unable to be estimated due to the restrictions put on the variance estimates. Fixing these variance estimates allows ASReml to reach a log-likelihood convergence, and as such the estimates are the best point estimate given within the parameter space. Also, similar to the trait DTUREA, the HRV model has the permanent environmental variance approaching zero.

The DTUBF estimates obtained from the LRRR model (Table 6.15) show the presence of standard errors associated with each of the genetic variance estimates. Perhaps, the modeling of the error covariance with the LRRR model better allows ASReml to estimate genetic variance of the random regression within the parameter space. Another difference between the two models is the estimates obtained from the LRRR model are much smaller in magnitude than those obtained from the HRV model suggesting that some of the variation attributed to genetics in the HRV model may have been partitioned to permanent environment in the LRRR model.

Observed DTUBF genetic, permanent environmental and phenotypic variance estimates are presented below in Figures 6.15, 6.16, and 6.17, respectively. Estimates of observed heritability are presented in Figure 6.18. Here, the LRRR model tended to give lower estimates of genetic variance when compared to the heterogeneous residual variance model.

Table 6.14. Variance estimates (SE) obtained from a quadratic random regression model for days to ultrasound back fat area using heterogeneous residual variance.

	1[1]	2[1]	3[1]	4[1]	5[1]	6[1]
LogL[2]		-7838.95	-7838.87	-7817.23	-7787.05	-7786.65
Intercept[3]		546.6 (0.0)	557.5 (0.0)	526.5 (0.0)	634.9 (0.0)	634.1 (0.0)
Int, Lin[3]		-349.6 (0.0)	-365.9 (0.0)	-359.6 (0.0)	-448.5 (0.0)	-448.8 (0.0)
Linear[3]		284.9 (0.0)	304.1 (0.0)	305.4 (0.0)	380.0 (0.0)	379.5 (0.0)
Int, Quad[3]		-254.2 (0.0)	-263.2 (0.0)	-246.0 (0.0)	-271.4 (0.0)	-271.6 (0.0)
Lin, Quad[3]		148.4 (0.0)	175.4 (0.0)	149.6 (0.0)	184.2 (0.0)	183.6 (0.0)
Quadratic[3]		133.6 (0.0)	156.4 (0.0)	135.8 (0.0)	156.2 (0.0)	155.7 (0.0)
PE Intercept[4]		278.4 (58.5)	238.6 (58.9)	260.3 (58.9)	0.7 (0.0)	0.7 (0.0)
R11[5]		544.2 (20.4)	544.9 (20.5)	529.3 (19.9)	520.9 (19.6)	521.3 (19.6)
R22[5]		628.7 (23.6)	629.4 (23.6)	625.3 (23.4)	631.7 (23.7)	632.6 (23.7)
R33[5]		519.0 (19.7)	518.8 (19.7)	525.4 (20.0)	527.5 (20.1)	526.9 (20.1)
R44[5]		326.4 (12.9)	326.1 (13.0)	324.2 (12.9)	315.5 (12.5)	315.3 (12.5)

[1]Order of the polynomial used as the mean regression of age on ultrasound back fat.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept, linear and quadratic random regression terms.
[4]Permanent environmental variance.
[5]Residual variance estimates corresponding to each of the four ultrasound rib eye area quartiles.

Table 6.15. Variance estimates (SE) obtained from a quadratic random regression model for days to ultrasound back fat using linear residual random regression.

	1[1]	2[1]	3[1]	4[1]	5[1]	6[1]
LogL[2]		-7921.25	-7920.68	-7896.52	-7874.15	-7873.93
Intercept[3]		296.1 (4.9)	309.7 (5.1)	296.6 (4.9)	322.4 (5.3)	321.5 (5.3)
Int, Lin[3]		-299.4 (5.0)	-300.8 (5.0)	-308.0 (5.1)	-273.3 (4.5)	-273.0 (4.5)
Linear[3]		498.3 (8.2)	470.7 (7.8)	367.6 (6.1)	343.4 (5.7)	343.5 (5.7)
Int, Quad[3]		-181.2 (3.0)	-181.1 (3.0)	-163.3 (2.7)	-117.8 (2.0)	-117.8 (2.0)
Lin, Quad[3]		316.2 (5.2)	295.7 (4.9)	175.8 (2.9)	168.1 (2.8)	169.2 (2.8)
Quadratic[3]		311.4 (5.2)	297.4 (4.9)	202.7 (3.4)	227.1 (3.8)	230.2 (3.8)
PE Intercept[4]		273.5 (4.5)	277.2 (4.6)	300.0 (5.0)	300.5 (5.0)	300.8 (5.0)
PE Int, Lin[4]		-47.3 (0.8)	-48.1 (0.8)	-55.2 (0.9)	-60.2 (1.0)	-60.4 (1.0)
PE Lineaar[4]		30.0 (0.5)	30.9 (0.5)	32.1 (0.5)	34.3 (0.6)	34.3 (0.6)
Residual[5]		499.6 (8.3)	499.7 (8.3)	498.9 (8.3)	495.0 (8.2)	494.9 (8.2)

[1]Order of the polynomial used as the mean regression of age on ultrasound back fat.
[2]REML log-likelihood obtained from ASReml.
[3]Direct genetic (co) variance for the intercept, linear and quadratic random regression terms.
[4]Permanent environmental (co) variance for the intercept and linear random regression terms.
[5]Residual variance.

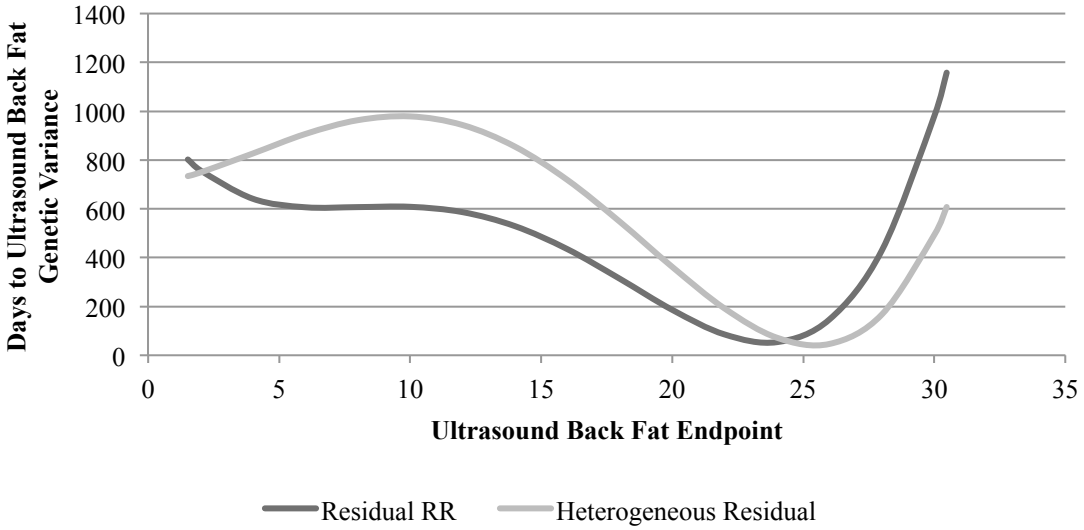


Figure 6.15. Plot of observed days to ultrasound back fat genetic variance obtained from the quadratic random regression of age on ultrasound back fat using models containing both heterogeneous residual variances as well as linear residual random regression on residuals.

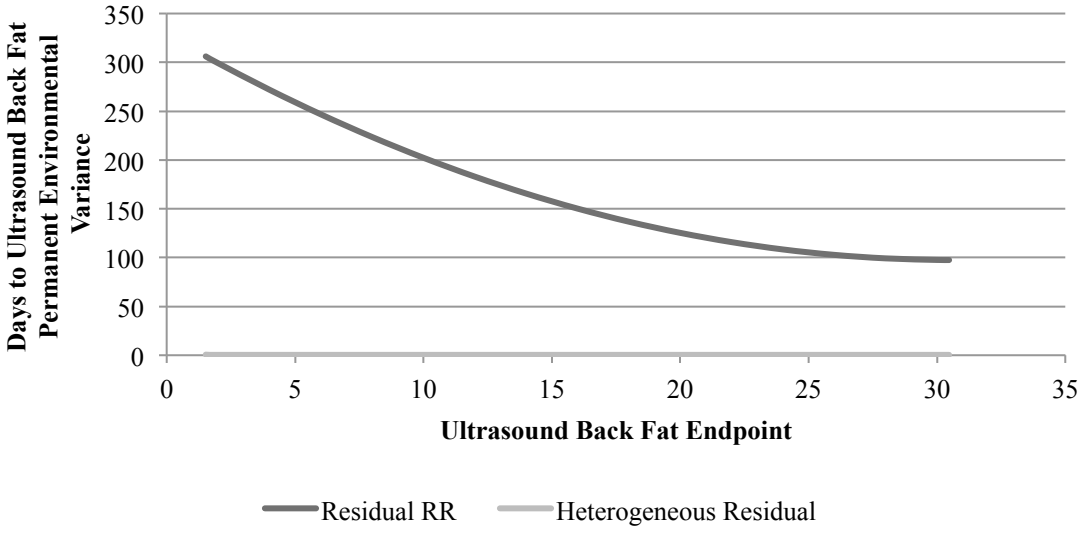


Figure 6.16. Plot of observed days to ultrasound back fat permanent environmental variance obtained from the quadratic random regression of age on ultrasound back fat using models containing both heterogeneous residual variances as well as linear residual random regression.

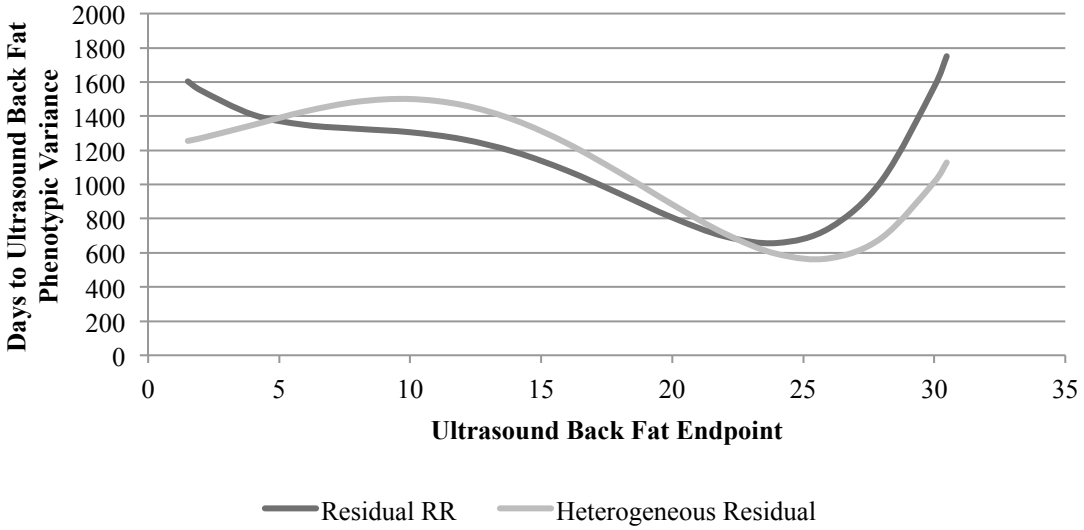


Figure 6.17. Plot of observed days to ultrasound back fat phenotypic variance obtained from the quadratic random regression of age on ultrasound back fat using models containing both heterogeneous residual variances as well as linear residual random regression.

