PREPARTUM SELECTION OF HOLSTEIN HEIFERS TO MAXIMIZE PROFIT OF MICHIGAN DAIRY PRODUCERS

By

Brian R. Radke

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1998

ABSTRACT

PREPARTUM SELECTION OF HOLSTEIN HEIFERS TO MAXIMIZE PROFIT OF MICHIGAN DAIRY PRODUCERS

By

Brian R. Radke

Based on the culling optimization and simulation literature, it can be concluded that the Michigan dairy industry has an excessive cull rate. The corollary is that producers have an excess of heifers, dictating the need for a method of prepartum heifer selection. Selecting heifers on the basis of lifetime profit corrected for the opportunity cost of postponed replacement will maximize producer profit. While this profit measure can not be calculated prior to death, it has moderate correlation with first lactation milk production. Estimated breeding values (EBV) which are functions of the heifers' parents' predicted transmitting abilities (PTA) are the only widely available data on which to base heifer selection decisions. While theoretically and intuitively parents' PTA should predict subsequent daughter milk production, and therefore, lifetime profit corrected for the opportunity cost of postponed replacement, empirical support for this relationship was lacking. Based on the data of 5,123 Michigan Holstein heifers, parents' PTA of milk, fat, and protein available just prior to the heifers' calvings account for under 7% of the within herd variation in milk production. While diffuse, the distribution of within herd rank correlations between predicted mature equivalent milk production, based on genetic evaluations, and the actual mature equivalent milk production, was skewed to the left suggesting the evaluations may be useful in prepartum heifer selection. The omission ofheifers which were culled prior to first lactation, did not appear to cause selection bias in

the sample data. It also appeared herd could be modeled as a random effect. Despite evidence that first lactation milk production did influence producers' culling decisions, parents' PTA of milk, fat and protein estimated after the birth of 5,619 Michigan Holstein heifers were unable to predict heifer culling during rearing or first lactation. This also suggested Michigan producers were not basing their culling decisions on genetics. There appeared to be little difference in the voluntary nature of voluntary versus involuntary cullings casting doubt on the existence or utility of such a culling classification scheme. Modified Box-Complex, a search algorithm, was used to determine optimal heifer selection rules under three scenarios. Separate data sets of 58 and 57 Michigan herds were used to derive and test the rules, respectively. The objective of the rule was to maximize average profit per heifer, summed over all herds in the data set, subject to a given proportion of heifers being selected. Profit was a function of the returns to heifer sales which varied across the four raising periods the rule operated over and lifetime profit corrected for opportunity cost of each heifer estimated from first lactation milk production. Genetic data consisted of PTA milk, fat, protein, and associated reliabilities of each parent in each of the four periods. Considerable sampling error was encountered in terms of rule performance between the samples used to derive and test the selection rules. Regardless of scenario, selection based on EBV of milk in the most profitable period to sell heifers performed comparably to the rule developed by Complex. Use of either of these rules resulted in a $20 per heifer improvement in profit over random selection. While heifer selection based on EBV milk is more profitable than random selection, only a portion of the profit available in heifer selection is being captured.

This research is dedicated to my family who, although I'm not sure they understood, supported my leaving practice to return to academia.

Specifically, this is for Natalie, who understands all too well why I left practice. Who, during the last four and a half years, endured my worst – studying nights and weekends, and my best- windsurfing, while she toiled away. Your turn?

ACKNOWLEDGEMENTS

I am indebted to my major professor James W. Lloyd, who treated me as a respected colleague. Our relationship was the envy of many of the other Agricultural Economics graduate students. Jim supported me throughout my studies and research. He further contributed to my development by opening doors for my participation in a number of presentations and workshops to varied audiences, as well as providing consulting opportunities. And when the time was right, Jim gave the nudge out of the nest to encourage me to stretch my wings. On a personal note, Natalie and I are grateful to the entire Lloyd family for their support and friendship.

I also thank my guidance committee: J. Roy Black, Stephen Harsh, Jeff M. Wooldridge, John B. Kaneene, and Robert J. Tempelman for providing guidance in their respective areas of expertise.

Finally, the financial support of the Department of Agricultural Economics through the Animal Industry Initiative funds provided by the Agricultural Experiment Station at Michigan State University while the majority of this work was conducted is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

STATEMENT OF THE RESEARCH ISSUE AND INTRODUCTION

Statement of the Research Issue

There is a lack of thought, much less knowledge, in the area of prepartum Holstein heifer selection. Numerous optimization and simulation models suggest the culling rate of cows in Michigan dairy herds is excessive, resulting in a 2% to 40% decrease in producer profit. These studies indicate a method of heifer selection is required.

Given the need to select only a portion of the available heifers, it would be desirable to select those animals which will subsequently be most profitable. This contrasts the naïve method of random heifer selection employed in culling optimization studies. Genetic information, including milk, fat, and protein estimated breeding values (EBV) and their associated reliabilities, are the most viable data on which to base heifer selection decisions. An EBV, the estimation of an animal's genetic merit for the respective trait, is the sum of its parents' predicted transmitting abilities (PTA) for the trait.

However, the utility of a profit maximizing decision rule for heifer selection based on genetic information is not clear for a number of reasons. Firstly, the relationship between profitability and the genetic information is not known. Secondly, it is not clear how the value of the genetic information updated every six months over the two year period of heifer rearing will compare to the rearing costs of those periods. Finally, the decision rule is complicated by its dynamic nature -- heifer selection in previous rearing periods will alter the subsequent heifer population on which selection is practiced, and this must be considered by the decision rule.

Random heifer selection should improve producer profitability and cash flow versus a scenario where heifers are not selected, but may not be intuitively appealing to producers. A profit maximizing method of heifer selection will not only further improve producer profitability, but may also be more acceptable to managers thereby encouraging them to utilize the management technique of heifer selection.

Introduction

Michigan milk producers face an increasingly competitive market, not only nationally but as well internationally. Least cost and efficient production are necessary to survive in commodity markets. The call for attention to the economically important issue of culling has been made (Nelson, 1994). As well "additional research is required to determine how dairy replacement systems reflect the lifetime productivity and profitability of the dairy cow" (Heinrichs, 1993).

Briefly, Chapters 2 through 5 of this dissertation consist of literature review. Chapter 6 contains a summary of the literature review and the three hypotheses that were generated as a result. The following three chapters (7, 8, and 9) each evaluate one of the hypotheses. Chapter 10 summarizes the dissertation and discusses potential areas of further research.

While much of the literature review uses mathematical explanations, intuitive explanations of the concepts are also provided. Chapters 2 and 3 consider the economics of asset replacement theory, also known as capital budgeting. While capital budgeting is not directly applied in this dissertation, the current research is highly predicated upon studies which have applied this theory to dairy herds. As this wealth of work on optimal culling rates in dairy herds has been largely overlooked, it is important to review this literature along with its associated methods. Given an understanding and acceptance of the results from this large body of literature, the logic of the necessity of a method of heifer selection will become clear.

Following a brief introduction to neoclassical production economics, Chapter 2 reviews standard replacement theory using Net Present Value analysis. While some models of dairy herd culling have employed this technique, it is not particularly appropriate to the current problem for various reasons including the replacements are not identical to the animals they replace. These models are however useful for gaining an intuitive understanding of the replacement issue, as well as, stressing the importance of time horizon, model specification, and opportunity costs. Despite the lack of sophistication, these models suggest dairy cow culling, at least in the Netherlands, may be greater than that to maximize profit.

Chapter 3 considers dynamic programming optimization and simulations of dairy herd culling. A number of studies will be considered to stress the diversity of these models, in terms of modeling approach, and assumptions. Despite the diversity of these models (including the country modeled), the results are remarkably similar - culling rates to optimize profitability are less than those currently employed by the dairy industry. This

literature is tediously reviewed because despite the consensus, it is unfortunately not well understood by the dairy industry or its advisors.

The simulation studies help to provide some understanding of why maximum culling rates do not maximize income. The excessive culling rates used by the dairy industry may maximize milk production and genetic gain. However, the industry and its advisors have failed to consider the cost of these gains, namely the opportunity cost of the heifers.

Chapter 3 concludes with a discussion of why the dairy industry as a whole has too high a culling rate. This is not a trivial concern as the research presented in Chapters 2 and 3 suggests the industry is acting inefficiently and irrationally from an economic perspective. Included in the discussion of the potential motivations of the dairy industry is my argument that involuntary culls simply rarely exist.

The conclusion to be drawn from Chapters 2 and 3 is that many Michigan dairy producers have too high a culling rate. This directly suggests these producers have an excess of heifers, necessitating a method of heifer selection.

The weakness of these culling models is their modeling of heifer selection, which was often random or arbitrarily based on some genetic parameter. Given that all replacement heifers are not required, Chapter 4 considers the potential methods of ranking heifers for maximum profit. Lifetime profit corrected for the opportunity cost of postponed replacement is revealed to be the economically correct objective in heifer selection. This value is shown to be moderately correlated with first lactation milk production.

Currently estimated breeding value, based on parents' PTA, is one of the few indicators of first lactation milk production, and by association, lifetime profit corrected for the opportunity cost of postponed replacement. Chapter 5 critically evaluates the Selection Index estimation of genetic value.

Chapter 6 summarizes the major points from the literature review. It is apparent that estimates of genetic value currently represent the only readily available means of heifer selection. Chapter 6 then presents the three hypotheses which were evaluated in this research.

While parents' PTA should have some predictive ability for subsequent first lactation milk production and therefore be useful in heifer selection for maximum profit, surprisingly this theoretical relationship has not been empirically evaluated to any degree. In keeping not only with the standard methodology of economics, but as well, the scientific method of logical positivism it was deemed important to empirically examine the theory. So prior to generating a heifer selection rule based on genetic estimates it was deemed important to return to the basics of evaluating how well genetics predicted subsequent milk production. This research is presented in Chapter 7.

With some indication that genetics may be useful in heifer selection, the research turned to exploring a method of heifer selection. Producers are likely making the correct decisions as to which animals they are culling, the problem is the late timing of the decision. By waiting until after freshening to decide which animals to cull, producers are paying too high a price to make this correct decision - the value of the decision is less than its cost. If genetics could predict which animals the producers would subsequently cull, a

very good heifer selection rule with a powerful motivational influence to lower dairy producers' culling rates would result. Chapter 8 considers the ability of genetic indices to predict culling during rearing or the first lactation. As well, some light is shed on the voluntary nature of what has traditionally been deemed voluntary versus involuntary culls.

Following the use of econometric methodology to elicit a method of selecting heifers using parents' genetic estimates, Chapter 9 turns to an operations research approach. Modified Box-Complex, a search algorithm, was used to derive a potentially dynamic method of heifer selection to maximize producer profit. This approach incorporated the multi-period nature of heifer selection, where every six months during the rearing process new improved genetic estimates were produced and there was an economic consequence, either positive or negative, associated with raising the heifers for that period. Given the proportion of heifers to be selected, the rules consisted of the proportion of heifers and the weightings for the genetic information in each period. Using a sample of herds which was used solely for testing the rules, Complex's rule was tested against the profits associated with random heifer selection and a rule based on genetic theory.

Chapter 10 provides a summary of this dissertation, and discusses potential avenues of further research.

BIBLIOGRAPHY

Heinrichs, A. J. 1993. Raising dairy replacements to meet the needs of the 21st century. J. Dairy Sci. 76:3179-3187.

Nelson, A. J. 1994. Information needs of the dairy industry for health and nutrition management. J. Dairy Sci. 77:1984-1991.

Chapter 2

PRESENT VALUE MODELS

Livestock culling is fundamentally a replacement issue. This chapter will briefly introduce the standard neoclassical production economics approach to input selection. Finding this technique inadequate in the assessment of durable assets, the Net Present Value (NPV) method of capital budgeting will be employed to optimize increasingly complex replacement models. The flexibility of these models will be showcased by addressing the issue of the optimal calving interval, before turning to the animal replacement decision. It will become apparent that the dynamic and complex nature of replacement in general, and dairy cattle in particular, is frequently underestimated. For example, these results highlight that the approach of collecting milk production data on dairy cows with the goal of assessing the most profitable time for insemination is so simplistic as to be meritless.

While even the moderately complex analytical models presented in this chapter are not fully adequate to address replacement decisions in dairy herds these models and their applications do provide some insight into the problem. These models provide intuition into the understanding of capital asset replacement theory decisions including the importance of choosing the appropriate model and time horizon for a given problem, as well as, considering the opportunity costs of replacement. Applying standard replacement theory, I also explain why if genetics is useful in heifer selection, such a selection scheme should be associated with a higher culling rate among the cows than a herd which randomly selects

heifers. Turning to applications of these NPV models in the literature, numerical versions of the models which suggest that as of the mid-1980s, culling strategies in the Netherlands were undesirable and associated with excessive culling rates are reviewed. Applying an understanding of NPV models and replacement economics, the result that the profit maximizing culling rates were most sensitive to the price of the replacement relative to the cow and minimally influenced by milk price or rate of genetic improvement are explained. Finally, Van Arendonk's (1991) application of replacement theory revealing that neither lifetime profit nor profit/day correctly rank animals based on profit is reviewed along with the economically correct method of ranking cows.

Production economics is the area of neoclassical economics which considers the selection and application of constrained inputs to maximize an objective, commonly profit. Inputs used in production can be classified according to their functional life span: expendables are instantaneously and completely consumed in the production process, while durable inputs or assets have the potential to supply services for multiple periods (Robison and Barry, 1996). Production economics tends to focus on the comparative static analysis of instantaneous production functions utilizing expendable inputs; all dynamic or intertemporal components are ignored (Beattie and Taylor, 1993). Feeding a dairy cow to produce milk is an example of such a production function, and can be mathematically represented as

$$m = f^{\alpha} \tag{2. 1}$$

where m = milk production

f = level of feed supplied

α = the productivity of feed in milk production.

The profit (π) from such a production function is a function of the prices of milk (P) and feed (A).

$$\pi = Pf^{\alpha} - Af \tag{2. 2}$$

Equating the marginal cost of the feed with the marginal value product of the milk reveals the profit maximizing level of the feed input. As we will see in the next chapter, failure to understand or use this simple static profit maximizing decision rule has resulted in poor advice on optimal culling rates from dairy industry advisors.

Given the optimal input level, static analysis can then be conducted to assess the effects of changes in any of the parameters on the optimal level of feed.

However, even simple extensions of the above problem can not be adequately dealt within a standard production economics framework. Consider a situation where the amount of feed available is limited for a given period of time, for example the inventory of feed is fixed at level F until time period T when the next harvest will occur. Intuitively it can be reasoned that the production functions are now temporally related. Decisions made in any time period will have ramifications on subsequent periods. So the feed available in a given period (t) is a function of the total amount available and the amount previously fed.

$$0 \leq f_t \leq F - \sum_{i=0}^{t-1} f_i \qquad 1 \leq t \leq T \tag{2. 3}$$

Equations (2.1) and (2.2) then become

$$m_t = f_t^{\alpha} \tag{2. 4}$$

$$\sum_{t=1}^{T} \pi_t = \sum_{t=1}^{T} (Pf_t^{\alpha} - Af_t) \quad \text{s.t. Equation (2.3)} \tag{2.5}$$

The solution will then consist of a time path of optimum feed levels in each period. Static analysis of this problem is rarely appropriate given that changes in f_t will have ramifications on the amount fed in time periods greater than t. Comparative static analysis is apropos in a dynamic system only in the special situation in which the time periods are uncoupled. This requires, at a minimum, a nonbinding feed constraint which reduces the problem to solving Equation (2.2) T times.

Given that static analysis may not be suitable for solving situations involving expendable inputs, it is not surprising that this technique is inadequate for dealing with production using durable assets, which by definition have multiple period lives. Consider a feed wagon purchase decision, the initial cost of the wagon will be borne when purchased in the first period, while returns and expenses associated with operation of the wagon will be realized in subsequent periods. And finally, at the end of its useful life, a return or cost will likely be associated with retirement of the asset.

Capital budgeting is one technique used to analyze the economics of assets or inputs which generate costs and returns over multiple periods. While payback, accounting return on investment (i.e., return on equity, or return on assets), benefit-cost analysis (or profitability index), and internal rate of return have all been used to assess the profitability of capital assets, net present value is the method of choice. NPV is the only method which unambiguously considers all the cash flows generated by an asset, the time value of money, does not make overly rigid reinvestment assumptions, and provides a unique solution (Robison and Barry, 1996). The formula for NPV in discrete time is

$$NPV = -V_0 + \sum_{t=1}^{n} \frac{R_t}{(1+r)^t} + \frac{V(n)}{(1+r)^n} \qquad (2.6)$$

where V_0 = the initial investment in the asset
R_t = is the net cash flow in time period t
n = the life span of the asset
$V(n)$ = the (salvage) value of the asset in time period n
r = the discount rate of the challenger.

The challenger is the next best alternative use of the investment, while the defender is the asset explicitly being considered and for which the NPV is being calculated. The discount rate which is associated with a particular challenger and should be adjusted for risk and taxes (Robison and Barry, 1996). While the challenger may be a interest bearing investment, it is incorrect to generally define r as the interest rate. This incorrect definition of r in not uncommon. If nothing else such a definition begs the question which interest rate - that associated with a chequing account, savings account, certificate of deposit, treasury bill, grade of private or public bond?

Equation (2.6) can be equivalently expressed in continuous time as

$$NPV = -V_0 + \int_0^n R(t)e^{-rt}\,dt + e^{-rn}V(n) \qquad (2.7)$$

Now consider the simple decision of when to sell a cull dairy cow, assuming no replacement (Perrin, 1972). The decision rule is derived by setting the derivative of the objective function (*NPV* in Equation (2.7)) with respect to the choice variable (time) equal to zero (Beattie and Taylor, 1993).

$$\frac{dNPV}{dn} = R(n)e^{-rn} + V'(n)e^{-rn} - re^{-rn}V(n) = 0 \qquad (2.8)$$

Equation (2.8) can be simplified to

$$R(n) + V'(n) = rV(n) \qquad (2.9)$$

Negativity of the second derivative is required for a global maximum. Equation (2.9) reveals that the optimal time to sell is the period in which the current period return, *R(n)*, plus the change in salvage value, *V'(n)*, are equal to the opportunity cost of the salvage value, *rV(n)*. The original cost of the animal does not influence the decision of when to sell the animal. This is an application of the economic principle of ignoring fixed cost in subsequent decision making. Clearly then, optimizing the decision of when to cull the animal does not necessarily result in a non-negative net present value.

More complex replacement problems will now be considered. The vehicle for this analysis will be the selection of the optimal lactation length of a dairy cow. This problem is appropriate to consider not only because it is the focus of a large body of contemporary research, but as well, it will serve to highlight some other issues. Namely, that the solution technique must be customized to the problem being considered -- static analysis is not necessarily appropriate for all expendable inputs, and capital budgeting is not limited to only strict capital assets. The lactation length decision also makes apparent the complexity of even a seemingly simple problem, and the necessity of the optimal lactation length as an input to optimal culling decisions.

Consider first that, assuming a constant dry period length, the optimal lactation length can be equivalently stated as the optimal calving interval or time of conception. This problem then is one of optimal replacement, but the factor being considered for

replacement, the lactation, does not fit the strict definition of a capital asset, since it is not a separate physical entity with investment and salvage values. In contrast, the animal itself, is a strictly defined capital asset. As this optimal calving interval problem is explored it should become apparent that solving this problem is integral to solving the problem of actual cow replacement.

Consider a simple model of a lactation which accounts only for the daily net returns (*l(t)*) of a single lactation.

$$NPV = \int_{0}^{m} l(t)e^{-rt}\,dt \qquad \textbf{(2. 10)}$$

For a global maximum, strict concavity$\left(NPV' > 0, \text{ and } NPV'' < 0\right)$, is required in the neighborhood of the solution to the first order condition (Beattie and Taylor, 1993). If it can be assumed that daily returns, which are equivalent to net marginal profit, follow the same general shape as the lactation curve, a global maximum is assured.

The profit maximizing decision of when to end the only lactation will obviously be the equimarginal rule of static production -- continue milking the cow until the net return is zero (Equation (2.12)).

$$\frac{dNPV}{dm} = l(m)e^{-rm} = 0 \qquad \textbf{(2. 11)}$$

$$l(m)=0 \qquad \textbf{(2. 12)}$$

This simple single lactation model overlooks the reality that a lactation is often ended in preparation for a subsequent lactation. The problem is truly one of optimal replacement of a cyclical production function, namely lactations. A more realistic NPV model can then be respecified in Equation (2.13). It includes a constant dry period cost and arbitrarily assumes culling after the third lactation, and, similar to the problem above, assumes all lactations are identical. Notice that while multiple lactations of a single animal are considered, the fundamental decision is lactation length, as animal replacement is not considered.

$$NPV = -V_0 + \int_0^{n_1} l(t)e^{-rt}dt + e^{-rn_1}\left(\int_0^{n_2} l(t)e^{-rt}dt - c\right) + e^{-r(n_1+n_2)}\left(\int_0^{n_3} l(t)e^{-rt}dt - c\right) + e^{-r(n_1+n_2+n_3)}SV \qquad (2.\ 13)$$

where V_0 = the initial purchase price of the animal

$l(t)$ = the net lactational returns

c = the cost associated with the constant length dry period

SV = the salvage value of the cow at the end of the third lactation

n_i = the length of lactation i.

The cost of the dry period is accrued at the beginning of the subsequent lactation. Note that the cost of the dry period in the third lactation had been omitted since with a constant salvage value the animal would be sold prior to the dry period. Making c a function of t would permit the cost of the dry period to vary with the length of the lactation. Taking the derivative of NPV with respect to each n_i

$$\frac{dNPV}{dn_3} = e^{-r(n_1+n_2)}l(n_3)e^{-rn_3} - re^{-r(n_1+n_2+n_3)}SV = 0 \qquad (2.14)$$

$$\frac{dNPV}{dn_2} = e^{-rn_1}l(n_2)e^{-rn_2} - re^{-r(n_1+n_2)}\left(\int_0^{n_3} l(t)e^{-rt}dt - c\right) - re^{-r(n_1+n_2+n_3)}SV = 0$$

$$(2.15)$$

$$\frac{dNPV}{dn_1} = l(n_1)e^{-rn_1} - re^{-rn_1}\left(\int_0^{n_2} l(t)e^{-rt}dt - c\right) - re^{-r(n_1+n_2)}\left(\int_0^{n_3} l(t)e^{-rt}dt - c\right) - re^{-r(n_1+n_2+n_3)}SV = 0 \quad (2.16)$$

These first order conditions for a maximum can be restated as

$$l(n_3) = rSV \quad (2.17)$$

$$l(n_2) = r\left(\int_0^{n_3} l(t)e^{-rt}dt + e^{-rn_3}SV - c\right) \quad (2.18)$$

$$l(n_1) = r\left(\int_0^{n_2} l(t)e^{-rt}dt - c + e^{-rn_2}\left\{\int_0^{n_3} l(t)e^{-rt}dt - c + e^{-rn_3}SV\right\}\right) \quad (2.19)$$

Equation (2.17) suggests that the animal should be sold in that period of the third lactation in which the net profit is equal to the opportunity cost of salvage. Comparing the decision rules for the third lactation (Equation (2.17)) and the single lactation (Equation (2.12)) reveals that since the right hand side (RHS) of Equation (2.17) is positive, $l(n_3)>l(m)$. Recall that in the neighborhood of the optimal length of lactation the profit function must be increasing at a decreasing rate (i.e., the function must be strictly concave) so the marginal profit is decreasing. Therefore, a higher net marginal profit function ($l(n_3)$) corresponds to a shorter lactation length and unambiguously $n_3<m$. Intuitively this makes sense since the RHS is the opportunity cost of postponed culling in the last lactation. Once sold the salvage value could be invested in the challenger at rate r. So once marginal profit

drops to this level it is optimal to sell. Obviously this will occur prior to the net lactational returns dropping to zero. There is a cost associated with not culling the animal, and as a result the level of profit at which the animal is sold will be higher than if this cost is not considered.

Solving for *SV* in Equation (2.17) and substituting into Equation (2.18) results in Equation (2.20).

$$l(n_2) = r\left(\int_0^{n_3} l(t)e^{-rt}\,dt - c\right) + e^{-rn_3} l(n_3) \qquad \textbf{(2. 20)}$$

From Equation (2.20) it can be deduced that $n_2 < n_3$, for the RHS of Equation (2.20) includes not only the discounted salvage value opportunity cost of Equation (2.17), but also the opportunity cost of the third lactation less the cost of the dry period. Since the marginal profit of the last day of the second lactation must be greater than that of the third lactation, the second lactation must be shorter than the third.

Dividing Equation (2.18) by *r* and substituting into Equation (2.19) results in Equation (2.21) which reveals that $n_1 < n_2$, since again, the first lactation confronts all those opportunity costs that the second faced, discounted back another lactation, as well as the opportunity cost of the second lactation itself.

$$l(n_1) = r\left(\int_0^{n_2} l(t)e^{-rt}\,dt - c\right) + e^{-rn_2} l(n_2) \qquad \textbf{(2. 21)}$$

The optimal lactation length is then a function of the number of subsequent lactations with $n_1 < n_2 < n_3 < m$. Clearly then, the simple lactation decision rule of Equation (2.12), by ignoring the cost of the dry period and the opportunity cost of future lactations, overestimates the optimal length of all lactations, and therefore, calving intervals.

Even intuitively, it is apparent that the calving interval decision is dynamic. The rationale behind shorter calving intervals is to forego the lower returns associated with the end of lactation, in order to return the animal to peak production when returns are supposedly higher, ignoring the peri-parturient risks for the animal. Using this reasoning, it should be apparent that the consideration of a single lactation is an inappropriate method of problem formulation.

This simple model can provide insight into what factors should be considered when assessing when to breed a cow. Given identical milk and profit functions, which would approximate production of mature cows, the optimal calving interval is a function of the rate of decline of milk production and profit, in other words, persistency of milk production. Peak level of milk production, which is felt to be a good indicator of the area under the lactation curve, is relevant only when levels of production will vary in subsequent lactations as occurs with animals in their first or second lactation. In terms of first lactation animals the breeding decision is now a function of both persistency and total production -- when to end the generally highly persistent, but lower yielding first lactation to enter the higher production, lower persistency mature lactations. Notice that assuming a constant length dry period, the shorter the calving interval the greater the proportion of time spent in the dry period. So even with poorly persistent mature cows, very short lactations which maximize the amount of time the animal spends in or near peak production may not be economically optimal. Thus, the rationale for the argument that given the ability to maintain persistency, such as with bovine somatotropin, the optimal calving interval will be extended. (Of course this assumes that the use of bST is economically sound, and ignores issues such as the shift in the proportion of the herd milking and dry at any point.) The simple model indicates as the cost of the dry period increases, ceteris paribus, the lactation length increases. This cost could included not only obvious items such as feed, but as well, the potential cost of periparturient disease. It should also be apparent that the suggestion of a single optimal calving interval for all the animals in a herd is inappropriate, as is the "cost of a day open" concept.

Given this analysis, the inappropriateness of collecting empirical milk production data on cows to then determine the optimal insemination period is obvious. At a minimum for such an analysis, as the previous models showed, all future lactations would need to be considered. Practically such an analysis would need to account for the censored nature of the data due to culling.

Also, consideration of only a single animal's lactations is only appropriate if that animal is not replaced. In reality, most dairy herds operate under a binding herd size constraint, and the sale of a current asset is required for the introduction of a new asset. So ultimately the insemination decision includes not simply when to inseminate but as well who to breed, and at what point breeding should no longer be continued. To address the issue of asset replacement, infinite replacement models will now be considered. While these models can apply to any asset, the finite lives and therefore lactations of individual cows, are not particularly suitable for such examples, so attention will focus on animal replacement. Following the review of infinite models, the role of individual animal lactations in these models will the broached.

Assuming an infinite time horizon of identical replacement, this problem can be modeled as

$$NPV = \int_0^{n_1} g(t)e^{-rt}dt + e^{-rn_1}\int_0^{n_2} g(t)e^{-rt}dt + e^{-r(n_1+n_2)}\int_0^{n_3} g(t)e^{-rt}dt + \cdots \qquad (2.22)$$

Where $g(t)$ = net cash flow in period t

$$g(n) = -V_0 + \int_0^{n} g(t)e^{-rt}dt + e^{-rn}V(n)$$

and n_j = the life span of animal j, $j = 1, \ldots \infty$.

The optimal life span of each asset is again determined by taking the derivative of NPV with respect to each n_j.

$$\frac{dNPV}{dn_1} = g(n_1)e^{-rn_1} - re^{-rn_1}\int_0^{n_2} g(t)e^{-rt}dt - re^{-r(n_1+n_2)}\int_0^{n_3} g(t)e^{-rt}dt + \cdots = 0 \quad (2.23)$$

$$\frac{dNPV}{dn_2} = g(n_2)e^{-rn_2} - re^{-rn_2}\int_0^{n_3} g(t)e^{-rt}dt - re^{-r(n_2+n_3)}\int_0^{n_4} g(t)e^{-rt}dt + \cdots = 0 \ .$$

(2.24)

Due to the infinite time horizon all the first derivatives are the same and therefore $n_1 = n_2 = n_3 = \ldots = n_\infty$. Equation (2.22) can then be restated as

$$NPV = \left(1 + e^{-rn} + e^{-r2n} + e^{-r3n} \cdots\right)\int_0^n g(t)e^{-rt}\,dt \qquad (2.25)$$

$$= \left(\frac{1}{1-e^{-rn}}\right)\int_0^n g(t)e^{-rt}\,dt \qquad (2.26)$$

with first order condition

$$\frac{dNPV}{dn} = \left(\frac{1}{1-e^{-rn}}\right)g(n)e^{-rn} - \frac{re^{-rn}}{\left(1-e^{-rn}\right)^2}\int_0^n g(t)e^{-rt}\,dt = 0 \qquad (2.27)$$

$$g(n) = \frac{r}{\left(1-e^{-rn}\right)}\int_0^n g(t)e^{-rt}\,dt \qquad (2.28)$$

The integral in Equation (2.28) is simply the present value of a replacement, while $\frac{r}{\left(1-e^{-rn}\right)}$ is the annuity factor. As a result, the right-hand side is the annuity or average return from future replacements. Equation (2.28) then represents the standard decision rule of asset replacement -- replace when the marginal (annual) value of the defender (the LHS of Equation (2.28)) equals the average (annual) present value of the challenger (Faris, 1960). The maximum average value of a subsequent replacement can be considered the opportunity cost of postponed replacement. As alluded to in the lactation replacement model, as the number of future replacements increases so does the opportunity cost of replacement, so the optimal life span of an asset with infinite replacement will be shorter than if a finite number of replacements were available.

Finally consider the case of continuous improvement of the asset (Robison and Barry, 1996) as is expected in livestock, due to genetic improvement.

$$NPV_i = \int_0^{n_{1i}} g(t)e^{-rt}dt + (1+h)e^{-rn_{1i}} \int_0^{n_{2i}} g(t)e^{-rt}dt + (1+h)^2 e^{-r(n_{1i}+n_{2i})} \int_0^{n_{3i}} g(t)e^{-rt}dt + \cdots$$

(2. 29)

Where h represents the rate of improvement between successive assets. The optimal life span of each asset is again determined by taking the first derivative of *NPV* with respect to each n_{ji}, where the i subscripting simply serves to distinguish this *NPV* model with improvement from that of Equation (2.22) without improvement of replacements.

$$\frac{dNPV_i}{dn_{1i}} = g(n_{1i})e^{-rn_{1i}} - (1+h)re^{-rn_{1i}} \int_0^{n_{2i}} g(t)e^{-rt}dt - (1+h)^2 re^{-r(n_{1i}+n_{2i})} \int_0^{n_{3i}} g(t)e^{-rt}dt + \cdots = 0$$

(2.30)

$$\begin{aligned}\frac{dNPV_i}{dn_{2i}} = (1+h)g(n_{2i})e^{-rn_{2i}} - (1+h)^2 re^{-rn_{2i}} \int_0^{n_{3i}} g(t)e^{-rt}dt \\ - (1+h)^3 re^{-r(n_{2i}+n_{3i})} \int_0^{n_{4i}} g(t)e^{-rt}dt + \cdots = 0\end{aligned}$$

(2. 31)

Dividing both sides of Equation (2.31) by (1+h),

$$\frac{dNPV_i}{dn_{2i}} = g(n_{2i})e^{-rn_{2i}} - (1+h)re^{-rn_{2i}} \int_0^{n_{3i}} g(t)e^{-rt}dt - (1+h)^2 re^{-r(n_{2i}+n_{3i})} \int_0^{n_{4i}} g(t)e^{-rt}dt + \cdots = 0$$

(2. 32)

Comparing Equations (2.30) and (2.32), it is again apparent that the first order derivatives are identical and $n_{1i} = n_{2i} = n_{3i} = \cdots$.

Equation (2.29) can then be restated as

$$NPV_i = \left[1 + (1+h)e^{-rn_i} + (1+h)^2 e^{-r2n_i} + (1+h)^3 e^{-r3n_i} + \cdots\right] \int_0^{n_i} g(t)e^{-rt}dt \quad (2.33)$$

$$= \left(\frac{1}{1-(1+h)e^{-rn_i}}\right) \int_0^{n_i} g(t)e^{-rt}dt \quad (2.34)$$

with first order condition

$$\frac{dNPV_i}{dn_i} = \left(\frac{1}{1-(1+h)e^{-rn_i}}\right) g(n_i)e^{-rn_i} - \frac{r(1+h)e^{-rn_i}}{\left(1-(1+h)e^{-rn_i}\right)^2} \int_0^{n_i} g(t)e^{-rt}dt = 0 \quad (2.35)$$

$$g(n_i) = \frac{r(1+h)}{\left(1-(1+h)e^{-rn_i}\right)} \int_0^{n_i} g(t)e^{-rt}dt \quad (2.36)$$

This first order condition assumes the function is well behaved ($h>0$, $g'(t) > 0$ and $g'' < 0$). By comparing Equations (2.28) and (2.36), it is apparent that $n_i<n$. Having previously seen the effect of opportunity costs on the optimal length of asset life, it is not surprising that the length of each assets' life is shorter with improvement than without. And that the greater the rate of improvement the greater the decrease in length of asset life.

$$\frac{dg(n_i)}{dh} = \frac{r}{\left(1-(1+h)e^{-rn_i}\right)^2} \int_0^{n_i} g(t)e^{-rt}dt > 0 \quad \therefore \quad \frac{dn_i}{dh} < 0 \quad (2.37)$$

A constant optimal asset life (e.g., $n_{1i} = n_{2i} = n_{3i} = \cdots$) with constant improvement is not as intuitive a result.

The inverse relationship between the rate of improvement of the asset and the optimal asset lifetime bears emphasizing and will be used in Chapter 3. In Chapter 3, I review herd culling models which employ two different methods of heifer selection: random and based on genetics. Now due to genetic improvement from generation to generation even with random selection the potential replacement heifers are improving. However, if genetic indices are useful in heifer selection, then by definition the rate of improvement under this selection scheme should be greater than that under random selection. If this is the case, then the optimal length of the cows' lives should be shorter than under random selection. Ceteris paribus, failure to find a shorter cow life (i.e., higher culling rate) with genetic selection as opposed to random selection would suggest genetic indices are not useful in heifer selection.

Burt (1965) showed that under conditions of uncertain replacement (i.e., involuntary culling) and risk neutrality, the decision rule for the infinite horizon was simply to replace when the expected net current period return equaled the expected weighted average net return from the potential replacement.