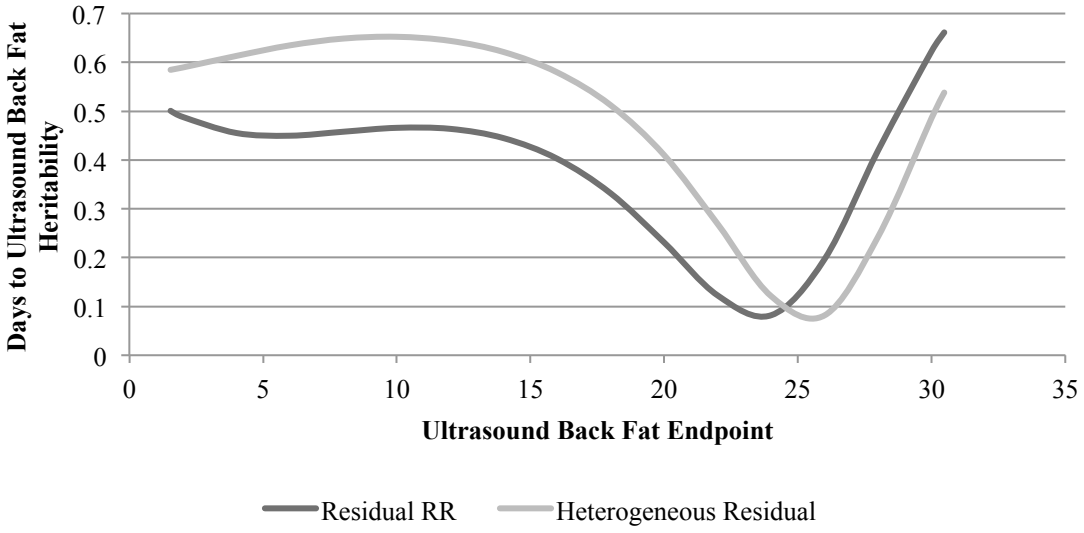


Figure 6.18. Plot of observed days to ultrasound back fat phenotypic variance obtained from the quadratic random regression of age on ultrasound back fat using models containing both heterogeneous residual variances as well as linear residual random regression.

Also, the permanent environmental variance tended to be much higher for the LRRR model as opposed to the HRV model. Phenotypic variance between the two models was fairly equal, suggesting again that the inclusion of the error covariance between observations re-partitions some of the variation from the genetic component of DTUBF. Figure 6.18 contains the observed heritability estimates and shows the LRRR model gives more reasonable estimates, even though both models result in a spike in heritability as UBF increases.

One thing interesting about these charts for DTUBF, is the drop in genetic and phenotypic variance as well as heritability that occurs between approximately 20 and 28 mm of UBF. The animals with UBF observations above 20 mm are all Angus animals. There were a total of 92 UBF observations with observations of 20 mm and greater, representing 71 individual animals averaging 1.2 observations per animal. A total of 52 animals have only one observation in this data range. Also, the distribution of UBF Figure 6.2 seems to be heavily skewed toward the left or toward smaller observations. This is the only trait where this trend occurs, which may be a contributor to this dip in variance. The lower observations are weighted fairly heavily while the higher observations are not and where this dip occurs, only Angus animals are represented.

Cubic and Quartic Random Regressions. Random regression models with an order higher than quadratic were only performed for the HRV model. As will be discussed in a subsequent section of this chapter, the highest significant random regression order for the models performing the LRRR was a linear regression. Considering the HRV model here, the highest significant order for DTUREA was the

highest order considered in this study, a quartic random regression. Therefore, discussion of DTUREA here will contain results from both the cubic and quartic random regressions.

Cubic random regression models for both DTUREA and DTUBF were implemented using the data set sifted to remove those animals with fewer than four observations per individual (see Table 6.1). This data set contained 7,340 age and ultrasound observations on 1,310 individual animals averaging 5.6 observations per individual. This data set contained 21 fewer observations on seven fewer animals than did that used for the quadratic random regressions. Contemporary groups were formed in the same manner as the previously described analyses. Here there are still 62 unique contemporary groups with the average number of animals per group of 21.1 being slightly smaller than that used in the quadratic random regression model. A 4-generation pedigree was built from this sifted final data file resulting in 5,398 individual animals with 1,385 unique sires and 2,703 unique dams.

A quartic random regression model was performed for DTUREA using the data set described in Table 6.1, which includes data from individual animals possessing five or more observations per animal. This data set consisted of 6,600 age and ultrasound observations representing 1,125 individual animals. This data set is much smaller than all of the previous data sets due to the observation requirement. An entire year's worth of data was removed because there are only four observations per animal.

Many of the issues occurring with the quadratic random regression model were also observed here in both the cubic and quartic random regressions. These issues include the inflated variance estimates at the upper end of the data for both DTUREA and DTUBF. Now, with higher order random polynomials, the lower ends of the data range

are seeing these inflated variance estimates. This problem is exacerbated with increasing the polynomial order to the quartic regression for DTUREA. As such, only a brief discussion of these models will be presented here. Results presented will be the observed variance and heritability estimates, with the actual random regression variance estimates excluded.

Beginning with DTUREA, observed genetic, phenotypic variance estimates along with heritabilities are shown below in Figure 6.19, Figure 6.20, and Figure 6.21 for both the cubic and quartic random regressions, respectively. The shape of the curve of observed genetic variance is nearly identical for both cubic and quartic random regression models. Differences between these two curves are observed in the tails of the data range with the quartic random regression model being more inflated than the cubic random regression model. Comparing these estimates to the quadratic random regression (Figure 6.11), all estimated variances are nearly identical where the observation density is at its greatest, with the only differences being the magnitude of variance inflation in the data extremes. As such, phenotypic variance (Figure 6.20) and heritability (Figure 6.21) are nearly identical in the two models as well suggesting that the cubic and quartic models are essentially predicting the same observed estimates of variance. Plots of observed permanent environmental variance were not included here because as with the quadratic random regression, these were estimated as essentially zero.

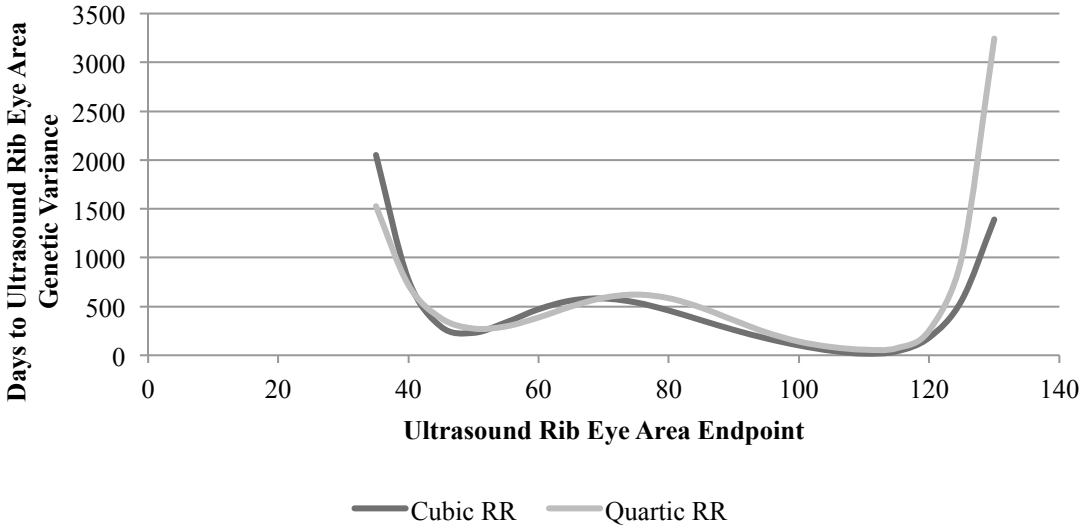


Figure 6.19. Plot of observed days to ultrasound rib eye area genetic variance obtained from both cubic and quartic random regressions of age on ultrasound rib eye area.

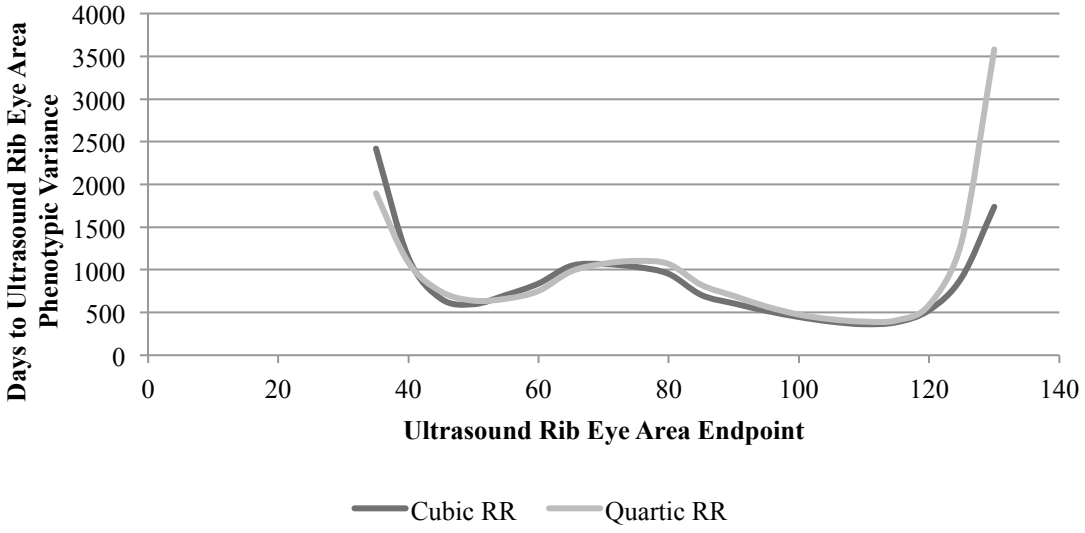


Figure 6.20. Plot of observed days to ultrasound rib eye area phenotypic variance obtained from both cubic and quartic random regressions of age on ultrasound rib eye area.

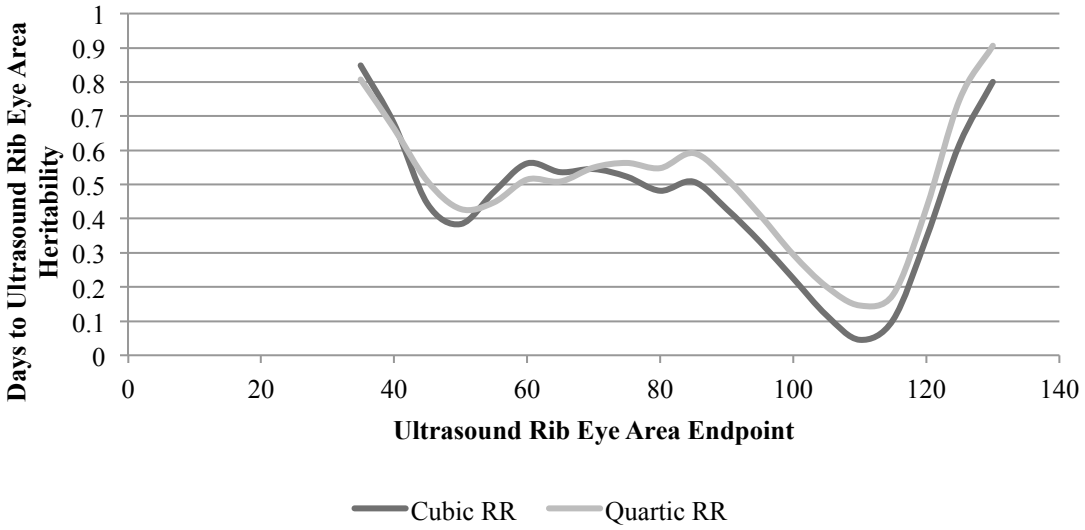


Figure 6.21. Plot of observed days to ultrasound rib eye area heritability obtained from both cubic and quartic random regressions of age on ultrasound rib eye area.

Plots of observed genetic, phenotypic variance as well as heritability for DTUBF are shown below in Figure 6.22, Figure 6.23 and Figure 6.24**,** respectively. Genetic variance is very similar to that estimated by the quadratic random regression model, with the exception of the inflation which occurred at the upper end of the UBF data range. The dip in genetic variance, which occurred at approximately 20 – 28 mm of UBF for the quadratic random regression model, is not present in the cubic random regression model. Permanent environmental variance was again estimated to be near zero, which is why the plot of observed variance is omitted from this discussion, and why the plot of observed phenotypic variance is a mirror of the genetic variance plot.

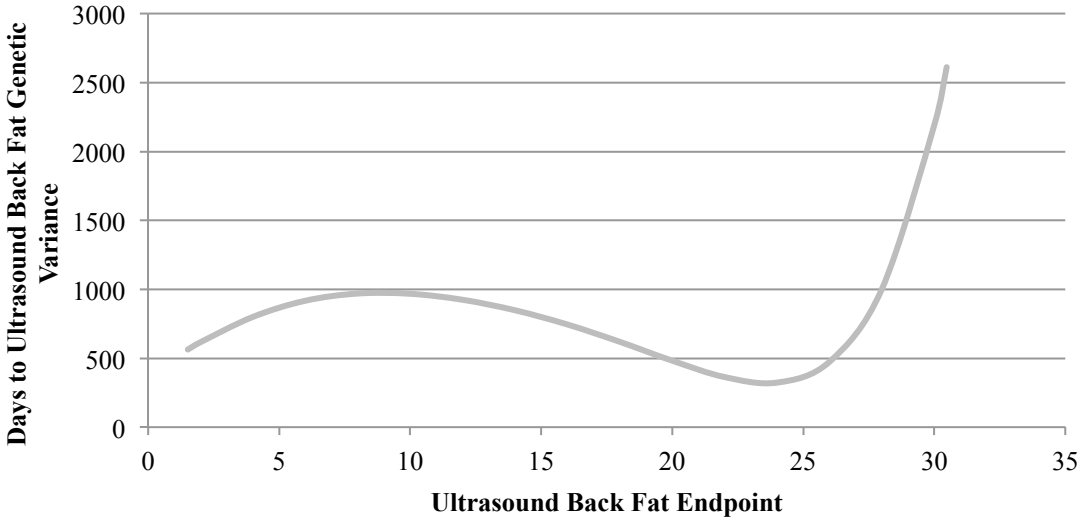


Figure 6.22. Plot of observed days to ultrasound back fat genetic variance obtained from a cubic random regression of age on ultrasound back fat.

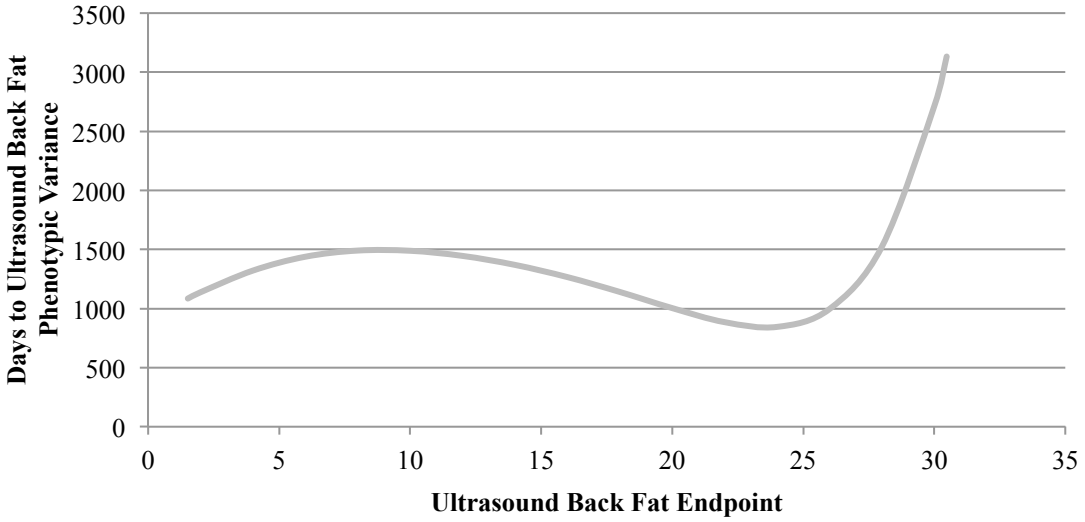


Figure 6.23. Plot of observed days to ultrasound back fat phenotypic variance obtained from a cubic random regression of age on ultrasound back fat.

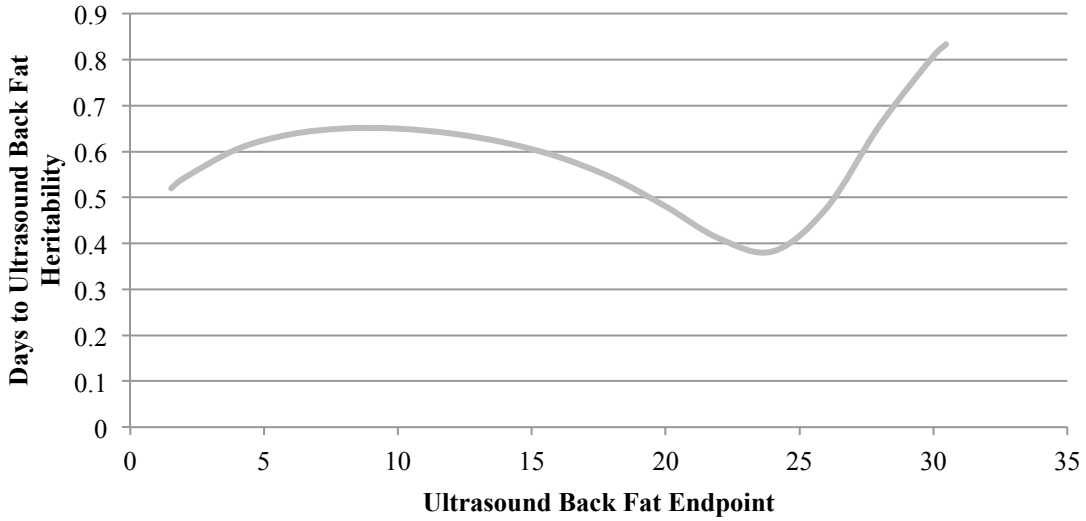


Figure 6.24. Plot of observed days to ultrasound back fat heritability obtained from a cubic random regression of age on ultrasound back fat.

Random Regression Model Selection. The above sections show differing fixed regression orders were required to properly account for the mean relationship between age and ultrasound traits. This order was dependent upon items such as the order of the random regression model and trait being analyzed.

Likelihood ratio tests were conducted to determine the statistically significant random regression order for both DTUREA and DTUBF with results from these tests presented below in Table 6.16 and Table 6.17. A requirement of LRT is that the two models being compared have equivalent fixed effect specifications with no differences in data. As such, both tables present the significant fixed regression order of the full or more complex model. All LRT comparisons were made between the more complex and simplified models using this fixed regression order.

Table 6.16. REML log likelihood (logL) estimates and associated significance values used to determine the random regression order for the days to ultrasound rib eye area evaluation for both the heterogeneous residual variance and linear residual random regression models.

	Quad v Linear		Cubic vs Quad	Quartic vs Cubic
	Heterogeneous Residual Variance	Residual Random Regression	Heterogeneous Residual Variance	Heterogeneous Residual Variance
Full Model logL[1]	-7053.09	-7043.30	-6931.31	-4183.24
Reduced Model logL[1]	-7087.86	-7041.00	-6941.21	-4194.74
DF[2]	3	3	4	5
LRT Test Statistic[2]	113.56	-4.52	19.80	23.00
P-value[2]	<0.0001	N/A[4]	0.0005	0.0003
Fixed Regression Order[3]	4	4	6	5

[1]Full and reduced models correspond to the more complex versus simpler models, respectively.
[2]Likelihood ratio test statistic and associated P-values obtained from a Chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the two models.
[3]Highest significant fixed regression order of the full model, resulting in LRT comparisons to be made with the same reduced fixed regression order.
[4]Log likelihood of the full model is lower than the reduced model resulting in a negative LRT test statistic.