While the preceding results have provided analytical results which aid in understanding the culling problem and solutions, they present limited application to solving actual culling decisions for two reasons. First, recall that the opportunity cost of postponed replacement is the average present value of potential replacements. This then requires optimization of all future replacements with respect to all choice variables. For

example Equation (2.13), the NPV for a single animal derived from it's three potential lactations, could then be substituted for $g(t)$ in Equation (2.25). Even with identical replacement, solving this problem would then require simultaneous consideration of four first order conditions, one each for n, n_1, n_2, and n_3. Analytical solutions then become tedious due to the large number of simultaneous equations. Second, the net revenue flows of a dairy cow are very variable throughout a given lactation and across lactations. At the extreme, consider the negative returns during the dry period, which are in fact preparatory for the subsequent lactation. This lack of identical net return or profit functions over time and between challenger and defender realistically removes the possibility of analytical solutions. As we shall see this also presents difficulties in numerical analysis of these problems.

The numerical technique of comparing the marginal profit of the defender versus the average annual profit of the challenger, also known as the marginal net revenue approach, has been modeled for replacement decisions. Renkema and Stelwagen (1979) developed the primary model used for NPV analysis of the culling decision.

In their model the earned income, under Dutch conditions, for a lactation was calculated as gross revenue from milk and calves, less cost of feed and sundries, the change in slaughter value was also considered. Assuming a 12 month calving interval and a constant mature equivalent milk production, the earned income was calculated for lactations 1 through 14. Incorporating the probability of disposal at the end of the next lactation, the expected marginal profit associated with the next year was calculated. With the cost of the replacement heifer and the marginal expected earned income for every

lactation, the non-discounted average annual income of the replacement heifer associated with successive lactations could have been calculated, permitting the determination of the optimal lifespan of the replacement. However, instead the replacement's lifespan was assumed to be 4.2 lactations, approximating the average productive life of the typical Dutch Friesian. The analysis was repeated with a higher level of milk, and as well, the assumption of identical replacement was relaxed by repeating the analysis using a 1% rate of genetic improvement in milk production.

Many of the assumptions of the model are unsatisfying and potentially biasing. For example, the assumption of culling at the end of the lactation versus at day 205 was subsequently found to overestimate profitability by 30% (Van Arendonk, 1984). Care is needed in extrapolating this model of a single challenger and defender to a herd situation. To make this extension all animals in the herd are required to be homogeneous with respect to the 12 month calving interval and mature equivalent milk production. While the model can compare a genetically superior replacement to a cow, it cannot represent a herd with multiple levels of genetic capability. This is a drawback of all such marginal revenue models (Van Arendonk, 1984).

Although the decision rule approximated that of the marginal revenue optimization model, note the lack of discounting, without having optimized the length of the replacement's life, it could easily be argued that this model is actually a simulation rather than an optimization. Simulation versus optimization will be more fully discussed in Chapter 3.

Given the shortcomings of this model, the result nonetheless are interesting. It was determined that increasing herdlife from an average productive life of 3.3 to 5.3 lactations increased annual earned income by 20%. This simplified model suggests the benefit of prolonged productive life, this benefit was minimally affected by genetic improvement. The economic return from longer life was much more sensitive to the net cost of replacement (i.e. cost of replacement heifer less salvage value of the culled cow) than changes in the milk price.

The realism of this model was increased by assuming culling at day 245 of lactation and used in a traditional simulation model (Korver and Renkema, 1979) which will be discussed under simulations in Chapter 3.

The model was also further updated and expanded to include the time value of money and optimal insemination decisions (Dijkhuizen, Renkema and Stelwagen, 1984). The potential calving interval was 345 to 525 days, and at 20 day intervals the replacement decision as well as the insemination decision for the current lactation was made. Replacement was possible in future lactations only at day 225. At each heat (20 day interval) the animal was to be inseminated if the sum of her expected future marginal revenues in the case of pregnancy, less the annuity of the average replacement, was greater than zero. The correlation between length of successive calving intervals was 0.6. This cumulative future difference between the defender and challenger is actually the decision rule used in the original model, in place of that stated. Full numeration of expected returns for each animal are then required.

If it was deemed uneconomical to inseminate, the animal was not culled until the point in her lactation where the expected returns for the remainder of the lactation were less than the annuity of the replacement. A 1% annual genetic increase in milk production was assumed.

While it is not clear, it appears that similar to the original model, the lifespan of the replacement has not been optimized, but rather arbitrarily chosen.

The results were in the form of critical production levels by lactation, and days at insemination below which it is not optimal to inseminate a cow. These guides indicate that a young cow which is average or above average in level of production can profitably be inseminated up to 8 to 9 months post partum, suggesting the culling strategy in the Netherlands was not optimal. While the results were not sensitive to the price of milk, since this affected the returns of both replacements and defenders, changes in the cost of a replacement did affect the critical production level. As expected from theory, the model found the greater the persistence of milk production, the longer it was profitable to inseminate open cows. The authors never calculated the distribution of productive lives in the herd to determine whether rational expectations were fulfilled (i.e., if the assumed length of the replacement's productive life equaled that of a herd resulting from following the guide). Failure to fulfill the self-expectation nullifies any apparent solution. This may have been what the authors were referring to as "more insight is desired into the economic consequences of making replacement decisions over a number of years with different herds on the above mentioned criteria". Also note that future replacements are a function of the animals currently in the herd and management, including culling. So the simple static

comparison of a defender with a genetically superior heifer is over simplified. The dynamic nature of the problem is clear - decisions made today affect future decisions.

While the realism of the model has been increased through the modifications some unrealistic assumptions, such as the timing of future culling still remain. These simplifications are a direct function of this marginal method of analysis. As was discussed previously, permitting net return or profit functions to vary over time and between challenger and defender would require calculation of all expected net revenue for all potential future decisions. As the number of potential decisions increases, the complexity and number of calculations quickly become overwhelming.

Kuipers (1980) also produced a marginal net revenue model to evaluate the expected marginal profit of a challenger over the remainder of the current lactation and first six months of the subsequent one. This poor choice in time horizon for consideration of the marginal value of the challenger underestimated the value of younger animals, since, as noted earlier, production and net revenue increase over the initial lactations.

It is only recently that the concepts from replacement theory models have been applied to other avenues such as cow ranking for profitability (Van Arendonk, 1991). A large amount of literature has focused on the contributions of such factors as increased length of herdlife and milk yield on cow profitability. The most common profit functions used to determine the dependent proxy variable for profit have been lifetime profit and profit per day of life. Van Arendonk realized that the former implies no replacement or implicitly a non-binding herd size constraint. Profit per day of life implies identical replacement, which is unrealistic due to genetic improvement and the lack of systematic differences between heifers available for replacement of different cows. The failure to account for the opportunity cost of future replacement has resulted in the possible nullification of a large amount of previous work. This area will be considered further in Chapter 4.

In summary, this chapter focused on the optimal calving interval decision, as well as, animal replacement. It was apparent that even these simple problems are complex and increasingly difficult to solve due to the number of choice variables, the intertemporal and often dynamic nature of the problem. The importance of considering model specification and time horizon also were evident. While these models provided some intuitive insight into some of the underlying principles of these solutions, such as the importance of opportunity costs, ultimately they were unable to answer the complex question of the optimal herd level culling and replacement strategies. Simplified numerical models suggested the potential for improvement in culling strategy, at least under Dutch conditions. The difficulty with numerical analysis was the requirement to consider all potential future decisions and returns. A subset of these dynamic problems, namely those that can be recursively stated, can be solved with greater ease through dynamic programming. This technique and the results of replacement models using this technique will be addressed in Chapter 3.

BIBLIOGRAPHY

Beattie, B. R., and C. R. Taylor. 1993. Page 1-8 in the Economics of Production. Kreiger Publishing Company, Malabar, Florida.

Burt, O. R. 1965. Optimal replacement under risk. J. Farm Econ. 47:324-346.

Dijkhuizen, A. A., J. A. Renkema, and J. Stelwagen. 1984/85. Economic aspects of reproductive failure in dairy cattle. II The decision to replace animals. Prev. Vet. Med. 3:265-276.

Faris, J. E. 1960. Analytical techniques used in determining the optimum replacement pattern. J. Farm Econ. 44:755-766.

Korver, S. and J. A. Renkema. 1979. Economic evaluation of replacement rates in dairy herds II. Selection of cows during first lactation. Livest. Prod. Sci. 6:29-37.

Kuipers, A. 1980. Development and comparison of selection criteria for cows and bulls with a dairy herd simulation model. Ph.D. Thesis, Univ. Wisconsin, Madison.

Perrin, R. K. 1972. Asset replacement principles. Am. J. Agr. Econ. 54:60-67.

Renkema, J. A., and J. Stelwagen. 1979. Economic evaluation of replacement rates in dairy herds I. Reduction of replacement rates through improved health. Livest. Prod. Sci. 6:15-27.

Robison, L. J., and P. J. Barry. 1996. Ch. 1, 2, 8, 11, 21 in Present Value Models and Investment Analysis. 1st ed. The Academic Page, Northport, Alabama.

Van Arendonk, J. A. M. 1984. Studies on the replacement policies in dairy cattle. Z. Tierzuchtg. Zuchtgsbiol. 101:330-340.

Van Arendonk, J. A. M. 1992. Use of profit equations to determine relative economic value of dairy herd life and production from field data. J. Dairy Sci. 74:1101-1107.

Chapter 3

DYNAMIC OPTIMIZATION AND CULLING MODELS

From Chapter 2 it was evident that while the traditional NPV approach to animal replacement lent some intuition to the understanding of the problem, it was not particularly well suited to actually solving these problems. In this chapter another solution technique, dynamic programming, will be reviewed. After defining the model through an example, analytical dynamic programming will be covered to gain familiarity with the model and its methods. Numerical dynamic programming solutions, a much more practical approach to complex dynamic problems such as livestock replacement will follow. Given this foundation the results of dairy cattle culling studies utilizing dynamic programming and optimization will be reviewed. While tedious, consideration of these models, in at least a superficial way, is necessary for two reasons. First, in order to draw proper conclusions about the culling issue based on the results of these models, the diversity of the models must be realized. And second, these models should be reviewed for this area of management tends to be overlooked by the industry despite the wealth of knowledge in this area. These studies are summarized in Tables 3.7 and 3.8 at the end of the chapter. The issues of voluntary and involuntary culling and simulation versus optimization will also be considered. The conclusion will be that the market is operating inefficiently through excessive culling, corroborating evidence of this inefficiency is presented. I then try to resolve why the market is operating in an economically irrational manner.

To introduce dynamic programming an analytical solution to a simple replacement model will be used. Analytical models tend to be quite general, utilizing symbols rather than numbers. The advantage of this method is that it will provide some insight into the solution, as the models in Chapter 2 did. Numerical analysis, where the values of the parameters are used, is easier to operationalize to get actual results especially in complex problems.

Consider the problem of when, in its lifetime, to sell an animal without replacement. This is an example of a simple optimal stopping problem. Reconsider Equation (2.6) from Chapter 2

$$NPV = -V_0 + \sum_{t=1}^{n} \frac{R_t}{(1+r)^t} + \frac{V(n)}{(1+r)^n} \tag{3. 1}$$

The objective is to maximize the profit given that the animal is present so the initial investment may be omitted

$$\max \sum_{t=1}^{n} \frac{R_t}{(1+r)^t} + \frac{V(n)}{(1+r)^n} \tag{3. 2}$$

The net cash flow in any lactation (π_t) can be denoted as

$$\pi_t = R(n_t) + d_t V_t \qquad 1 \le t \le T \tag{3. 3}$$

The notation has been altered slightly so $R(n_t)$ explicitly recognizes that the returns in a lactation from milk production are a function of the lactation number n_t. As was discussed in Chapter 2, in order to optimize the culling decision, the length of the lactation needs

also to be optimized. To maintain simplicity assume that the $R(n_t)$are already optimized with respect to lactation length.

Let d_t be a dichotomous choice or control variable with

$$d_t = \begin{cases} 0 & \text{if the animal is not culled at } t \\ 1 & \text{if the animal is culled at } t \end{cases} \tag{3. 4}$$

V_t remains the salvage value of the animal in period t, and T is the maximum potential lifespan of the animal, after which the animal has no value, $V_{T+1}=0$.

Equation (3.2), in dynamic programming termed the objective functional, can then be generically restated

$$\max \beta\pi_t(n_t,d_t) + \beta^{t+1}\pi_{t+1}(n_{t+1},d_{t+1}) + \cdots + \beta^T\pi_T(n_T,d_T) \tag{3. 5}$$

where $\beta = 1/1+r$, and $1 \le t \le n$.

In traditional dynamic programming terminology, π_t is the return function representing the immediate period return associated with n_t, the state variable, and d_t, the control variable. The state variable or vector, if more than one state variable, define the state of the system in period t. In period t, the decision maker cannot affect the state variables, they are taken as given. The decision maker has direct control over the control variable or vector, d_t, in period t. The transition equation, which links the state and control variables, for this problem is

$$n_{t+1} = n_t + 1 \qquad \text{if } d_t=0$$

$$n_{t+i} = 0 \ (1 \le i \le T-t) \qquad \text{if } d_t=1 \tag{3. 6}$$

This simple transition equation states that if the animal has not been culled, it will proceed to the next lactation. Obviously once the animal has been culled, there can be no subsequent returns. Notice that the current state variable is a function of the previous control variable

$$n_{t+1} = g_t(n_t, d_t) \tag{3. 7}$$

This then defines the problem as being recursive (i.e., the current controls are only a function of previous, not future, controls). To see this note that

$$n_t = g_{t-1}(n_{t-1}, d_{t-1}) \tag{3. 8}$$

$$n_{t-1} = g_{t-2}(n_{t-2}, d_{t-2}) \tag{3. 9}$$

$$\vdots$$

$$n_2 = g_1(n_1, d_1) \tag{3. 10}$$

Sequential substitution of Equations 3.8 through 3.10 into 3.7 reveals

$$n_{t+1} = \tilde{g}_t(n_1, d_1, d_2, \ldots, d_{t-1}, d_t) \tag{3. 11}$$

Lagging Equation (3.11) one period and substituting into Equation (3.5) reveals

$$\max \sum_{t=1}^{T} \beta^t \tilde{\pi}_t(n_1, d_1, d_2, \ldots, d_{t-1}, d_t) \tag{3. 12}$$

The return function in any period is then only a function of the current and previous controls. Since dynamic programming solves problems recursively, this specialized recursive problem structure is required for this technique. Calculus of variations and the maximum principle are other techniques appropriate for solving recursive dynamic optimization problems. The former requires the transition equation have the restrictive form of depending solely on the current control and not the current state (i.e., $n_{t+1}=g_t(d_t)$). The maximum principle is useful in continuous time formulations, where it is also known as control theory. Control theory utilizes the Hamiltonian as a method to solve differential equations. Solving stochastic infinite horizon problems using control theory is quite complex and not well suited to numerical analysis. In discrete time the Hamiltonian reduces to the Lagrangian, which gives the same results as dynamic programming.

Dynamic programming then reduces to selecting the optimum set of policy functions from a space of functions, $d_t=h_t(n_t)$. This formidable task is made easier through the use of the principle of optimality.

Suppose $\{h_t\}_{t=1}^{T}$ is a set of policy functions that solve the recursive problem Equation (3.5). Now consider the modified problem

$$\max\sum_{t=i}^{T}\beta^t\pi_t(n_t,d_t) \qquad \textbf{(3. 13)}$$

Let $\left\{h_t^*\right\}_{t=i}^{T}$ be the optimal policy functions for the modified problem. The principle of optimality states that for the periods i to T, $h_t=h_t^*$. Notice that it is the optimal policy function that is being chosen, not the optimal control choice. The principle of optimality then suggests that the optimal policy functions are invariant to the earlier time periods. The optimum function is selected such that it is unaffected by changing circumstances. Notice that as the state variables change the control variables will change but not the policy function. Also the principle of optimality does not necessarily suggest the policy function is time invariant (i.e., $h_t = h_{t+1}$). These points will become more clear with an example.

Based on the principle of optimality Equation (3.5) can be restated in Bellman's current period value function.

$$V_t(n_t) = \max_{d_t}\left\{\beta\pi_t(n_t, d_t) + \beta^2 V_{t+1}(n_{t+1})\right\} \tag{3. 14}$$

subject to the transition Equations (3.6), $1 \leq t \leq T$, and given n_1.

The principle of optimality and the recursive nature of the problem permit a simple technique to solve for the optimal policy functions over a space of functions. Dynamic programming starts at the terminal period and solves for the optimal policy function for each time period in a sequentially recursive fashion. The result will be a time path of control variables given an initial value of the state variable.

Substituting the specific form of the return function (Equation (3.3)) into Equation (3.14) reveals

$$V_t(n_t) = \max_{d_t}\left\{\beta R(n_t) + \beta d_t V_t + \beta^2 V_{t+1}(n_{t+1})\right\} \tag{3. 15}$$

subject to the transition Equations (3.6), $1 \leq t \leq T$, and given n_1.

The terminal period value function is denoted

$$V_T(n_T) = \max_{d_T}\left\{\beta R(n_T) + \beta d_T V_T + \beta^2 V_{T+1}(n_{T+1})\right\} \tag{3. 16}$$

T denotes the end of the animal's life, and as a result $V_{T+1}(n_{T+1})=0$. So the decision rule will be to cull the animal at T, if it has not occurred previously. The terminal period policy functions are:

$$d_T = \begin{cases} 0 & \text{if } n_T = 0 \\ 1 & \text{if } n_T = T \end{cases} \tag{3. 17}$$

Due to the categorical nature of the control variable, the solution technique to this problem does not involve the standard technique of differentiation, since due to its discontinuous nature the equation is not differentiable. This then presents another problem with analytical results, including NPV techniques, which was not addressed in the previous chapter.

Notice that the solution is a policy function; the control variable being a function of the state variable. As the state variable changes the optimal control variable will correspondingly change. This optimal policy function for the terminal period will, however, never change regardless of how far back into time the problem is worked.

Substituting Equation (3.17) into (3.16), the terminal period value function is

$$V_T(n_T) = \begin{cases} \beta R(T) + \beta V_T & \text{if } n_T = T \\ 0 & \text{if } n_T = 0 \end{cases} \tag{3. 18}$$

At $t=T-1$, Bellman's equation becomes

$$V_{T-1}(n_{T-1}) = \max_{d_{T-1}}\left\{\beta R(T) + \beta d_{T-1} V_{T-1} + \beta^2 V_T(n_T)\right\} \tag{3. 19}$$

If culling has not already occurred there are two possible options for the period $T\text{-}1$. If the animal is not to be culled, $d_{T-1}=0$, and culling occurs in the final period, so

$$V_{T-1}(n_{T-1})=\beta R(T)+\beta^2[R(T)+V_T] \qquad \textbf{(3. 20)}$$

Or the animal may be culled, $d_{T-1}=1$, and $V_T(n_T)=0$. The returns generated would then be $V_{T-1}(n_{T-1})=\beta R(T\text{-}1)+\beta V_{T-1}$.

The decision rule for $T\text{-}1$, then is

$$d_{T-1}=\begin{cases}0 & \text{if } V_{T-1}<\beta[R(T)+V_T]\\ 1 & \text{if } V_{T-1}\geq\beta[R(T)+V_T]\end{cases} \qquad \textbf{(3. 21)}$$

This decision rule is similar to those from Chapter 2, cull the animal now if the salvage value today is greater than or equal to the discounted returns from lactation and salvage value in the next lactation. Culling when the terms are equal has arbitrarily been assumed.

$$V_{T-1}(n_{T-1})=\begin{cases}\beta R(T)+\beta\max\{V_{T-1},\beta[R(T)+V_T]\} & \text{if } n_{T-1}=T-1\\ 0 & \text{if } n_{T-1}=0\end{cases} \qquad \textbf{(3. 22)}$$

Notice from Equation (3.18) that $\beta R(T)+\beta V_T=V_T(n_T)$ so all terms from Equation (3.19) are represented in Equation (3.22).

Similarly, for period $T\text{-}2$, the animal can be culled in which case the returns are

$$\beta R(T\text{-}2)+\beta V_{T-2} \qquad \textbf{(3. 23)}$$

or alternatively, culled later with returns

$$\beta R(T\text{-}2)+\beta^2\left(R(T-1)+\max\{V_{T-1},\beta[R(T)+V_T]\}\right) \qquad \textbf{(3. 24)}$$

which can be restated using Equation (3.22) as

$$\beta R(T\text{-}2)+\beta^2 V_{T\text{-}1}(T\text{-}1) \tag{3.25}$$

The decision rule is then

$$d_{T-2}=\begin{cases}0 & \text{if } V_{T-2}<\beta V_{T-1}(n_{T-1})\\ 1 & \text{if } V_{T-2}\geq \beta V_{T-1}(n_{T-1})\end{cases} \tag{3.26}$$

With the current period value function being

$$V_{T-2}(n_{T-2})=\begin{cases}\beta R(n_{T-2})+\beta\max\{V_{T-2},\beta V_{T-1}(n_{T-1})\} & \text{if } n_{T-2}=T-2\\ 0 & \text{if } n_{T-2}=0\end{cases} \tag{3.27}$$

In general then, Bellman's equation becomes

$$V_t(n_t)=\begin{cases}\beta R(t)+\beta\max\{V_t,\beta V_{t+1}(n_{t+1})\} & \text{if } n_t=t\\ 0 & \text{if } n_t=0\end{cases} \tag{3.28}$$

The optimal decision rule or policy function is characterized as

$$d_t=\begin{cases}0 & \text{if } V_t<\beta V_{t+1}(n_{t+1})\\ 1 & \text{if } V_t\geq \beta V_{t+1}(n_{t+1})\end{cases} \tag{3.29}$$

The rule is then to cull now if the optimized discounted returns from all future periods are less than or equal to the current salvage value. Despite the recursive nature of the solution technique the results are forward looking considering the possible future returns.

This analytical approach can be easily modified to include replacement and effects of uncertainty of the state variables. Uncertainty is introduced to dynamic programming through making the transition equation dependent on an identically and independently distributed random variable with a mean of zero, ε_{t+1}. In the context of the present example this error term could include the probability of unplanned death. Equation (3.7) would be modified to

$$n_{t+1} = g_t(n_t, d_t, \varepsilon_{t+1}) \tag{3.30}$$

Since the random shock is unknown when the control variable is chosen, the future path of the state variables is uncertain. The stochastic version of Equation (3.15), the current period value function with the specific return function is

$$V_t(n_t) = \max_{d_t}\left\{\beta R(n_t) + \beta d_t V_t + E_t \beta^2 V_{t+1}(n_{t+1})\right\} \tag{3.31}$$

where E_t is the expectation operator conditional on information available at time t. At t, $R(n_t)$ and V_t are known with certainty. Since the expectation of ε_{t+1} is zero, the stochastic aspect of the problem is eliminated (Bellman, 1957). The approach to solving the problem is again one of determining the set of optimal policy functions, $d_t = h_t(n_t)$. Notice however that since n_t is a stochastic variable and d_t is a function of the state variable, the time path of the control variables is itself a stochastic process.

Substituting the policy function into the transition equation reveals

$$n_{t+1} = g_t(n_t, h_t(n_t), \varepsilon_{t+1}) \tag{3.32}$$

The sequence of state variables then follows a Markov process, in that the current state contains all the relevant information for predicting future states. While this may seem quite restrictive, using vectors, autoregressive processes can also be represented as Markov processes.

It is a simple matter to introduce the returns from multiple animals into the objective functional to represent replacement. Iteration of Bellman's equation would simply continue for more periods. Solving dynamic programming problems through iterations requires some consideration for the time horizon of the problem. As the example illustrates the farther back in time one goes the less critical the generally arbitrary value of the terminal state becomes. As iterations continue into the past it is possible for the system to enter a steady state, where the policy functions are time invariant (i.e., $h_t = h_{t+1}$).

Analytical problems are not easily adjusted to solve for multiple state and control variables. Rather numerical analysis is more amicable to such complex problems. There are a number of methods for numerical analysis of dynamic programming problems. The full enumeration of future possible states will be examined here.

The terminal value function is a vector with the number of elements equal to the number of possible states. The state and control variables are members of a finite defined set. For numerical analysis, a finite state Markov chain is used. A probability matrix contains the probabilities of being in a certain vector of the state variables in $t+1$ given a present vector of the state variables. For a given state vector the terminal period value function is calculated through consideration of the return function and terminal value function for every possible control vector. The control vector associated with the maximum value of the terminal period value function is the optimal control policy for the given state vector at *T-1*. This is repeated for every value of the state vector, resulting in an optimal control vector for every possible state vector for the terminal period.

In this full enumeration approach iteration occurs backwards in time. The period value function for *T-1*, is calculated similarly to the original iteration. The terminal period (*T*) value function, is determined by using the finite state Markov chain to weight the possible value functions in *T* by the probability of the respective state vector occurring given the state vector in *T-1*. (This is simply an application of Burt's (1965) idea from Chapter 2 that the expected value is a function of the summation of each possible value weighted by it's probability of occurrence.) Summing the current return function, and the terminal value function gives the current period value function in *T-1*. The optimal policy function for period *T-1*, is a vector of state variables and the associated control variables which maximize the respective state variable's current period value function. The policy function from the first iteration is then the optimal policy function for the next period (i.e., *T*).

By assuming identical future replacement the marginal net revenue approach from Chapter 2 is a special case of dynamic programming (Burt, 1965).

The advantages of numerical solutions are the relaxation of the model requirements such as continuous differentiability. Also there is no need to solve numerous analytical simultaneous equations. The disadvantages are the lack of intuition provided by the model, which is important not only in gaining insight into the general or specific problem, but as well, greater difficulty in realizing errors in the problem formulation or solution technique. These models also require a large body of information which is commonly not available. For example transition probabilities between all possible combinations of states are required.

Highly complex dynamic optimization programs are available to solve for management decisions in dairy herds. Van Arendonk's model (Van Arendonk, 1985a; Van Arendonk and Dijkhuizen, 1985) will the be primary focus as it is the most sophisticated and widely used model. While components of the model have been validated at least for Dutch conditions (Van Arendonk, 1985), in general the complexity of dynamic optimization models make their evaluation quite difficult, for due to their large size full disclosure of the model is not possible. Furthermore, the accuracy of those parts presented is often not obvious, for example, there is little data on which to base evaluation of probabilities of conception for each combination of days in milk, and animal age. Model evaluation will then be based, not only on the inputs used in the model, but also on how well its description of the herd matches generally accepted parameters (i.e., income over feed cost). Some of the inputs and assumptions, as well as, outputs not necessarily directly related to the culling issue will then be reported. Sensitivity analysis will reflect the robustness of model results to input changes. Following this the last source of model validation - comparison of the results with other optimization and simulation models will be considered. Tables 3.7 and 3.8 contain selected results of some of the optimization and simulation models, respectively, that will be reviewed in this chapter.

The objective of Van Arendonk's stochastic dynamic programming model is to maximize the expected present value of the monthly net revenue stream for a replacement heifer and her replacements over a 15 year planning horizon. The control variables are the replacement and insemination decisions. The model operates recursively, starting in the final month of the planning horizon, where the terminal value of animals was carcass (salvage) value. The optimization then proceeds backwards, at each month making the decision to replace or keep the animal based on which decision coupled with the respective subsequent optimum decisions yielded the greater expected value of net revenue. In considering the insemination decision, the probability of conception, cost of insemination, and expected net revenues calculated from milk and fat yields, calf values, carcass values, feed costs, health costs and housing costs (Van Arendonk, 1985) were considered. The model has been updated numerous times.

One collaboration was with Rogers and McDaniel (1988) to adjust the model to US dairy conditions. Briefly, the state variables and the number of levels of each variable were:

lactation number	12
stage of lactation	16
milk yield in past lactations	15
milk yield in present lactations	15
time of conception	6

Cows were considered for insemination in months 2 through 7 of lactation. Voluntary replacement was defined as replacement due to the decision making of the model. Involuntary culling reasons included mastitis, other health disorders, milking speed, temperament, pendulous udders, other husbandry related disorders, and death. The monthly probabilities of involuntary culling and associated economic costs by parity were exogenous to the model. The proportion of involuntary culling ranged from 20% in the first month of lactation to 4% in the last month. First lactation animals had the lowest probability of involuntary culling at 10%, while the parity 12 animals had a 30% chance. Cows were not considered for voluntary culling prior to the second month of lactation, and sale of an animal was not necessarily concurrent with the decision to cull the animal.

The repeatability of milk yield was assumed to be 0.55 for adjacent records and 0.50 for those two years apart. Conception rates were assumed to be independent of the previous calving interval and milk yield in the previous and present lactation. The marginal probabilities of conception decreased with increased parity and increased from month two to five of a given parity before declining slightly in the sixth and seventh months of lactation. For example the marginal probabilities of conception in the second and fifth month for a first parity animal are respectively 0.30 and 0.52, whereas the comparable numbers for a fifth parity animal are 0.29 and 0.49. Genetic improvement in first lactation cows was assumed to be $10/year. Improvement in older cows was calculated based on a 0.9 genetic correlation between first and later lactations, and age factors for milk yield.

The annual discount rate of 4% was used in determination of the expected net present value from a replacement heifer over the 15 year planning horizon, in earlier use of the model the optimum replacement policy was not influenced by a reduction in the planning horizon from 20 to 15 years (Van Arendonk, and Dijkhuizen, 1985). This present value was then converted to a yearly annuity for purposes of comparison among various management alternatives. Strictly speaking, reinvestment at the discount rate and an infinite time horizon are required for an annuity to be an appropriate measure of profitability (Robison, and Barry, 1996). Prices and parameters were chosen to be

representative of American Holstein dairy farms in 1986. Milk price was $24.64/100 kg with a $0.3696 fat differential. The base price of a mixed ration with 16% crude protein and 1.72 Mcal NE_l/kg DM was $0.084/Mcal NE_l. Energy requirement, after accounting for mobilization of body reserves, limited DMI. Average 305 d mature equivalent milk yield was 8000 kg with 3.6% fat test. Price per unit of semen was $13.

All calves were sold at one week of age, with heifers being valued at $75 more than bull calves. Carcass values were calculated from body weight, dressing percentage (46.4 to 49.2%), and price per kilogram of carcass weight ($1.793 to $3.032) and varied from $571 to $521. An unlimited supply of freshening 24 month old heifers, at a real cost of $1100 each, was assumed.

Of the 259,200 potential combination of state variables, profit after conception was calculated for 14,940 combinations and future profitability for 174,300 (Van Arendonk, 1988). Cost of management and labor for milk harvest were not included.

The optimum average culling rate was 25.1% (8.6% voluntary and 16.5% involuntary), which was lower than the estimated national average of 31%. Following first lactation, where 35% of replacements were voluntary, the marginal voluntary culling rate declined until the sixth lactation. Only high producing cows were kept for many lactations because otherwise the net revenues from the lower producing cows were sufficient to compensate for the decline in fertility and increased probability of involuntary culling. The optimum policy resulted in an average calving interval of 377 days, with 38% of the cows having intervals of greater than 13 months. Table 3.1 is a reproduction from the paper.

Table 3.1. Production, as a percent of the mean production for that parity, below which insemination would not be optimum by parity for the base scenario (8000 kg ME).

	Month in lactation		
Parity	3	5	7
1	78	82	86
2	78	82	86
3	78	82	90
4	82	86	90
5	86	90	94
6	90	94	102
7	94	98	106
8	98	106	110
9	106	110	118
10	110	118	126
11	122	126	...
12	...	...	...

Feed costs represented 42% of the value of milk yield for an average cow, and marginal feed costs represented about 24% of any additional amount of milk yield for a cow.

Table 3.2. Selected results of sensitivity analysis of altering base[1] scenario assumptions on the optimum culling policy.

Scenario	Yearly Annuity ($)	Culling Rates (%) Overall	Involuntary	Voluntary
Base	443	25.1	16.5	8.6
9600 kg ME	732	26.8	16.3	10.5
$19.71/100 kg milk	73	23.5	16.8	6.7
$29.57/100 kg milk	814	27.6	16.2	11.4
$0.067/Mcal NE_l	599	25.8	16.4	9.4
$0.101/Mcal NE_l	286	24.8	16.6	8.2
$1000/freshening heifer	476	28.0	16.0	13.0
$1200/freshening heifer	412	23.7	16.8	6.9

[1] 8000 kg ME, $24.64/100 kg milk, $0.084/Mcal NE_l, $1100/freshening heifer.

The effect of the sensitivity analysis on the yearly annuity is as expected with increased output or output price, or decreased input prices being associated with increased profitability and vice versa. The overall optimum culling rates maintain a narrow range and never attain the estimated current national average.

The unexpected increased voluntary culling rate associated with increased milk production is due to an increased absolute difference in production between younger and more mature animals as a result of estimating production on a proportional basis relative to the mature equivalent (Rogers, Van Arendonk, et al., 1988b). Nevertheless, the difference in the optimal policy is minimal. This insensitivity of optimal policy to herd level of production is expected, contrary to conventional wisdom, by simply reflecting on the earlier replacement models. The choice is between the current animal and the replacement, increasing the production and profit of each has no effect on the relative difference and therefore little effect on the replacement policy.

Similarly, altering milk price or feed cost had little effect on the optimal policy. In contrast, the culling rates were highly sensitive to the cost of the replacement. Lowering heifer costs by 9% increased voluntary culling in all lactations, but none more so than the doubling in first lactation.

The authors concluded "culling rates in the US may be higher than optimum. Our model suggests they should be near 25% in most situations. Perhaps farmers who manage well could eliminate rearing of replacements with undesirable pedigrees or purchase less replacements and improve their profits....increased culling to improve yield per cow may not be justified due to an increased number of replacements needed."

These results are consistent with the optimum policies generated by the same model under Dutch conditions (Van Arendonk, and Dijkhuizen, 1985) which tend to have lower beef prices, calf values and feed costs than the US. The model suggested the optimal Dutch culling rate was 28%, and sensitive to a change in the difference between heifer price and carcass value, but insensitive to changes in milk, feed or calf prices or conception rate. This study compared modeling animals with 15 levels of milk production on a mature equivalent basis versus simply assuming average production, corrected for age, of all animals. The variation in production between present and replacement cows was important when considering the effects of changes in production or price situations on optimum replacement policies. Doubling then quadrupling the rate of genetic improvement in milk revenues minus feed costs resulted in minimal effects on culling rate. This failure for genetic progress to affect the optimal policy is not surprising given Equation (2.37). For given a low rate of improvement (i.e., $h \cong 1\%$) it can be seen that the effect on the optimal life will be minimal. Also, as expected based on the explanation in Chapter 2, insemination continued longer with higher persistency of milk production but again the effect on herd life was slight.

A sister paper provides some interesting insight into the dynamics of total culling rate, involuntary and voluntary culling (Rogers, Van Arendonk, et al., 1988b). As is evident in Table 3.3, increased voluntary culling partially compensated for the lowered involuntary culling. The increase in yearly annuity associated with a lowered involuntary culling rate was due to lower net replacement costs (cost of heifer less carcass value), decreased frequency of low yielding cows and increased productive life of high yielding cows. The latter two effects were responsible for over 50% of the increase in the annuity indicating that forced reduction in the intensity of voluntary culling is a major cost associated with excessive involuntary culling. This fact is further highlighted by the last two rows of Table 3.3 where the economic loss associated with involuntary cows was removed. Of the $113 gain in yearly annuity associated with reducing the base level of involuntary culling to only infertile cows, only $32 was associated with the direct financial losses of those involuntarily culled, while the remainder was associated with reductions in net revenues due to limited opportunities for voluntary culling. Once the coefficient of variation for milk yield was made consistent, altering the involuntary culling probabilities using the 9600 ME level of production as the base had similar results as the 8000 kg ME base. Lowering involuntary culling by one cow per year in a 100-cow herd improves net revenue by approximately $750 to $900/yr (a 1.6% to 1.3% increase). This finding that lowering the involuntary culling rate resulted in a considerable increase in the annuity, and a lesser decrease in the overall culling rate due to a somewhat compensatory rise in voluntary culls, was also produced under Dutch conditions (Van Arendonk, 1985a). This suggests that management should concentrate on the lowering of involuntary culls as a method of achieving lower culling rates. As with any model, prior to embracing recommendations the model and its assumptions must be compared with reality. The evaluation of whether management should focus on decreasing involuntary culls will be postponed until a general discussion of voluntary and involuntary culling which follows later in the chapter.

Table 3.3. Resulting yearly annuity and culling rates when average ME was 8000 kg, including and not including the financial losses[1] associated with cows involuntarily culled.

Percent of base involuntary culling probabilities	Yearly Annuity ($)	Culling Rates (%) Overall	Involuntary	Voluntary
120	420	27.2	19.4	7.8
100	443	25.1	16.5	8.6
80	465	23.4	13.6	9.8
0	556	16.8	1.6[2]	15.2
100[1]	475	24.5	16.8	7.7
80[1]	492	23.9	13.8	9.1

1. All financial losses from reduced carcass value, idle production factors, lowered milk yield prior to disposal, and health costs associated with involuntary culling.
2. Cows still not pregnant at the eighth month of lactation.

The results of the model appear valid, and the general conclusion that the optimal culling rate should be under 30% for US conditions can be made, and the potential for sizable economic loss is present due to non-optimal culling policy.

Van Arendonk (1987) subsequently investigated the effects of ignoring variation of variables in the optimal culling rate as determined by his dynamic programming model. The base culling rate was 23.8%. The resulting optimal culling rates, in parentheses, follow the respective omitted variables. It was determined that excluding the variation in carcass value due to age, stage of lactation, and reason for replacement (25.1%), or ignoring the feed cost (27.8%) variation increased culling especially among the first lactation animals. The results were similar when the objective was maximization of milk revenues (26%), or the scenario where all cows were inseminated and kept if they conceived (16.8%).

Congleton and King (1985) estimated future incomes from an extremely sophisticated biological simulation model (Congleton, 1984) which only assumed culling at 180 days if open and a maximum of 4 services. All other culling reasons were simulated.

Voluntary culling occurred if the NPV of a heifer was greater than that of the cow currently in the herd. The NPV of a cow was the sum of her own discounted cash flows for the time horizon weighted by their probability of occurrence, and the sum of her potential replacements' discounted cash flows for the time horizon weighted by their probability of occurrence. Similarly, the NPV of the potential replacement heifer included her and all her potential replacements' weighted discounted cash flows. Using a dynamic program they found relatively few cows (21%) were culled compared to most commercial dairy herds. Their results suggest culling of first lactation animals should be limited to that due to involuntary culling. They also found that increasing the time horizon beyond two calving intervals increased the culling rate to its maximum of 21% at the maximum six lactation time horizon.

A simple dynamic program of Irish conditions (Killen and Kearney, 1978) included a twenty-year time horizon, no variation in production within a lactation was modeled and culling was assumed to occur prior to the initiation of a lactation. The optimal culling rate was found to vary from 1/5 to 1/6 as the price of milk, level of herd production, and involuntary culling rates were varied to represent conditions in the industry over the previous two decades.

A Canadian dynamic programming culling model also illustrated the effects of model time horizon, as well as, modeling animal weight (Stewart, Burnside et al., 1977). The state variables consisted of lactation number (7 levels), estimated 305-day milk yield (11 levels), average 305-day milk fat % (7 levels), and body weight (5 levels). A 12 month calving interval was used, culling occurred at 60 days post partum. The probability of

culling and some of the transition probabilities were based on Canadian cow records. Annual genetic improvement of milk production of 0.9% per year was included.

Heavier cows required higher milk production to be kept. Increasing milk price or decreasing feed price resulted in more intensive culling of lighter cows, although subsequently this effect was deemed to have such a minor impact on profit that it could be omitted from the model. In contrast to most other studies the optimal policies were insensitive to net cost of replacement (replacement cost less salvage value). Use of a 5 year rather than 10 year time horizon resulted in more intensive culling. Altering the feeding regimen from corn silage and hay providing the roughage to haylage as the sole source altered the optimal policy for only 6% of the states. The actual optimal culling rates were not mentioned.

A subsequent paper (Stewart, Burnside et al., 1978) also found that the policies did not change when the effect of genetic improvement in production was removed. The simplification of an average replacement heifer was overcome by incorporating a probability distribution of possible replacement heifers into the model. The probabilities of weight, milk production and fat test of the heifer were included. The authors found that the specification of an average replacement tends to underestimate the replacement value, but the effect appeared minor as the optimal policies changed for only 3.3% of the states.

This issue of a standard replacement is not new. Ben-Ari et al. (1983) recognized that unlike traditional replacement theory heifers cannot be considered standard replacement items. Kristensen (1987), in his dynamic programming replacement model for dairy cows, incorporated this concept by including the breeding value of the sire as a state variable. The advantage of using each parents' genetic information was suggested. The model predicted heavier culling among animals of lower genetic value.

There is general agreement among the optimization models that the optimal culling rate is between 20 to 30% of the herd. This assumes that culls are sold for beef and not dairy. This suggests that most herds will not need all their potential replacement heifers. The optimal culling rates were sensitive to the relative cost of a replacement versus that of the culled animal, but was not sensitive to most other parameters including genetic trend or herd level of milk production. The greatest economic gain from reduced culling is lowering that due to involuntary reasons.

Initially it may be curious that all the models, except for Congleton and King (1985), assumed a base rate of involuntary culling when in fact the purpose of the models was to optimize culling decisions. This assumption is pragmatic. While all culling has an economic component, in order to optimize involuntary culling it must be modeled. The modeling of the reasons for involuntary culling require a very detailed, data intensive model. For example, consider the disease of mastitis. To incorporate this disease occurrence one would need to know the transition probability from each state vector to all the potential state vectors. The later would reflect the varying severity and duration of potential mastitis episodes. While survival analysis can assist with determination of the transition probabilities the economic impacts in terms of treatment costs are also required. While dynamic optimization models including the stochastic occurrence and optimal culling decisions related to single diseases such as mastitis are starting to be developed (Houben, Huirne, et al., 1994) they are not very prevalent. The biological and economic data required for these models are generally not available. The issue of voluntary versus involuntary culling as it applies to actual dairy farms will be addressed further in the chapter.

The model of Rogers et al. did not include the effect of seasonality. Two models suggest the potential impact of season on the optimal culling policies. Under Dutch conditions (Van Arendonk, 1986) the optimum policy for inseminating and replacing cows was greatly affected by the seasonal variation in production, and to a much smaller extent, in calf prices and feed costs. Variation in prices of replacement heifers and salvage animals, as well as, conception rate played a minimal role in causing the new policies under the effect of seasonality. Under southern US conditions (DeLorenzo, Spreen et al., 1992) seasonality of milk price, replacement cost, and conception rate were included in the model. The resulting decision rules were significantly different than those generated without seasonality. Use of the new policies resulted in monthly milk over and underproduction that mimicked that actually found in Florida. The seasonal effects on conception rates and lactational performance are the main factors determining the seasonal pattern and optimal culling rate nearer 40% (McCullough and DeLorenzo, 1996).

Another management problem which should ideally be solved with the replacement problem is the optimal feeding regime (Kennedy, 1986). If feed prices are not stochastic, the optimum ration need only be optimized once for each possible state vector.

The dairy dynamic programming models only considered optimization of insemination and culling policy ignoring the effects of other management decisions. Similar simplified models are the norm in beef cow replacement as well. A recent beef cow replacement paper (Frasier and Pfeiffer, 1994), however, also included the decisions of breeding season length and winter nutrition. A sensitivity analysis of long-term capital gains exclusions, feed costs, cost of replacement females and returns from weaned calves and cull cows, showed very stable feeding and breeding policies. Similar to the results from the dairy models, the optimal culling policy varied drastically with a unilateral change in the cost of replacement, while simultaneous changes in the prices for both cows and calves had little effect on the policy. The study also highlighted the effects the relative easing of capital gains taxes can have on optimal replacement, since the returns from home-raised replacements culled from the beef herd are considered capital gains.

Tronstad and Gum (1994) in their study, cited the use of stochastic rather than deterministic livestock prices as the reason that their model of beef cow culling found it may be profitable to keep nonpregnant cows in contrast to previous models.

Congleton (1988b) considered the effect of uncertainty of future production and milk price state variables in estimation of future income. It is possible that the optimal decisions will vary with the realized future state. The example used is that with a drop in future milk price the mature cow, with higher production in the next few lactations, would be preferred over the younger animal. In a highly simplified analysis the advantage of increased predictive value of animal performance far outweighed the benefit of more accurate prediction of future milk prices.

Another problem with dynamic programming is that unless producers have the capacity to run these programs for their specific situation, the ability to apply the results is questionable. For McCullough and DeLorenzo (1996) showed that while varying individual inputs rarely caused meaningful changes in optimal policy, varying multiple inputs consistently resulted in meaningful changes. While with the advancement of computer technology, repeating the computer runs is not as great an impediment as it was a decade ago, this approach still provides little into insight into the reasons for the culling decisions or the cost of not following the recommendations. Tronstad and Gum (1994) used Classification and Regression Trees (CART), to provide advice on culling decisions. The process generates two binary decision trees, one each for open and pregnant cows, which provides an easy to understand set of culling rules, with the associated cost of ignoring the recommended action. While the percent of animals correctly classified by CART, as compared to the DP, at the terminal node of the decision tree was as low as 64%, the overall accuracy of CART was 95% and 99.5% of correctly classified "open" and "pregnant" cows respectively. Using the decision rule of CART yielded over 99% of the expected returns from the DP. The use of CART in conjunction with spring and fall calving increased expected wealth by 10% over the simple rule of thumb of culling all open cows in the fall with spring-only calving.

Finally, a difficulty with optimization models is that only a point estimate of the returns from the policy is given, with no indication of the variation in the potential returns. Repeated simulation is a method to evaluate the distribution.

This is an appropriate place to make the distinction between simulation and optimization. Notice that all optimization models contain simulations. Generally a mathematical model is being used to simulate some phenomenon (i.e., the profits generated by a Holstein cow in her third lactation.) However not all simulation models are able to optimize. While this may seem so obvious as to be trivial, it is often overlooked. Some of the confusion may be due simply to that fact that computers are often used to solve both types of problems.

The advantages of simulations are their ability to compare results between optimal policies and with those currently used, validate dynamic models and describe the transition to optimized structure of the herd (DeLorenzo, Spreen et al., 1992).

Consider a recent (1996) paper by Galton, Everett, Van Amburgh and Dauman concerning the economics of extended calving intervals with use of bST. The result of this study is "a 18.0 month calving interval is recommended for improved animal performance and herd profitability". The simple fact that this was a simulation, not an optimization study, appears to have escaped the authors. At best, the correct conclusion from this study is that under the conditions and assumptions of the simulation the bST cows with an extended calving interval had greater returns than the traditional management group with a 13.2 month calving interval. (Even this result is suspect due to a number of fundamental flaws in the study design, including, but not limited to, ignoring the increased physical and economic pressure bST treatment places on the milking cow facilities, using a single point estimate for length of herd life rather than a distribution, assuming different culling rates between treatments, and a time horizon bias that favors the bST animals.) Assuming that bST does impart an economic benefit, it is entirely possible that a 17, 19, 20, or even a 13.2 month calving interval is superior to one of 18 months. Consider that if the goal of the study was not simply to promote bST but rather increase producer profitability there are a myriad of other potential management changes that could be considered, some of may have

formation a

ately c

simplist

, but i

Other si

the rol

among

prod

body we

over the

ation be

50, 70,

condition

decision

rate

, 198

returns

A deterr

insight

involun

which may have less associated risk. Such a biased and misinterpreted study only provides disinformation and misleads those without the expertise or time to assess the study. Unfortunately confusion over simulation and optimization and the correct interpretation of equally simplistic studies are not uncommon. Obviously the appeal of such models is their simplicity, but it is not clear what is being accomplished by oversimplifying a complex problem.

Other simulations have been more fruitful. Dijkhuizen and Stelwagen (1988) revealed the robustness of the culling guide generated by Van Arendonk's dynamic program among varied reproductive efficiencies. Their simulation model included stochastic production, heat detection, conception rate and involuntary disposal. Feed intake, body weight and prices were deterministic. The culling guide produced a higher return over the policies of retaining all cows which conceive within 240 days, or no insemination beyond 160 days of lactation, under all combinations of heat detection rate (HDR) (50, 70, 90%) and conception rate (40, 60, 80%). But the difference was small under conditions of poor reproductive performance due to the limited ability to make culling decisions. Updating the guide which was created assuming 70% HDR and average conception rates per lactation that varied from 0.47 to 0.63 (Van Arendonk and Dijkhuizen, 1985) to reflect the poor reproductive performance had minimal improvement on the returns.

A deterministic model of Irish conditions (Allaire, and Cunningham, 1980) provides insight into the effects of voluntary culling based on level of production. Three levels of involuntary culling (15, 20, 25%), and three levels of replacement cost relative to beef price (0.94, 1.26, 1.57) were simulated for 16 years. Maximum allowable levels of voluntary culling were 26, 19, and 11%, for involuntary levels of 15, 20, and 25%, respectively. Replacement occurred at the beginning of the following year. Results presented are for the final year. The greater the involuntary culling level the lower the annual milk production per cow, as mature high producing cows are replaced with lower producing heifers. For a given level of involuntary culling the increase in milk production per cow rises then plateaus and may then decrease as voluntary culling increased. Also as the involuntary culling decreased, the voluntary level to maximize production increased.

Table 3.4. Simulated level of total culling to maximize objectives under varying conditions.

	Level of involuntary culling		
	15%	20%	25%
Maximum culling	42%	39%	36%
Culling to maximize milk production	33%	38%	36%[1]
Culling to maximize annual NPV per cow, low rearing costs[2]	42%[1]	39%[1]	36%[1]
Culling to maximize annual NPV per cow, medium rearing cost[3]	30%	34%	36%[1]
Culling to maximize annual NPV per cow, high rearing cost[4]	23%	27%	30%

1. Maximum allowable.
2. Ratio of heifer cost to cull value of 0.94.
3. Ratio of heifer cost to cull value of 1.26.
4. Ratio of heifer cost to cull value of 1.57.

The authors conclude that the intensity of voluntary culling should at most be 3 to 8% in addition to the involuntary replacement rate when the market price for a replacement exceeds 1.5 times their cull value. Under the high rearing cost, the maximum culling rate lowered NPV by about 10% from its maximum.

Allaire (1981) subsequently updated the model to study the effect of genetic improvement. Six policies were modeled:

Replacements enter with no change in genetic merit for yield.

(1) Involuntary culling only at a rate of 20%.

(2) Voluntary culling above the 20% base rate of involuntary culling.

Replacements enter at genetic merit for yield reflecting expected response from selection.

(3) Voluntary culling, replacements from random sample of dams, no change in sire merit.

(4) Voluntary culling, replacements from best of potential dams, no change in sire merit.

(5) Voluntary culling, replacements from best of potential dams, sire merit increases 0.5% annually (≅ 28 kg milk).

(6) Voluntary culling, replacements from best of potential dams, sire merit increases 1.0% annually (≅ 56 kg milk).

All voluntary culling was for milk yield based on estimated real producing ability. Potential replacements were selected based on their estimated breeding values using only dam information, as for a given policy sire merit was constant. Replacement rates of 20 to 39% were considered over a 20 year time horizon. The rearing cost of a replacement was 1.3 times her beef value. All replacements entered the milking herd at 2 years of age, all culling occurred at the end of each year.

The 30 to 35% replacement rate which maximized milk production exceeded that of the 25 to 27% rate which maximized NPV per cow. A 40% culling rate lowered NPV by approximately 15 to 20% versus the rate to maximize NPV per cow. The greater the genetic progress the higher the NPV, but "cow replacement rates necessary to maximize NPV of the herd are insensitive to the influence of a continuously genetic trend in milk between 0% to 1.0% per year". The results appeared insensitive to the relative cost of replacements. The author noted that in 1979, Ohio Dairy Herd Improvement herds consisted of 30% first lactation cows.

The net marginal revenue model reviewed in Chapter 2, was used to compare various strategies of selecting cows during the first lactation (Korver and Renkema, 1979). Given a particular strategy, the culling decisions were optimized. It was assumed that potentially 40% of the herd could annually be replaced by heifers. Four culling policies were considered:(1) Ill-health or insufficient production are reasons for disposal. (2) Removal only for ill-health. (3) Low producing cows are removed at the highest possible rate during the first lactation. (4) Similar to 3 but no calves are raised from cows that leave the herd due to low production. Policy one was considered to reflect typical Dutch culling policy. The effect of the culling policies on future genetics of the herd was ignored. The results are presented in Table 3.5.

Table 3.5. Results of Dutch simulation of four selection policies during first lactation.

	Culling Policy			
	1	2	3	4
Percent of Available Heifers Used	58%	51%	100%	85%
Ave. Culling rate	23%	20%	40%	34%
Ave. Productive Life (lact.)	4.3	4.9	3.5	3.9
Ave. Annual Earned Income per Cow (Dfl)	1363	1351	1246	1315
Percent Increase in Income Over Policy 3	9.4%	8.4%	0.0%	5.5%

While the authors correctly note that these results do not imply policy one is optimal (since this is a simulation study), the disadvantage of heavy culling policies (3 and 4) is apparent, at least under the assumptions and conditions of the study. The higher milk production of the older animals is insufficient to compensate for the higher cost of rearing young stock and the unfavorable age distribution (more younger animals in the herd). Also the critical production levels, relative to the average first lactation production for the herd, below which first lactation cows should be culled were calculated. Using this information and assuming a normal distribution of milk production with a 15% standard deviation of phenotypic intra-herd milk yield, about 20% of the first lactation cows should be culled based on production.

The authors suggested that instead of assuming replacement heifers have average productive capacity, they could be ranked according to their estimated breeding value. This ranking, as well as, lower rearing cost or higher salvage values will increase the culling of first lactation cows.

Pearson and Freeman (1973) used a deterministic model to simulate the effect of four culling systems on the milk production of US commercial Holstein operations. The four cow selection systems were: (1) only enough females were raised to replace involuntary losses, estimated breeding value (EBV) was used to select those calves to raise; (2) selection on EBV based only on the cow's own record; (3) selection of the youngest cows; (4) all voluntary selection during the first lactation based on EBV. Under systems 2 to 4 all females were allowed to freshen. Each herd was simulated for twenty calvings with 13 month calving intervals. Profit functions were quite complete including returns from milk, bull and cull heifer calves, cull dairy cows and those sold for dairy, fixed costs per herd, interest costs for livestock, expenses associated with artificial insemination, feed, rearing of heifers and an additional cost of milk for higher production based on rolling herd average. While the study included the effects of different herd sizes and levels of herd production, the rank of selection systems remained unchanged.

Table 3.6. Results of US simulation of four selection policies[1].

	Culling Policy			
	1	2	3	4
Genetic gain/period (kg)	56	57	53	56
Actual production gain/period (kg)	67	68	62	69
Ave. Profit with low rearing costs[2,3]	\$211	\$229	\$192	\$218
Ave. profit with medium rearing costs[2,4]	\$189	\$188	\$151	\$218
Ave. profit with high rearing costs[2,5]	\$167	\$147	\$110	\$136

1. Using intense sire selection (equivalent to PD>363 kg).
2. Salvage values held constant.
3. Estimated ratio of replacement cost to salvage value of cull is 1.3.
4. Estimated ratio of replacement cost to salvage value of cull is 1.5.
5. Estimated ratio of replacement cost to salvage value of cull is 1.9.

Selection can be characterized as: (1) high intensity low accuracy; (2) accuracy increases with each lactation, but intensity decreases; (3) selection pressure is zero; (4) intensity and accuracy of selection are high. Policies 1, 2, and 4 result in genetic improvement at nearly the same rate. The higher production of 2 and 4 is due to the positive environmental effect resulting from selecting on the cows' own records. With sire selection intensity equal to the 2% theoretical rate of genetic improvement, policy 1 had the slowest rate of genetic and phenotypic improvement.

System 1 was more profitable than expected by its lower production due to the smaller cost of replacements. The advantage of system 1 increased with the cost of replacements as this overshadowed the increased returns associated with higher milk production. The low and medium rearing costs appear inappropriate for application to the present day. The effect of the sire selection system (2% genetic gain/year, sires with PD > 363 kg, or sires 91 kg ≥ PD ≥ -91 kg) on profits were greater than that of the cow selection. The authors note that if calf selection is required, some selection should be reserved to remove the mistakes made earlier and allow for an adverse sex ratio.

A simulation model estimated milk yield based on dams' and sires' breeding values, the effects of health problems including reproductive disorders and mastitis were included (Congleton and King, 1984). It is not clear why culling was based on cost per unit of production during present lactation including the preceding dry period. This basis for the culling decision would likely be biased against first lactation animals and increase their culling. Pregnant cows could be culled at the end of lactation, while infertile cows could be culled after 305 days of lactation. The time horizon was 30 years. Using this questionable decision criterion the optimum culling rate was 20% to maximize discounted returns. Increasing herd life from 3.4 lactations to 4.2 linearly increased discounted income by $518.96 per lactation.

Congleton (1988a) conducted a simulation on the effect of culling on predicted production or income on herd profitability and genetic trend. Future production was estimated using previous performance and repeatability of complete lactation production. Culling on profitability was a two stage affair. First the herd was screened with Kuiper's short-term MaxAMR to identify potential cull candidates (estimate returns from cows through the next ten months of the next lactation). These animals were then ranked on long-term income using cow attrition data. The order of the use of the two screens seems intuitively backwards.

One of these two criteria were used to select a cow for culling when a heifer freshened (20% as many bred heifers as milking cows were targeted for cow replacement resulting in a constant herd life regardless of culling method, heifers were selected on breeding values). Since this would entail culling many cows after breeding, the third

culling option was not to breed the lowest producing cows and remove them when maximum income had been produced for the lactation. Herds were simulated with varying intensities of nonbreeding on production and then culled on either production or income. A 20 year time horizon was simulated.

Culling on projected income increased herd income by 4.3 to 4.8%, and decreased the genetic gain to 65.3 to 86.2% of that observed with cow selection on production. (Breeding only the highest producing cows increased herd income by 4.3% to 5.7%).

Even the simulation model of Kuipers (1980) which, as discussed in Chapter 2, underestimated the value of younger animals by using an inappropriately short time horizon, found selling 30% of the heifer calves, rather than raising them all, gave higher returns to management of about $50/cow/year for all selection methods. This simulation model considered culling of cows based on their genetic index.

The results of these simulation studies agree with those of the dynamic programming models. The cost of replacements is such that maximizing genetic progress or milk production through increased culling is not economical. The realization that the use of all replacements does not maximize profit is comforting from an economic perspective. These models are simply portraying the fundamental, at least under certainty, decision rule of profit maximization which appeared trivial in the previous chapter since it was so obvious. Namely continue to use more of an input (heifers) until the cost of doing so exceeds the return (i.e. use an input until the marginal input cost exceeds the marginal value product).

Marginal value product is the product of the physical marginal product and the additional gross revenue generated by the employment of the extra unit of input. In a competitive market the cost of inputs and outputs is constant, so the marginal physical product essentially drives the profit maximizing decision. These studies then suggest that as more and more heifers are used their returns decrease, this is simply a statement of the well accepted theory of diminishing marginal productivity of inputs.

By focusing on using all potential replacements the industry is reflecting its output bias. To maximize output, inputs will be used as long as they contribute anything to output (i.e. as long as the marginal physical product is greater than zero). Obviously this decision rule will result in more use of the input than the profit maximizing level of input. This is supported by the studies which found that the replacement rate which maximizes milk production exceeds that which maximized economic returns. The two levels of input will be equal only if the price of the input is zero. The traditional biological focus of herd management then has implicitly assumed zero cost of inputs.

The optimal culling rates are sensitive to the cost of the replacements relative to the salvage value of the cull. However, if replacement rates are excessive it appears the greatest gain to be made is through lowering involuntary culling.

Prior to considering the actual culling rates in Michigan and whether all replacements are being used, a more careful assessment of the definition of culling rate is required. Up until now culling rate has been loosely defined. Following this, the partitioning of culling into voluntary and involuntary will be critically considered.

Many of the studies mentioned previously focused on average productive herd life as a measure of culling rate. Productive herd life is the average length of time an animal spends in the herd starting with the first calving. Average productive life can also be considered the expectation of life (Rendel and Robertson, 1950). They state "if one-quarter of the population is replaced each year, then obviously the expectation of life is 4 years."

The definition of culling rate (*C*) is not explicitly defined in the optimization papers. Rather the studies report the culling rate (i.e. 25%) and the optimal average herd life (i.e. 47.8 months). The average herd life (*H*) is actually the average productive life of the animals in the herd. This is calculated as the summation of the proportion of the herd in each respective lactation, assuming the distribution of ages is in a steady state.

$$H = \sum_{i=1}^{m} n_i P_i \qquad \textbf{(3. 33)}$$

where n_i = the length of lactation number i in months,

P_i = proportion of herd in lactation i.

P_i is equal to the marginal culling rate for lactation i (C_i) multiplied by those animals that have not been previously culled.

Equation (3.33) can then be restated as

$$H = 1C_1 + 2C_2(1 - C_1) + 3C_3(1 - C_1)(1 - C_2) + \dots\dots\dots + nC_n \prod_{i=1}^{n-1}(1 - C_i) \qquad \textbf{(3. 34)}$$

To calculate the average culling rate based on the average herd life, the marginal propensity of culling in each lactation is assumed equal.

$$H = 1C + 2C(1-C) + 3C(1-C)^2 + 4C(1-C)^3 + \dots\dots\dots + nC(1-C)^{n-1} \quad \textbf{(3.35)}$$

$$H = \sum_{n=1}^{\infty} nC(1-C)^{n-1} \quad \textbf{(3.36)}$$

Equation (3.35) reveals the sum to be an infinite linear geometric series. Using the notation of Robison and Barry (1996, pg 95) $R = 0$, and $\alpha = r = C/(1-C)$ so $(1+r) =$

$(1-C)^{-1}$.

$$H(1-C) = 1C(1-C) + 2C(1-C)^2 + 3C(1-C)^3 + \dots\dots\dots + nC(1-C)^n \quad \textbf{(3.37)}$$

Subtracting (3.37) from (3.35),

$$H - H(1-C) = C + C(1-C) + C(1-C)^2 + C(1-C)^3 + \dots\dots\dots + C(1-C)^{n-1}$$

$$HC = C\left[1 + (1-C) + (1-C)^2 + (1-C)^3 + \dots\dots\dots + (1-C)^{n-1}\right] - nC(1-C)^n$$

$$= \frac{C}{1-C}\left[(1-C) + (1-C)^2 + (1-C)^3 + \dots\dots\dots + (1-C)^n\right] - nC(1-C)^n$$

Dividing by C, and substituting b for *(1-C)*,

$$\mathrm{H} = \frac{1}{b}\left[b + b^2 + b^3 + \dots\dots + b^n\right] - nb^n$$

$$= \frac{1}{b} US_o \left[\frac{1-b}{b}, n \right] - nb^n$$

where $\frac{1-b}{b} = \frac{1-(1-C)}{1-C} = \frac{1-1+C}{1-C} = \frac{C}{1-C}$

As $n \Rightarrow \infty$, $US_o(1-b/b,n) \Rightarrow b/1-b$.

Since $C > 0$, $(1-C) < 1$, as $n \Rightarrow \infty$, $n(1-C)^n \Rightarrow 0$. Therefore as the potential number of lacations approaches infinity, the value of the series converges to

$$H = \frac{1}{1-C} \bullet \frac{1-C}{C} = \frac{1}{C}$$

If the average culling rate is 25%, the average herd life should then equal the inverse of the culling rate which is 4.0 years or 48 months which is that predicted. The 47.8 in the article differs slightly because not all culling is done at the end of a year and the marginal culling rates per lactation are not all equal. The age structure of the herd is a function of the marginal culling rate each lactation or year, so herds with identical culling rates can vary in average age (Sattler and Dentine, 1989). This then establishes the definition of culling rate to be the number of animals culled in a year divided by the average herd size.

The average culling rate among Michigan herds on official DHIA test was 37% in 1996 (Michigan Dairy Herd Improvement Association, 1996). However, few Michigan DHIA herds will have this high a number reported on their production records. For Michigan DHIA calculates the culling rate as the number culled divided by all the animals that were present in the herd at anytime through the year (Ferris, 1987), this is commonly referred to as the herd turnover rate (*TO*).

$$TO = \frac{S}{I+S} \quad \text{versus} \quad C = \frac{S}{I}$$

where S is the number of animals sold, and I is the average herd inventory.

$$\frac{1}{TO} - 1 = \frac{I+S}{S} - 1 \text{ or } \frac{1 - TO}{TO} = \frac{I}{S}$$

Therefore, the culling rate, C, equals $\frac{TO}{1 - TO}$

So a herd turnover of 30% is equivalent to a culling rate of 43%. It is likely that Rogers et al. (1988a) in stating the average US culling rate was 31%, and thus not significantly different than their determined optimal culling rate, assumed this was directly comparable to their definition of culling rate (i.e. C). Whereas, this 31% truely represented the herd turnover, with a resulting C of 45% which is considerably higher than the optimal culling rate as determined by the dynamic program.

The proportion of first lactation animals in a herd is not an appropriate indicator of herd culling. For consider that while at any one time only a single first calf heifer may be in the herd, one could actually have culled 52 in a year if they were, on average, kept in the herd for only 1 week. It does, however, represent the minimum culling rate.

Evidence from Ontario (Ontario DHI Corporation, 1997) suggests the distribution of culling rates across herds is symmetrical. So it is quite possible that over half the Michigan dairy herds have culling rates significantly in excess of the suggested optimum.

Only a high death loss or sales of heifer calves would prevent the average producer from possessing an excessive heifer inventory. The presence of excess heifers is supported by a survey of heifer management practices of producers comprising 78% of the US dairy cow population and calf rearing population (Heinrichs, Wells et al., 1994). This survey found that only 25.8% of herds sold any heifers prior to parturition.

If producers are interested in selecting heifers to lower their culling rate, it must be recognized that this culling rate is dynamic. In making the transition to freshening fewer heifers, the culling rate among the heifer population will decrease as the number of heifer calves produced by first calf heifers drops.

Assuming herd size is constant and no heifer culling in practiced, O_0, is the number of offspring (as a percent of herd size) in period 0 due to calve in the next year, and C_0 is the excessive nonoptimal herd culling rate in period 0. In period one a level of heifer culling, C_{h1}, is required to achieve the desired adult herd culling rate C^* in period 1 (i.e., $C_1 = C^*$)

$$C^* = O_o\left(1 - C_{h_1}\right)$$

$$C_{h_1} = 1 - \frac{C^*}{O_o}$$

As a result of heifer culling prior to their calving, there will be a smaller subsequent heifer population, O_1.

$$O_1 = O_o - \left(\frac{C_0 - C^*}{2}\right) \qquad \textbf{(3. 38)}$$

The denominator accounts for the sex ratio.

Because after the first period the culling rate in the adult herd equals the optimal rate, if the optimal herd culling rate remains constant, so too will the number of offspring and the heifer culling rate.

$$O_2 = O_1 - \left(\frac{C_1 - C^*}{2}\right) = O_1 - \left(\frac{C^* - C^*}{2}\right) = O_1, \quad \therefore O_1 = O_2 = O_3 = \dots\dots\dots = O_n$$

$$C_{h_2} = 1 - \frac{C^*}{O_1} = 1 - \frac{C^*}{O_2} = 1 - \frac{C^*}{O_3} = \dots\dots\dots = 1 - \frac{C^*}{O_n}, \quad \therefore C_{h_2} = C_{h_3} = C_{h_4} = \dots\dots\dots = C_{h_n}$$

Heifer culling settles into a lower steady state rate after the initial period of higher culling, noting that in this context a period is the length of time for a heifer to freshen.

If $C_0 = 43\%, C^* = 30\%$ and $O_o = .5$

$$C_{h_1} = 1 - \frac{C^*}{O_o} = 1 - \frac{0.3}{0.5} = 0.40 = 40\%$$

$$O_1 = O_o - \left(\frac{C_0 - C^*}{2}\right) = 0.5 - \left(\frac{0.43 - 0.30}{2}\right) = 0.44 = 44\%$$

$$C_{h_2} = 1 - \frac{C^*}{O_1} = 1 - \frac{0.3}{0.44} = 0.32 = 32\%$$

Algebraic manipulation reveals the relationship between the initial heifer culling rate and the subsequent steady state level.

$$\frac{\left(1 - C_{h_1}\right)}{\left(1 - C_{h_2}\right)} = 1 - \left(\frac{C_0 - C^*}{2O_o}\right) \qquad \textbf{(3. 39)}$$

As discussed previously the culling models assumed a base level of involuntary culling for the pragmatic reason of not having to model the underlying biological processes. Consideration of Table 3.1 is disturbing because it suggests that many producers have been culling animals for infertility, generally accepted as an involuntary cull, when the optimal choice would have been to continue to inseminate. In analyzing culling records of Canadian dairy cows, Dohoo and Dijkhuizen (1993) determined that reproductive culling was occurring too early, and that it was economically justifiable to persist in breeding dairy cows much later into lactation than is often considered feasible. The recommendations of the optimization and simulation models were to lower culling rate preferably by decreasing the rate of involuntary culls. It would appear that this is possible, simply by reconsidering the definition of voluntary and involuntary culling.

Webster (1994) defines involuntary as "not done of one's own free will, not done by choice" or "not consciously controlled, automatic", an example given for the latter is sneezing. I propose that in the use of the term involuntary with regards to culling, the first definition reflects the accepted meaning of the adjective. However, the second definition actually portrays involuntary culling in the US dairy industry.

It is apparent that with the exception of death and absolute sterility there is no such thing as an involuntary cull. Culling is solely an economic decision. To recognize this consider the following question: given a four year old cow rated excellent in conformation, with a 30,000 lb ME would you cull her for reason X? Where reason X is any reason for which an animal may be culled. If she has mastitis or infertility, commonly cited reasons for involuntary culling, and yet you would not sell her then by definition it can not be an involuntary cull.

The recognition that culling is fundamentally an economic issue has been proposed (Fetrow, 1987). However, this distinction purely for the purpose of semantics serves little purpose. The distinction is important if the involuntary/voluntary culling paradigm is resulting in "not consciously controlled" culling decisions. Stressing the economic nature of culling decisions may be a method to encourage the dairy industry to critically evaluate their culling decisions.

The presence of producers who purchase the culls of others at beef price and then milk these cows is empirical evidence that many in the dairy industry are not correctly valuing their culls. This incorrect valuation could be due to their failure to critically evaluate their assumptions and beliefs concerning the optimal culling rate and involuntary culls.

I am troubled by my suggestion that the market is operating inefficiently by having too high a culling rate. I base this irrationality on the following arguments: as the past few chapters have highlighted, the optimal culling rate is highly complex and not intuitively obvious, furthermore its relation to profit is even less clear. The behavior of producers is consistent with bounded rationality, where knowledge of the optimal cull lies outside the bounds. Milk output is easier to measure and maximization of milk production is often used as a proxy for profit maximization. The suggestion that the higher than optimal industry culling rates are the result of attempting to maximize milk production is supported by some of the culling studies. Then we need to resolve why we never realized that at these higher culling rates the marginal value product of a heifer was less than her marginal cost. I believe this is due to overestimation of the value of genetic improvement in the heifer and her potential future offspring. As the models show, the rate of genetic improvement has little effect on the optimal culling rate and vice versa. Finally, the focus by veterinarians and animal scientists on herd health and heifer rearing has resulted in increased survivability of heifers and thereby provided the potential fuel to fire the excessive culling rates due to increased heifer pressure.