Likelihood test results for DTUREA shown above in Table 6.16 indicate some differing results. First, with comparison between the quadratic and linear random regressions, the inclusion of the quadratic random term resulted in a significantly better fit over the model containing the linear term for the HRV model. This term was no longer significant once the error covariance was added to the model in the LRRR. In the LRRR model, the REML logL estimate obtained from ASReml is actually lower for the quadratic model than it is for the linear model suggesting that the addition of the quadratic random regression term actually makes the model poorer. Continuing to higher random regression orders containing HRV estimates, all of the higher order random

regression models were significant, suggesting that the higher order random regressions are picking up additional variation in days most likely resulting from the covariance between residuals that is not being accounted for. Another problem with these higher order random regressions is they tend to become uncontrollable at the extremes resulting in inflated variance estimates as suggested by Meyer (2005). This behavior of random regressions was also observed here, and was illustrated in Figure 6.19 through Figure 6.21. Additionally, as the order of the random regression model is increased above the cubic polynomial to the quartic, the data point requirement results in an entire year's worth of data to be removed from the evaluation. This caused the removal of 174 individual animals from the test year 2002 to 2003 because there were only four observations per animal.

Likelihood ratio test results for DTUBF are given below in Table 6.17. Here, similar results to those presented above for DTUREA were observed. Considering the quadratic versus linear random regressions, the HRV model shows the quadratic term accounted for significantly more variation in days than the linear random regression. When error covariance was added to the LRRR model, the linear random regression was sufficient in describing the variation in days. One difference in DTUBF observed here, in comparison to the DTUREA model is for DTUBF, the quadratic random regression was the highest significant order. The addition of the cubic term to the quadratic heterogeneous residual variance model did not significantly account for any additional variation in days.

Table 6.17. REML log likelihood (logL) estimates and associated significance values used to determine the random regression order for the days to ultrasound back fat evaluation for both the heterogeneous residual variance and linear residual random regression.

	Quadratic vs. Linear		Cubic vs Quadratic
	Heterogeneous Residual Variance	Residual Random Regression	Heterogeneous Residual Variance
Full Model logL[1]	-7787.05	-7874.15	-7707.00
Reduced Model logL[1]	-7798.07	-7877.29	-7706.77
DF[2]	3	3	4
LRT Test Statistic[2]	22.04	6.28	-0.46
P-value[2]	0.0002	0.18	N/A[4]
Fixed Regression Order[3]	5	5	5

[1]Full and reduced models correspond to the more complex versus simpler models, respectively.
[2]Likelihood ratio test statistic and associated P-values obtained from a Chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the two models.
[3]Highest significant fixed regression order of the full model, resulting in LRT comparisons to be made with the same reduced fixed regression order.
[4]Log likelihood of the full model is lower than the reduced model resulting in a negative LRT test statistic.

Given the significance of the higher random regression orders observed in the HRV models for both DTUREA and DTUBF, observed variance estimates from each of the significant random regression orders was plotted versus increasing target endpoint. Figure 6.26, Figure 6.26 and Figure 6.27 contain the genetic variance, phenotypic variance and heritability estimates from the linear, quadratic, cubic and quartic DTUREA random regressions. The observed differences resulting from increasing the order of the random regression appears to be in the tails of the data distribution. Where UREA observations are most dense, genetic and phenotypic variances estimates were very similar across all random regression orders. As random regression order was increased, variance estimates became inflated in the tails of the data distribution. The linear random

regression is the order least susceptible to these Legendre polynomial estimation problems therefore the linear random regression is probably sufficient in describing the data using the HRV sub-classes. Differences in heritability estimates appear to exist, but they are more due to the scaling of the charts than any thing else.

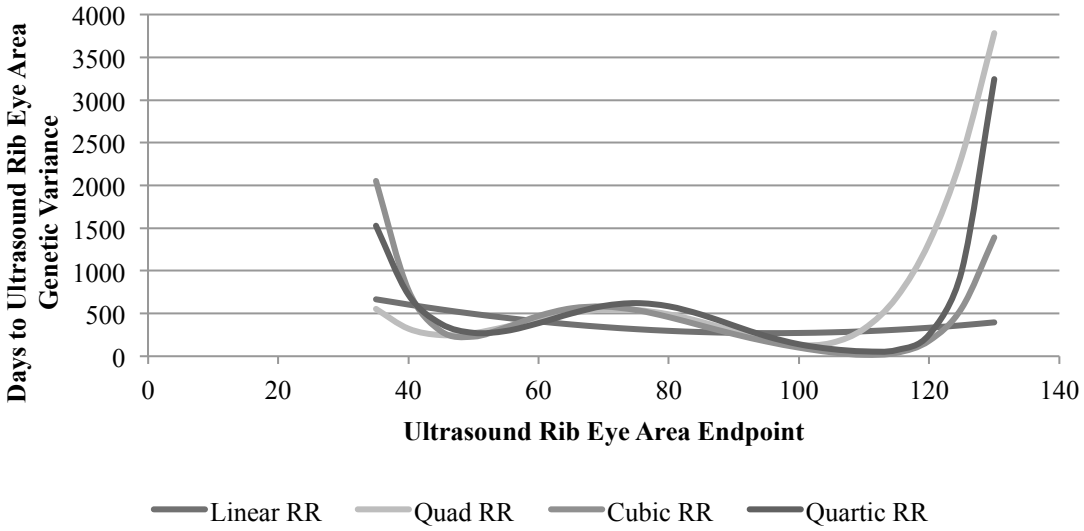


Figure 6.25. Plot of observed days to ultrasound rib eye area genetic variance obtained from the linear, quadratic, cubic and quartic random regressions of age on ultrasound rib eye area using heterogeneous residual variance sub-classes.

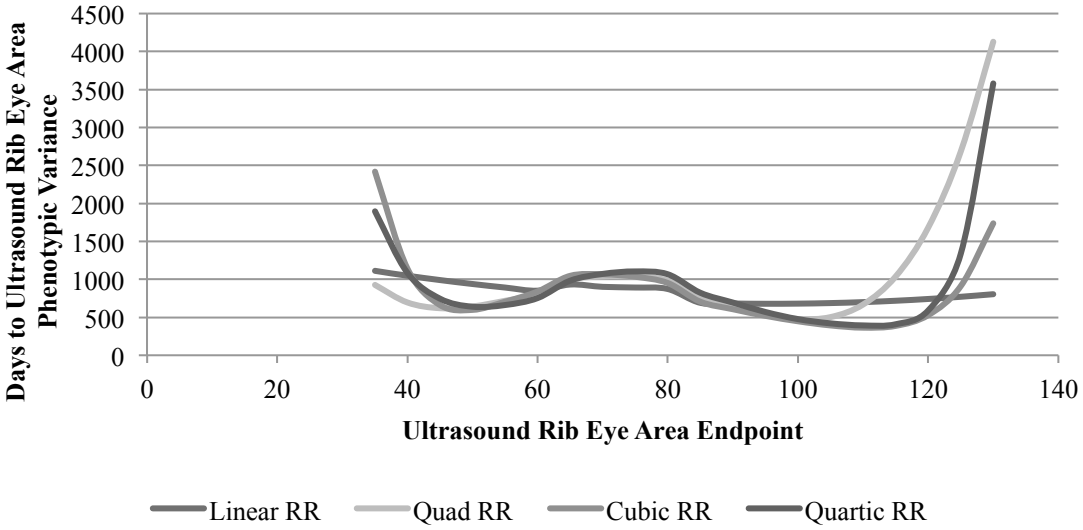


Figure 6.26. Plot of observed days to ultrasound rib eye area phenotypic variance obtained from the linear, quadratic, cubic and quartic random regressions of age on ultrasound rib eye area using heterogeneous residual variance sub-classes.

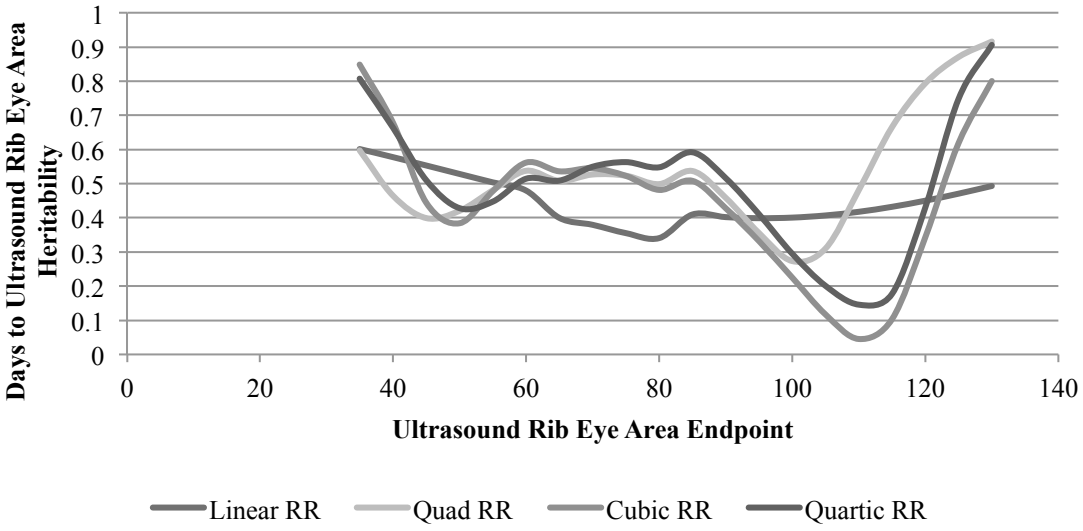


Figure 6.27. Plot of observed days to ultrasound rib eye area heritability obtained from the linear, quadratic, cubic and quartic random regressions of age on ultrasound rib eye area using heterogeneous residual variance sub-classes.

Figure 6.28, Figure 6.29 and Figure 6.30 contain the plots of observed genetic variance, residual variance and heritability obtained from the linear, quadratic and cubic random regressions for DTUBF. Here, trends similar to DTUREA are observed. Variance estimates for the linear random regression tend to be more conservative than those obtained from higher orders. The estimates obtained from the cubic and quadratic regressions are very similar to one another with the cubic random regression inflating the estimates of variance in the upper end of the range of UBF observations. An important observation to make is that the heritability estimated from the linear random regression does not trend toward zero in the range of 20 to 30 mm of UBF as the estimates from the quadratic and cubic do. Also, the heritability estimates from the linear random regression look to be more conservative, in the 0.40 to 0.50 range, than those from the higher order random regressions. They are also much less variable being more constant across the entire range of UBF observations.