In summary, the evidence is overwhelming that the optimal culling rates for Michigan dairy producers are 20 to 30%, assuming culls are sold for salvage value. This assumption is appropriate for most commercial herds. Despite varying models with varying assumptions and conditions, as well as, different solution techniques these results are very robust. Perhaps the greatest criticism of these models, in terms of truely optimizing the culling rate, is in the widespread use of the industry involuntary culling rate as a baseline parameter. Despite the uniformity of the studies' results the industry has been slow to adopt the recommendation of lower culling rates. In order to lower the culling rate some potential replacements must be culled. The models described assumed random selection or arbitrary selection of heifers based on some genetic estimate of potential for milk production. Selected results of the culling optimization and simulation studies are presented in Tables 3.7 and 3.8, respectively. Chapter 4, will explore the issue of heifer selection.

BIBLIOGRAPHY

Allaire, F. R. 1981. Economic consequences of replacing cows with genetically improved heifers. J. Dairy Sci. 64:1985-1995.

Allaire, F. R., and E. P. Cunningham. 1980. Culling on low milk yield and its economic consequences for the dairy herd. Livest. Prod. Sci. 7:349-359.

Bellman, R. 1957. Chapter 1 (page 40) in Dynamic Programming. Princeton University Press, Princeton, New Jersey.

Ben-Ari, Y., I. Amir, and S. Sharar. 1983. Operational replacement decision models for dairy herds. J. Dairy Sci. 66:1747-1759.

Burt, O. R. 1965. Optimal replacement under risk. J. Farm Econ. 47:324-346.

Congleton, W. R. 1984. Dynamic model for combined simulation of dairy management strategies. J. Dairy Sci. 67:644-660.

Congleton, W. R. 1988a. Dairy cow culling decisions. 2. Profitability and genetic trend in herds with culling on production versus income. J. Dairy Sci. 71:1905-1915.

Congleton, W. R. 1988b. Dairy cow culling decisions. 3. Risk of culling on predicted income (An application of Bayes criterion) J. Dairy Sci. 71:1916-1925.

Congleton, W. R., and L. W. King. 1984. Profitability of dairy cow herd life. J. Dairy Sci. 67:661-674.

Congleton, W. R., and L. W. King. 1985. Culling cows on predicted future profitability with a variable planning horizon. J. Dairy Sci. 68:2970-2983.

DeLorenzo, M. A., T. H. Spreen, G. R. Bryan, D. K. Beede, and J. A. M. Van Arendonk. 1993. Optimizing model: Insemination, replacement, seasonal production, and cash flow. J. Dairy Sci. 75:885-896.

Dijkhuizen, A. A., and J. Stelwagen. 1988. An economic comparison of four insemination and culling policies in dairy herds, by method of stochastic simulation. Livest. Prod. Sci. 18:239-253.

Dohoo, I. R., and A. A. Dijkhuizen. 1993. Techniques involved in making dairy cow culling decisions. Compend Contin Educ Pract Vet 15:515-520.

Ferris, T. 1987. Michigan DHI records manual. Guide to record use, herd evaluation and explanation of report forms. Cooperative Extension Service, Michigan State University.

Fetrow, J. 1987. Culling dairy cows. Proc 20th Annu Meet AABP:102-107.

Frasier, W. M., and G. H. Pfeiffer. 1994. Optimal replacement and management policies for beef cows. Amer. J. Agr. Econ. 76:847-858.

Galton, D. M., R. Everett, M. Van Amburgh, and D. Dauman. 1996. Economics of extended calving intervals with use of bST. Proceedings 1996 NE Dairy Producers Conference.

Heinrichs, A. J., S. J. Wells, H. S. Hurd, G. W. Hill, and D. A. Dargatz. 1994. The national dairy heifer evaluation project: A profile of heifer management practices in the United States. J. Dairy Sci. 77:1548-1555.

Houben, E. H. P., R. B. M. Huirne, A. A. Dijkhuizen, and A. R. Kristensen. 1994. Optimal replacement of mastitic cows determined by a hierarchic markov process. J. Dairy Sci. 77:2975-2993.

Howard, R. A. 1960. Dynamic Programming and Markov Processes. M.I.T. Press, Cambridge, Massachusetts.

Kennedy, J. O. 1986. Chapter 7 Livestock Management, pg 194 in Dynamic Programming: Applications to Agriculture and Natural Resources. Elsevier, London.

Killen, L., and B. Kearney. 1978. Optimal dairy cow replacement policy. Ir. J. Agric. Econ. Rur. Sociol. 7:33-40.

Korver, S., and J. A. Renkema. 1979. Economic evaluation of replacement rates in dairy herds II. Selection of cows during the first lactation. Livest. Prod. Sci. 6:29-37.

Kristensen, A. R. 1987. Optimal replacement and ranking of dairy cows determined by a hierarchic process. Livest. Prod. Sci. 16:131-144.

Kuipers, A. 1980. Development and comparison of selection criteria for cows and bulls with a dairy herd simulation model. Ph.D. Thesis, Univ. Wisconsin, Madison.

McCullough, D. A., and M. DeLorenzo. 1996. Effects of price and management level on optimal replacement and insemination decisions. J. Dairy Sci. 79:242-253.

Michigan Dairy Herd Improvement Association. 1996. Annual summary of production records.

Neufeldt, V. (ed). 1994. Webster's New World Dictionary of American English, 3rd College Edition. Macmillan, New York, New York.

Ontario DHI Corporation. 1997. Provincial management centre benchmarks - 1996.

Pearson, R. E., and A. E. Freeman. 1973. Effect of female culling and age distribution of the dairy herd on profitability. J. Dairy Sci. 56:1459-1471.

Rendel, J. M., and A. Robertson. 1950. Some aspects of longevity in dairy cows. Empire Journ. of Exper. Agric. 18:49-56.

Robison, L. J., and P. J. Barry. 1996. Present Value Models and Investment Analysis. 1st ed. The Academic Page, Northport, Alabama. Ch 1, 2, 5, 8, 11, 21.

Rogers, G. W., J. A. M. Van Arendonk, and B.T. McDaniel. 1988a. Influence of production and prices on optimum culling rates and annualized net revenues. J. Dairy Sci. 71:3453-3462.

Rogers, G. W., J. A. M. Van Arendonk, and B.T. McDaniel. 1988b. Influence of involuntary culling on optimum culling rates and annualized net revenues. J. Dairy Sci. 71:3462-3469.

Sattler, C. G., and M. R. Dentline. 1989. Trends in herd age structure and the relationship with management characteristics in Wisconsin Holstein herds. J. Dairy Sci. 72:1027-1034.

Stewart, H. M., E. B. Burnside, and W. C. Pfeiffer. 1978. Optimal culling strategies for dairy cows of different breeds. J. Dairy Sci. 61:1605-1615.

Stewart, H. M., E. B. Burnside, J. W. Wilton, and W. C. Pfeiffer. 1977. A dynamic programming approach to culling decisions in commercial dairy herds. J. Dairy Sci. 60:602-617.

Tronstad, R., and R. Gum. 1994. Cow culling decisions adapted for management with CART. Amer. J. Agr. Econ. 76:237-249.

Van Arendonk, J. A. M. 1985. A model to estimate the performance, revenues and costs of dairy cows under different production and price situations. Agric. Syst. 16:157-189.

Van Arendonk, J. A. M. 1985a. Studies on the replacement policies in dairy cattle. II. Optimum policy and influence of changes in production and prices. Livest. Prod. Sci. 13:101-121.

Van Arendonk, J. A. M. 1986. Studies on the replacement policies in dairy cattle. IV. Influence of seasonal variation in performance and prices. Livest. Prod. Sci. 14:15-28.

Van Arendonk, J. A. M. 1987. Contribution of variables to the optimum policy for insemination and replacement of dairy cows. J. Anim. Breed. Genet. 104:35-43.

Van Arendonk, J. A. M. 1988. Management guides for insemination and replacement decisions. J. Dairy Sci. 71:1050-1057.

Van Arendonk, J. A. M. and A. A. Dijkhuizen. 1985. Studies on the replacement policies in dairy cattle. III. Influence of variation in reproduction and production. Livest. Prod. Sci. 13:333-349.

Table 3.7. Selected results of culling optimization studies.

Basic assumptions: constant herd size, culls sold for slaughter not dairy, random heifer selection, heifers always available.

Author, Journal	Year & Country Modeled	Time horizon (years)	State Variables	Timing of Insemination & Culling	Heifer/ Cull Cow Prices	Optimal Culling rate (Involuntary, Voluntary)	Arc Elasticities of Optimal Culling rate: Increase, Decrease
McCullough & DeLorenzo J. Dairy Sci.	1996, Florida, U.S.	5	Number of lactation 12 Stage of lactation 16 Milk yield in present lact. 15 Timing of conception 7 Month fresh 12	-monthly from 3rd - 8th month of lactation - monthly after 3rd month lact.	~2.3:1	**38%** Without seasonality **27%**	Heifer Price: -1.73, Conception Rate: -0.5, -0.5
Rogers, Van Arendonk & McDaniel J. Dairy Sci.	1988, U.S.	15	Number of lactation 12 Stage of lactation 16 Milk yield past lactations 15 Milk yield in present lact. 15 Timing of conception 6	-monthly from 2nd - 7th month of lactation - monthly after 2nd month lact.	2:1	**25%** (16.1, 8.6)	Heifer Price:-0.6, -1.2, Milk Price: 0.5, 0.3, Milk Yield: 0.3, 0.3, Involun. Cull rate[1]: 0.4, 0.3
Van Arendonk & Dijkhuizen Livestock Prod. Sci.	1985, Holland	15	Number of lactation 12 Stage of lactation 16 Milk yield past lactations 15 Milk yield in present lact. 15 Timing of conception 6	-monthly from 2nd - 7th month of lactation - monthly after 2nd month lact.	~ 0.9:1	**20%** (13.9, 6.1)	Heifer Price:-1.4, -2.2 Persistency: 0.3, 0.2 Pregnancy Rate: -0.3, -0.2
Van Arendonk Livestock Prod. Sci.	1985, Holland	20	Number of lactation 12 Stage of lactation 16 Milk yield past lactations 15 Milk yield in present lact. 15	- not explicitly modeled (12 m CI) - monthly after 2nd month lact.	~ 0.9:1	**25.5%** (21.0, 4.5)	Heifer Price: -1.0, -1.7 Milk Price: 0.5, 0.6 Milk Yield: 0.1, 0.6 Involun. Cull rate: 0.5, 0.3 Genetic Progress: 0.6, 0.5

Table 3.7. (cont'd).

Author, Journal	Year & Country Modeled	Time horizon (years)	State Variables	Timing of Insemination & Culling	Heifer/ Cull Cow Prices	Optimal Culling rate (Involuntary, Voluntary)	Arc Elasticities of Optimal Culling rate: Increase, Decrease
Congleton & King J. Dairy Sci.	1985, U.S.	10 CI	Discounted net income from simulation model adjusted for age of cow 11 Milk yield past and present lactation			**21%** (13.8, 7.2)	
Killen & Kearney Ir. J. Ag. Econ. Rur. Sociol.	1978, Ireland	20	Number of lactation 9		~ 1.4:1	**20%** (17, 3)	

[1] One less involuntary cull in a 100-cow herd increases net income by **$750 to $900/year** (a 1.6% to 1.3% increase)

Table 3.8. Selected results of culling simulation studies.

Author, Journal	Year & Country Modeled	Time horizon (years)	State Variables	Timing of Insemination & Culling	Heifer/ Cull Cow Prices	Scenarios and Results
McCullough & DeLorenzo J. Dairy Sci.	1996, Florida U.S.	10	Using model of McCullough & DeLorenzo, 1996 with 10 Timings of conception	-monthly from 3rd - 8th month of lactation - monthly after 3rd month lact.	~2.3:1	Use of optimal decisions, rather than not inseminating any cow after 7th month of lactation or any cow in the lowest 7 production levels **increased net revenues 40%** for a 14% pregnancy rate, versus a **6%** increase in net revenues for a 46% pregnancy rate.
Dijkhuizen & Stelwagen Livestock Prod. Sci.	1988, Holland	15	Lactation length 10 Stochastically simulate in 20 day intervals: milk yield, reproductive performance, involuntary culling	-2nd - 7th month in 20 day intervals	0.9:1	Culling Rates under various strategies and reproductive performance:
						Culling Strategies: — Pregnancy Rate 20% / 42%
						- Cull if not pregnant by day 160 — 45% / 24%
						- Use of simplified culling rule from Van Arendonk & Dijkhuizen, 1985 — **35%** / **22%**
						Increased profit with rule use — **5%** / **2%**
Congleton J. Dairy Sci.	1988, U.S.	20	Heifer selection on pedigree index			Culling on projected income versus production **increased herd income by 4 to 5%**, and decreased genetic trend 65 to 86%.
Congleton & King J. Dairy Sci.	1984, U.S.	30	Culling based on cost/unit of production of present lactation	culling end of lactation		Maximum profit at 5.0 lactations. Increasing herd life from 3.3 to 5.3 lactations **increased profit by 9%**. Intensity of cow culling had limited impact on herd genetic merit.
Allaire J. Dairy Sci.	1981, U.S.	20	Deterministic Milk production is function of age, genetic merit.	culling end of lactation	1.3:1	Involuntary culling rate 20%, Total cull rate 20 to 39% - Maximum milk production per cow with use of all replacements (39%) - **Profit increased 15 to 20% with 26% culling rate** versus the 40% culling rate

Table 3.8. (cont'd).

Author, Journal	Year & Country Modeled	Time horizon (years)	State Variables	Timing of Insemination & Culling	Heifer/ Cull Cow Prices	Scenarios and Results
Allaire & Cunningham Livestock Prod. Sci.	1980, Ireland	16	Deterministic Culling on low milk yield ME		1.6:1	Culling rates: Involuntary To max. milk **To max. profit** 15% 33% **23%** 20% 39% **27%** 25% >37% **30%**
Korver & Renkema Livestock Prod. Sci.	1979, Holland		Deterministic 12 month calving interval, 14 lactations	-1st lactation culling monthly, thereafter at 2 months after calving and before insemination actually removal at end of 8th month		Culling strategies: Culling Rate - maximum culling for low production in first lactation 40% -removal for illness or sometimes low production 23% **Increased profit** with lower rate **9.4%**
Pearson & Freeman J. Dairy Sci.	1973, U.S.	22	Deterministic 13 month calving interval	-no culling prior to day 305 of lactation	1.9:1	Replacing only involuntary losses, and selecting calves on estimated breeding value (BV), rather than freshening all females and selecting on BV resulted in: 5% less milk, very similar genetic progress, and a **21% increase in profit**

Chapter 4

HEIFER SELECTION

In order to lower the replacement rate, not all heifers can enter the milking herd. The models reviewed in Chapter 3 generally assumed random selection of heifers or selection based on genetic parameters. The failure of producers to accept random selection may be one reason for the industry's failure to reduce the culling rate. The discovery of a method to predicting culling or production of heifers in the first lactation may encourage producers to select heifers and thereby reduce their culling rate. This chapter will review the goal of heifer selection, namely profitability, and the parameters available to estimate animal profit. Methods of selection for profitability will then be reviewed.

The goal of heifer selection is to maximize profitability. Given a choice among heifers, those with the highest profit are desirable. Unfortunately it is not a simple matter to assess which animals are most profitable. Absolute profit is not available for any animals, rather the dairy industry uses profit functions to estimate the profit of a given animal so the importance of various traits in the explanation of animal profit can be determined. These profit functions vary in method of estimation and completeness (Pearson and Miller, 1981). As will be discussed in Chapter 5, these equations are also used to weight the genetic value of traits used in multiple trait selection indices.

Tigges et al. (1984) found relative net income (RNI), calculated using information available from dairy herd improvement associations, to be highly correlated with total profit (r = 0.97) and profit per day of herd life (r = 0.92) calculated using more extensive economic information such as returns from protein yield and cost of mastitis and breeding. In comparing varying measures of profitability Cassell et al. (1993) found such high correlations between relative net income and discounted relative net income, the latter was of little utility.

Aside from the issue of the completeness of the profit function, the measure of profitability is a contentious issue. As discussed in Chapter 2, Van Arendonk (1991) noticed that use of lifetime profit implied no replacement while profit per day implicitly assumed identical replacement. He noted that each of these ignore the opportunity cost of replacement and therefore neither is an appropriate measure of animal profitability. Van Arendonk showed these erroneous assumptions have drastic ramifications on the value estimates of the two traits, production and herd life (Weigel et al., 1995a), most important in the determination of profit. He showed that using total lifetime profit as the dependent variable, the regression coefficient for first lactation milk production was correct, while using profit per day of productive life as the dependent variable correctly estimated the herd life coefficient. But in each case the value of herd life relative to production was greatly overestimated. Weigel et al. (1995b) further refined this point by estimating an opportunity cost for each lactation, removing the assumption of a constant opportunity cost.

Considerable confusion exists in the literature over results comparing measures of profit that don't include opportunity cost and those that do. Weigel et al. (1995a) determined the genetic correlation between RNI and relative net income corrected for opportunity cost (RNIOC) to be 0.97. The phenotypic correlation between RNI and

RNIOC has been reported as 0.95 (Cassell et al., 1993), and 0.92 (Weigel et al., 1995a). Some (Norman et al., 1996) have incorrectly interpreted these results as evidence for ignoring the opportunity cost adjustment. This despite Cassell et al.'s (1993) conclusion that results of herd management or selection practices using RNI corrected for opportunity cost could differ from those using RNI. This is not surprising given that the variance of the RNI has been reported as 2.5 (Cassell et al., 1993), and 3.5 (Weigel et al., 1995a) times that of the corrected RNI. Correcting the profit increased its heritability to 0.17 from 0.12 (Weigel et al., 1995a). Since the validity of the literature using profit equations not corrected for opportunity cost is questionable, and studies using the adjusted equations are available, the results of the former will not be considered.

Another source of confusion exists in the literature concerning Van Arendonk's paper (1991). He showed the regression coefficient between profit not including opportunity cost, and herd life was overestimated. Since this result was in the context of multiple regression, including milk production as an independent variable, this implies that the partial correlation between profit not including opportunity cost and herd life, given production, was overestimated. It does not necessarily imply that the simple correlation between profit not including opportunity cost and herd life was overestimated as has been reported (Cassell et al., 1993).

Given, that lifetime profitability is not apparent until the end of an animal's life, it is not a useful method of selection prior to the end of that life. The literature has tended to focus on indicators of lifetime profitability that are apparent early in the productive life of milk cows. No work has considered indicators apparent prior to productive life.

The phenotypic correlations between first lactation actual milk and total lifetime milk are 0.54 and 0.88 for RNIOC, and 0.52 and 0.96 for RNI (Cassell, Smith and Pearson, 1993). First lactation milk production was found to have a 0.86 genetic correlation, and a 0.57 phenotypic correlation with lifetime production among Canadian Holsteins (Jairath et al., 1995). Weigel et al. (1995a) determined the genetic correlation between the RNI corrected for opportunity costs and first lactation ME milk production to be 0.80, and similar to the previous study, a phenotypic correlation of 0.51. Some physical traits such as dairy form and udder depth had reasonably high correlations with RNIOC (Weigel et al., 1995a). Age at first calving has a small negative correlation with RNIOC (Cassell, Smith and Pearson, 1993). Both Weigel et al. (1995a) and Cassell et al. (1993) found herd life had a correlation of approximately 0.70 with RNIOC.

Weigel et al. (1995a) considered the reliability of indirect prediction of sires' RNIOC using type, first lactation yield, and proxies for herd life. Table 4.1 is a partial reproduction of a table in the paper.

Table 4.1. Approximate reliabilities of RNIOC from combination of mature equivalent yield of milk and fat during first lactation (yield), 14 linear type traits and final score (type), and total months in milk to 84 months of age (MIM), and MIM predicted from 36 months of age (PMIM).

	Approximate reliability of RNIOC	
Variables	Maximum	80 EPN[1]
Type	.48	.32
Yield	.67	.59
Yield and type	.75	.64
Type and PMIM	.70	.36
Yield and PMIM	.75	.60
Yield, type and PMIM	.80	.65

Table 4.1 (cont'd).

Yield and MIM	.90	.70
Yield, type and MIM	.91	.72

1. Effective progeny number, represents the accuracy of a first-crop evaluation of a well-sampled AI bull.

Type characteristics provide little additional information in the prediction of RNIOC, and are unsuitable as a method of heifer selection as only animals that have freshened are evaluated.

Herd life, or a proxy, is an important variable in the determination of profit. However, it is unsuitable as a method of heifer selection. Its heritability is low, and while the USDA is producing productive life genetic evaluations for cows and bulls, heifer evaluations of productive life are not produced.

To restate, ideally heifer selection would be based on lifetime profit corrected for the opportunity cost of postponed replacement. While RNIOC can not be measured and is not recorded, first lactation milk production is widely available and moderately correlated to RNIOC. This then suggests that selection of the highest producing first lactation heifers will likely result in selection of the most profitable animals. We then turn to prediction of first lactation milk production.

Compared to herd life, the yield traits have higher heritability. And studies have been completed on the ability to predict future production. These studies fall into general types of predictors - genetic and physiological. As few studies have considered the ability of genetic estimates from the USDA's current Animal Model to predict subsequent heifer performance, performance of estimates from the earlier Modified Contemporary Comparison will also be considered.

Dams' genetic index for milk production from the Modified Contemporary Comparison was found to have a 0.18 correlation with daughter first lactation ME production deviated from herdmates. This result was slightly less than the 0.20 expected (McGilliard and Freeman, 1976). Linear regression of the measure of daughter first lactation production on dam index was 1.15 which differed from the expected unity.

Powell et al. (1977) determined the average correlation between actual daughter performance for milk production, deviated from contemporaries, and that predicted from sire pedigree index using the Modified Contemporary Comparison was 0.40 for Holsteins, slightly below the expected 0.45. The regression coefficient between the two was near the expected 1.0 suggesting selection based on sire predicted differences should be successful.

Powell and Norman (1988) found an average correlation of 0.19 between the daughter's deviation of milk yield from contemporaries and the dams' genetic index for milk production from the Modified Contemporary Comparison, which was lower than expected. The correlation increased as the repeatability of the dams' indices increased, as the number of daughter records rose, and with a higher level of herd milk production. Selection was expected to be effective given a regression coefficient of approximately one between daughters' deviation and the dams' index.

A problem with each of the preceding studies is that a considerable period of time was used to collect the data, so parent indices were calculated years after a daughter's production. In heifer selection, it is the estimate of the parents' indices at the time of selection that is relevant.

Since 1989, when the USDA began using the Animal Model for estimating genetic merit, only two domestic empirical studies have considered the ability of parents' PTAs to predict a heifer's subsequent performance. Estimated genetic indices are produced for all dairy animals with identified parents who are on a production test program. The indices, which are updated every six months, and are generally provided at no extra cost to those enrolled in a production test program.

Other empirical studies (Ferris and Wiggans 1990, 1991; Samuelson and Pearson, 1995) have considered the ability of parents' PTAs to estimate the performance of artificial insemination sires, through the performance of the sires' daughters. Since these sires tend to have genetically elite parents, with generally high reliabilities, the results of these studies are not directly applicable to daughter performance.

Based on 76,871 daughters, VanRaden et al. (1989) found the regression coefficients of daughter performance on EBVs "were near the (theoretically) expected value of 1.0" with correlations that ranged from 0.20 to 0.49. Performance of daughter a was measured as yield deviation (YD_a) which according to Wiggans and VanRaden (1989) is defined as

$$YD_a = y_a - M_a - PE_a - (sire_a \cdot herd_a)$$

such that y represents standardized milk yield; M, PE and $sire \cdot herd$ are the effects of management group, permanent environment and herd-sire interaction, respectively for a. Management group includes effects for herd, year and season of calving, parity and registration status, as well as their interaction. This equation for YD can be compared with the USDA animal model of milk yield standardized for season and age at first calving,

$$y_a = G_a + M_a + PE_a + (sire_a \cdot herd_a) + U_a + e_a$$

where G is EBV, U is unknown parent group, and e is the error for a.

In heifer selection the producer is interested in predicting y_a. While the use of YD as a dependent variable serves to decrease the error variance, the results are not directly applicable to heifer selection. For a producer is paid based on is unlikely to know the value of M_a, PE_a, $(sire_a \cdot herd_a)$ and therefore, can not derive YD_a from y_a. The prediction of y_a, not YD_a, is of primary importance to the farmer because this is the basis on which the producer is paid.

On a small sample of 37 Holstein heifers, Mao et al. (1991) measured the serum β-lactoglobulin from 16 to 1 week prepartum. They found consistent positive correlations (r ≅ 0.3 to 0.6) between the serum protein over this 16 week period and first lactation, second lactation or the sum of the two, ME milk yield. In a subsequent paper (Mao and Bremel, 1995), the 0.034 coefficient of determination for prediction of first lactation 305d milk yield (i.e., y_a) of 26 Holstein heifers using only parent average of genetic merit for milk production was improved to 0.211 with the inclusion of the serum β-lactoglobulin level at the 26th week of pregnancy. Stating the residual (u_a) from the Mao and Bremel's regression in terms of effects accounted for in the VanRaden study, the lower multiple coefficient of determination of the former is expected.

$$u_a = M_a + PE_a + (sire_a \cdot herd_a) + U_a + e_a - herd_a$$

The correlations between 305d milk yield and the log of β-lactoglobulin, and parent averages for milk, protein and fat were, respectively, 0.456, 0.185, 0.149, and -0.003. The authors note "The PA (parent average) is normally used as a selection criterion for heifer replacements." Similar suggestions have been made to sell heifer calves from inferior parents (Jones and Rogers, 1995; Franck, 1994) however no specific recommendations as to what method or when to conduct selection is presented.

While physiological measurements present potential methods for heifer selection they are not routinely collected on heifers, and β-lactoglobulin is not produced until late in the rearing period when the animal is pregnant.

Weight gain from 350 to 462 days of age had a genetic and phenotypic correlations of 0.33 and 0.67 with first lactation 308 day milk yield (Lin, McAllister et al., 1985).

The ability of at least the genetic indices for milk to predict production, which is a major component of animal profitability, suggests their potential utility in heifer selection. However, two issues are not resolved. Firstly, there is no proof that the heifers' milk index which equally weights the PTA milk of each parent is optimal. Furthermore, given the potential 5 genetic indices for each parent in each period it is not clear how each of these should be used. Lastly, the genetic indices are simply estimates which are updated over time. It is not clear when in a heifer's life selection should occur. Specifically, which of the 10 pieces of genetic information, available every six months during rearing, should be used and how they should be used in heifer selection.

There are also other benefits from heifer selection besides increasing herd profitability. Aside from short-term improvement in cash flow, heifer selection will also reduce the nutrient loading on the land, this issue will become increasingly important as environmental restrictions become more prevalent and strict (Schmit and Knoblauch, 1995).

The importance of opportunity cost has resurfaced in this chapter. Likely, due to the industry's ignoring the work on the optimal culling rate, and therefore the need for heifer selection, no work has been completed on heifer selection for maximum profit. From the literature it can be determined that production is important to animal profitability. Genetic indices for milk have been shown to hold some promise for the selection of higher producing animals. The current availability and ease of access make genetic indices a potential method of heifer selection. While the use of genetic indices for heifer selection has been suggested, details and empirical results to support these statements are lacking. Chapter 5 will consider the estimation of genetic values of animals.

BIBLIOGRAPHY

Cassell, B. G., B. B. Smith, and R. E. Pearson. 1993. Influence of herd-life opportunity and characteristics of cows and herds on different net income functions. J. Dairy Sci. 76:1182-1190.

Franck, R. 1994. Cashing in on your cull rate. Dairy Herd Management 31:57-58.

Ferris, T. A., and G. R. Wiggans. 1990. Accuracy of animal model parent evaluations in predicting PTA for AI sampled bulls. J. Dairy Sci. 73(Suppl. 1):141.(Abstr.)

Ferris, T. A., and G. R. Wiggans. 1991. Accuracy of animal model parent evaluations in predicting Daughter Yield Deviations for AI sampled bulls. J. Dairy Sci. 74(Suppl. 1):265.(Abstr.)

Jairath, L. K., J. F. Hayes, and R. I. Cue. 1995. Correlations between first lactation and lifetime performance traits of Canadian Holsteins. J. Dairy Sci. 78:438-448.

Jones, L. R., and G. W. Rogers. 1995. Management strategies: Culling. Proc 27th Am Assoc Bovine Pract 93-96.

Lin, C. Y., A. J. McAllister, and A. J. Lee. 1985. Multitrait estimation of relationships of first-lactation yields to body weight changes in Holstein heifers. J. Dairy Sci. 68:2954-2963.

Mao, F. C., R. D. Bremel, and M. R. Dentine. 1991. Serum concentrations of the milk proteins α-lactalbumin and β-lactoglobulin in pregnancy and lactation: Correlations with milk and fat yields in dairy cattle. J. Dairy Sci. 74:2952-2958.

Mao, F. C., and R. D. Bremel. 1995. Prediction of milk yields from serum β-lactoglobulin concentrations in pregnant heifers. J. Dairy Sci. 78:291-295.

McGilliard, M. L., and A. E. Freeman. 1976. Predicting daughter milk yield production from dam index. J. Dairy Sci. 59:1140-1146.

Norman, H. D., R. L. Powell, J. R. Wright, and R. E. Pearson. 1996. Phenotypic relationship of yield and type scores from first lactation with herd life and profitability. J. Dairy Sci. 79:689-701.

Pearson, R. E., and R. H. Miller. 1981. Economic definition of total performance, breeding goals, and breeding values for dairy cattle. J. Dairy Sci. 64:857-869.

Powell, R. L., and H. D. Norman. 1988. Accuracy of cow indexes according to repeatability, evaluation, herd yield, and registry statue. J. Dairy Sci. 71:2232-2240.

Powell, R. L., H. D. Norman, and F. N. Dickinson. 1977. Relationships between bulls' pedigree indexes and daughter performance in the modified contemporary comparison. J. Dairy Sci. 60:961-966.

Samuelson, D. J., and R. E. Pearson. 1995. Accuracy of predicting genetic merit from pedigree information for bulls entering stud sampling programs. J. Dairy Sci. 78:2057-2066.

Schmit, T. M., and W. A. Knoblauch. 1995. The impact of nutrient loading restrictions on dairy farm profitability. J. Dairy Sci. 78:1267-1281.

Tigges, R. J., R. E. Pearson, and W. E. Vinson. 1984. Use of dairy herd improvement variables to predict lifetime profitability. J. Dairy Sci. 67:180-184.

Van Arendonk, J. A. M. 1991. Use of profit equations to determine relative economic value of dairy herd life and production from field data. J. Dairy Sci. 74:1101-1107.

VanRaden, P. M., C. A. Ernst, G. R. Wiggans, and H. D. Norman. 1989. Ability of animal model and Modified Contemporary Comparison evaluations of parents to predict future progeny performance. J. Dairy Sci. 72(Suppl. 1):72.(Abstr.)

Weigel, D. J., B. G. Cassel, I. Hoeschele, and R. E. Pearson. 1995a. Multiple trait prediction of transmitting abilities for herd life and estimation of economic weights using relative net income adjusted for opportunity cost. J. Dairy Sci. 78:639-647.

Weigel, D. J., B. G. Cassel, and R. E. Pearson. 1995b. Adjustment of a net income function for opportunity cost of postponed replacement on a lactation basis. J. Dairy Sci. 78:648-654.

Wiggans, G. R., and P. M. VanRaden. 1989. Developments in the USDA animal model system for genetic evaluation of yield traits. J. Dairy Sci. 72(Suppl. 1):71.(Abstr.)

Chapter 5

SELECTION INDEX THEORY

From Chapter 4 it is apparent that genetic evaluations currently present the only potentially viable methods on which to base selection of heifers prior to calving. A familiarity with the methods and assumptions made in estimating an animal's genetic value is then warranted. This chapter will start with a brief review of genetic effects and the assessment of genetic variance and covariance. The goals, methodology, and assumptions of the Selection Index (SI) model for a single trait are examined. The effects of the various assumptions on the model are explored. Multiple trait selection is then reviewed, followed by miscellaneous issues which further cast doubt on the optimality of SI. By the conclusion of this chapter it will be apparent that, due to the assumptions made, SI evaluation of an animal's value is not optimal, even more critical to heifer selection, it is not clear that heifers will even be correctly ranked. Based on this inability to correctly rank heifers for selection, combined with the effect of Mendelian sampling which greatly increases the prediction error of the index, the value of genetic indices in heifer selection is questionable.

The approach to genetic evaluation and SI in this chapter is heavily based on Selection Index and Introduction to Mixed Model Methods by Van Vleck (1993), supplemented with relevant information from the literature. Given the context of this chapter is understanding the estimation of a heifer's value based on her parents' information, examples will appropriately focus on such estimations.

The theoretical composition of genotype, as well as, the determination of genetic covariance between individuals is fundamental to understanding selection theory and its assumptions. So each of these issues will be addressed prior to reviewing selection theory.

The genotype of an animal can be broken down into linear genetic effects and epistatic or nonlinear genetic interaction effects. An additive genetic effect is the aggregate effect of allelic substitution at a particular locus. The sum of all additive genetic effects over all loci is the additive genetic value (G_A) of the animal, also known as the estimated breeding value (EBV).

Let $A_1 = 5$, $A_2 = 2$ where A_i is the i^{th} allele of gene A. If only additive effects exist, the phenotypes resulting from each allelic combination are:

$$A_1A_1 = 5 + 5 = 10$$

$$A_1A_2 = 5 + 2 = 7$$

$$A_2A_2 = 2 + 2 = 4$$

A dominance genetic effect is the average effect of allelic substitution at a particular locus less the additive genetic effects. The sum of all dominance genetic effects over all loci is the dominance genetic value (G_D) of the animal.

Suppose, similar to the above example, $B_1 = 5$, $B_2 = 2$ but a dominance effect of +1 occurs then

$$B_1B_1 = 10$$

$$B_1B_2 = 7 + 1 = 8$$

$$B_2B_2 = 4$$

An additive by additive genetic effect is the average effect of substituting a pair of alleles at different loci less the additive genetics effects. The sum of all specific effects of non-allelic gene pairs is the additive by additive genetic value (G_{AA}) of the animal.

Expanding the first example, let $A_1 = 5$, $A_2 = 2$, $C_1 = 1$, $C_2 = 0$, and an additive by additive effect of +3 occurs when A_1 and C_1 are in common.

	A_1A_1	A_1A_2	A_2A_2
C_1C_1	12+6=18	9+3=12	6
C_1C_2	11+3=14	8+3=11	5
C_2C_2	10	7	4

An additive by dominance genetic effect is the average effect of substituting an allele at one locus and an allelic pair at another locus. It is calculated as the difference from the additive, dominance, and additive by additive genetic effects. The sum of all such effects is the additive by dominance genetic value (G_{AD}) of the animal.

As in the second example, let $B_1 = 5$, $B_2 = 2$, $C_1 = 1$, $C_2 = 0$, with a dominance effect of +1 on the B locus, and an additive by dominance effect of -4 when the genotype $B_1B_2C_1$ is present.

	B_1B_1	B_1B_2	B_2B_2
C_1C_1	12	10-4=6	6
C_1C_2	11	9-4=5	5
C_2C_2	10	8	4

Further higher order epistatic effects can be defined. Given the number of genes and alleles in any mammal's genome, the number of potential effects quickly approaches infinity.

The total genetic value of an animal can then be represented by

$$G = G_A + G_D + G_{AA} + G_{DD} + G_{AD} + G_{AAA} + G_{AAD} + G_{ADD} + \dots \qquad \textbf{(5. 1)}$$

Selection theory also relies on the genetic variance or covariance. With the reasonable assumption of independence among the different type of genetic effects, the genetic variance can be represented by

$$\mathrm{E}[G^2] = \sigma_G^2 = \sigma_{G_A}^2 + \sigma_{G_D}^2 + \sigma_{G_{AA}}^2 + \sigma_{G_{AD}}^2 + \sigma_{G_{AAA}}^2 + \sigma_{G_{AAD}}^2 + \sigma_{G_{ADD}}^2 + \cdots \qquad \textbf{(5. 2)}$$

where E is the expectation operator, and the genetic value is defined to have zero mean.

Now consider the genetic covariance between individuals *i* and *j*.

$$\begin{aligned}\mathrm{E}[G_iG_j] = \sigma_{G_iG_j} &= a_{ij}{}^1 d_{ij}{}^0 \sigma_{G_A}^2 + a_{ij}{}^0 d_{ij}{}^1 \sigma_{G_D}^2 + a_{ij}{}^2 d_{ij}{}^0 \sigma_{G_{AA}}^2 + a_{ij}{}^0 d_{ij}{}^2 \sigma_{G_{DD}}^2 a_{ij}{}^1 d_{ij}{}^1 \sigma_{G_{AD}}^2 \\ &+ a_{ij}{}^3 d_{ij}{}^0 \sigma_{G_{AAA}}^2 + a_{ij}{}^2 d_{ij}{}^1 \sigma_{G_{AAD}}^2 + a_{ij}{}^1 d_{ij}{}^2 \sigma_{G_{ADD}}^2 + \cdots\end{aligned} \qquad \textbf{(5. 3)}$$

where a_{ij} = the fraction of additive gene effects in common between individuals *i* and *j*,

and d_{ij} = the fraction of dominance effects in common by descent.

These coefficients are calculated using simple probability theory. Equation (5.4) is the model of genetic covariance between a parent (*P*) and an offspring (*O*).

$$\begin{aligned}\sigma_{G_OG_P} = &\left(\frac{1}{2}\right)^1 0^0 \sigma_{G_A}^2 + \left(\frac{1}{2}\right)^0 0^1 \sigma_{G_D}^2 + \left(\frac{1}{2}\right)^2 0^0 \sigma_{G_{AA}}^2 + \left(\frac{1}{2}\right)^0 0^2 \sigma_{G_{DD}}^2 + \left(\frac{1}{2}\right)^1 0^1 \sigma_{G_{AD}}^2 + \\ &\left(\frac{1}{2}\right)^3 0^0 \sigma_{G_{AAA}}^2 + \left(\frac{1}{2}\right)^2 0^1 \sigma_{G_{AAD}}^2 + \left(\frac{1}{2}\right)^1 0^2 \sigma_{G_{ADD}}^2 + \cdots\end{aligned} \qquad \textbf{(5. 4)}$$

Given that the offspring receives half of its genes from each parent there is a $(0.5)^n$

probability that the two share *n* non-allelic genes in common. Notice the fraction of dominance effects in common by descent is zero between this pair. In fact this null coefficient for a biallelic locus is used for all but full siblings for whom it is 0.25. Given that at least one parent has a heterozygous genotype and exhibits dominance effects, the probability of an offspring having the same genotype at this locus is 0.50 not zero, regardless of the other parent's genotype.

The unexpected null coefficient is the result of the model used to determine the relationship between individuals. These coefficients consider only genes in common which source from a common ancestor. This is certainly a reasonable assumption for additive genes which are simply linear interactions. The null coefficient is appropriate when using the identical by descent criteria, since heterozygosity is required for dominance effects and one allele of each locus is transmitted from parent to offspring there is no opportunity for the dominance effect to be identical by descent. Considering the purpose of Equation (5.3) is to quantify the similarity of the genetic effects of the two individuals, regardless of the source of that similarity, ignoring the identical in state genes, those not from a common ancestor, misrepresents the true composition of genetic covariance and hence the interest in DNA sequencing.

It is also clear that as the number of non-allelic genes and/or allelic pairs increases the coefficients decrease in magnitude. This is the frequently cited reason for routinely ignoring any higher order epistatic effects. However, in determining whether a right hand side element of Equation (5.3) is so small as to be ignored, the variance of the genetic effect itself should also be considered.

The result of this assumption of no interaction effects of order two or higher, and the assumption of identical by descent is to reduce Equation (5.4) to

$$\sigma_{G_O G_P} = \frac{1}{2}\sigma^2_{G_A} \tag{5.5}$$

This simplification is used for all but full sib genetic covariances for whom a dominance effect may be included.

Reliable estimation of variance/covariance components for genetic parameters is one of the major tasks for any livestock improvement scheme. Maximum likelihood is a method of estimation that empirically results in asymptotically unbiased estimates in the face of selection (Rothschild, Henderson, and Quaas, 1978; Schaeffer, and Song, 1978). A major drawback of this method is the variance estimates of a mixed model are biased since the fixed effects are treated as known, resulting in overestimation of the degrees of freedom. According to Meyer (1983) and Harville (1972) restricted maximum likelihood (REML) overcomes this problem by maximizing only that part of the likelihood that is independent of the fixed effects. Empirically REML also appears to remove any selection bias from the estimates (Meyer, 1983). In the genetics literature it is common to ignore the potentially complex distributional properties of the estimator. This may be due to the work by Harville (1972) where it was suggested that the REML estimates derived under normality may be reasonable estimators even when the form of the distribution of the estimate is unspecified.

The REML covariance estimate is then set equal to the product of the additive genetic variance and the appropriate a_{ij}. The result is a potentially exaggerated additive genetic variance for all relationships, since all the elements of the right hand side of Equation (5.3) are allocated to either the additive genetic variance or the residual term.

Tempelman and Burnside (1990) showed that this exaggeration can be sizable. Their study revealed the identical in state dominance variation for Canadian Holsteins' first lactation milk and fat were, respectively, 15 and 75 percent of the additive genetic variance. Despite similar standard errors for each of the estimates, the former was not statistically significant due to a high coefficient of variation (i.e., standard error divided by the mean).

A second Canadian paper (McAllister, Lee, Batra, et al., 1994) indirectly investigated the presence of dominance effects between dairy breeds. In acceptance with previous studies, this experiment showed a significant heterotic effect in Holstein x Aryshire F_1 females. Heterosis is defined as the realized phenotype less the expected phenotype calculated as the average additive effect of the two breeds. The necessary conditions for a heterotic or crossbreeding effect are the presence of dominance effects and different gene frequency between the two breeds.

This paper is of interest for two reasons. Firstly, while heterotic effects were generally accepted to occur, conventional wisdom among animal breeders held that the dominance effects of crossbreeding were insufficient to compensate for the much lower production of the Aryshire parent. As a result it was accepted that, despite heterosis, the F_1 would have lower production than the higher producing Holstein parent. This study found no significant difference between the Holstein x Aryshire F_1 female and the purebred Holstein line for lifetime milk production. This suggests the presence of a much stronger dominance effect than generally accepted by the industry. This paper also gave evidence of how heterosis may vary with the period over which a given trait is considered, thereby bringing into question the ability to extrapolate the aforementioned first lactation results of Tempelman and Burnside to production in general. For example the heterosis effect was 3.7% for first lactation yield and 16.5% for lifetime milk production. Lifetime heterosis effects were 20.0, 17.2, 17.9 and 20.6% for lifetime fat yield, protein yield, milk value and annualized discounted returns.

Clearly it is extremely difficult if not impossible to estimate dominance effects resulting from identical in state genes. Estimation of other variance effects would require supplemental equations to solve for the extra parameters. It is only recently that inroads have been made in the inclusion of dominance effects in genetic evaluations (Misztal, 1997). Furthermore, the number of equations required to estimate higher order genetic variance components is overwhelming. This model, as all others is a tractable simplification of reality, and should not be judged on its realism. As with any model it should be judged on its performance, this may include theoretical results or the ability of the model to mirror reality. It is in this assessment of the model that its assumptions become critical. For failing to recognize the assumptions, and their effects on the results, leads to the model being accepted as reality, and therefore beyond reproach.

Given these theoretical models, the fundamental challenge in selection is to choose those individuals with the greatest genetic merit for the trait(s) of choice to be used as parents of the next generation (μ_{G_s}), i.e.,

$$\max \Delta G = \mu_{G_s} - \mu_G \tag{5. 6}$$

where μ_G is the average genetics of the total population.

This challenge is heightened by the fact that rarely can the genetic effects be directly observed, and this is especially common in economically important production traits. Rather the breeder must attempt to deduce the genetic value of the animal when only phenotype (P) is apparent.

$$P_{ij} \equiv \mu + G_i + PE_i + TE_{ij} \tag{5. 7}$$

Where P_{ij} = the phenotype of individual i in period j

μ = the population mean

G_i = the total genetic value of i deviated from the population mean

PE_i = the permanent environmental effect on P_i deviated from the population mean

TE_{ij} = the temporary environmental effect of period j on P_i deviated from the population mean.

E_i the total environmental effect on P_i is equal to the sum of TE_{ij} and PE_i.

Heritability in the narrow (h^2) and the broad (h_b^2) sense can now be defined, the former is the most common definition,

$$h^2 = \frac{\sigma^2_{G_A}}{\sigma^2_P} \tag{5. 8}$$

$$h_b^2 = \frac{\sigma^2_G}{\sigma^2_P} \tag{5. 9}$$

where σ^2_P is the phenotypic variance in the population, and is easily estimated.

Notice that overestimation of the additive genetic variance results in overestimation of narrow heritability.

Armed with an understanding of the various genetic effects and an understanding of the assumptions made in estimating genetic covariance we can proceed to review selection theory. Two models have been used for animal selection: Selection Index (SI) and the Animal Model, the latter is the current method of genetic analysis used by the USDA. Hazel et al. (1994) and Philipsson et al. (1994) provide reviews of the past, present and future of SI. SI employs all the same assumptions as the Animal Model. For SI all phenotypic and genetic variances and heritabilities, genetic and phenotypic covariances for using multiple traits are assumed known (Vandepitte, and Hazel, 1977). The former also requires all fixed factors be known. Fixed factors include all those nongenetic factors, excluding temporary and permanent environmental effects, which affect phenotype. Fixed factors, represented by μ in Equation (5.7), include effects due to year and age of the animal. Despite this one difference in the practical application of the models, all the theoretical assumptions and results of each model are identical. Since these assumptions are more explicit in the SI, it will be the method reviewed here.

The goals of the SI are:

1) Minimize the prediction error variance or $E[T-I]^2$ where T is the true value and I is the index. This is equivalent to maximizing r_{TI}, the correlation between the true value and its prediction.

2) Maximize the probability of correctly ranking animals.

3) Maximize the true value of the selected group.

In the development of selection theory the main emphasis, rightly so, was on goals 2 and 3. However, the solution to even a simple formulation of this problem was mathematically intractable without the assumption that the P_i, the phenotype of individual i, and T must have a bivariate normal distribution (Williams, 1962). Under these conditions, goals two and three can be maximized through maximizing r_{TI}. Clearly selection will result in a truncated distribution if the $r_{TI} > 0$, and as a result the normality assumption will be violated to some degree, and goals 2 and 3 may not be realized. Potential violation of this normality assumption is not a trivial problem, from the perspective of individual animal breeders it is the correct parametric ranking of animals which is of importance, and will permit maximum genetic progress. Under certain conditions the SI indices are unbiased when the population consists of individuals who were themselves selected (Goffinet, 1983).

Geneticists have considered the effect of ignoring the changing population parameters as selection occurs. This change is frequently ignored, and the selection coefficients are not updated as the population changes. Under conditions of an infinitesimal genetic model and an infinite population size, the effect of updating the selection coefficients was found to be negligible with less than a 1.5% loss of aggregate genetic progress per generation (Villanueva and Kennedy, 1993).

The only data available to predict an animal's true genetic value is the phenotype of relatives, and perhaps its own phenotype. By definition, relatives are expected to have more genes in common with the individual than the population at large. The general linear selection index model for individual α (I_α) can then be expressed as dependent on the phenotypes of a number of animals ($X_1, X_2, \ldots, X_N$), where α is not necessarily a member of N.

$$I_\alpha = b_1X_1 + b_2X_2 + \ldots + b_NX_N \qquad \textbf{(5. 10)}$$

where $X_i = P_i - \mu$.

All that remains then is to determine the values of the bs.

$$\max_{\{b_i\}_{i=1}^{N}} r_{TI} = \frac{\sigma_{TI}}{\sqrt{\sigma_T^2 \sigma_I^2}} \qquad \textbf{(5. 11)}$$

A logarithmic transformation results in a linear objective function.

$$\max_{\{b_i\}_{i=1}^{N}} \log(r_{TI}) = \log(\sigma_{\mathrm{TI}}) - \frac{1}{2}\log(\sigma_T^2) - \frac{1}{2}\log(\sigma_I^2) \qquad \textbf{(5. 12)}$$

where

$$\mathrm{E}[TI] = \sigma_{TI} = b_1\sigma_{TX_1} + b_2\sigma_{TX_2} + \cdots + b_N\sigma_{TX_N} \qquad \textbf{(5. 13)}$$

$$\begin{aligned}\sigma_I^2 = {} & b_1^2\sigma_{X_1}^2 + 2b_1b_2\sigma_{X_1X_2} + \cdots + 2b_1b_N\sigma_{X_1X_N} + b_2^2\sigma_{X_2}^2 + 2b_2b_3\sigma_{X_2X_3} \\ & + \cdots + 2b_2b_N\sigma_{X_2X_N} + \cdots + b_N^2\sigma_{X_N}^2\end{aligned} \qquad \textbf{(5. 14)}$$

This unconstrained maximization problem is solved by taking the first derivative of the objective function with respect to each b, recognizing that the variance of T is independent of the current bs. (Notice, however, that σ_T^2 is a function of the index, assuming the index is at all useful (r_{TI}>0) and has been used for selection, the variance of T is then dependent on previous bs.)

$$\frac{\delta \log(r_{TI})}{\delta b_1} = \frac{\sigma_{TX_1}}{\sigma_{TI}} - \frac{b_1\sigma_{X_1}^2 + b_2\sigma_{X_1X_2} + \cdots + b_N\sigma_{X_1X_N}}{\sigma_I^2} = 0$$

$$\frac{\delta \log(r_{TI})}{\delta b_2} = \frac{\sigma_{TX_2}}{\sigma_{TI}} - \frac{b_1\sigma_{X_1X_2} + b_2\sigma^2_{X_2} + \cdots + b_N\sigma_{X_2X_N}}{\sigma^2_I} = 0 \qquad (5.\,15)$$

$$\vdots \qquad \vdots \qquad \qquad \vdots$$

$$\frac{\delta \log(r_{TI})}{\delta b_N} = \frac{\sigma_{TX_N}}{\sigma_{TI}} - \frac{b_1\sigma_{X_1X_N} + b_2\sigma_{X_2X_N} + \cdots + b_N\sigma^2_{X_N}}{\sigma^2_I} = 0$$

Redefining this system of simultaneous equations in standard linear notation, with the primed symbol designating a transpose, reveals

$$\mathbf{Pb} = k\mathbf{c} \qquad (5.\,16)$$

where

$\mathbf{P}$ = variance - covariance matrix of the $X_i = \mathrm{E}(\mathbf{XX}')$
$\mathbf{X} = [X_1 \quad X_2 \quad \cdots \quad X_N]'$
$\mathbf{b} = [b_1 \quad b_2 \quad \cdots \quad b_N]'$
k = the scalar $\frac{\sigma^2_I}{\sigma_{TI}}$
$\mathbf{c}$ = covariance vector of T and X_i

Since the scalar k does not affect the relative magnitude of **b**, it can arbitrarily be set equal to unity. Equation (5.10) can then be restated in linear notation as

$$\hat{\mathbf{T}} = \mathbf{I} = \mathbf{b}'\mathbf{X} \qquad (5.\,17)$$

with

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{c} \qquad (5.\,18)$$

Since the **b** vector is a function of **P** and **c** it is important to consider how each of these matrices are estimated. Once an understanding of the assumptions made in estimating **b** are gained, we will turn to actual use of an index. First derivation of **c**, the covariance between the true value and the phenotypes, is reviewed. This covariance is a function of the desired true value which is being selected for. The most commonly used true value in the dairy cattle industry is estimated breeding value (*EBV*) or the additive genetic effect of individual α, then the first element of **c** is

$$\begin{aligned}\sigma_{TX_1} &= \sigma_{G_{A_\alpha}X_1} \qquad \textbf{(5. 19)}\\ &= \mathrm{E}[G_{A_\alpha}X_1]\\ &= \mathrm{E}[G_{A_\alpha}G_1]+\mathrm{E}[G_{A_\alpha}E_1]\\ &= \mathrm{E}[G_{A_\alpha}G_{A_1}+G_{A_\alpha}G_{D_1}+G_{A_\alpha}G_{AA_1}+G_{A_\alpha}G_{AD_1}+G_{A_\alpha}G_{DD_1}+G_{A_\alpha}G_{AAA_1}\\ &\qquad +G_{A_\alpha}G_{AAD_1}+G_{A_\alpha}G_{ADD_1}+\cdots]+\mathrm{E}[G_{A_\alpha}E_1]\end{aligned}$$

$$\sigma_{TX_1} = a_{\alpha 1}\sigma^2_{G_A} \qquad \textbf{(5. 20)}$$

where $a_{\alpha 1}$ is defined analogously to that used in Equation (5.3).

In arriving at the final expression of Equation (5.20), the phenotype of individual 1 has been decomposed into a genetic and environmental effect. Using the results of Equations (5.2) and (5.3), the covariance of the true breeding value and 1's genetic effect is estimated to be represented by Equation (5.20). Since the strong assumption of zero expected additive genetic-environmental effect covariance was made, the genetic covariance equals the covariance between the *EBV* and X_1. Clearly this later assumption abstracts from reality. It is not hard to imagine additive genetic-environmental interaction. Intuitively, consider a genetic trait such as nervousness or shyness, individuals possessing such a trait will perform better with a quiet gentle manager (environment) than not. The independence between the genetics and environment is then refuted and furthermore, a role of management is to capitalize on this potential interaction. Preferential treatment, which will be discussed more later in the chapter, can be considered selective exploitation of a environment-genetic interaction. While a model to account for this genetic-environment covariance has been developed by Schaeffer and Henderson (1983), it is not employed by the USDA.

Even ignoring the assumption of zero additive genetic-environmental covariance, Equation (5.20) does not accurately represent the covariance between the additive genetic value of α and the genotype of individual 1 for the desired trait. Recall that earlier in the discussion of estimation of genetic covariance it was shown that the additive genetic covariance was overestimated because, it in effect, included some genetic covariance other than the additive portion. Tempelman and Burnside (1991) showed that omitting the dominance portion from an animal model resulted in a 4% overestimation of the additive variance for milk. If the goal of selection is to maximize progress in the additive genetic effect of a trait, the net effect on **c** is ambiguous, since ignoring the genetic-environment covariance underestimates **c**, while the additive genetic covariance is slightly overestimated.

The diagonal elements of **P** are phenotypic variances while the remaining elements are covariances. Each type of element may be calculated using individual records or the average of multiple records. The estimation of each will be assessed in turn, starting with estimations using single records.

Since the phenotypic variance of an individuals single record is the expectation of the population variance, the variance elements of **P** based on single records per individual are readily available from the population being studied and therefore no extra estimation is required for SI. However, this assumes homogenous variance among all lactations. Van der Werf et al. showed that ignoring variance heterogeneity resulted in sire evaluations for milk based on progeny records being 157 kg lower than that expected based the sires' EBVs.

The nondiagonal elements of **P** estimated from the data are again unbiased, and are represented by Equation (5.21). However, as can be seen from (5.22), in SI the assumptions made in the estimation of the non-diagonal elements of **P** are similar to those used in determining **c**. Specifically, SI assumes that the only source of phenotypic covariance is additive genetic variance, and as a result the additive genetic variance is overestimated.

$$\begin{aligned}\sigma_{X_1X_2} &= \mathrm{E}[(G_1 + E_1)(G_2 + E_2)] \\ &= \mathrm{E}[G_1G_2 + G_1E_2 + G_2E_1 + E_1E_2]\end{aligned} \qquad \textbf{(5. 21)}$$

$$= a_{12}\sigma^2_{G_A} \qquad \textbf{(5. 22)}$$

Equation (5.22) then does not accurately reflect the additive genetic covariance. Also, many strong assumptions are made concerning the environmental interactions. If, as argued previously, the environment of an animal may interact with its genotype, given that the genotypes of the animals are related, it would not be surprising to find that the genotype of one animal covaries with the environment of the other. The assumption of no covariance between the environmental effects of the two animals may also be strong.

These assumptions will result in the underestimation of these non-diagonal elements of P.

The elements of P may themselves be an average if multiple records are available. The source of these multiple records include an individual with more than a single record, or records from a homogenous group of relatives (i.e., 20 daughters all with comparable single records). Taking the first example of multiple records on an individual animal, the phenotypic variance of this individual is then actually the squared standard error (i.e. the variance of a mean), which will be less than the variance of a single record. The initial assumptions made in estimating the standard error are: the variance of each record is the same, and the covariance between each pair of records is identical.

$$\sigma^2_{\bar{X}_i} = \frac{n\sigma^2_X + n(n-1)\sigma_{X_iX_i'}}{n^2} = \frac{\sigma^2_X + (n-1)\sigma_{X_iX_i'}}{n} \tag{5. 23}$$

The covariance between records of a single animal, X_i, can be represented as

$$\begin{aligned}\sigma_{X_iX_i'} &= \mathrm{E}\left[(G_i + PE_i + TE_{ij})(G_i + PE_i + TE_{ij'})\right] \\ &= \sigma^2_G + \sigma^2_{PE} + \sigma_{TE_{ij}TE_{ij'}} + \sigma_{GTE_{ij'}} + \sigma_{PETE_{ij'}} + \sigma_{GTE_{ij}} + \sigma_{PETE_{ij}} + 2\sigma_{GPE}\end{aligned} \tag{5. 24}$$

Making the strong assumption of zero covariance among all terms leads to

$$\sigma_{X_iX_i'} = \sigma^2_G + \sigma^2_{PE} \equiv \sigma^2_A \tag{5. 25}$$

Where A, the animal effect, is the sum of all permanent effects on the animal (i.e., G + PE). While Equation (5.23) can then accurately be denoted as

$$\sigma^2_{\bar{X}_i} = \frac{\sigma^2_X + (n-1)\sigma^2_A}{n} \tag{5. 26}$$

geneticists commonly employ the concept of repeatability (R) in this calculation where

$$R \equiv \frac{\sigma_G^2 + \sigma_{PE}^2}{\sigma_G^2 + \sigma_{PE}^2 + \sigma_{TE}^2} = \frac{\sigma_A^2}{\sigma_X^2} \tag{5. 27}$$

By ignoring the interaction effects in Equation (5.24), Equation (5.23) is then underestimated through the use of

$$\sigma_{\bar{X}_i}^2 = \frac{\sigma_X^2 + (n-1)R\sigma_X^2}{n} \tag{5. 28}$$

When X_i is the average of a single group of equally related relatives then Equation (5.24) becomes

$$\begin{aligned} \sigma_{\bar{X}_i \bar{X}_i'} &= \mathrm{E}\left[(G_i + PE_i + TE_{ij})(G_{i'} + PE_{i'} + TE_{i'j'})\right] \\ &= \sigma_{G_i G_{i'}} + \sigma_{PE_i PE_{i'}} + \sigma_{TE_{ij} TE_{i'j'}} + \sigma_{G_i PE_{i'}} + \sigma_{G_i TE_{i'j'}} + \sigma_{PE_i TE_{i'j'}} \\ &\qquad + \sigma_{G_{i'} PE_i} + \sigma_{G_{i'} TE_{ij}} + \sigma_{PE_{i'} TE_{ij}} \end{aligned} \tag{5. 29}$$

The strong assumption that all terms, other than the first, are zero is generally made. Note this implies no environmental covariance even though animals may be in the same herd (i.e., environment). The inclusion of herd in the model would account for this environmental covariance. Given that genetic-environmental interactions occur for a given animal, that the environments of two animals aren't independent, and that genetic covariance exists it logically follows that the genetic-environmental interactions in Equation (5.29) are not zero either. The assumption of importance only of additive genetic effects is invoked to arrive at

$$\sigma_{\bar{X}_i \bar{X}_i'} = a_{ii'} \sigma_{G_A}^2 \tag{5. 30}$$

It is then apparent that $\hat{\mathbf{P}}$, is an unbiased estimate of **P**. However, in using these estimates of phenotypic variance, SI makes simplifying assumptions. Given the bias in $\hat{\mathbf{c}}$, the estimate of **c**, and that many of the elements of $\hat{\mathbf{P}}$ are not equivalent to that required for SI theory, $\hat{\mathbf{b}}$ is not equal to the optimal **b**, however the degree of inoptimality is indeterminant. Progress in selection will then be less than potentially possible, but more importantly, it is likely that the level of bias in the relative values of the $\hat{b}$ s (the elements of $\hat{\mathbf{b}}$) will vary across individuals so the paramount rankings of individuals may be incorrect and less reliable than suggested by SI theory.

Having assessed the RHS and LHS of Equation (5.16), **b** for EBV can now be derived. Consider estimating the EBV of a heifer (H) for milk production given the single milk production record of her dam (D) and the effect of the sire. Since the sire cannot produce milk, his effect is represented by his relatives. For this simple example, we consider that the sire has a daughter who has completed a record. This daughter is a half sib (S) to H, as they are assumed to have different dams. This example can be extended to consider multiple records on multiple relatives of each of H's parents. The index can then be represented as

$$I_H = \begin{bmatrix} b_D & b_S \end{bmatrix} \begin{bmatrix} X_D \\ X_S \end{bmatrix} \qquad \textbf{(5. 31)}$$

From Equation (5.16)

$$\mathbf{Pb} = \mathbf{c}$$

where

$$\mathbf{P} = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_X^2 \end{bmatrix}$$

$$\mathbf{c} = \begin{bmatrix} 0.5\sigma_{G_A}^2 \\ 0.25\sigma_{G_A}^2 \end{bmatrix}$$

The off-diagonal elements of **P** are zero since the dam and half sib, are assumed to be unrelated (i.e. $a_{DS} = 0$). From Equation (5.20), the elements of **c** are expressions of the additive genetic variance weighted by one half for D as the heifer receives half of her genes from each parent. The weight for S is one quarter, reflecting that she and H have different dams.

$$\begin{aligned} \mathbf{b} &= \mathbf{P}^{-1}\mathbf{c} \\ &= \begin{bmatrix} \sigma_X^{-2} & 0 \\ 0 & \sigma_X^{-2} \end{bmatrix} \begin{bmatrix} 0.5\sigma_{G_A}^2 \\ 0.25\sigma_{G_A}^2 \end{bmatrix} = \begin{bmatrix} 0.5h^2 \\ 0.25h^2 \end{bmatrix} \end{aligned} \tag{5. 32}$$

Upon reviewing the definition of heritability in Equation (5.8), it is not surprising that the *b*s will always be functions of the heritability of the trait (in this case, milk production) since **P** is the phenotypic variance-covariance matrix, while **c** contains the additive genetic variance. Equation (5.33) represents the more general index coefficients for a parent, as derived by Van Vleck, where *n* is the number of records of the parent, and *R* is repeatability.

$$b_{Dam} = \frac{0.5nh^2}{1 + (n-1)R} \tag{5. 33}$$

The relative size of b_l is proportional to the strength of the relationship between the individual being indexed and individual 1. Ceteris paribus, the size of b_l will also increase as the number of 1's records increase. In reality, each parent would likely have more than a single record so the *b*s would vary among parents. In general, sires would be expected to have more records and therefore contribute more information to their daughters' EBV.

The estimation of the average additive genetic value that will be transmitted to progeny is widely used in the dairy industry. Every six months the USDA calculates the equivalent of $b_D X_D$, and $b_S X_S$ and labels them predicted transmitting abilities (PTAs). In other words, PTAs are one half the estimated additive value of an individual. So to determine the milk EBV of a heifer all that is required is to sum the PTA milk of the dam and sire. PTAs are routinely published for pounds of milk, fat, and protein, as well as, dollar value of milk and fat yield in cheese production.

The EBV of an animal is an estimate of the mean additive genetic value. We are, therefore, not only interested in the variance of that estimate, but as well, its accuracy, or the degree of correlation between the estimate and the true value.

Since X is a random variable, *I* is also a random variable with a variance. Recalling the linear notation for I used in Equation (5.17) and the definitions of **P** and **c** from Equation (5.16),

$$\sigma_I^2 = \mathrm{E}[\mathbf{b'XX'b}] = \mathbf{b'}\mathrm{E}[\mathbf{XX'}]\mathbf{b} = \mathbf{b'Pb} \tag{5.34}$$

$$\sigma_{TI} = \mathrm{E}[\mathbf{Tb'X}] = \mathbf{b'}\mathrm{E}[\mathbf{TX}] = \mathbf{b'c} \tag{5.35}$$

As a result the variance of the index and the covariance between the trait and index can be shown to be equal.

$$\sigma_I^2 = \mathbf{b'Pb} = \mathbf{c'P^{-1}PP^{-1}c} = \mathbf{c'P^{-1}c} = \mathbf{b'c} = \sigma_{TI} \tag{5.36}$$

$$r_{TI} = \frac{\sigma_{TI}}{\sqrt{\sigma_I^2 \sigma_T^2}} = \frac{\mathbf{b}'\mathbf{c}}{\sqrt{\mathbf{b}'\mathbf{c}\sigma_T^2}} = \sqrt{\frac{\mathbf{b}'\mathbf{c}}{\sigma_T^2}} \tag{5.37}$$

However, since $\hat{\mathbf{b}} \neq \mathbf{b}, \hat{\sigma}_I^2 \neq \sigma_I^2$, and $\hat{\sigma}_{TI} \neq \sigma_{TI}$ and therefore, $\hat{r}_{TI} \neq r_{TI}$. The variance of I and the covariance of IT are also functions of h^2, the relationship weighting and the additive variance, while accuracy is a function of h^2. For a given trait, heritability is constant, as is the additive genetic variation. Accuracy is then a function of the strength of the relationship between the 1 and α, with increased records helping to increase accuracy. As a pedagogical device in deriving the accuracy of the heifer's EBV index, it will be assumed that sires have records. Then the accuracy of the heifer's EBV index based on her parents' records is

$$\begin{aligned} r_{TI} &= \sqrt{\frac{\mathbf{b}'\mathbf{c}}{\sigma_T^2}} = \sqrt{\frac{\left[\frac{0.5 n_D h^2}{1+(n_D-1)R_D} + \frac{0.5 n_S h^2}{1+(n_S-1)R_S}\right]\begin{bmatrix} 0.5\sigma_{G_A}^2 \\ 0.5\sigma_{G_A}^2 \end{bmatrix}}{\sigma_{G_A}^2}} \\ &= \sqrt{\frac{1}{4}\left(\frac{n_D h^2}{1+(n_D-1)R_D} + \frac{n_S h^2}{1+(n_S-1)R_S}\right)} \end{aligned} \tag{5.38}$$

where n_j and R_j are respectively the number of records and repeatability of parent j. The previously determined values of the **b** and **c** vectors were inserted into Equation (5.37), and since the true value being estimated is additive genetic value, $\mathrm{E}\left[T^2\right] = \mathrm{E}\left[\mathrm{G}_\mathrm{A}^2\right] = \sigma_{G_A}^2$.

Notice that Equation (5.38) can be explicitly stated as a function of the accuracy of the parents' own estimated additive genetic value. The index to predict parent j's additive genetic value using only its own records is simply twice its PTA or b_j. And the equivalent of Equation (5.37) to determine the accuracy of *j*'s PTA is

$$r_{TI}^{j} = \sqrt{\frac{\mathbf{b'c}}{\sigma_T^2}} = \sqrt{\frac{\left(\frac{n_j h^2}{1+(n_j-1)R_j}\right)\sigma_{G_A}^2}{\sigma_{G_A}^2}} = \sqrt{\frac{n_j h^2}{1+(n_j-1)R_j}} \qquad \textbf{(5. 39)}$$

where r_{TI}^{j} is the accuracy of parent *j*'s own additive genetic value.

The coefficient on **c** is one since, in the absence of inbreeding, this is the relationship of an animal to itself.

Equation (5.38) is then equivalent to

$$r_{TI} = \sqrt{\frac{1}{4}\left(r_{TI}^{S^2} + r_{TI}^{D^2}\right)} \qquad \textbf{(5. 40)}$$

As will be shown, while sires don't have phenotypic records, their squared accuracies are widely reported and available.

The accuracy of a progeny's record is considerably less than that of her parents'. Even if the parents' values are known with certainty, the EBV of progeny would explain only 50% of the variation in the animals' true additive genetic value. This perhaps surprising finding is the result of Mendelian sampling. Even if the genetic effects of the parents are known with absolute accuracy and certainty, a progeny is the summed effects of its entire genome. In the production of gametes a distribution of genotypes is produced, with a mean equal to the respective parents' genotype.

Now consider the prediction of this daughter's first record. Returning to Equation (5.18), since the previous relatives, in this example the dam and half sib, are still the only individuals upon which the index is based the phenotypic variance-covariance matrix **P** and its inverse will be the same as that used in the estimation of EBV. **c** represents the covariance between each relative and the trait of interest, namely the heifer's phenotype.

$$\mathbf{c} = \begin{bmatrix} \sigma_{X_H X_{Dam}} \\ \sigma_{X_H X_{Sib}} \end{bmatrix} \tag{5.41}$$

Making the standard selection theory assumptions of:

-zero expected covariance between genetic effects;

-only the expected genetic variance due to the additive portion is non zero;

-no expected genetic-environmental covariance;

-no environmental covariance despite the fact that heifer and dam may be in the same herd;

-if either parent has multiple records no covariance between animal effects and TE, and zero TE variance;

the first element of **c** in Equation (5.41) can be expanded as

$$\begin{aligned} \sigma_{X_H X_{Dam}} &= \mathrm{E}\left[\left(G_{A_H} + PE_H + TE_H\right)\left(G_{A_D} + PE_D + TE_D\right)\right] \\ &= 0.5\sigma^2_{G_A} \end{aligned} \tag{5.42}$$

The elements of $\mathbf{P}^{-1}$ and **c**, and therefore also **b**, are all the same for estimation of a heifer's EBV or next record. This is the result of the assumptions made as well as the inability to predict the random variable of environmental effects. Since the indices are identical so are their variances.

Now consider the variances of the true values. The variance of EBV_{milk} is simply the additive genetic variance, while for the first milk record it is the phenotypic variance.

The ratio of these two variances is then the heritability of milk production. The USDA uses a heritability of milk production for all lactations of 0.25 (personal communication, George Wiggans, Animal Improvement Programs Laboratory, Agricultural Research Center, USDA).

$$\frac{r_{record_{milk}}}{r_{BV_{milk}}} = \frac{\sqrt{\frac{0.5h^2\sigma^2_{G_A}}{\sigma^2_X}}}{\sqrt{\frac{0.5h^2\sigma^2_{G_A}}{\sigma^2_{G_A}}}} = \frac{\sqrt{\sigma^2_{G_A}}}{\sqrt{\sigma^2_X}} = \sqrt{h^2}$$

(5. 43)

$$r_{record_{milk}} = \sqrt{h^2_{milk}}\,r_{BV} = \sqrt{0.25}\,r_{BV} = 0.50\,r_{BV_{milk}}$$

So while the accuracy of estimating an offspring's additive genetic value was low, the accuracy of estimating a heifer's first record is only half as large.

Had the assumptions made in calculating **P** and **c** held true, *I* would have been an unbiased estimate of *T*, but the index would still have prediction error. The variance in the true value given a particular index, or the prediction error is

$$\sigma^2_T\Big|_{I=I_O} = \left(1 - r^2_{TI}\right)\sigma^2_T \tag{5. 44}$$

The USDA calculates reliability of PTAs which is defined as the squared correlation of an animals predicted transmitting ability and its true transmitting ability variance (Van Raden and Wiggans, 1991). Reliability is then equivalent to our previous definition of accuracy (see Equation (5.38)) squared. The confidence interval for an index is then $\sqrt{1 - \text{Reliability}} * \sigma * Z$, combining this with Equation (5.40) the confidence interval can be expressed as a function of parent reliabilities

$\sqrt{\frac{4 - \text{Reliability}_{\text{Dam}} - \text{Reliability}_{\text{Sire}}}{4}} * \sigma * Z$, where σ is the population standard deviation of the index, and Z is the appropriate standard normal probability for the specified confidence interval width.