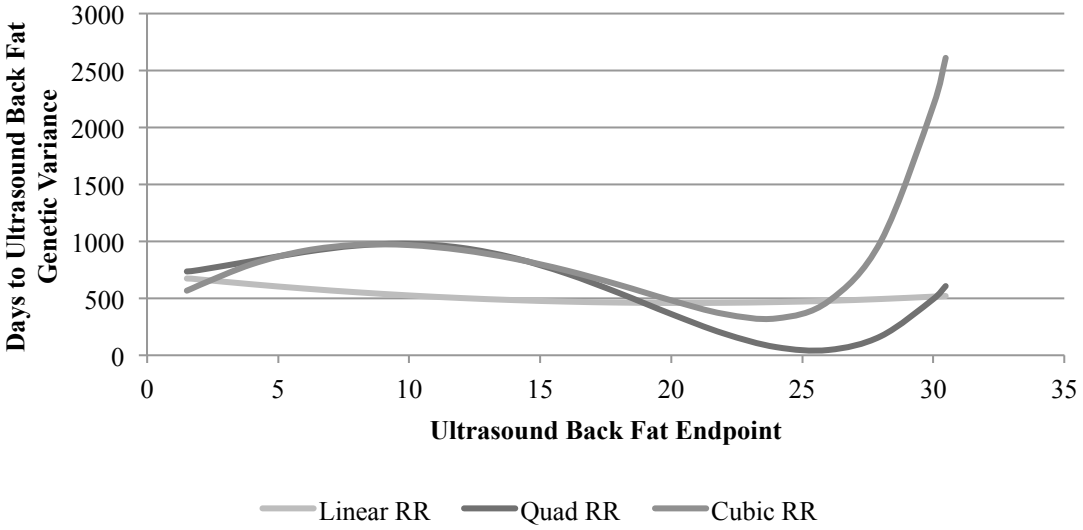


Figure 6.28. Plot of observed days to ultrasound back fat genetic variance obtained from the linear, quadratic and cubic random regressions of age on ultrasound back fat using heterogeneous residual variance sub-classes.

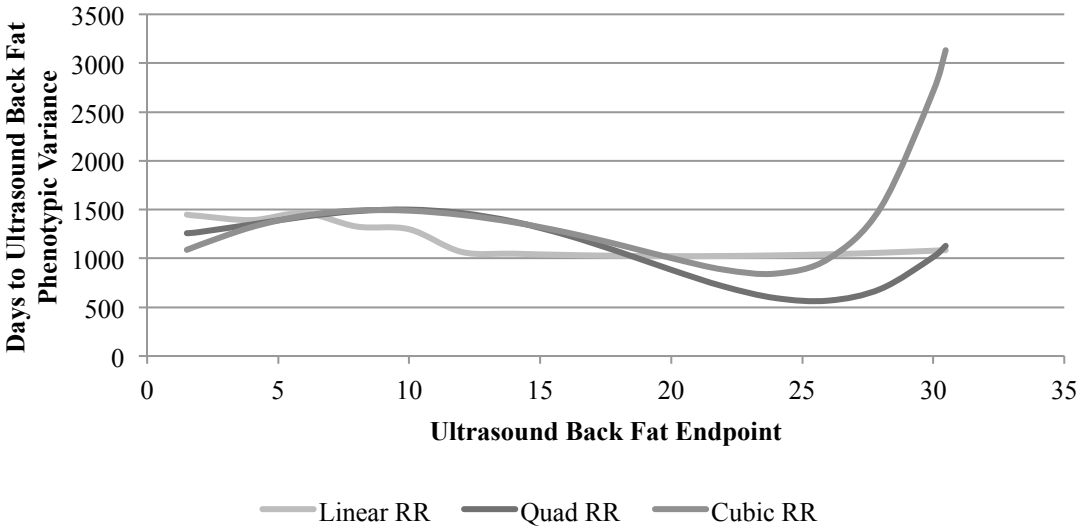


Figure 6.29. Plot of observed days to ultrasound back fat phenotypic variance obtained from the linear, quadratic and cubic random regressions of age on ultrasound back fat using heterogeneous residual variance sub-classes.

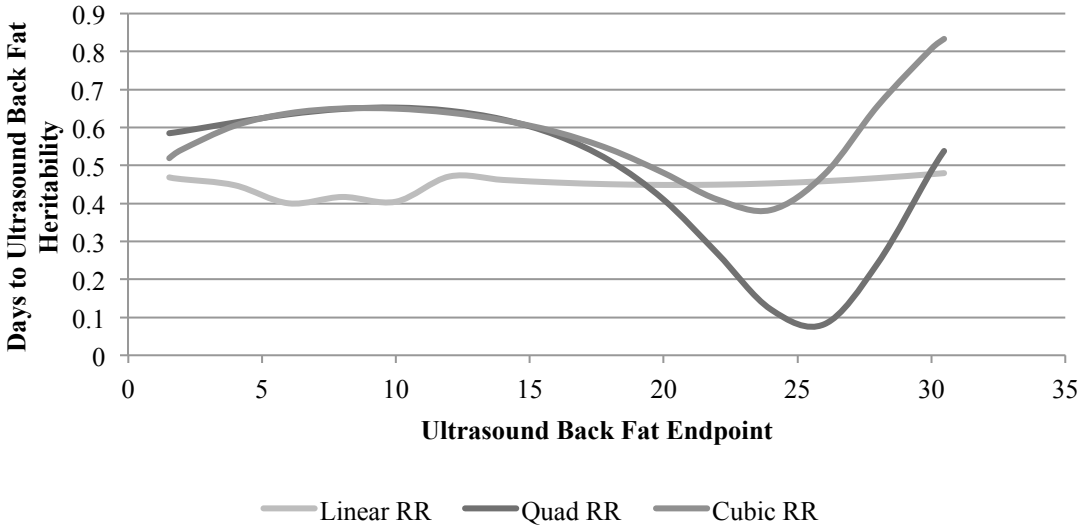


Figure 6.30. Plot of observed days to ultrasound back fat heritability obtained from the linear, quadratic and cubic random regressions of age on ultrasound back fat using heterogeneous residual variance sub-classes.

Given the behavior of the HRV linear random regression variance estimates when compared to the significant higher order random regressions, the simplicity of the linear regression is attractive for the implementation of these traits in a national cattle evaluation scheme. As mentioned earlier, the highest significant random regression order from the models containing the LRRR model was the linear polynomial. Observed variances obtained from the linear random regression from both residual variance sub-models were plotted versus one another for DTUREA in Figure 6.3, Figure 6.4, Figure 6.5, and Figure 6.6 and for DTUBF in Figures 6.7, Figure 6.8, Figure 6.9 and Figure 6.10. For both traits, DTUREA and DTUBF, the linear random regression model produced similar heritability estimates for both residual variance sub-models.

Next, sire EBV were calculated for both DTUREA and DTUBF. Sire EBV for DTUREA were calculated from both residual variance models for each 5 cm^2 increment in the range of UREA observations (36.77 cm^2 to 129.54 cm^2). EBV corresponding to

UREA endpoints of 36.77, 71.77 and 129.54 cm^2 representing minimum, median and maximum UREA endpoints, respectively, were correlated to one another with the results presented in Table 6.18. Likewise, sire EBV for DTUBF were calculated from both residual variance models for each 2 mm increment in the range of UBF observations (1.53 mm to 30.47 mm). EBV corresponding to UBF endpoints of 1.53, 7.53 and 30.47 mm representing minimum, median and maximum UBF endpoints, respectively, were correlated to one another with results presented in Table 6.18 below. For DTUREA, correlations remained high (> 0.992) across the entire range of UREA observations. Regression coefficients representing the regression of HRV EBV on LRRR EBV are nearly one, indicating that for each unit increase in residual random regression EBV, a corresponding unit increase in heterogeneous residual variance EBV is observed.

Table 6.18. Correlation coefficients along with the EBV regression coefficient from the regression of EBV obtained from the heterogeneous residual variance model on those EBV obtained from the residual random regression model for each of three weight endpoints representing minimum, median and maximum observations.

	Days to Ultrasound Rib Eye Area		
Endpoint	36.77	71.77	129.54
Pearson Correlation	0.997	0.998	0.992
Spearman Rank	0.994	0.995	0.988
Regression	0.951	1.03	1.001
	Days to Ultrasound Back Fat		
Endpoint	1.53	7.97	30.47
Pearson Correlation	0.995	0.997	0.935
Spearman Rank	0.99	0.993	0.906
Regression	1.143	1.133	0.885

For the trait DTUBF, the results are more puzzling. At UBF endpoints of 1.53 mm and 7.97 mm, correlations are high, and nearly unity (> 0.993). Also, the regression coefficients are slightly higher than one meaning that for every increase in LRRR EBV, the corresponding change in HRV EBV is around 1.13 to 1.14. As the upper end of the range of UBF EBV is approached, the correlations fall off and the regression coefficient is much lower (0.885). Perhaps this is a result of the severe lack of data density in this upper region (see Figure 6.2), or it may be a result of the nature of the UBF data where only Angus animals are represented in the upper end of the UBF range of observations. The data density of UBF tends to be skewed to the left, meaning the majority of the UBF observations are located at the lower end of the range of observations. The differences seen here are the result of how the different residual variance sub-models handle the lack of data density in the upper range of UBF.

Days to ultrasound rib eye area sire EBV were calculated from both the HRV model and the LRRR model for every 5 cm^2 UREA increment across the entire range of observations. Figure 6.31 contains the EBV plots for each of the five most used sires in the pedigree while Figure 6.32 contains the plots for each of the five least used sires in the pedigree. The most heavily used sires averaged 38.4 progeny per individual. Looking at the plots of their corresponding EBV, the type of residual variance model has virtually no effect on the prediction of their DTUREA EBV. Looking at the five least used sires' EBV in Figure 6.32, small differences between the residual variance sub-models are observed here as well. Here, the range of EBV is not as large as observed for the five most used sires, mostly due to the fact that sires with very few progeny have EBV that are more conservatively estimated than from sires with larger amounts of data.