Table 5.1 presents for EBV milk and first lactation production indices the 95% confidence intervals of the indices based on various combinations of parents' repeatabilities of milk PTA. These combinations of reliabilities are what can be expected in the population. The mean of all indices is 20,000 lb.

Table 5.1. Estimated breeding value, and first lactation production 95% confidence intervals for the range of parent reliabilities of milk production genetic value.

Milk Production (lb)			20,000						
Repeatability			95% Confidence Intervals						
			Breeding Value Milk			1st Lactation Production			
Sire	Dam					Within Herd		Between Herds	
			Variance[1]:	8,880,400		Variance[2]:	12,882,870	Variance[3]:	18,404,100
0.99	0.6		15,466	24,534		14,539	25,461	13,473	26,527
0.8	0.6		15,291	24,709		14,328	25,672	13,221	26,779
0.6	0.6		15,113	24,887		14,114	25,886	12,965	27,035
0.4	0.6		14,942	25,058		13,908	26,092	12,718	27,282
0.99	0.5		15,373	24,627		14,427	25,573	13,339	26,661
0.8	0.5		15,201	24,799		14,220	25,780	13,092	26,908
0.6	0.5		15,027	24,973		14,010	25,990	12,841	27,159
0.4	0.5		14,858	25,142		13,807	26,193	12,598	27,402
0.99	0.4		15,282	24,718		14,317	25,683	13,208	26,792
0.8	0.4		15,113	24,887		14,114	25,886	12,965	27,035
0.6	0.4		14,942	25,058		13,908	26,092	12,718	27,282
0.4	0.4		14,776	25,224		13,708	26,292	12,479	27,521

1) From INTERBULL Centre Routine Genetic Evaluation, August 1996. Table 2. Estimated sire standard deviation and genetic correlations between countries considered in the INTERBULL Centre evaluation of August 1996. The sire standard deviation multiplied by four results in the genetic standard deviation (personal communication, George Wiggans, Animal Improvement Programs Laboratory, Agricultural Research Center, USDA).

2) Within herd variance assumed to be 70% of between herd variance.

3) Jairath et al. (1995) phenotypic variation first lact production 1950 kg with a mean of 5155.
or Weigel et al. phenotypic variance first lactation milk ME 1686 kg mean 8500.

The phenotypic variance of milk production in the population was taken from the literature. Weigel et al.'s (1995) estimate of the variance of first lactation ME milk was 13,758,165 lb^2. Notice that the genetic variance is only 48% of the phenotypic variance, while a heritability of 0.25 implies that it should be 25%. The confidence intervals reported may then be too narrow.

Even the most accurate confidence intervals are quite wide. With the range of the expected or mean EBV milk being ± 1,000 lb in the population there will be considerable overlap of the prediction intervals for all animals in the population. While the intervals do narrow as the reliabilities of the parents increase, given the still considerable width of the interval for the progeny of even the most reliable parents, it is not clear whether this increase is practically significant.

Despite SI or the equivalent animal model evaluations being theoretically superior to previous methods of predicting genetic merit, this did not stop Samuelson and Pearson (1995) from recently investigating this very issue. They explored whether the daughter yield deviation and EBV of young AI bulls were more accurately predicted through the use of the parents' PTAs (i.e. EBV as defined earlier in the chapter), use of sire PTA, sire PTA and dam PTA or a pedigree index based on the PTA of the young sires' sire and maternal grandsire. EBV was found to explain the most variation in both the young sires' PTA based on daughter information, as well as, his daughters' actual yield deviations. Although, the regression coefficients were less than the expected 0.5, with sire coefficients generally slightly greater than dam coefficients, and "accuracy of prediction appeared to be declining erratically with time". They also investigated the change in the EBV over 3 periods prior to the realization of daughter information. Over this time, the R^2 between

EBV and young sires' PTA and daughter yield deviation increased, with the increase being due to improving accuracy of the dams' PTAs.

The genetic improvement per year under normality and other assumptions is

$$\Delta G/yr = {(r_{TI} D \sigma_G)}\Big/{L} \tag{5. 45}$$

Where D is a measure of selection intensity and L is the generation interval.

The rate of genetic gain has consistently been less than theoretically possible and expected using SI (Van Tassell, and Van Vleck, 1991). While numerous reasons for this shortfall have suggested, incorrect assumptions in the genetic models has not been included. Given the assumption violations, resulting in unoptimal indices, it is not surprising that frequently actual genetic progress lags that which is expected from the model. However, the method of calculating the gain must also be questioned. From Equation (5.37), the estimated r_{TI} is not equal to the true correlation, and this is assuming correct estimation of σ_T^{-2}. The inverted variance of the true value will likely be unbiased if the trait is observable, while if unobservable, such as additive genetic effect, this estimate may be biased as well. Inaccurate estimation of population parameters can lead to high overprediction of expected gains and loss in the efficiency of the index (Villanueva, and Kennedy, 1993; Hill, and Myer, 1984). The appropriateness of Equation (5.45) as a measure of genetic improvement is then unclear.

Until now only selection for a single trait using that single trait has been considered. Selection using multiple traits will now be briefly reviewed. Selection for multiple traits presents an analogous problem to that encountered previously, but now we are concerned not only with estimating the weights for different individuals, but also, how to weight the different traits in the index. Again Van Vleck's development of the solution is followed.

Let $\mathbf{g}$ = vector of genetic values of $\mathbf{m}$ economically important traits for animal α,

$\mathbf{v}$ = vector of economic values for the traits in $\mathbf{g}$,

$\mathbf{G} = E(\mathbf{gg'}) = E(\mathbf{Xg'})$ = the genetic variance-covariance matrix for $\mathbf{m}$ traits,

Consider the derivation of $\mathbf{m}$ single trait index equations analogous to Equation (5.17),

$$\hat{\mathbf{T}}_j = \mathbf{I}_j = \mathbf{b}'_j\mathbf{X} \qquad (5.46)$$

If $\hat{\mathbf{T}}_j$ is additive genetic values, $\mathbf{g}_j$ can replace $\mathbf{c}$ in Equation (5.18), where

E[X(genetic value of trait j)] = $\mathbf{g}_j$.

Equation (5.18) can then be restated as

$$\mathbf{b}_j = \mathbf{P}^{-1}\mathbf{g}_j \qquad (5.47)$$

where j = 1, ..., m.

Notice that $\mathbf{b}'_j = \left(\mathbf{P}^{-1}\mathbf{g}_j\right)' = \mathbf{g}'_j\left(\mathbf{P}^{-1}\right)' = \mathbf{g}'_j(\mathbf{P}')^{-1} = \mathbf{g}'_j\mathbf{P}^{-1}$, recalling since it is a symmetric matrix $\mathbf{P}' = \mathbf{P}$.

The overall index (I) is the sum of the $\mathbf{I}_j$ weighted by v,

$$\mathbf{I} = \mathbf{v}'\begin{pmatrix}\mathbf{I}_1\\ \mathbf{I}_2\\ \vdots\\ \mathbf{I}_m\end{pmatrix} = \mathbf{v}'\begin{pmatrix}\mathbf{b}'_1\\ \mathbf{b}'_2\\ \vdots\\ \mathbf{b}'_m\end{pmatrix}\mathbf{X} = \mathbf{v}'\begin{pmatrix}\mathbf{g}'_1\mathbf{P}^{-1}\\ \mathbf{g}'_2\mathbf{P}^{-1}\\ \vdots\\ \mathbf{g}'_m\mathbf{P}^{-1}\end{pmatrix}\mathbf{X} = \mathbf{v}'\mathbf{G}\mathbf{P}^{-1}\mathbf{X} = \mathbf{B}'\mathbf{X} \qquad (5.48)$$

Van Vleck shows that

$$\sigma_I^2 = \mathbf{B'PB} = \sigma_{TI}$$
$$\sigma_T^2 = \mathbf{v'Gv}$$
$$r_{TI} = \sqrt{\frac{\mathbf{B'Gv}}{\mathbf{v'Gv}}}$$

$$\mathbf{T} = \mathbf{v'g} \qquad \textbf{(5. 49)}$$

The true value is now expressed in economic units, while only relative economic value of the traits is required for selection. Applying Equation (5.45), if progress (ΔT) is to be measured in dollars, absolute economic values are required.

$$\Delta T = r_{TI} D\sigma_T = D\sigma_I \qquad \textbf{(5. 50)}$$

The response of a trait *j*, whether included in the index or not, to selection can be measured, by regressing G_j on I.

$$\Delta G_j = \frac{\sigma_{G_jI}}{\sigma_I^2} D\sigma_I = \frac{\sigma_{G_jI}}{\sigma_I} D \qquad \textbf{(5. 51)}$$

where

$$\begin{aligned} \sigma_{G_jI} &= \mathrm{E}\left(G_j, B_1X_1 + \cdots + B_NX_N\right) \\ &= B_1\sigma_{G_jG_1} + B_2\sigma_{G_jG_2} + \cdots B_N\sigma_{G_jG_N} \end{aligned} \qquad \textbf{(5. 52)}$$

As portrayed in Equation (5.53), selection can still be practiced for a single trait by setting the economic value of all other traits equal to zero.

$$\begin{aligned} &\text{if } v_i = 0 \forall i \neq j, \text{ then let } v_j = 1 \\ &\sigma_{G_jI} = b_{j1}\sigma_{G_jG_1} + b_{j2}\sigma_{G_jG_2} + \cdots b_{jN}\sigma_{G_jG_N} \end{aligned} \qquad \textbf{(5. 53)}$$

The use of other traits will result in more progress being made than if selection was conducted using only the trait of interest. This incremental progress is a function of the heritability of the other traits as well as the genetic correlation between the trait of interest and the other traits

Intuitively, in order to make progress in selection, variance is required in the trait of interest, such that individuals can be differentiated and selection can occur. If other traits are heritable and genetically correlated with the trait of interest, they can aid in the assessment of an individual's genetic value of the trait of interest and therefore aid selection. As shown in Table 5.2, first lactation dairy production traits of interest tend to be highly correlated.

Table 5.2. Heritabilities (on diagonal), phenotypic (above diagonal) and genetic (below diagonal) correlations among first lactation yield traits (Jairath et al., 1995).

Trait	Milk	Fat	Protein
Milk	.25	.95	.98
Fat	.89	.23	.97
Protein	.96	.92	.22

To make economic progress through SI, the economic weights must be equi-proportional to their net marginal value per unit per trait. To measure selection progress in actual dollar units the absolute values are required. Generally the national per unit price of a trait (i.e., $/pound of milk) is used as the respective weight. Prices resulting from perfectly competitive markets meet the above requirements for measurement of selection progress. For in such markets, price equals the marginal cost of production, with the latter including an allowance for "normal profit".

The criteria for a perfectly competitive market are:

1) complete markets;

2) more is preferred to less;

3) numerous consumers and producers;

4) no increasing returns to scale;

5) no externalities;

6) perfect information;

The market for raw milk meets the first two criteria, the last four are questionable. Firstly, there is concern in the dairy industry that the dairy processors, who purchase the milk from dairy producers, are few in number and possess potential market power. Recently Kraft Foods Inc. has come under investigation for manipulation of the exchange price of cheddar cheese. This price forms the basis of that received by farmers for their product.

A topic of much debate among the dairy industry has been whether increasing returns to scale exist. Increasing returns to scale implies that marginal costs are decreasing. The literature generally does not reject the null hypothesis that the industry possesses constant returns to scale. Continued nation wide expansion trends among producers appear to support the concept of increasing returns to scale.

Externalities are prevalent in agriculture in general. Externalities may be defined as production by a firm affecting the preferences of a consumer or vice versa. It is important to recognize that this is a direct effect, not one transmitted through the price mechanism. Examples of externalities in dairy production are neighbor displeasure with the smell of manure, and animal welfare concerns.

Lastly, for this market to be perfectly competitive all market participants must be omniscient. Consumers must be fully aware of the value of dairy products and all others, and producers must have perfect knowledge about their own operations as well as that of others. Given that a significant number of producers are not even aware of their cost of production per hundred weight of milk, much less their marginal cost of production for milk fat, milk protein, and milk less the components, the perfect information criteria can be refuted. Clearly then the dairy market is not perfectly competitive. What is not so clear is whether the violations are significant enough to suggest the withdrawal of prices as the economic weights, even to the point of not using any weights under the auspices that more harm may be ensuing under this inappropriate plan than without.

A simulation revealed that a ± 50% error in any single economic weight in a seven trait swine selection index caused less than a one percent loss in relative efficiency of the index (Vandepitte, and Hazel, 1977). The study showed that underestimation of economic weights generally caused greater errors than overestimation, and the effect were nonlinear, and could be sizable with greater than 50% error. As the authors point out the results are specific to the **P** and **G** matrices and economic weights used in the example and any generalization must be approached cautiously. It is not clear how well these results extrapolate to stableness of an index such as PTA$ with fewer traits, different covariances, and differing relative magnitudes of the economic weights.

Lin and James also found that errors in relative economic weights result in loss of efficiency with selection index, but the predicted losses are not highly sensitive to this type of error (Villanueva, and Kennedy, 1993).

While SI is based on linear estimations, Dekkers, Van Erp, and Schukken, (1996) showed that the economic weight for savings in penalty costs by reducing somatic cell count in Ontario were not linear. The authors recognized that penalty costs are only a part of the savings associated with improved SCC, however, there is no mention of the importance of considering the cost of attaining lowered somatic cell counts. The authors' recommendation of a target SCC appears quite arbitrary. Similarly, the use of the average marginal value of SCC reduction across herds is suggested as the proper economic weight for use in a population wide breeding program. It is not clear what objective function this suggestion is intended to maximize or the basis on which this decision was made. Given the availability of the marginal economic value of lowered SCC for a given herd SCC level, and the advent of personal computer it is surprising that the creation of individually customized indexes based on farm customized economic weights was not proposed.

Sampling error variance in estimating economic weights has been noted to also increase the estimated genetic gain, overestimate the attainable genetic gain, as well as, the realized gain (Vandepitte, and Hazel, 1977).

Other pragmatic problems also arise with the use of price as the economic weights. Many traits of interest do not have readily available prices, for example somatic cell count and type traits. The weights for these traits appear to be arbitrarily chosen. Clearly the net marginal returns per trait is a function of individual operations. This is not however reflected in the price. Furthermore, national price estimates are used despite many regions of the country increasingly having separate milk pricing formulas, especially with the advent of component pricing. Hazel (1943), a pioneer in SI development, recognized that the proper economic values may vary by region. The relationship of national prices to the true net economic values appears ambiguous. He also suggested that since the results of current selection are not apparent until the future that the economic values should represent the anticipated future values.

A number of technical difficulties are also encountered by the USDA in doing the calculations. The USDA Animal Model includes a sire x herd interaction term. Recent work (Dimov, Albuquerque, Keown, et al., 1995) suggests that the estimate used in the model may overestimate the true effect of this interaction term by a magnitude of five to ten. Overestimation of this effect may lead to underestimation of genetic merit of superior sire and dams, and therefore incorrect animal ranking. Differences in the heritability of milk yield (0.21 versus 0.28) were also recorded among different regions of the country.

Preferential treatment of animals also appears to be a problem in the US. Preferential treatment can include any management practice that increases production and is applied to one or several cows but not contemporaries (i.e. $\sigma_{G_A E} \neq 0$, for some individuals). Assuming preferential treatment among dams with high cow indexes, Powell and Norman (1988) found the milk yield deviations of daughters of these dams were higher than expected. This underestimation of daughter performance was thought to have occurred due to preferential treatment of this group of daughters. There is also evidence that such effects occur in the animal model estimation of cow PTAs (Kuhn, Boettcher, and Freeman, 1994). The bias of PTAs resulting from preferential treatment can be sizable, with 0.06 to 0.39 of the increased production resulting from the special treatment being transferred to her estimate of PTA. Preferential treatment is also considered to occur in genetic evaluations of cows for selection of their sons as sires (Murphy, Everett, and Van Vleck, 1982). Such work has resulted in the suggestion to consider only the first record of a sire's dam be used in predicting a bull's evaluation.

In estimating the EBV of animals, an adjustment must be made to remove the variation in yield records due to systematic environmental effects. "Genetic trend in milk yield was sensitive to the choice of adjustment procedure"(Khan, and Shook, 1996). The current method of adjustment by USDA biases the estimates of poor sires downwards and good sires upward, although it is not clear whether the rankings of cows or bulls are changed.

Up until this point only nuclear genetics have been considered. While the embryo receives half of its nuclear genetic material from each the sperm and ovum, the latter contributes all the cytoplasm. This maternal cytoplasm which contains mitochondrial DNA can contribute to the offspring's genome. While the effects of cytoplasmic inheritance on dairy production traits is disputed in the literature, failing to account for cytoplasmic variance, if present, causes overestimation of additive genetic variance and underestimation of permanent environmental variance, as a result heritability is overestimated (Boettcher, Kuhn, and Freeman, 1996). This also causes slight errors in the accuracy of selection of cows. The USDA Animal Model does not currently include cytoplasmic effects.

The purely tactical application of SI can also present some problems. There are a number of commonly used indices in which may include such factors as production, type traits, and possibly genetic evaluations of somatic cell counts and herd life. Even overlooking the caveats from chapter 2 on the effect of ignoring opportunity costs on the relative value of herd life and production, it is not clear how the value of herdlife can be assessed when, based on the information on optimal culling rates in Chapter 3, it would appear that in general herdlifes are less than optimal. There are still other pragmatic problems with the determination of which of many indices to use. Despite similar correlations between Canadian indices, the reranking of animals was found to be substantial (Dekkers, Kolstad, Jairath et al., 1996).

It has been noted that the average gain expected from use of an index is not appropriate for comparison of indexes since sampling is only occurring once. Rather the variance and expected gain should be considered when comparing indexes (Williams, 1962). Expected gain is an average of realized gain over all samples and is unaffected by the sample-to-sample index variance. The probability of correctly ranking individuals is a function of the error variance of the index. So while two indices can vary in expected genetic gain there may be only negligible differences in the probability of correctly ranking of individuals, and therefore little real difference between indices. In terms of heifer selection there is little to indicate what index would best rank animals on future profit.

The dynamics of selection have not been totally ignored by geneticists. Weigel, Cassel, and Pearson (1995) considered what proportion of each decile of potential AI bulls, ranked by EBV$ at 1 and 3 years of age, subsequently proved to be above the 85th percentile of their sampling group. The analysis was simplistic and unsophisticated often relying on inaccurate charts. The final recommendations of percentages of bulls to cull at each age are arbitrary. Conclusions were made as to the economic impact on the AI centers, this despite the fact that neither the costs of raising the bulls, nor the costs of the incorrect classifications were considered. All incorrect classifications were considered equal despite the great range in returns generated by the bulls.

Fortunately, more sophisticated and appropriate techniques have been used to assess the dynamics of animal selection. Interest has stemmed from the potential economic savings of sequentially selecting animals as traits of interest are revealed over time, rather than the traditional selection index approach of waiting until all traits of interest can be measured before selection occurs. Given the multivariate normally distributed traits of interest and their economic weights(v), **P**, **G**, the number of stages over which selection is to occur, and the total proportion to be selected numerical integration can be used to optimize an objective function such as aggregate breeding value. The information on those individuals which haven't been culled in previous periods is accumulated and used in selection at subsequent periods. Conventional independent culling, multistage index selection, multistage selection, and selection index updating are equivalent terms used to denote this technique which will solve for the for the optimal truncation points at each selection period.

An orthogonal transformation procedure of the traits has recently been developed (Xu and Muir, 1992). Since subsequent distributions are no longer a function of previous selection, numerical integration is not needed, with the concomitant saving in CPU time. However, the required computer time with numerical integration was not inhibitive. The relative efficiency of this transformation compared to numerical integration in maximizing aggregate breeding value for various scenarios was reported at 87% to 300% (Xu and Muir, 1991). Given that the transformation places a restriction on the solution space, it is not clear how it can result in increased efficiency. The cited references do not support the explanation provided in the paper.

Given the cost of measuring the various traits, the objective function of multistage index selection can include indicators of profit. For a particular example it was shown that when the objective function included the cost of the increase in aggregate breeding value, the results of multistage index selection varied greatly from the results where the focus was purely on increasing aggregate breeding value without regard to cost (Xu and Muir, 1992). In particular, both the aggregate breeding value and costs were lower when cost was included in the function to be maximized. However, it should be noted that the profit function used in the study was incorrect.

There is no record of these techniques having been used for selection. It would seem that an approach that lies between the lax method of analysis used by Weigel et al. and the restrictiveness of multistage index selection would be appropriate. Note that the latter maintains the hypotheses of multivariate normal distributions and uses estimates of **P**, **G**, and **v**, which as discussed earlier in the chapter will not result in the optimal indices in each period, and the level of bias across individuals may well vary. Also multistage index selection requires the intertemporal phenotypic and genetic correlations among the traits used for selection. These correlations are currently unknown and genetic theory is silent on potential values.

In summary the methodology of selection index for single and multiple traits has been considered. Even assuming optimal weights for the indices, the proportion of explained variance of an animal's EBV, much less first record, is very low due to Mendelian sampling. Furthermore, based on all the assumptions made in deriving the variances, covariances, and economic values, the index weights clearly are not optimal. But even more worrisome is the fact that, the level of inoptimality may vary across individuals resulting in inconsistent rankings. SI will also fail to correctly rank individuals if the genetic and phenotypic values of the trait of interest are not normally distributed.

Given these difficulties, it is not clear whether any genetic indices will be of greater value in heifer selection than their cost. Furthermore, should they be of some value, it is unknown which indices should be used and at what time periods in a heifer's prepartum life they should be applied to permit heifer selection to maximize producer profit.

BIBLIOGRAPHY

Boettcher, P. J., M. T. Kuhn, and A. E. Freeman. 1996. Impacts of cytoplasmic inheritance on genetic evaluation. J. Dairy Sci. 79:663-675.

Dekkers, J. C. M., B. Kolstad, L. K. Jairath, and Z. Liu. 1996. 1996 Dairy research report, University of Guelph, Publ. No. 0396:15-18.

Dekkers, J. C. M., T. Van Erp, and Y. H. Schukken. 1996. Economic benefits of reducing somatic cell count under the milk quality program of Ontario. J. Dairy Sci. 79:396-401.

Dimov, G., L. G. Albuquerque, J. F. Keown, L. D. Van Vleck, and H. D. Norman. 1995. Variance of interaction effects of sire and herd for yield traits of Holsteins in California, New York, and Pennsylvania with an animal model. J. Dairy Sci. 78:939-946.

Goffinet, B. 1983. Selection on selected records. Genet. Sel. Evol. 15:91-98.

Hazel, L. N., G. E. Dickerson, and A. E. Freeman. 1994. Symposium: selection index theory; the selection index-then, now and for the future. J. Dairy Sci. 77:3236-3251.

Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics. 28:476-490.

Harville, D. A. 1972. Maximum likelihood approach to variance component estimation and to related problems. J. Am. Stat. Assoc. 72:320-338.

Hill, W. G., and K. Myer. 1984. Effects of errors in parameter estimates on efficiency of restricted genetic selection of indices. In: Hinkelmann, K (ed). Experimental Design, Statistical Models, and Genetic Statistics. Essays in honor of O. Kempthorne. Marcel Dekker, New York, pp 345-367.

Jairath, L. K., J. F. Hayes, and R. I. Cue. 1995. Correlations between first lactation and lifetime performance traits of Canadian Holsteins. J. Dairy Sci. 78:438-448.

Khan, M. S., and G. E. Shook. 1996. Effects of age on milk yield: Time trends and method of adjustment. J. Dairy Sci. 79:1057-1064.

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

McAllister, A. J., A. J. Lee, T. R. Batra, C. Y. Lin, G. L. Roy, J. A. Vesely, J. M. Wauthy, and K. A. Winter. 1994. The influence of additive and nonadditive gene action on lifetime yields and profitability of dairy cattle. J. Dairy Sci. 77:2400-2414.

Meyer, K. 1983. Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. J. Dairy Sci. 66:1988-1997.

Misztal. I. 1997. Estimation of variance components with large-scale dominance models. J. Dairy Sci. 80:965-974.

Murphy, P. A., R. W. Everett, and L. D. Van Vleck. 1982. Comparison of first lactations and all lactations of dams to predict sons' milk evaluations. J. Dairy Sci. 65:1999-2005.

Philipsson, J., G. Bangos, and T. Arnason. 1994. Present and future uses of selection index methodology in dairy cattle. J. Dairy Sci. 77:3252-3261.

Powell, R. L., and H. D. Norman. 1988. Accuracy of cow indexes according to repeatability, evaluation, herd yield, and registry status. J. Dairy Sci. 71:2232-2240.

Rothschild, M. F., C. R. Henderson, and R. L. Quaas. 1978. Comparison of maximum likelihood and Henderson's method 1 estimates of variances and covariances for two traits when selection is present. J. Dairy Sci. 61 (Suppl. 1):91. (Abstr.)

Samuelson, D. J., and R. E. Pearson. 1995. Accuracy of predicting genetic merit from pedigree information for bulls entering stud sampling programs. J. Dairy Sci. 78:2057-2066.

Schaeffer, L. R., and C. R. Henderson. 1983. Best linear unbiased prediction when error vector is correlated with other random vectors in the model. Genet. Sel. Evol. 15:395-400.

Schaeffer, L. R., and H. Song. 1978. Selection bias and REML variance-covariance component estimation. J. Dairy Sci. 61 (Suppl. 1):91. (Abstr.)

Tempelman, R. J., and E. B. Burnside. 1990. Additive and nonadditive genetic variation for production traits in Canadian Holsteins. J. Dairy Sci. 73:2206-2213.

Tempelman, R. J., and E. B. Burnside. 1991. Additive and dominance genetic variation for production traits under an animal model. J. Anim. Breed. Genet. 108:330-342.

Van der Werf, J. H. J., T. H. E. Meuwissen, and G. De Jong. 1994. Effects of correction for heterogeneity of variance on bias and accuracy of breeding value estimation for Dutch dairy cattle. J. Dairy Sci. 77:3174-3184.

Van Raden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737-2746.

Van Tassel, C. P., and L. D. Van Vleck. 1991. Estimates of genetic selection differentials and generation intervals for four paths of selection. J. Dairy Sci. 74:1078-1086.

Van Vleck, D. L. 1993. Selection Index and Introduction to Mixed Model Methods. CRC Press, Boca Raton

Vandepitte, W. M., and L. N. Hazel. 1977. The effect of errors in the economic weights on the accuracy of selection indexes. Ann. Genet. Sel. Anim. 9:87-103.

Villanueva, B., and B. W. Kennedy. 1993. Index versus tandem selection after repeated generations of selection. Theor. Appl. Genet. 85:706-712.

Weigel, D. J., B. G. Cassel, I. Hoeschele, and R. E. Pearson. 1995. Multiple trait prediction of transmitting abilities for herd life and estimation of economic weights using relative net income adjusted for opportunity cost. J. Dairy Sci. 78:639-647.

Weigel, D. J., B. G. Cassel, and R. E. Pearson. 1995 Relative genetic merit and effectiveness of selection of young sires for artificial insemination. J. Dairy Sci. 78:2481-2485.

Williams, J. S. 1962. The evaluation of a selection index. Biometrics 18:375-393.

Xu, S., and W. M. Muir. 1992. Selection index updating. Theor. Appl. Genet. 83:451-458.

Xu, S., and W. M. Muir. 1991. Multistage selection for genetic gain by orthogonal transformation. Genetics 129:963-974.

Chapter 6

LITERATURE SUMMARY AND HYPOTHESES

Literature Summary

The literature review can be summarized by the following points:

1) Optimal culling rates in dairy production are a complex economic question. This issue is best structured as an infinite capital asset replacement model with nonidentical replacement. The question is best solved with stochastic dynamic programming.

2) Stochastic dynamic programming and simulation models suggest the optimal culling rate is between 20 and 30%, assuming risk neutrality of producers, a constant herd size, culls are sold for beef not dairy, and no strong seasonal effects on milk production or reproduction. Higher culling rates are associated with increased milk production but not increased profit. Optimal culling rates are not impacted by the assumed rate of genetic improvement. With the cost of replacements being high relative to the cost of culls, as occurs in North America, maximizing genetic progress or milk production through maximum culling rates is not economically optimal.

3) The excessive culling rate in Michigan indicates producers have excess heifers, and a means of heifer selection is required. To maximize profit, heifers should be selected for lifetime profit corrected for the opportunity cost of postponed replacement. First lactation milk is a proxy for lifetime profit corrected for the opportunity cost of postponed replacement.

4) Genetic evaluations by the USDA are the only widely available means of selecting

heifers. Theoretically, animal model genetic evaluations should be useful in predicting first lactation milk production, and therefore, lifetime profit corrected for the opportunity cost of postponed replacement. However, there is little empirical support for the ability of Holstein parent PTAs to predict subsequent daughter milk production.

5) Given that parent PTAs are intuitively useful in heifer selection, no consideration has been given to a selection scheme for heifers. This scheme would need to address not only which of the potential genetic evaluations (i.e., milk, fat, protein, reliabilities) of each parent would be used, and how they would be used, but also, in which period(s) would they be used, for the dynamic nature of the selection process needs to be accommodated. New genetic estimates are produced every six months providing 4 complete sets of estimates over the course a heifer being raised.

Based on the literature review the following hypotheses were generated:

Statement of Hypotheses

Hypothesis I: Parents' PTAs from the last rearing period will not predict first lactation milk.

Hypothesis II: Genetic indices based on parents' PTAs from the first rearing period will be of little value in estimating whether culling occurred prior to, or during, the first lactation.

Hypothesis III: Prepartum heifer selection based on their parents' PTAs and associated reliabilities, of milk, fat, and protein from throughout the rearing process, will be more profitable than random heifer selection.

Hypothesis I, along with various estimation methods are evaluated in Chapter 7. Chapter 8 considers hypothesis II, again evaluating various methods to estimate the relationship between culling and parents' PTA. And the value of genetic evaluations in prepartum heifer selection is addressed in Chapter 9.

Chapter 7

PREDICTION AND RANKING OF FIRST LACTATION MILK USING PARENT PTA

ABSTRACT

The objectives of this study included determining the ability of parents' PTA for milk, fat and protein to predict subsequent first lactation milk production of daughters; the ability of the predictions of first lactation milk production to correctly rank heifers based on actual milk production; and finally whether, in estimating this relationship, the herd effect needed to be estimated as fixed effects. The data consisted of first lactation mature equivalent milk production of 5,123 Michigan Holstein heifers, and the heifers' parents' animal model evaluations immediately prior to the heifers' freshenings. While statistical comparison of the estimates resulting from treating herd as a fixed or random effect revealed they were different, the practical differences were not as clear. Regardless of modeling technique, parents' genetic evaluations explained less than 7% of the variation in subsequent daughter first lactation milk production. A nonparametric comparison of the within herd rankings of the predictions from the random and fixed effect models suggested the rankings were substantially similar. While diffuse, the distribution of within herd rank correlations between predicted mature equivalent milk production, based on parents' genetic evaluations, and the actual mature equivalent milk production, was skewed to the left and suggested the evaluations may be useful in prepartum heifer selection. The omission of heifers which were culled prior to first lactation, did not appear to cause

selection bias in the sample data.

(**Key words:** random effect, fixed effect, prediction first lactation mature equivalent)

Abbreviation key: FE = fixed effect, **MEM** = mature equivalent milk, **OLS** = ordinary least squares, **RE** = random effect.

INTRODUCTION

Given a constant herd size, no seasonal effects on reproduction and production, and culled cows being sold for meat (not dairy) purposes, optimization (5, 18, 23, 28, 33) and simulation (1, 2) studies were unanimous in their conclusions that a culling rate of 20% to 30% optimizes producer profit. While other simulation studies (4, 19, 20, 25), which considered subjectively selected culling rates or strategies for comparison, indicated that profit would be increased with culling rates below the average 37% culling rate[1] in the Michigan dairy industry in 1996 (24). Comparability of the actual culling rate and those in the studies assumes no herd contraction or internal expansion. The results of these multiple studies are compelling because, not only did the various models employ different assumptions and analytical techniques, but as well, the models operated under economic conditions representative of a number of countries (i.e., the U.S.A, England, Ireland, Netherlands). In the models, the increased profit associated with use of the lowered culling rates varied from 2% to 40%.

Comparison of only the mean profit by the studies implies risk neutrality, whereas a majority of Michigan and Dutch dairy producers are risk adverse (17). The risk

---

[1] Culling rate, annually defined as, the number of animals leaving the herd due to sale or death divided by the average herd size.

preferences of decision makers can have significant effects on their choices, so the optimal rule assuming risk neutrality cannot necessarily be extrapolated to those who are risk averse.

At least under risk neutrality, one direct ramification of Michigan's excessive culling rate, assuming herd sizes have at least been constant, is that Michigan dairy producers do not need all potential replacement dairy heifers to enter the lactating herd, necessitating a method of heifer selection.

This need for heifer selection is in direct contrast to the present situation in which most producers rear and freshen the majority of heifers produced on the farm (6, 14). The simulation studies revealed that while this practice maximizes genetic progress and milk production, the economic performance is lower due to failure to consider the opportunity costs of higher culling rates.

While the culling models suggest the need for heifer selection, they lend little insight into how selection should be conducted. Most optimization studies employed random replacement selection. Some simulation studies arbitrarily selected replacements based on their simulated EBV of milk. While this approach to selection has intuitive appeal and indeed theoretical justification (15), empirical support for heifer selection on this basis is lacking. If a heifer selection rule more profitable than those used in the optimization and simulation models can be developed, producer profit may increase more than 2% to 40%.

While use of genetics to select heifers is appealing, application of finance theory may suggest otherwise. A standard result of capital asset replacement theory is that as the expected profitability of the potential replacements rises relative to the existing assets, the optimal length of the utilized assets' lives decrease (27) which is equivalent to an increase in the replacement rate. Therefore, if EBV were valuable in selecting more profitable heifers, the simulation model which utilized this method of heifer selection should have had higher optimal culling rates than those studies which used random heifer selection. However, the study (1) which based replacement selection on EBV had a lower optimal culling rate than all of the studies which randomly selected heifers (2, 5, 18, 28, 33), except one (23). One possible explanation for the similarity between the most profitable culling rates of the studies, despite the use of selection based on EBV milk versus random replacement selection, is that the genetic information is not highly valuable in selecting heifers for profitability. An alternative explanation is that differences among the models may preclude the recognition of the value of EBV in heifer selection.

Given that heifer selection is required, the first step in determining a method of selection is defining the objective. The objective of heifer selection is to choose the most profitable animals. Van Arendonk (32) showed that lifetime profit corrected for the opportunity cost of postponed replacement is the correct method for ranking animals in terms of profitability. Unfortunately, corrected lifetime profit of animals is generally not available and furthermore, can not be determined prior to the culling of the animals. However, the literature has focused on indicators of lifetime profitability that are apparent early in the productive life of milk cows. No studies have considered the relationship between information available prior to productive life and any measure of subsequent lifetime profit, much less lifetime profit corrected for the opportunity cost of postponed replacement.

The phenotypic correlation between first lactation actual milk and lifetime profit corrected for opportunity cost has been reported at 0.54 (3). Weigel et al. (35) determined the phenotypic and genetic correlations between corrected lifetime profit and first lactation mature equivalent milk production to be 0.51, and 0.80, respectively. For purposes of heifer selection, first lactation milk production may then be a suitable proxy for corrected lifetime profit.

Presently, alternative predictors of subsequent milk production, such as the concentration of serum β-lactoglobulin 26 weeks prepartum (22) and genetic molecular markers (7, 8), are not widely available. From Lush and Hazel (13) to the more recent mixed model genetic prediction by Henderson (15), there is a strong theoretical basis for using EBV in predicting most probable producing ability. EBV provide estimates of the genetic merits for milk production which should on average manifest as phenotypic performance. Prior to utilizing any management technique, including heifer selection, managers reasonably require, if possible, some evidence of the technique's expected performance.

Since 1989, when the USDA began using the animal model for estimating genetic merit, only two domestic empirical studies have, in a temporally correct sense, considered the ability of parents' PTA to predict a heifer's subsequent milk production (22, 31). These studies are temporally correct for the purpose of heifer selection in that the parents' PTA were those available prior to the heifers' production. Contrast this with the situation in which the parents' PTA were generated some years after the heifers calved and therefore were based on information that was unavailable at the time of prepartum heifer selection. Notice this information unavailable at the time of selection would likely include the very production records of the daughters which the parents' PTA were to predict. Neither of the two temporally correct studies included EBV fat or EBV protein, along with EBV milk, in the prediction of daughter performance.

Based on 76,871 daughters, VanRaden et al. (31) found the correlations of average parent animal model evaluations with daughter milk and fat performance ranged from 0.20 to 0.49. Daughter performance was measured as yield deviation of standardized milk yield.

In heifer selection the producer is interested in predicting standardized milk yield. While the use of yield deviation as a dependent variable serves to decrease the error variance, the results are not directly applicable to heifer selection for a producer is unlikely to know the value of management group, permanent environment and herd-sire interaction effects for each heifer prior to parturition, and therefore, could not derive predicted milk yield from predicted yield deviation. Given that producers are paid based on milk yield, their interest would lie in prediction of milk yield, perhaps standardized for age and season of calving. Even if the expectation of yield deviation was identical to the expectation of milk yield, the accuracy of the milk yield predictions would affect producers' acceptance of this method of heifer selection.

The second study (22), based on a small group of 26 heifers in a single herd, found that the EBV milk had negligible value ($R^2 = 0.034$) in prediction of first lactation 305d milk yield.

With heifer rearing, and the associated potential costs and revenues, occurring over the span of many months and new genetic information provided by USDA quarterly, heifer selection utilizing genetic evaluations is potentially a dynamic issue. As suggested by Quinton and Smith (26), there has been little empirical checking of the efficacy of genetic evaluations. They found that the rank correlations with progeny phenotype were 36% and 27% higher for parental BLUP genetic evaluation than for parental phenotypes of fat depth and days to 100 kg, respectively. Prior to developing a dynamic heifer selection rule to maximize profit based on EBV, the empirical ability of EBV to predict milk production should be ascertained.

A producer, prior to making an informed decision about implementing a heifer selection scheme based on parents' PTA, would prefer to see empirical evidence of the scheme's performance in the real world. This preference is only reasonable especially since some selection has already occurred in terms of which sires and dams to breed. Managers are faced with limited time to complete a potentially nearly infinite number of management tasks. With some understanding of the rule's potential performance, the producer could evaluate whether the performance was sufficient to warrant foregoing some other management task in order to employ the selection scheme.

Given the high genetic correlations between milk and fat, and between milk and protein, inclusion of parents' PTA fat and PTA protein, along with PTA milk, would increase the ability to predict first lactation milk production of daughters. The disadvantage of selecting for the real producing ability of a single trait (i.e., milk) based on multiple traits (i.e., milk, fat, and protein) is the theoretical weights for the index are

unique for each combination of the number and source of records that contributed to each parents' PTA. Specifically, the nondiagonal elements of the variance-covariance matrix of the multiple traits are functions of the number and source of records that contribute to the records of the traits. Furthermore, given that the weights for any type of record (i.e., milk, fat, protein) would rarely be the same for sire and dam, there is no theoretical basis to equally weight the same type of record of each parent and base multitrait analysis of MEM on heifers' EBV.

This inability to determine the theoretical weights is not critical in an empirical analysis where the focus is on the ability of information, available to producers, to predict heifer milk production. Furthermore, this recognition that there is not a single optimal vector of theoretical weights for an analysis, while explicit, is not unprecedented in Selection Index or mixed models. Relaxing assumptions in genetic evaluations can easily result in multiple optimal values for any single theoretical weight. For example, the bias in breeding values resulting from assuming homogeneity of variances at a minimum dictate the need for herd level corrections, if not animal level corrections in the weights (Van der Werf). The assessment of whether to base milk production prediction on a single trait versus multiple traits reduces to whether the theoretical ability to derive a single vector of optimal weights is preferred over the additional information provided by the additional traits.

In selection schemes the actual prediction of performance is not as important as the ability to correctly rank the animals in relative terms. Parents' PTA would be useful in heifer selection if models using this information resulted in within herd heifer rankings identical to those based on actual first lactation milk production.

Finally, the pragmatic issue of appropriate estimation technique must be considered. The apparent paradox of estimating herd effect, which intuitively can be considered a random variable, as a fixed effect (FE) is explained by Henderson (16). He stated that, should the other independent variables be correlated with the herd effect, the bias in estimation can be eliminated through estimating the herd effects as fixed, though at the price of larger variance of the estimates.

It can be inferred that Henderson recognized that estimates resulting from treating the unobserved effect, or categorical variable, as random or fixed were not necessarily different (21). Specifically, as herd size increases, or the variance of the herd effect increases relative to the variance of the error term, the estimates from random effects (RE) and fixed effects estimation converge, even if the herd effect is correlated with the independent variables. A priori, it is difficult to assess whether RE will converge to FE. So given a likely correlation between genetics of a herd and the herd effect, and small herd sizes, Henderson reasonably opted for the more conservative FE estimation. A few years after Henderson's publication, Hausman (12) developed a statistical test of the equivalence of RE and FE estimates, permitting statistical evaluation of whether RE is consistent, or if FE is required.

The objectives of this study were to determine: 1) the ability of parents' PTA for milk, fat and protein to predict subsequent first lactation milk production of daughters; 2) the ability of the predictions of first lactation milk production to correctly rank heifers based on actual milk production; and 3) whether, in estimating the relationship in objective one, the herd effect needed to be estimated as fixed effects.

MATERIALS AND METHODS

Data Collection

A total of 10,854 Holstein heifer records were collected from Michigan DHI. Only Holstein heifers that met all the following criteria were included: 1) as of June 30, 1995, identified as prepartum with at least one parent's PTA for milk, fat, or protein recorded; 2) born between July 1, 1992 and December 31, 1992; 3) a calving date between July 1, 1994 and December 31, 1994; 4) first lactation mature equivalent milk (MEM) production of greater than zero.

Genetic data were the heifers' parent PTA for milk, fat and protein from the July 1994 sire and dam animal model evaluations calculated by the Animal Improvement Programs Laboratory, USDA (Beltsville, MD). Heifer information including birthdate and calving date were collected from Michigan DHI animal identification records. The production data, which consisted of MEM production, were collected from Michigan DHI cow production records. MEM, standardizes production for age and season of at calving.

After deleting heifers that were culled or missing the genetic information of either parent, records of 5,123 heifers were available for estimation of the empirical selection index weights. In effect then this data sample consisted of, for each of the 716 herds, a cohort of heifers which were born within a six month window, and then, two years later, calved in a six month window. The number of heifers in a cohort for a given herd varied in size from 1 to 165 heifers with median and mean sizes of 4 and 7.2, respectively.

Estimation of Empirical Model

The unbalanced general population model to empirically estimate the relationship between parents' PTA and standardized (i.e., deviated from the mean and divided by the standard deviation) first lactation MEM was:

$$y_{ha} = \mathbf{x}_{ha}\beta + c_h + u_{ha}$$

where

y_{ha} = standardized first lactation MEM of animal a in herd h;

$\mathbf{x}_{ha}$ = 1 x 6 row vector. The six elements of $\mathbf{x}_{ha}$ are the standardized milk, fat and protein PTA of individual a's parents, from the July 1994 animal model evaluations calculated by the Animal Improvement Programs Laboratory, USDA;

β = vector of empirical regression coefficients;

c_h = unobserved random variable representing the effect of herd h;

u_{ha} = the idiosyncratic error term.

Three estimation techniques were considered to determine the nature of the unobserved herd effect. The first method of analysis was ordinary least squares (OLS) pooled over all observations. In the OLS regression, MEM was regressed on parents' PTA ignoring the herd effect. This technique results in the herd effect being included with the idiosyncratic error term, to create a composite error term. Standard OLS regression mechanics indicates the estimates of the empirical selection weights will be consistent if, for each heifer, neither component of the composite error term is correlated with the PTA of the heifers' parents. A variance estimator robust to correlation among the residuals was required to account for the presence of the unobserved herd effect in the composite error term.

Making the assumption that the herd effects all originated from the same distribution, and therefore had a common variance, generalized least squares was used to estimate a RE model. The estimates will be asymptotically consistent and more efficient than OLS if the unobserved herd effect and the idiosyncratic error term for any heifer in a herd were uncorrelated with the parents' PTA of all heifers in the same herd.

Finally, the herd effect was estimated as a FE. Fixed Effects estimation or within estimation in the econometrics literature (10), can be viewed as conducting OLS estimation on transformed data. Because the value of the herd effect was not of interest, the data were transformed to eliminate the unobserved effect. The transformation, also known as herdmate deviation or within herd deviation, involved herd demeaning each observed variable in the regression (e.g., y_{ha}, x_{ha}). For example, the value of sire PTA milk averaged over all heifers in a given herd was subtracted from the sire PTA milk of each heifer in the herd. Because the mean of a constant is itself, and by definition the herd effect is constant for all heifers in the herd, this transformation removed the unobserved herd effect. The resulting FE estimates were identical to those from a regression where a dummy variable was inserted for each herd. These estimates will be consistent assuming the idiosyncratic error of each heifer in a herd is uncorrelated with the parents' PTA of all heifers in the same herd.

The RE estimator lies within the continuum spanned by OLS and FE (21). The RE estimates approach the OLS estimates as the ratio of the unobserved effect variance to the idiosyncratic error variance approaches zero. In contrast, the RE estimates approach those of the FE as the ratio of the variances approaches infinity, or as the number of animals in each herd gets large.

Statistical Comparison of Empirical Regression Coefficients

Hausman (12) developed a general specification test for the null hypothesis that an asymptotically efficient estimator must have zero asymptotic covariance with its difference from a consistent but asymptotically inefficient estimator. The Hausman test is a Wald based test. The Hausman test can be applied in the comparison of the RE estimates with either the FE or OLS estimates. Maintained hypotheses of the Hausman test are that the idiosyncratic errors are homoskedastic, are not correlated with themselves, and are not correlated with the independent variables.

If the unobserved effect and the idiosyncratic error are uncorrelated with the independent variables, the OLS and RE estimates should be asymptotically consistent, with RE being efficient, and the Hausman test would not be expected to reject the null hypothesis. Should either of these assumptions be false, each set of estimators would be inconsistent with generally different probability limits, and the Hausman test should have some power to detect this.

The more traditional use of the Hausman test is in comparison of the RE and FE estimates. If herd is uncorrelated with PTA, both sets of estimates would be asymptotically consistent, and RE estimators efficient, again assuming the independent variables are uncorrelated with the idiosyncratic error. However, should herd be correlated with PTA, the FE estimates would still be consistent while the RE estimates would not, and the Hausman statistic should reject the null hypothesis.

To assess the practical significance of different empirical weights in prediction of milk production, a nonparametric approach was used. For with the focus of using these weights to select animals, it is the resulting within herd rankings of animals, not the prediction error, that is paramount. Spearman's within herd rank correlations, compared the predicted first lactation MEM based on the RE and FE estimators to each other, as well as, to the rankings based on actual MEM.

Statistical Evaluation for Potential Selection Bias

Of the total 10,854 Holstein heifer records collected from Michigan DHI, it was determined that 1,546 heifers with complete parent PTA from the USDA July 1994 animal model evaluations were culled prior to first lactation. A heifer was identified as culled based on the following criteria: 1) as of June 30, 1995, identified as prepartum with at least one parent's PTA for milk, fat, or protein recorded; 2) born between July 1, 1992 and December 31, 1992; 3) if culled, the culling date was prior to July 1, 1994; 4) if culled, a culling code was recorded; 5) as of January 4, 1996 no first lactation MEM record.

Culled animals were then born within the same time window as the 5,123 heifers which had a first lactation. The culled heifers were either identified as culled prior to first lactation, or had not initiated their first lactation by at least three years of age. Heifer information including birthdate, calving date, culling date and culling code were collected from Michigan DHI animal identification records. The MEM values were collected from Michigan DHI cow production records.

With 23% of the heifers with parent PTA being culled prior to first lactation, the potential for selection bias existed. Selection or sample bias exists if the presence of MEM is based on the idiosyncratic error (i.e., the portion of MEM not explained by parent PTA). For example, selection bias could exist if producers assessed heifers' milk production, then kept and identified to Michigan DHI for record evaluation only the higher producing heifers. In the presence of selection bias, the model results are applicable only to the sample, rather than to the population in general.

To be representative of the population in the presence of selection bias, the model based on the selected sample has another term, denoted the inverse Mills ratio (36). The inverse Mills ratio reflects that only the selected animals have been included in the sample. The addition of this term, in effect, transforms the population model into a tobit model or censored model (Greene) which, in the canonical form, explicitly recognizes that the dependent variable is observed only if it exceeds some value (e.g., $y_{ha} > 0$).

Following the approach of Wooldridge (36), the sample was examined for evidence of selection bias using a standard two step estimation procedure. In the first step the inverse Mills ratio was estimated using a pooled probit, accounting for the herd effect as a FE. In this first step, the data sample consisted of 6,669 animals, the 1,546 culled and the 5,123 not culled. Specifically, the dependent variable for the probit model was whether or not the heifer was culled prior to first lactation. Technically, the inverse Mills ratio is the linear prediction, in the Z metric, from this estimated probit selection model evaluated at the standard normal probability density function divided by the linear prediction evaluated at the cumulative normal distribution function.

The second step of the procedure used only the data of the 5,123 heifers which had a first lactation. In the second step the linear FE model, including the inverse Mills ratio term, was re-fitted. Statistical significance of the inverse Mills ratio in the second step, suggests selection bias is present, otherwise the sample can be considered to be representative of the general population.

In the presence of selection bias with an unobserved effect, the estimates of the second stage model would not be consistent for the population parameters. Therefore, this procedure is appropriate only for testing, not correcting, sample selection bias. The FE analysis was chosen over the RE in each of the two steps due to consistency of the estimates in the presence of arbitrary correlation between the unobserved effect and either selection or the independent variables.

All models and tests were conducted with Stata Statistical Software (30), except for the Spearman's rank correlations which were computed by SAS (29).

RESULTS AND DISCUSSION

Means, standard deviations, minima and maxima for parent genetic information, including reliabilities, and first lactation MEM are in Table 7.1.

Regression coefficients, and standard errors for each of the three regressions of standardized parent PTA of milk, fat and protein on standardized MEM are in Table 7.2.

The addition of PTA fat and protein of each parent as independent variables increased the predictive ability for first lactation milk. Though the inclusion of these variables likely resulted in multicollinearity, best linear unbiased estimates were still produced by regression analysis (9). While the variance of the estimates would have been smaller if correlation among the independent variables was not present, the point estimates and their variance estimators remain unbiased. The large standard errors of the estimates did, however, cause larger *P* values.

In comparing the estimates from the OLS and RE models, the relative change in the PTA fat of sire coefficient was considerably larger than for the other independent variables, but the estimate was still not significantly different than zero. As expected, the standard errors of the RE estimates were less than the respective OLS estimates. In the RE analysis the coefficient on PTA fat of dam was statistically significant at $P \leq 0.05$.

The FE estimates were similar to those from RE. The standard errors of the FE model were much closer to those from the RE model, as opposed to the considerably higher OLS standard errors. The level of significance among the FE estimates was similar to those of RE except PTA fat and protein of dam were now significant at the 0.01 level.

As indicated by the R^2 listed in Table 7.2, parents' PTA explained 6.86%, 6.80%, and 6.69% of the within herd variation in first lactation MEM using the OLS, RE, and FE models respectively. The low coefficient of determination of all models was consistent with sampling error in genetic evaluations and the large environmental variance relative to genetic variance in first lactation milk production. Estimating an OLS model which included the unobserved herd effect as a FE by including a dummy variable for each herd resulted in an R^2 of 0.4692. This indicates that parents' PTA and herd explain 46.92% of the variation in first lactation MEM. In other words, within the herds parents' PTA explain less than 7% of the variation. Then conditional on parent PTA, herd explains roughly 40% of the variation in first lactation MEM among heifers from different herds.

The Hausman test of the OLS and RE estimates was rejected at the 0.0154 level, rejecting the null hypothesis that the two vectors of estimates varied only due to sampling error. The presence of a unobserved effect which is correlated with the independent variables, would cause inconsistency among the OLS estimates and the RE estimates and would be compatible with the rejection, implying that the OLS and RE estimates varied significantly from each other.

The Hausman test of the RE and FE estimates was rejected at less than the 0.0001 level. Thus statistically rejecting the null hypothesis that the asymptotically efficient RE estimates had a zero asymptotic covariance with its difference from the asymptotically inefficient FE estimates. This would be compatible with the herd effect being correlated with the independent variables, causing inconsistency among the RE estimates. Either of the Hausman test rejections could also be due to the idiosyncratic error being correlated with PTA which would result in the OLS, RE and FE estimates being inconsistent.

Statistically, the results of each of the above tests suggest herd is correlated with parents' PTA, necessitating a FE model. However, it is not as clear that the RE and FE estimates differ in practically important ways, rather the rejection may be a vestige of the large sample size. The closeness of the estimates is surprising given the median and average number of animals in each cohort was only four and 7.2, respectively. Also, the ratio of the variance of the herd effect to the variance of the idiosyncratic error was only 2.2 and 1.2 for the RE and FE models, respectively.

To investigate the practical effect of these different weights on heifer selection, the within herd rankings of the heifers based on their predicted MEM by the RE and FE models were compared using Spearman's rank correlation. Based on the similarity of the RE and FE estimates, it was hypothesized that, within herd, heifers would be ranked the same, resulting in a rank correlation of one.

Figures 7.1 through 7.3 present histograms of the rank correlations. The distributions are bounded by plus and minus unity, the minimum and maximum potential values of the Spearman's rank correlation. A rank correlation could not be computed for herds of a single heifer and thus 139 herds were excluded from the analysis for this reason. A herd which consisted of a pair of identical twins was also excluded as the tie in the rankings precluded calculating the correlation. The within herd rank correlations between the predicted MEM based on the RE and FE models are presented in Figure 7.1. The four herds with a correlation of -1 each consisted of two heifers, and as a result, the possible values of rank correlation were limited to -1 or +1. All of the rank correlations less than 0.70 were associated with herd sizes of four or less. Based on the within herd rank correlations, the RE and FE estimates appear much more similar in a pragmatic sense than indicated by the statistical Hausman test.

Figures 7.2 and 7.3 present the within herd rank correlations between actual MEM and its estimates based on RE or FE, respectively. The distributions of rank correlations were diffuse in each case. Each of the distributions were skewed to the left, with approximately 65% of the herds having a positive rank correlation between the actual MEM rankings and the predicted MEM rankings. This suggests that, while the amount of MEM phenotypic variation explained by genetic evaluations is low, the evaluations may, nevertheless, be useful in heifer selection. The utility of genetic estimates may be increased by evaluating the estimates as they are updated over the course of heifer rearing.

Finally, no evidence of selection bias was found as the level of significance of the inverse Mills ratio in the FE equation of the second stage of the test for sample selection bias was 0.440. This suggests these results are applicable to those heifers which were culled prior to first lactation.

CONCLUSIONS

Estimation of the empirical selection weights for each parent PTA of milk, fat, and protein using a RE model resulted in more efficient estimates than OLS as expected. While the RE and FE empirical weights appeared quite similar numerically, this was strongly rejected statistically. However, an investigation of the comparability of within herd heifer rankings based on predictions from the RE and FE models supported the similarity of the estimates with over 96% of the herds having a Spearman's rank correlation between the rankings of predicted MEM from the models of greater than 0.90. Given the similar performance of the two models, consideration should be given to modeling the herd effect as a random effect due to the corresponding increased efficiency of the estimates.

It is important for producers to temper any expectations of heifer selection for

MEM based on genetic evaluations with the realization that parents' PTA explain only a small amount of the variation in MEM. However, the within herd rank correlations between predicted MEM, based on genetic evaluations, and the actual MEM suggest the evaluations may be useful in prepartum heifer selection.

The omission of heifers which were culled prior to first lactation, did not appear to cause selection bias in the sample data. As a result, these results were applicable to all heifers enrolled in Michigan DHI.

ACKNOWLEDGMENTS

The authors thank Al Thelen, Terry Speer and Lyle Smith at Michigan DHI, as well as, Leigh Walton and H. D. Norman at the Animal Improvement Laboratory, USDA for assistance with data collection. J. M. Wooldridge's econometric comments were highly appreciated. The financial support of the primary author by the Department of Agricultural Economics through the Animal Industry Initiative funds provided by the Agricultural Experiment Station at Michigan State University while this work was conducted is gratefully acknowledged.

REFERENCES

1 Allaire, F. R. 1981. Economic consequences of replacing cows with genetically improved heifers. J. Dairy Sci. 64:1985-1995.

2 Allaire, F. R., and E. P. Cunningham. 1980. Culling on low milk yield and its economic consequences for the dairy herd. Livest. Prod. Sci. 7:349-359.

3 Cassell, B. G., B. B. Smith, and R. E. Pearson. 1993. Influence of herd-life opportunity and characteristics of cows and herds on different net income functions. J. Dairy Sci. 76:1182-1190.

4 Congleton, W. R. 1988. Dairy cow culling decisions. 2. Profitability and genetic trend in herds with culling on production versus income. J. Dairy Sci. 71:1905-1915.

5 Congleton, W. R., and L. W. King. 1985. Culling cows on predicted future profitability with a variable planning horizon. J. Dairy Sci. 68:2970-2983.

6 Connor, L. J., L. G. Hamm, S. B. Nott, D. Darling, W. Bickert, R. Mellenberger, H. A. Tucker, O. B. Hesterman, J. A. Partridge, and J. H. Kirk. 1989. Summary of the 1987 Michigan State University dairy farm survey. Research Report 498, Michigan Ag. Exper. Station, East Lansing.

7 Cowan, C. M., M. R. Dentine, R. L. Ax, and L. A. Schuler. 1990. Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. Theor. Appl. Genet. 79:577-582.

8 Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao, J. E. Womack, and I. Hoeschele. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920.

9 Greene, W. H. 1993. 9.2 Multicollinearity. Page 266 *in* Econometric Analysis, 2nd ed., Macmillan Publishing Company, New York.

10 Greene, W. H. 1993. 16.4. Longitudinal Data. Page 464 *in* Econometric Analysis, 2nd ed., Macmillan Publishing Company, New York.

11 Greene, W. H. 1993. 22.3.2. The Censored Regression Model - Tobit Analysis. Page 694 *in* Econometric Analysis, 2nd ed., Macmillan Publishing Company, New York.

12 Hausman, J. A. 1978. Specification tests in econometrics. Econometrica 46:1251-1270.

13 Hazel, L. N., and J. L. Lush. 1942. The efficiency of three methods of selection. Jour. Hered. 33:393.

14 Heinrichs, A. J., S. J. Wells, H. S. Hurd, G. W. Hill, and D. A. Dargatz. 1994. The national dairy herd evaluation project: A profile of heifer management practices in the United States. J. Dairy Sci. 77:1548.

15 Henderson, C. R. 1963. Selection index and expected genetic advance. Page 141 *in* Statistical Genetics and Plant Breeding ; a symposium and workshop. W. D. Hanson and H. F. Robinson, ed. Natl. Acad. Sci.-NRC, Washington, DC.

16 Henderson, C. R. 1973. Sire evaluation and genetic trend. Page 10 *in* Proc. Anim. Breeding Genet. Symp in Honor of Dr. J. L. Lush, Am. Soc. Anim. Sci., Am. Dairy Sci. Assoc., Champaign, IL.

17 Huirne, R., A. Dijkhuizen, R. King, and S. Harsh. 1993. Goal, critical success factors, and information systems needs on swine and dairy farms – Identified by workshops. Zeitschrift für agrainformatik 3:61-65.

18 Killen, L., and B. Kearney. 1978. Optimal dairy cow replacement policy. Ir. J. Agric. Econ. Rur. Sociol. 7:33-40.

19 Korver, S., and J. A. Renkema. 1979. Economic evaluation of replacement rates in dairy herds II. Selection of cows during the first lactation. Livest. Prod. Sci. 6:29-37.

20 Kuipers, A. 1980. Development and comparison of selection criteria for cows and bulls with a dairy herd simulation model. Ph.D. Thesis, Univ. Wisconsin, Madison.

21 Maddala, G. S. 1971. The use of variance components models in pooling cross section and time series data. Econometrica 39:341-358.

22 Mao, F. C., and R. D. Bremel. 1995. Prediction of milk yields from serum β-lactoglobulin concentrations in pregnant heifers. J. Dairy Sci. 78:291-295.

23 McCullough, D. A., and M. DeLorenzo. 1996. Effects of price and management level on optimal replacement and insemination decisions. J. Dairy Sci. 79:242-253.

24 Michigan Dairy Herd Improvement Association. 1996. Michigan dairy production herd evaluation report.

25 Pearson, R. E., and A. E. Freeman. 1973. Effect of female culling and age distribution of the dairy herd on profitability. J. Dairy Sci. 56:1459-1471.

26 Quinton, V. M., and C. Smith. 1997. An empirical check on best linear unbiased prediction genetic evaluation using pig field recording data. Can. J. Anim. Sci. 77:211-216.

27 Robison, L. J., and P. J. Barry. 1996. Chapter 21, Replacement Principles in Present Value Models and Investment Analysis. 1st ed. The Academic Page, Northport, Alabama.

28 Rogers, G. W., J.A.M. van Arendonk, and B. T. McDaniel. 1988. Influence of production and prices on optimum culling rates and annualized net revenue. J. Dairy Sci. 71:3453-3462.

29 SAS® User's Guide: Statistics, 6th ed. 1989. SAS Inst., Inc., Cary, NC.

30 StataCorp. 1997. Stata® Statistical Software: Release 5.0. Stata Press, College Station, TX.

31 VanRaden, P. M., C. A. Ernst, G. R. Wiggans, and H. D. Norman. 1989. Ability of animal model and Modified Contemporary Comparison evaluations of parents to predict future progeny performance. J. Dairy Sci. 72(Suppl. 1):72.(Abstr.)

32 Van Arendonk, J. A. M. 1991. Use of profit equations to determine relative economic value of dairy herd life and production from field data. J. Dairy Sci. 74:1101-1107.

33 Van Arendonk, J. A. M. and A. A. Dijkhuizen. 1985. Studies on the replacement policies in dairy cattle. III. Influence of variation in reproduction and production. Livest. Prod. Sci. 13:333-349.

34 Van der Werf, J. H. J., T. H. E. Meuwissen, and G. De Jong. 1994. Effects of correction for heterogeneity of variance on bias and accuracy of breeding value estimation for Dutch dairy cattle. J. Dairy Sci. 77:3174-3184.

35 Weigel, D. J., B. G. Cassel, I. Hoeschele, and R. E. Pearson. 1995. Multiple trait prediction of transmitting abilities for herd life and estimation of economic weights using relative net income adjusted for opportunity cost. J. Dairy Sci. 78:639-647.

36 Wooldridge, J. M. 1995. Selection corrections for panel data models under conditional mean independence assumptions. J. Econometrics. 68:115-132.

Table 7.1. Parent PTA of milk, fat and protein, reliabilities for milk-fat, and first lactation mature equivalent milk production for 5,123 animals.

Variable[1]	$\bar{X}$	SD	Minima	Maxima
PTA Milk of Sire, kg	916.5	261.1	-947.1	1518.2
PTA Fat of Sire, kg	28.4	9.8	-39.9	57.2
PTA Protein of Sire, kg	24.7	8.1	-29.9	42.6
PTA Milk of Dam, kg	409.0	298.6	-819.7	1579.9
PTA Fat of Dam, kg	12.2	11.4	-32.2	64.9
PTA Protein of Dam, kg	11.4	8.7	-25.4	47.6
REL of Sire Milk-Fat, %	93.7	10.8	17	99
REL of Dam Milk-Fat, %	49.3	6.0	14	85
MEM, kg	11084.7	2451.6	1362.6	18875.2

[1] Mature equivalent milk production (MEM), reliability (REL).

Table 7.2. Regression coefficients, and standard errors for each of the three regressions of standardized first lactation mature equivalent milk production on standardized[1] parent PTA of milk, fat and protein for 5,123 heifers.

Variable	Regression Coefficients and SE					
	OLS[2]		Random Effects		Fixed Effects	
	β_i	SE[3]	β_i	SE	β_i	SE
PTA Milk of Sire	0.1155***	0.0235	0.0917***	0.0184	0.0811***	0.0187
PTA Fat of Sire	-0.0021	0.0218	-0.0114	0.0162	-0.0121	0.0162
PTA Protein of Sire	0.0381	0.0276	0.0340	0.0208	0.0261	0.0208
PTA Milk of Dam	0.3055***	0.0435	0.2838***	0.0312	0.2850***	0.0310
PTA Fat of Dam	-0.0342	0.0260	-0.0428*	0.0217	-0.0563**	0.0215
PTA Protein of Dam	-0.0916*	0.0435	-0.0853*	0.0351	-0.0952**	0.0346
R^2	0.0686		0.0680[4]		0.0669	

[1] Deviated from the mean and divided by the standard deviation.

[2] Ordinary least squares pooled over all observations.

[3] Robust to heteroskedasticity and correlation among the errors.

[4] Reported R^2 is from a model which included an intercept.

*$P \leq 0.05$.

**$P \leq 0.01$.

***$P \leq 0.001$.

Table 7.3. Frequency distributions of within herd Spearman's rank correlations comparing predictions and actual first lactation milk production for 576 herds.

Midpoint of category	Variables Compared with Spearman's Rank Correlations		
	RE[1], FE[2]	RE, Actual[3]	FE, Actual
-0.95	2	44	45
-0.85	0	4	4
-0.75	0	4	3
-0.65	0	1	3
-0.55	0	5	5
-0.45	0	28	26
-0.35	0	12	7
-0.25	0	18	17
-0.15	0	27	28
-0.05	0	23	25
0.05	0	41	45
0.15	0	39	43
0.25	0	47	41
0.35	1	42	44
0.45	0	43	39
0.55	4	48	48
0.65	0	25	27
0.75	3	20	20
0.85	12	23	26
0.95	554	82	80

[1] Predicted MEM from the RE model.

[2] Predicted MEM from the FE model.

[3] Actual MEM.

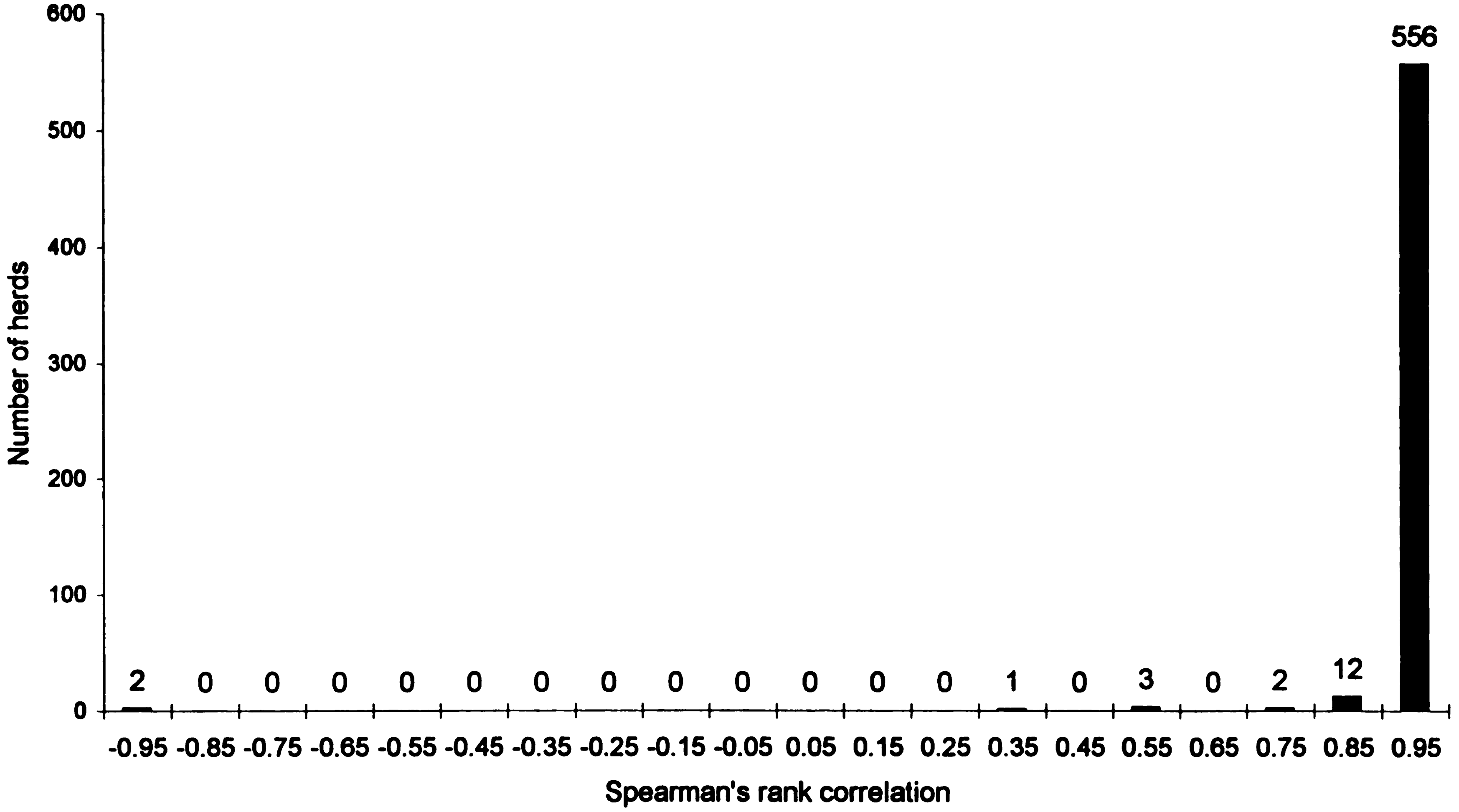

Figure 7.1. Spearman’s rank correlations between RE MEM and FE MEM predictions for 576 herds.