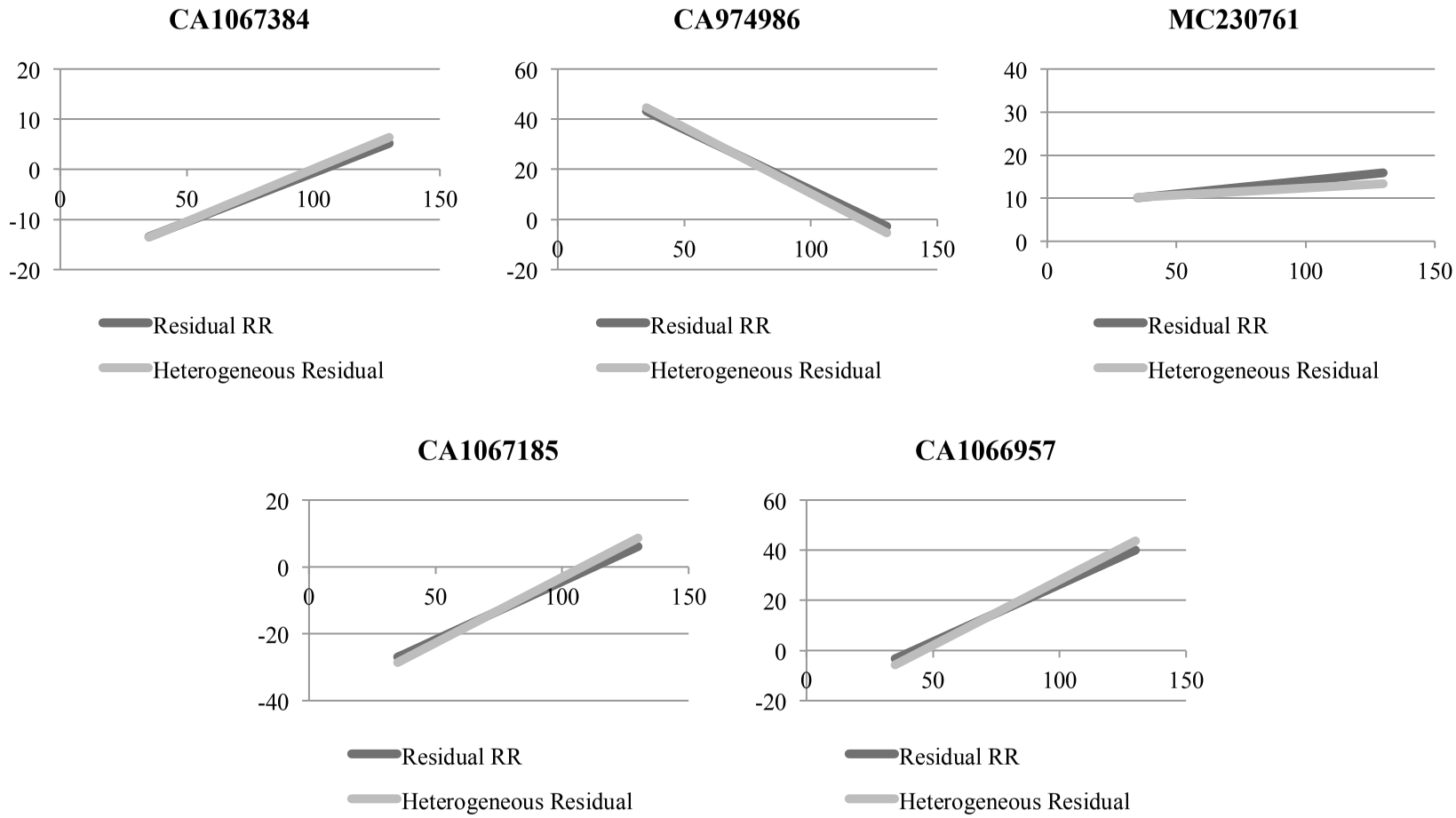


Figure 6.31. Days to ultrasound rib eye area estimated breeding values obtained from the linear random regression model using both linear residual random regression and heterogeneous residual variance for the five most used sires represented in the ultrasound data set. For the figures above, the vertical axis represents the estimated breeding values for days to ultrasound rib eye area while the horizontal axis is the ultrasound rib eye area endpoint.

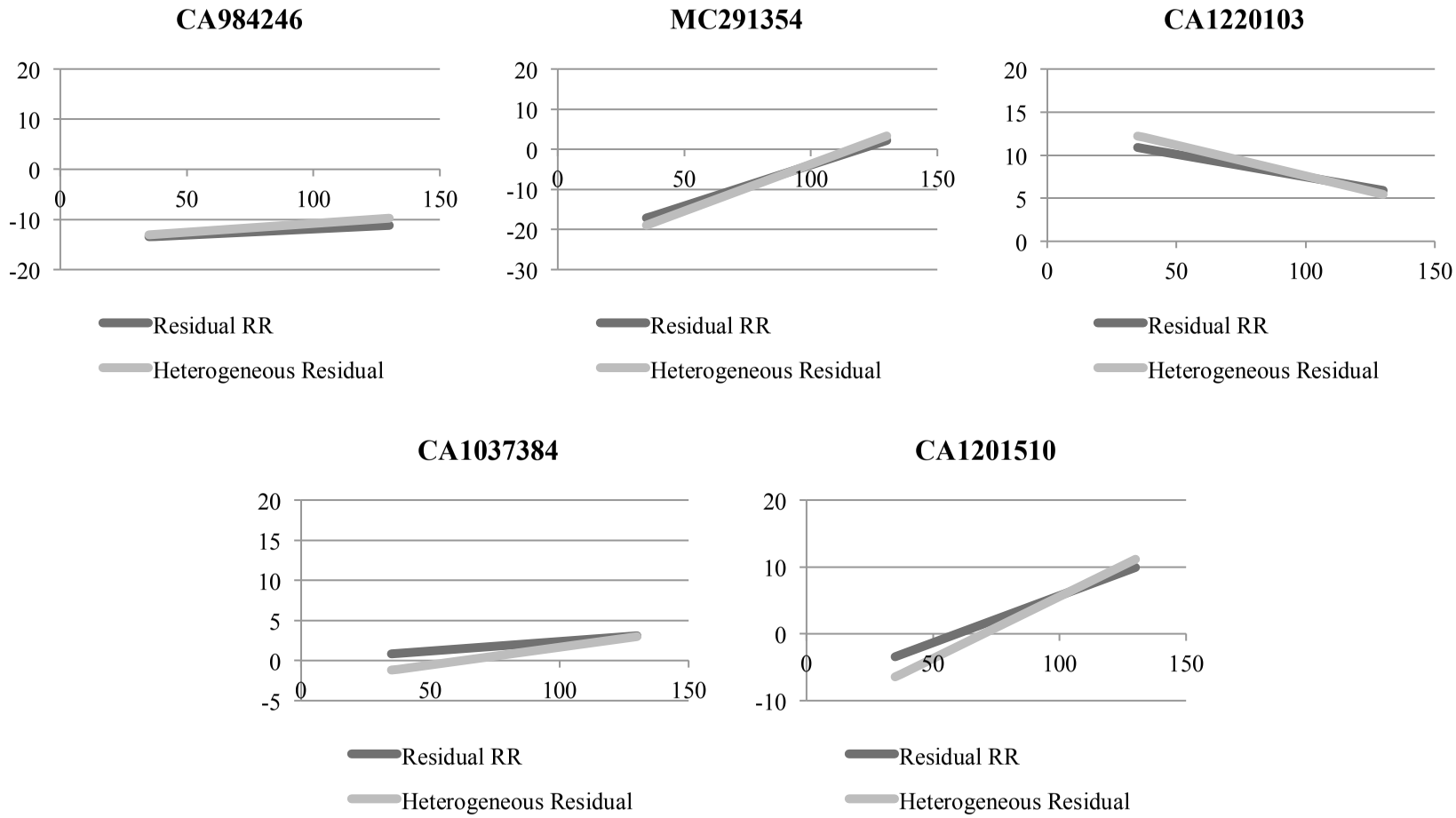


Figure 6.32. Days to ultrasound rib eye area estimated breeding values obtained from the linear random regression model using both linear residual random regression and heterogeneous residual variance for the five least used sires represented in the ultrasound data set. For the figures above, the vertical axis represents the estimated breeding values for days to ultrasound rib eye area while the horizontal axis is the ultrasound rib eye area endpoint.

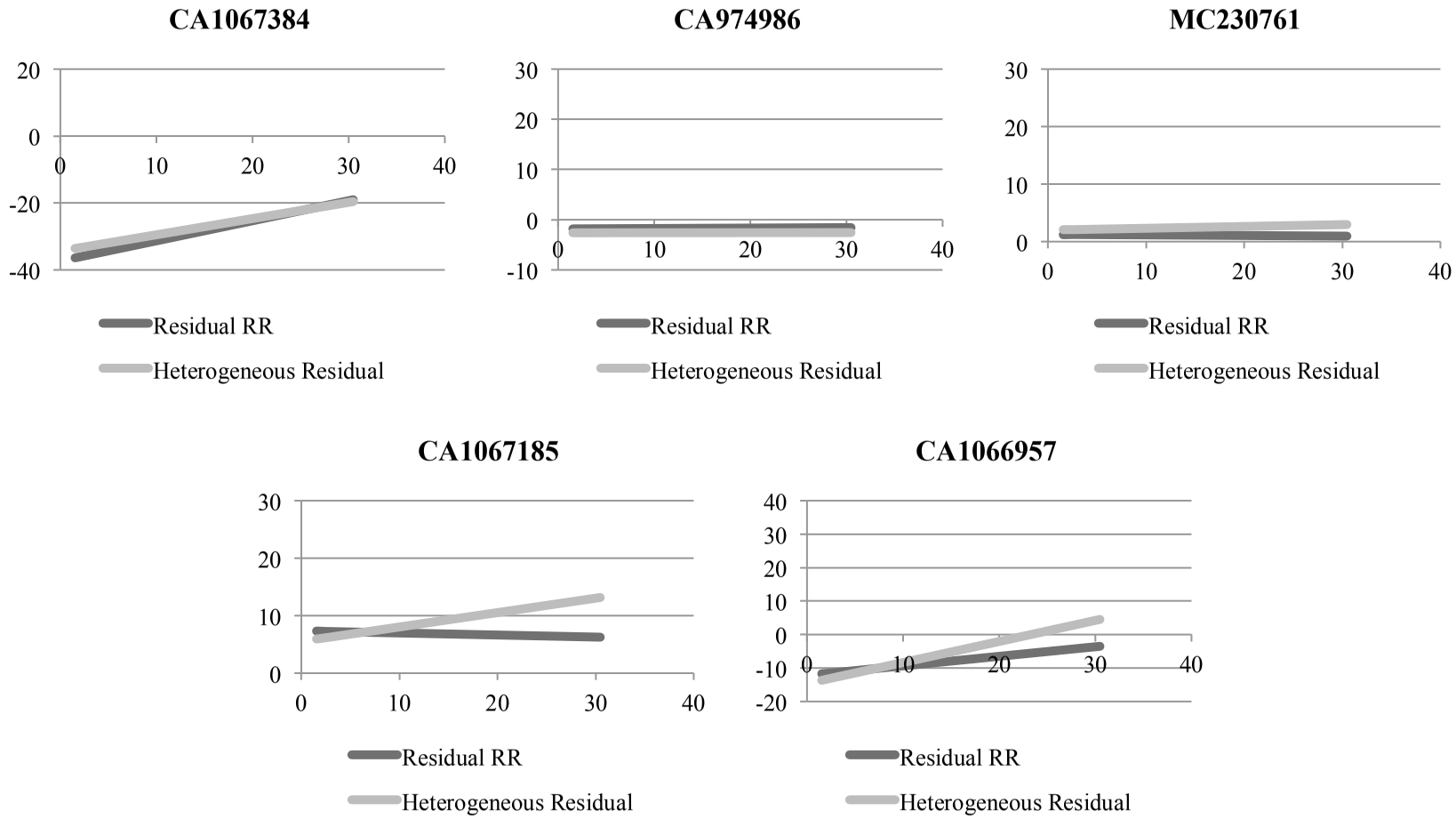


Figure 6.33. Days to ultrasound back fat estimated breeding values obtained from the linear random regression model using both linear residual random regression and heterogeneous residual variance for the five most used sires represented in the ultrasound data set. For the figures above, the vertical axis represents the estimated breeding values for days to ultrasound back fat while the horizontal axis is the ultrasound back fat endpoint.

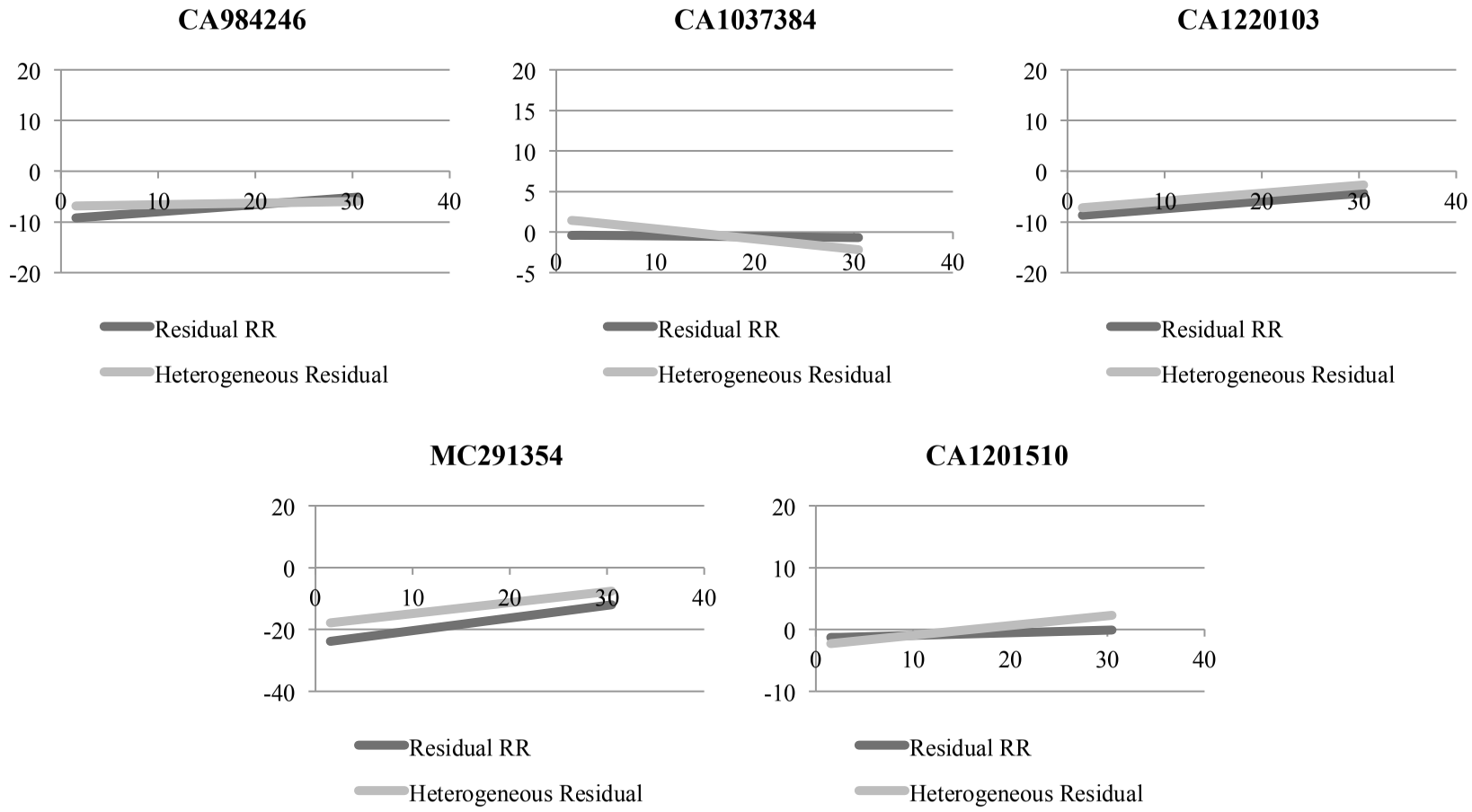


Figure 6.34. Days to ultrasound back fat estimated breeding values obtained from the linear random regression model using both linear residual random regression and heterogeneous residual variance for the five least used sires represented in the ultrasound data set. For the figures above, the vertical axis represents the estimated breeding values for days to ultrasound back fat while the horizontal axis is the ultrasound back fat endpoint.

Sire EBV for DTUBF were calculated for every 2 mm increment within the range of UBF data. For the five most heavily used sires (Figure 6.33), the LRRR model did just as good as the HRV model in predicting sire EBV for three of the sires. For two of the sires (CA1067185 and CA1066957), the EBV from both models tend to spread apart as UBF increases. In each instance, EBV from the residual random regression models predicted fewer days to finish than the heterogeneous residual variance model. Perhaps this is a function of the disproportionate number of observations in each of the four residual variance sub-classes from these sires. CA1067185 has 171 UBF observations in this data set, with only four in the leaner UBF residual variance category and 139 in the latter two categories. Sire CA1066957 has 153 UBF observations in this evaluation, and all but 8 fall in the latter half of the four residual variance categories. This sire in particular has 107 observations in the fattest category. The residual random regression model is appearing to handle breeding value predictions more appropriately in this instance, by accounting for the covariance between residuals than the heterogeneous residual variance model does. For the five least used sires in the data set, both residual variance models estimated sire EBV similarly.

Comparison to Repeated Measures Model. In an effort to compare the results for the genetic evaluation of both DTUREA and DTUBF using random regression methodologies to more traditional models, repeated measures analyses were performed using the same ultrasound data set that was used for the linear random regression model.

Age observations for DTUREA and DTUBF were adjusted to a constant 66 cm^2 and 7 mm, respectively. These endpoints were chosen because they represented

approximately the middle of the UREA and UBF distributions (see Figure 6.1 and Figure 6.2). These points represent the points in the distribution of observations where data density was highest.

Beginning with DTUREA, genetic variance and heritability obtained from the linear random regression model using HRV were 372 $days^2$ and 0.40, respectively. Similarly, genetic variance and heritability from the linear random regression model using LRRR were 392 $days^2$ and 0.44, respectively. The same estimates obtained from the repeated measures model were 369 $days^2$ and 0.36. Genetic variance estimates across all three models were very similar. Heritability obtained from the repeated measures model was slightly lower than the same obtained from the random regression models. This reduced heritability estimate can be attributed to the fact that both residual and permanent environmental variance obtained from the repeated measures models were slightly higher than those from the random regression models, resulting in slightly larger estimates of phenotypic variance.

For DTUBF, genetic variance and heritability obtained from the HRV model were 570 $days^2$ and 0.41. From the LRRR model they were 706 $days^2$ and 0.51. Corresponding estimates obtained from the repeated measures model were 2543 $days^2$ and 0.36, a much larger difference. There appears to be a rather large scaling issue for the repeated measures evaluation. This may be a function of the nature of the UBF data. Data used in the DTUBF evaluation have a rather small scale and range of the predictor variables (UBF observations) and a much larger scale and range of the response variable (age). This seems to be inflating the estimate of genetic variance obtained from the DTUBF evaluation. Heritability estimates from the repeated measures model seem to be

in range, and appear to have similar relationships as the repeated measures estimates of heritability did to those obtained for the traits DTUREA and days to weight (Chapter V).

Summary

This chapter presented the results from the development of both a DTUREA and a DTUBF genetic evaluation using random regression methodology with Legendre polynomials as the base random polynomial function. The required order of the random polynomials was tested using likelihood ratio tests to determine whether or not the addition of higher order polynomials accounted for significantly more variation in the trait days than did the more simple model. These likelihood ratio tests were performed within each residual variance sub-model (LRRR versus HRV).

In the models that used the LRRR, the highest random polynomial order statistically significant in terms of the likelihood ratio test was the linear random polynomial. This was the case for both DTUREA and DTUBF. In the models that used the HRV, higher order random polynomials were required to account for the variation in the trait days. For the trait DTUREA, the highest significant random order was the quartic random polynomial. For the trait DTUBF, the quadratic random polynomial was the highest significant order.

As was observed in the days to weight genetic prediction presented in the previous chapter, as the order of the random polynomial was increased above a linear polynomial, the estimates of genetic and phenotypic variance as well as heritability appear to have become artificially inflated, particularly for the tails of both the UREA and UBF distributions (See Figures 6.15 through 6.24).

In this data set, the models containing the LRRR are most attractive for many reasons. First, they allow for simplicity, as the linear polynomial was the highest order required for accounting for the genetic variation in days. Linear random regressions are not only simpler to understand they also have the smallest data requirement (two data points per animal). In an industry where data density continues to be a problem, the fewer data points needed the better. Second, given the apparent data stratification between the Charolais and Angus animals in this data set, particularly for UBF, the LRRR models handle the residual covariance more appropriately. It is for this reason I am recommending the linear random regression model using linear residual random regression for implementation in a national evaluation.

Literature Cited

Gilmour, A. 2009. Random regression – maternal effects. Available at: http://www.vsni.co.uk/forum/viewtopic.php?t=205&sid=dd107da8ac86898f99dae5f0c7580f8d Accessed Feb. 7, 2010.

Golden, B. L., D. J. Garrick, S. Newman, and R. M. Enns. 2000. Economically relevant traits, a framework for the next generation of EPDs. Proc. Beef Improvement Federation, Wichita, KS. pp. 2-13.

Harris, D. L., and S. Newman. 1994. Breeding for profit: Synergism between genetic improvement and livestock production (A Review). J. Anim. Sci. 72:2178-2200.

Johnston, D. J., L. L. L. Benyshek, J. K. Bertrand, M. H. Johnson, and G. M. Weiss. 1992. Estimates of genetic parameters for growth and carcass traits in Charolais cattle. Can. J. Anim. Sci. 72:492-499.

Meyer, K. 2005. Advances in methodology for random regression analyses. Aust. J. Exp. Agric. 45:847-858.

McWhir, J. and J. W. Wilton. 1987. Heritabilities of predicted performance measures in beef bull tests and their relationships with feedlot efficiency measures. J. Anim. Sci. 64:1323-1331.

Schaeffer, L. R. 2003. Random regression models. ANSC637 Course Notes – Quantitative genetics and animal models. Available at: http://www.aps.uoguelph.ca/%7Elrs/ABModels/NOTES/RRM14a.pdf. Accessed Jan. 8, 2009.

STAGES. 2006. Swine Testing and Genetic Evaluation System. Available at: http://www.ansc.purdue.edu/stages. Accessed Jan. 15, 2009.

Stewart, T. S., D. L. Lofgren, D. L. Harris, M. E. Einstein, and A. P. Schinckel. 1991. Genetic improvement programs in livestock: swine testing and genetic evaluation system (STAGES). J. Anim. Sci. 69:3882-3890.

CHAPTER VII

CONCLUSIONS AND IMPLICATIONS

Results from the development of the three "days to" genetic predictions (days to weight, days to ultrasound back fat and days to ultrasound rib eye area) show the existence of genetic and phenotypic variability, and that genetic progress in reducing the number of days to reach a finish endpoint can me made selection.

First, random regression models were shown to be equivalent to their multivariate counterparts in situations where the order of the random polynomial was equal to the number of traits in the multivariate model. In situations where the number of observations prohibits the use of multiple trait models, random regressions can be used to reduce the overall size of the equation set. Also, the predictions resulting from the use of these models allow the calculation of EBV for any given point along the regression line, where as with multiple trait models, resulting EBV are limited to trait specific endpoints.

The trait days to weight was estimated using both a random regression model and a more traditional repeated measures model. Two different methods were used to account for the residual variation in the random regression model, the first of which divided the weight observations into four groups based on their quartiles. The second method used a linear regression on the residuals, to allow for a changing covariance structure as observations became farther apart from one another.

Results from this model building exercise suggested that a linear random polynomial was sufficient for describing the genetic variation in days. When the different methods for modeling the residual variance were considered, the model containing the linear regression on residuals provided more realistic heritability estimates than the model dividing the observations into four distinct residual variance sub-groups. This is mostly due to the ability of the linear residual random regression model to appropriately handle the changing residual covariance structure, which in turn resulted in larger phenotypic variance estimates. Genetic variance and heritability estimates from both residual variance sub-models were compared to those obtained from a repeated measures model. The results from those comparisons yield similar estimates of heritability and genetic variance.

Following the days to weight genetic evaluation, evaluations for days to ultrasound back fat and days to ultrasound rib eye area were also conducted. Considering the random regression models for both traits, the model containing the linear residual random regression accounted for enough of the variation in days to allow a linear random polynomial to be sufficient. When the observations were broken into their residual variance sub-groups, higher order random polynomials were needed to fully describe the genetic variation in days. For days to ultrasound back fat, a quadratic polynomial was needed to fully describe the genetic variation in days, while for the trait days to ultrasound rib eye area a quartic polynomial was needed.

A detractor to the use of random regression models, especially those using Legendre polynomials are the inflated estimates of variance obtained with higher order random polynomials. This trend was observed with these higher order models,

particularly in the tails of the data distribution where the variance and heritability estimates appeared to be artificially inflated. The linear random polynomial did not result in these inflated estimates.

For all three traits, a linear random regression was deemed sufficient in describing the genetic variation in days. The linear random regression allows for simplicity in the calculation of EBV with respect to a particular endpoint as well as reducing the order of the equation sets needing to be solved. Given these simplifications of the linear random polynomial over some of the higher order polynomials, considerable re-ranking among the ten most used sires and the ten least used sires in the data set was observed for all three days to traits.

These predictions for the three days to finish traits are only a beginning. Further research is needed, particularly with larger single breed data sets to determine if these linear random regression models will still be sufficient. The data set used in the evaluation consisted of Angus, Charolais, and Charolais cross cattle. Angus and Charolais are are two different types of cattle. Stratification in the data was observed between these two breeds and was largely evident in the ultrasound back fat data where the Angus animals had greater fat deposition than the Charolais animals. How will the days to ultrasound back fat evaluation change if the same biological type of animals were evaluated? Further analysis also needs to be performed using a data set containing intra-muscular fat observations. Evaluations using this data could result in a days to marbling score or days to quality grade genetic prediction.

APPENDIX I

INCIDENCE MATRICES USED IN THE MULTIVARIATE VERSUS RANDOM REGRESSION EQUIVALENCY EXAMPLE

Below is the incidence matrix that relates the traits for weight observations measured on test day two to the corresponding animals in the pedigree on which the data were observed.

$\mathbf{Z}_2 =$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1	731
0	0	0	0	0	0	0	0	1	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	1	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	1	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	1	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	1	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	1	14	27	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	62	972
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	90	1058
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	119	1118

Below is the incidence matrix that relates the traits for weight observations measured on test day three to the corresponding animals in the pedigree on which the data were observed.

$\mathbf{Z}_3 =$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	27	861
0	0	0	0	0	0	0	0	1	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	1	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	1	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	1	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	1	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	1	14	62	972
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	90	1058
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	119	1118

Below is the incidence matrix that relates the traits for weight observations measured on test day four to the corresponding animals in the pedigree on which the data were observed.

$\mathbf{Z}_4 =$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	27	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	62	972
0	0	0	0	0	0	0	0	1	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	1	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	1	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	1	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	1	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	1	14	90	1058
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	119	1118

Below is the incidence matrix that relates the traits for weight observations measured on test day five to the corresponding animals in the pedigree on which the data were observed.

$\mathbf{Z}_5 =$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	*ID*	*TD*	*Weight*
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	627
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	1	712
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	1	632
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	1	604
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	1	630
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	27	732
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	27	855
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	27	728
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	27	731
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	27	758
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	27	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	62	828
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	62	952
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	62	861
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	62	869
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	62	972
0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	90	927
0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	90	1039
0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	90	924
0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	90	940
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	90	957
0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	90	1058
0	0	0	0	0	0	0	0	1	0	0	0	0	0	9	119	969
0	0	0	0	0	0	0	0	0	1	0	0	0	0	10	119	1111
0	0	0	0	0	0	0	0	0	0	1	0	0	0	11	119	1007
0	0	0	0	0	0	0	0	0	0	0	1	0	0	12	119	1051
0	0	0	0	0	0	0	0	0	0	0	0	1	0	13	119	1042
0	0	0	0	0	0	0	0	0	0	0	0	0	1	14	119	1118

APPENDIX II

ASREML COMMAND FILES FOR RUNNING LINEAR AND QUADRATIC RANDOM REGRESSION MODELS FOR DAYS TO WEIGHT

Below is the ASReml Command file for running a linear random regression model for the prediction of days to weight.

```
Lethbridge Weight Data
 anim                !P      #1
 sire                !A      #2
 dam                 !A      #3
 year        *       !A      #4
 pen         *       !A      #5
 breed       *       !A      #6
 bdate               !A      #7
 sbdate                      #8
 strt_date                   #9
 weigh_date                  #10
 td                          #11
 on_feed_date                #12
 on_feed_td                  #13
 weight                      #14
 aod                         #15
 age                         #16
 cg          *       !A      #17
 assess                      #18

ped.stk !make !alpha
fin2 !maxit 1000 !mvremove !DDF !dopart $1

!part 1     #Model for a linear random regression with four separate residual variances

age ~ cg leg(weight,1) !r leg(weight,1).anim leg(weight,0).ide(anim)

4 1 2 !STEP 0.01
1822 0 0 !S2=29.1021        #Starting value for the first quartile (1822 records)
1662 0 0 !S2=43.5097        #Starting value for the second quartile (1662 records)
1717 0 0 !S2=55.6037        #Starting value for the third quartile (1717 records)
1757 0 0 !S2=43.1105        #Starting value for the fourth quartile (1757 records)

leg(weight,1).anim 2
2 0 US !GP
 1407.87
 618.216        417.918
anim 0 AINV

leg(weight,0).ide(anim) 2
1 0 US !GP
 243.695
```

```
ide(anim)
!end


!part 2      #Model for a linear random regression accounting for residual covariance.

age ~ cg leg(weight,1) !r leg(weight,1).ide(anim) leg(weight,1).anim

1 1 2 !STEP 0.01
7632              #Number of records in the final data file.

leg(weight,1).ide(anim) 2
2 0 US !GP
  3.1212
 -0.1778  0.1326
ide(anim)

leg(weight,1).anim 2
2 0 US !GP
 23.7477
  7.5948  7.5215
anim 0 AINV
!end
```

Below is the ASReml Command file for running a quadratic random regression model for the prediction of days to weight.

```
Lethbridge Weight Data
 anim                !P      #1
 sire                !A      #2
 dam                 !A      #3
 year          *     !A      #4
 pen           *     !A      #5
 breed         *     !A      #6
 bdate               !A      #7
 sbdate                      #8
 strt_date                   #9
 weigh_date                  #10
 td                          #11
 on_feed_date                #12
 on_feed_td                  #13
 weight                      #14
 aod                         #15
 age                         #16
 cg            *     !A      #17
 assess                      #18

ped.stk !make !alpha
fin2 !maxit 1000 !mvremove !DDF !dopart $1

!part 1
age ~ cg leg(weight,1) !r leg(weight,2).anim leg(weight,0).ide(anim)

4 1 2 !STEP 0.01
1822 0 0 !S2=28.6053        #Starting value for the first quartile (1822 records)
1662 0 0 !S2=43.2429        #Starting value for the second quartile (1662 records)
1717 0 0 !S2=55.4351        #Starting value for the third quartile (1717 records)
1757 0 0 !S2=42.6780        #Starting value for the fourth quartile (1757 records)

leg(weight,2).anim 2
3 0 US !GP
 1294.03
 540.983        387.251
-34.0863        6.99767        27.6220
anim 0 AINV

leg(weight,0).ide(anim) 2
1 0 US !GP
 223.659
```

```
ide(anim)
!end

!part 2      #Model for a quadratic random regression accounting for residual covariance.

age ~ cg leg(weight,1) !r leg(weight,1).ide(anim) leg(weight,2).anim

1 1 2 !STEP 0.01
7632              #Number of records in the final data file.

leg(weight,1).ide(anim) 2
2 0 US !GP
  3.1212
 -0.1778  0.1326
ide(anim)

leg(weight,2).anim 2
3 0 US !GP
 29.5797
  6.3392  6.7089
  3.7282  0.8356  4.8659
anim 0 AINV
!end
```