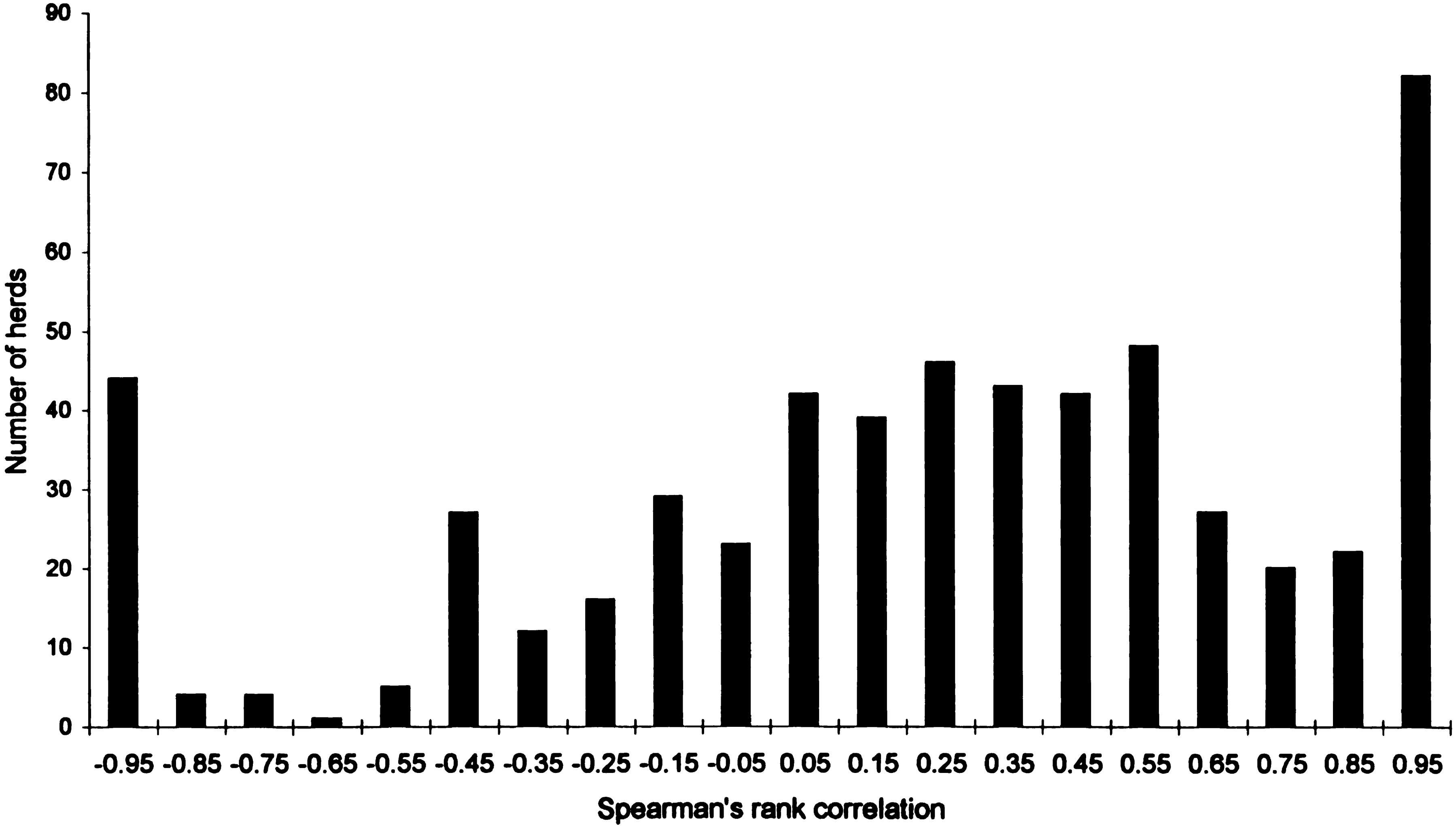

Figure 7.2. Spearman's rank correlations between Actual MEM and RE MEM predictions for 576 herds.

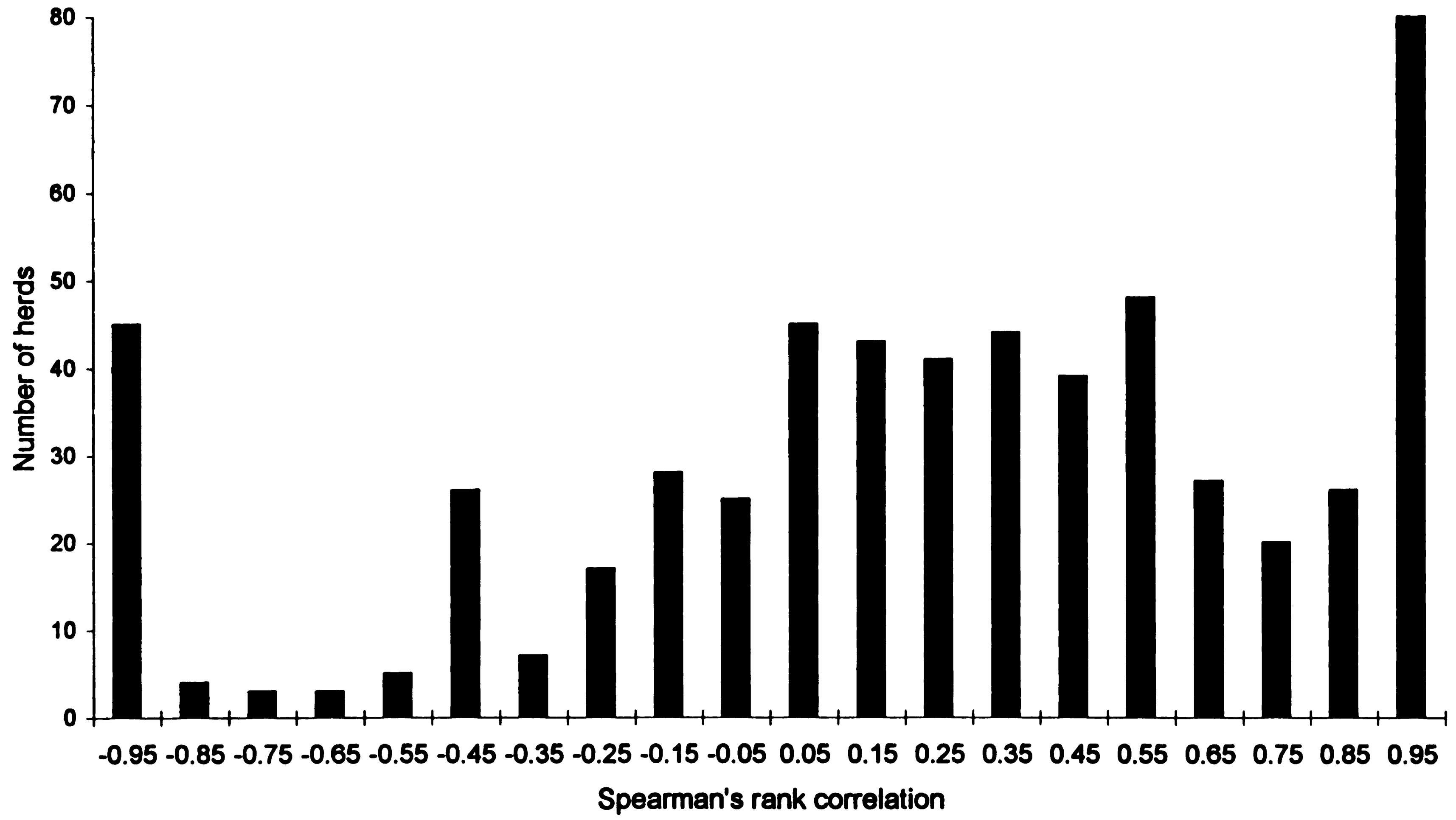

Figure 7.3. Spearman's rank correlations between Actual MEM and FE MEM predictions for 576 herds.

Chapter 8

PREDICTION OF CULLING PRIOR TO SECOND LACTATION BASED ON PARENTS' PREDICTED TRANSMITTING ABILITIES

Abstract

The objective of this study was to determine the ability of parents' PTA for milk, fat and protein to predict subsequent culling of heifers during rearing or first lactation. The data consisted of the culling outcome of 5,619 Michigan Holstein heifers during rearing or first lactation, and the heifers' parents' animal model evaluations from the period in which the heifer was born. To ensure asymptotic unbiasedness of the estimates the unobserved herd effect was modeled as a fixed effect. Despite evidence that first lactation milk production did influence producers' culling decisions, the PTA of milk, fat, and protein for each heifers' parents did not predict subsequent culling during rearing or first lactation. This result was not surprising given that first lactation milk production explains less than 20% of the variation in culling and from Chapter 7 is was apparent that parents' genetic estimates explained less than 7% of the variation in first lactation milk production. This poor predictive ability was consistent regardless of the distributional specification. Weak evidence for the necessity of modeling the herd effect as fixed was present. Furthermore, the inability to correctly predict culling reason broken down into voluntary and involuntary reasons, coupled with the counterintuitive effects of parents' PTA on the probability of being culled for each of the reasons raised questions as to the existence and utility of classifying culling reasons according to the voluntary nature of the culling decision. Also,

it can be concluded that Michigan producers are not using parents' genetics in heifer culling decisions.

Keywords: dairy culling, prediction, fixed effect

1. Introduction

Given a constant herd size, no seasonal effects on reproduction and production, and culled cows being sold for meat (not dairy) purposes, optimization (McCullough and DeLorenzo, 1996; Rogers et al., 1988; Congleton and King, 1985; Van Arendonk and Dijkhuizen, 1985; Killen and Kearney, 1978) and simulation (Allaire, 1981; Allaire and Cunningham, 1980) studies were unanimous in their conclusions that a culling rate of 20% to 30% optimizes producer profit. While other simulation studies (Congleton, 1988; Kuipers, 1980; Korver and Renkema, 1979; Pearson and Freeman, 1973), which considered subjectively selected culling rates or strategies for comparison, indicated that profit would be increased with culling rates below the average 37% culling rate[1] in the Michigan dairy industry in 1996 (Michigan Dairy Herd Improvement Association, 1996). Comparability of the actual and optimal culling rate assumes no herd contraction or internal expansion. The results of these multiple studies are compelling because, not only did the various models employ different assumptions and analytical techniques, but as well, the models operated under economic conditions representative of a number of countries (i.e., the U.S.A, England, Ireland, Netherlands). In the models, the increased profit

---

[1] Culling rate, annually defined as, the number of animals leaving the herd due to sale or death divided by the average herd size.

associated with use of the optimal culling rates varied from 2% to 40%.

Comparison of only the mean profit by the studies implies risk neutrality, whereas a majority of Michigan and Dutch dairy producers are risk averse (Huirne et al., 1993). The risk preferences of decision makers can have significant effects on their choices, so the optimal rule assuming risk neutrality cannot necessarily be extrapolated to those who are risk averse.

At least under risk neutrality, one direct ramification of Michigan's excessive culling rate is that Michigan dairy producers do not need all potential replacement dairy heifers to enter the lactating herd, necessitating a method of heifer selection.

This need for heifer selection is in direct contrast to the present situation in which most producers rear and freshen the majority of heifers produced on the farm (Connor et al., 1989; Heinrichs et al., 1994). Animals are then selected largely on the basis of the level of first lactation and expected future milk production. The simulation studies revealed that while this practice maximizes genetic progress and milk production, the economic performance is lower due to failure to consider the opportunity costs of higher culling rates.

The ultimate objective of heifer selection is to choose the most profitable animals. Van Arendonk (1991) showed that lifetime profit corrected for the opportunity cost of postponed replacement is the correct method for ranking animals in terms of profitability. Unfortunately, corrected lifetime profit of animals is generally not available and furthermore, can not be determined prior to the culling of the animals. However, the literature has focused on indicators of lifetime profitability that are apparent early in the productive life of milk cows. No work has considered indicators apparent prior to productive life.

The phenotypic correlation between first lactation actual milk and lifetime profit corrected for opportunity cost has been reported at 0.54 (Cassell et al., 1993). Weigel et al. (1995) determined the phenotypic and genetic correlations between corrected lifetime profit and first lactation mature equivalent milk production to be 0.51, and 0.80, respectively. For purposes of heifer selection, first lactation milk production may then be a suitable proxy for corrected lifetime profit. Then by choosing animals based on first lactation milk production producers have selected the correct animals.

However, this strategy of milking all heifers and heavily culling the herd to select the highest producing animals is costly. Given the large price differential between a heifer just prior to entering milk production and her value as a cull cow, the increased profit from even randomly selling some heifers prior to milking, and thereby lowering the culling rate, more than compensates for the decreased ability to identify and select the highest producing animals. In addition, with fewer heifers in the milking herd the proportion of older, more profitable cows in the herd rises.

Currently the information provided to Michigan producers upon which to base culling predictions is limited to the heifers' estimated breeding values (EBV) of milk, fat and protein production. The EBV are the average of the parents' predicted transmitting abilities (PTA). Because EBV milk provide estimates of the genetic merits for milk production which should on average manifest as phenotypic performance, EBV milk may have a role in heifer selection. While the culling models suggest the need for heifer selection, they lend little insight into how selection should be conducted. The optimization studies employed random replacement selection, and many simulation studies selected replacements based on their simulated EBV milk. While this latter approach to selection has intuitive appeal and indeed theoretical justification (Henderson, 1963), empirical support for heifer selection on this basis is lacking. Working with swine, Quinton and Smith (1997) found the within herd rank correlation with progeny performance was higher for parents' genetic estimates than for parents' phenotypes. However they warn the benefits of genetic estimates over phenotype in selection may be less in practice than in theory.

Due to the positive correlation between EBV fat and protein with EBV milk, genetic selection index theory, as well as statistical theory, suggest prediction of milk production will be improved by incorporating information on the components. Because selection is based on milk production, it is possible that the EBV or, alternatively, parents' PTA of milk, fat and protein can explain subsequent culling in first lactation.

The objective of this study was to determine the ability of parents' PTA of milk, fat and protein to explain culling prior to, or during first lactation.

2. Materials and methods

2.1 Data collection

A total of 15,116 Michigan Holstein heifer records were collected from Michigan DHI. Only Holstein heifers that met the following criteria were included in the data set to evaluate the ability of parents' genetic estimates to predict culling: 1) as of January 1, 1995, identified as prepartum; 2) born between July 1, 1992 and December 31, 1992; 3) at least one parent's PTA for milk, fat, or protein recorded; 4) either a calving date between July 1, 1994 and June 30, 1995, or if no calving date, a culling date prior to July 1, 1994; 5) if culled, a culling code was recorded.

The 15,116 heifer records were also used to generate a second data set to evaluate the ability of first lactation mature equivalent milk production (MEM) to predict culling in first lactation. To be included in this second data set, records were required to meet criteria 1, 2, 3 and 5 from above as well, the heifer must have calved between July 1, 1994 and June 30, 1995.

Genetic data were the heifers' parent PTA of milk, fat and protein from the January 1993 sire and dam animal model evaluations calculated by the Animal Improvement Programs Laboratory, USDA (Beltsville, MD).

Culled heifers included those that died or were culled prior to second lactation. Therefore, heifers considered culled included those animals that were sold or died during first lactation, as well as, those heifers without a reported calving date.

After deleting heifers that were missing the genetic information of either parent, records of 5,619 heifers were available for statistical analysis of the relationship between parents' genetic estimates and culling. These heifers were present in 781 herds. In effect then, this data sample consisted of, for each herd, a cohort of heifers which were born within a six month window, and then either calved in a six month window or were culled in the two years that separated the windows. The number of heifers in a cohort for a given herd varied in size from 1 to 157 heifers with median and mean sizes of 4 and 7.2, respectively.

The second data set of MEM and culling outcome in the first lactation contained 12,128 heifers in 1,100 herds. Again, each herd was represented by a single cohort of heifers which were born and calved within the prescribed time frames. The cohorts of heifers varied from 1 to 244 heifers, with a median size of 7 and mean of 11.0.

2.2 Estimation technique

2.2.1 Population Model

Consider the unbalanced general population panel data or cluster probability model

$$P[\text{Culled}_{ha} = 1 \mid \text{PTA}_h, \text{Herd}_h] = G(\mathbf{PTA}_{ha}\beta + \text{Herd}_h) \qquad (1)$$

$$a = 1, \ldots, A_h;\ h = 1, \ldots, H.$$

where

$P[\text{Culled}_{ha} = 1 \mid \text{PTA}_h, \text{Herd}_h]$ is the response probability of being culled conditional on PTA and the unobserved herd effect. Culled_{ha} is equal to one if animal *a* in herd *h* was culled during the rearing period or first lactation, and zero otherwise;

G is a binary response function with values in the open unit interval;

$\mathbf{PTA}_{ha}$ is a 1 x 6 row vector. The six elements of $\mathbf{PTA}_{ha}$ contain the first milk, fat and protein PTA of individual *a*'s parents, generated following *a*'s birth;

and Herd_h is an unobserved random variable representing the effect of herd *h*.

Given the need to empirically estimate the relationship between culling and parent PTA, the pragmatic issue of appropriate estimation technique must be considered. This consideration should include the distribution of the response probability as well as the nature of the unobserved herd effect.

As is often the case, it was unclear, a priori, whether a normal or logistic cumulative distribution function (cdf) best represented the probability of culled conditional on the independent variables. The normal cdf would imply a probit model is appropriate, while a logistic cdf would lead to a logit model. Pragmatically, the probability density functions of these two distributions differ only in the thickness of the tails and are generally considered to be sufficiently similar that either is appropriate. Given the ambiguity over the a priori appropriate distributional assumption, probit and logit models were each considered.

However, decisions with regard to modeling the herd effect are potentially not so innocuous. The herd effect, more generally termed an unobserved effect, cluster effect, or repeated measure indicator, denotes that this effect captures similarity among some observations in the data set. A priori, the magnitude of this effect is unobservable or unquantifiable. With regard to the herd effect, the issue to be resolved is whether it should be modeled as a random effect (RE) or a fixed effect (FE), with the former being used if the unobserved effect is assumed to have a distribution and the latter if the effect has no distribution. But this classification of an unobserved effect based on its theoretical distributional assumption ignores the pragmatically more important issue of whether the unobserved effect is uncorrelated with the independent variables. Notice that, by definition, if an unobserved effect is thought to be fixed (i.e., not have a distribution) it can not be correlated with anything, including the independent variables.

Modeling unobserved effects in linear probability models is relatively straight forward. If the effect is considered random, it can be accounted for through generalized least squares estimation, or least squares with a variance estimator robust to serial correlation among the error term. But notice that, if the RE is correlated with any of the independent variables, the resulting coefficients are biased (Greene, 1993), due to an omitted variable, namely the RE. If the unobserved effect is considered to be distributionless, or of more pragmatic importance, to be correlated with the independent variables, it can be modeled in a variety ways as a FE resulting in unbiased regression coefficients. Finally, the assessment of the preferred modeling technique has been aided through the test developed by Hausman (1978) to determine if FE estimation is required.

Unfortunately, for maximum likelihood estimation (MLE) of models with a dichotomous dependent variable, no such general test is available to determine whether an unobserved effect should be modeled as a RE or FE. While tests comparing the parameter estimates from RE and FE specifications may be available for specific models, the modeler's judgement must often guide the estimation.

Modeling an unobserved effect as a RE in a nonlinear probability model assumes the unobserved effect has a distribution and is independent of the explanatory variables (Hsiao, 1986, pp. 164-167). The FE approach does not require the latter assumption. RE estimation of nonlinear probability models generally involves integrating the unobserved effect out of the response function. Intuitively, if the unobserved effect is a function of the independent variables, performing such an integration must also affect the independent variables. Given that it is unclear what the independent variables would then represent, it is even less clear how to interpret the resulting regression coefficients.

In considering the nature of the herd effect, it must be recognized that this effect includes a cohort effect in addition to a herd management effect, which in turn selects the herd's genetics. It is therefore plausible that the herd effect is correlated with the independent variables. This suggests a FE estimation technique is appropriate, for the results from a RE model would be asymptotically biased.

While the unobserved effect can technically be modeled as a RE or FE, modeling it with a dummy variable should generally be avoided. This assumes that software that can add the potentially large number of dummy variables is available. To model an unobserved effect, herd effect, cluster effect or repeated measures indicator as a dummy variable results in all parameter estimates, including those associated with the independent variables, being asymptotically biased. This result is not dependent on the unobserved effect being correlated with the explanatory variables.

In a limited set of Monte Carlo experiments, Heckman (1981) showed that inclusion of a dummy variable to represent an unobserved effect in a probit specification resulted in asymptotically biased estimates for the coefficients of the independent variables. While the bias varied with the value of the coefficient, with 8 observations contributing to each dummy variable and as long as no lagged values of the dummy variables were included in the model, the bias was never more than ten percent. The bias was towards zero, and increased as the variation of the unobserved effect increased. Of course to permit predictions an estimate of the coefficient of the dummy variables is required. This may be problematic for out of sample predictions.

Abrevaya (1997) generalizing the work of Hsiao (1986, pp. 159-161) and Chamberlain (1984, pp. 1275-1276) showed that, for the case of two observations contributing to each dummy variable, the maximum likelihood logit estimates of the coefficients associated with the independent variables are asymptotically twice their true value. While a closed form proof for the case of greater than two observations contributing to an unobserved effect has not been discovered, there is evidence that the bias may be minimal for sufficiently sized clusters. Monte Carlo simulation revealed that when 20 observations contributed to each dummy variable the estimates were virtually unbiased compared to those from FE logit (Hsiao, 1986, pp. 228).

In general, the use of dummy variables does not necessarily result in asymptotic bias of the estimators. For example, inclusion of a gender dummy variable in a nonlinear model will not adversely affect the properties of the estimators, as the gender coefficient is based on the entire sample. If MLE is being used, this implies the sample size is sufficiently large for asymptotic analysis to be appropriate.

The bias originates with the fact that there are a limited number of observations associated with any given unobserved effect. In a panel data set, the asymptotic properties of the estimates are derivable only if the number of observations per unobserved effect are held fixed, while the number of unobserved effects gets large. As a result, any estimation of the coefficient associated with a dummy variable for an unobserved effect is dubious given that maximum likelihood results in large sample estimators. While statistically the number of observations per unobserved effect can be large with the number of effects fixed, a data structure traditionally dealt with in time series analysis, other issues arise such as the stationary properties of the variables. Taking the example where the unobserved effect is a herd effect, letting herd size tend to infinity is intuitively objectionable. It is much more natural to assume the number of animals per herd is fixed and the number of herds is large, than the reverse.

All analyses discussed will be asymptotic, making the reasonable assumptions that herd sizes are fixed, while the number of herds is large, and the herd effect is potentially correlated with the independent variables.

2.2.2 Probit model

While the FE probit can be conducted using full maximum likelihood which requires specifying the joint distribution of ($\mathrm{Culled}_{h1}, \ldots, \mathrm{Culled}_{hA}$), a common alternative is that proposed by Chamberlain (1984). In general, he suggested assuming the relationship between the unobserved effect and independent variables was such that the expectation of Herd_h conditional on $\mathbf{PTA}_h$ was linear. A parsimonious form of this assumption is

$$\mathrm{Herd}_h = \psi + \overline{\mathbf{PTA}}_h \lambda + d_h \tag{2}$$

where it is further assumed d_h, the error term, is distributed Normal($0, \sigma_d^2$). $\overline{\mathbf{PTA}}_h$ represents, for herd *h*, the herd average values of each of the six elements of the $\mathbf{PTA}_h$ vector.

If it is assumed that G, the binary response function from Equation (1), is the normal cdf the assumed conditional expectation of the unobserved effect (2) can be substituted into the population model (1) resulting in

$$P[\text{Culled}_{ha} = 1 \mid \text{PTA}_h, \text{Herd}_h] = G(\, \mathbf{PTA}_{ha}\beta_d + \psi_d + \overline{\mathbf{PTA}}_h\, \lambda_d) \qquad (3)$$

where $\beta_d = \beta/(1+\sigma_d^2)^{0.5}$, $\psi_d = \psi/(1+\sigma_d^2)^{0.5}$, and $\lambda_d = \lambda/(1+\sigma_d^2)^{0.5}$.

With this approach, only the probability of an individual animal being culled, and not the joint probability distribution of all the animals in a herd being culled, is assumed to be correctly specified. For this partial ML function to be correctly specified, P[Culled$_{ha}$ =1] conditional on the genetics of all the animals in the herd must be uncorrelated with all functions of the independent variables for all other animals in the same herd, a reasonable assumption. If the unobserved effect truly is a RE then $\lambda_d = 0$. This can be tested with a likelihood ratio test.

The normality assumption of the probit specification is met since the addition of the two normally distributed variables (i.e., the error terms from the population model and d_h) results in a normally distributed composite variable (i.e., error term). Equation (3) can then be estimated by running probit on all the observations pooled together. However, as a result of using pooled probit, the residuals will necessarily be conditionally correlated within herd due to the presence of d_h, necessitating a robust variance estimator.

The presence of d_h also causes attenuation bias in the resulting coefficients. As shown above, the β_d are biased away from β towards zero. In considering the interpretation of β and β_d, the attenuation biased estimates are likely preferable. β, if they were possible to estimate, would reflect the effect of changing $\mathbf{PTA}_{ha}$ evaluated at the mean herd effect which is assumed to be zero. As few, if any of the herds, actually have the mean herd effect, the representativeness of such a result is unclear. In contrast, β_d represents the effect of a change in $\mathbf{PTA}_{hs}$ averaged across the distribution of the herd effect in the population, a more appropriate evaluation for a random heifer.

To further evaluate performance of the probit model, alternative functional forms were considered. First, quadratic terms of each element of $\mathbf{PTA}_{hs}$ were added. A likelihood ratio test was used to test the hypothesis that the coefficients of the quadratic terms are jointly zero. Second, the assumption that the binary response function, G, in the population model of Equation (1) was symmetrical was relaxed through specifying it to be a log-log link function. Specifically the negative natural logarithm of one minus the probability of being culled is equal to the exponential of $\mathbf{PTA}_{hs}\beta$. The herd average values were excluded as this approach to modeling the herd effect is specific to G being the normal distribution. By ignoring the herd effect, the estimates that resulted from this link function may have been biased.

2.2.3 Logit model

If it is assumed that the idiosyncratic errors have a logistic distribution, a FE logit approach has the advantage of not imposing any restrictions about the nature of the arbitrary relationship between Herd_h and $\mathbf{PTA}_h$. It does however, require that within a herd the outcomes are independent. In the present analysis this amounts to assuming that after controlling for genetics, the culling of each heifer is independent of the culling of all other heifers in a given herd. Violations of this assumption would include a producer predetermining the number of members of the cohort to be culled, or a common environmental shock among some, but not all, heifers increasing their probability of being culled. Of course, the errors are also assumed to follow a logistic distribution.

Beginning with the joint distribution of (Culled_{h1}, ..., Culled_{hA}) which is then conditioned on the genetics of all animals in a herd, the herd effect, and $\sum_{a=1}^{A_h} \text{Culled}_{ha}$, the FE logit model can be derived. The resulting estimates of β are asymptotically unbiased. The outcomes, for a given herd size, are the different possible combinations of culled heifers that result in the total number of heifers sold. Those animals in herds in which all or none of the animals in the sample were culled do not contribute information to the estimation since their likelihood function is one. As Chamberlain (1984) shows, due to the logistic function, for each combination of herd size and number of animals culled the result is a multinomial logistic distribution where the independent variables are differences of the right hand side variables of equation (1). As a result, the unobserved herd effect cancels, leaving the joint distribution of the outcomes conditional on the data and number of successful outcomes. This approach is also termed conditional logistic, presumably due to the conditioning on the herd outcomes. See Appendix A for a fixed effect logit (conditional logistic) proof for a cluster of 3 with one success.

The FE logit estimates are for the observations evaluated at the level of the average herd. The average herd effect where, for a given PTA_{ha}, the estimate is averaged across the distribution of Herd_h in the population, requires a distributional assumption on the unobserved herd effect.

In summary, the FE probit has the disadvantage of requiring strong assumptions about the relationship between the unobserved effect and the independent variables, it does however result in the average treatment effects. In contrast, the FE logit specification, while not requiring any assumption about the relationship between the Herd_h and $\mathbf{PTA}_{hs}$, or the distribution of the herd effect, does demand conditional independence of the outcomes. The FE logit discards all data for clusters with invariant outcomes, and the coefficients are evaluated at the average unobserved effect. Finally, to be useful in prediction the FE logit requires a priori knowledge of the herd size and number to be culled.

2.2.4 Multinomial logit and ordered probit models

The ability to correctly classify culls conditional on $\mathbf{PTA}_{hs}$ and the fact that the animals had been culled was evaluated with a multinomial logit model and an ordered FE probit model. These models are extensions of their respective logit or FE probit models with a dichotomous dependent variables generalized to encompass multiple positive outcomes (Greene, 1993) which in the case of ordered models follow a logical ordering. For these models the dependent variable was recoded into culling classifications from most to least valuable animal where 0, 1, 2, 3 corresponded to not culled, culled during first lactation for reasons other than low production, culled during first lactation for low production, and culled prior to first lactation, respectively. A value of 1 then represented what has traditionally been considered an involuntary cull, and a 2 corresponded to a voluntary cull.

Theoretically, the ordered model, by utilizing the information contained in the ordering of the outcomes, should result in performance superior to the multinomial specification. Among the ordered models, the probit was selected over the logit because the probit permitted modeling of the fixed herd effect. Failure to account for a herd effect which was correlated with the independent variables would result in biased estimators. Assuming the herd effect existed, even if it is not correlated with PTA, it is unclear how a the logistic distributional assumption could be true for a general model such as Equation (1) extended to categorical responses. For there are no known distributions for the population model error term and the herd effect which when summed will result in a logistic distribution. Instead an approximate two step analysis of random effects ordered logit estimation with an extra distributional assumption has been developed (Stiratelli et al., 1984). In the first stage the within cluster variation is modeled assuming a logistic distribution, followed by modeling of the assumed normally distributed random effect in the second stage.

2.2.5 Prediction of first lactation culling based on first lactation MEM

The assumption that producers based selection on milk production can be represented by the following model

$$P[\text{CulledLact}_{ha} = 1 \mid \text{MEM}_h, \text{Herd}_h] = G(\rho \text{MEM}_{ha} + \text{Herd}_h) \qquad (4)$$

where

$P[\mathrm{CulledLact}_{ha} = 1 \mid \mathrm{PTA}_{h}, \mathrm{Herd}_{h}]$ is the response probability of being culled conditional on MEM and the unobserved herd effect. $\mathrm{CulledLact}_{ha}$ is one if animal *a* in herd *h* was culled during first lactation;

MEM_{ha} is the first lactation mature equivalent milk production for animal *a* in herd *h*;

and Herd_{h} is an unobserved random variable representing the effect of herd *h*.

Using an approach similar to that discussed earlier in the estimation of equation (1), G in equation (4) was assumed to be either a normal or logistic distribution and the unobserved herd effect was modeled as a fixed effect.

2.2.6 Model evaluation

Prior to considering culling predictions, evaluation of the marginal effects of parent PTA on the probability of being culled serves as another measure of model performance. To evaluate the effect of a change in one element of PTA_{ha} on the probability of being culled requires taking the derivative of the probability with respect to the element of interest. This resulting expression will be a function of all the elements of PTA_{ha}. As is typical when dealing with continuous variables, the function was evaluated at the mean of each of these elements, including the mean heifer cohort size for the probit estimates.

As the primary focus of these analyses were to ascertain the ability of parent PTA to predict subsequent daughter culling during rearing or the first lactation, the predictive value of each model was of interest. For each model the predicted probability of being culled was generated. However, for the FE logit the predicted probabilities were limited to those animals in cohorts where a single animal was culled. This limitation is due to the fact that the calculation of predicted probabilities, which is a function of the all the potential combinations of culling outcomes for all animals in the cohort that results in the total number of animals culled, is difficult for cases where more than a single animal was culled in each cohort. Specifically, as is apparent from simple extension of Appendix A to larger cohort sizes, the predicted probabilities are functions of different independent variables for each combination of cohort outcome. The 95% confidence intervals of the FE probit predicted probabilities of being culled were also calculated and evaluated.

A threshold of 0.5 was used in assigning the predicted probabilities to predicted outcomes. The predicted outcomes of each model were then compared with the actual culling outcome.

3. Results and Discussion

Means, standard deviations, minima and maxima for parent genetic information are in Table 8.1. Notice that the mean PTA for fat and protein is less than 30 kg for sires and under 12 kg for dams. Of the 5,619 animals, 9.0% were culled prior to first lactation and 15.4% were culled during first lactation for a total culling rate prior to second lactation of 24.4%.

Table 8.2 reports the results of the regression analyses with dichotomous outcomes of culled on parent PTA. The Chi-square near the bottom of the table reveals each model explains a statistically significant amount of the total variation. However, the pseudo R^2 values suggest the models explain little of the variation in culling prior to second lactation. This result of model statistical significance, in the face of poor explanatory nature of the models, is consistent with large sample size driving the statistical significance. As explained earlier, one drawback of the logit specification over the probit in a FE framework is the potential loss of observations. In the present analysis, 967 observations could not be used in the logit analysis as these animals belonged to cohorts which were invariant with respect to culling prior to second lactation.

The likelihood ratio test of $\lambda_d = 0$ resulted in a Chi-square statistic of 10.64, with 6 degrees of freedom or a level of significance of 10.0%. While this provides only weak evidence that the herd effect needed to be modeled as a FE, this conservative approach is appropriate given the alternative of potentially biased estimates from the RE specification.

The absolute size of the respective model coefficients reported in Table 8.2 are not directly comparable due to scale differences, attenuation bias in the probit estimates, and evaluation of the herd effect at different values. The coefficients of the FE logit model indicate the direction of the change in the probability of being culled given an increase in the independent variable. Care must be taken in making the same assessment in the FE probit models due to the presence of the herd mean variables. To assess the effect of an independent variable in the FE probit model on the response probability, its indirect influence through the herd mean of the respective variable must also be considered. Sire protein, dam milk and dam protein pose potential problems because their signs differ from their respective herd mean coefficients. The sire protein coefficient dominates the herd mean effect, so the direction of this effect will remain negative. And since the effect mediated through the herd mean effect will be divided by the size of the heifer cohort, for the average cohort size of 7.2 and for all but the smallest cohorts, the signs on dam milk and dam protein will not change. The signs of the coefficients between the FE probit and FE logit models are then identical, suggesting the qualitative results are not sensitive to the differing assumptions made between the models. Ignoring the effects of the herd mean PTA, the coefficients' levels of statistical significance were also similar between the two models, thereby supporting the robustness of the results. The high degree of correlation between PTA for milk, fat and protein likely contributed to the lack of statistical significance of many of the variables. In the presence of multicollinearity the coefficients remain unbiased (Greene, 1993, pp. 266) although the standard errors are increased over the situation where multicollinearity is not present.

Three of the PTA, namely sire fat, dam fat and dam protein, each possess a positive sign, suggesting increases in any of these, keeping all other independent variables the same, resulted in an increased probability of being culled prior to second lactation. This selection against fat, while not significantly different from zero at a *P* value of 5%, is consistent with component milk pricing. However, component milk pricing which places higher relative value on protein should have resulted in sire and dam protein PTA having a protective effect. The weak statistical significance of the positive sign of dam PTA protein is then unexpected. Rational expectations, or the ability of Michigan producers to survey the operating environment and surmise component pricing would be forthcoming, could explain why selection may have been consistent with such a pricing scheme although the scheme was not in effect at the time of selection.

To evaluate the practical significance of a 100 kg change in a parent's PTA on the probability of being culled, the derivative of the respective model was taken with respect to the PTA of interest. This in effect rescales the logit and probit coefficients from their logistic and Z scales, respectively. Table 8.3 represents the mean effect and 95% confidence intervals for a 100 kg change in a parent's PTA on the probability of being culled prior to second lactation.

As expected, based on the discussion of the signs of the coefficients in Table 8.1, the signs of the derivatives are identical between the two models. The size of the derivatives, for a given PTA, are the same order of magnitude for the FE probit and FE logit models. The attenuation bias evident in Equation (3) caused the FE probit estimates averaged over the herd effect distribution to be smaller in absolute size than estimates evaluated at the average value of the herd effect, as were the FE logit estimates. With the exception of sire milk, the FE probit estimates were smaller in absolute size than the respective FE logit derivatives. Given the difference in the evaluation of the herd effect, along with the other variant assumptions including the distributional assumption, and the relationship between the herd effect and the independent variables, the effect of a change in a given PTA on the probability of being culled is strikingly similar between the two models. This similarity suggests the results of the models are robust to the assumptions made.

Increasing the milk PTA of sire and dam by 100 kg decreased the probability of being culled by less than one percent and less than three percent respectively with narrow 95% confidence intervals. While changing the fat and protein PTA of either parent appears to have a large effect on the probability of being culled, reconsideration of Table 8.1 reveals that with respective means of less than 50 kg, the effect of changes in PTA on probability of being culled is minimal. The evaluation of the derivative for what amounts to a very large change in the independent variable explains why the upper bound of the 95% confidence interval of the FE logit derivative with respect to dam protein extends beyond unity. The confidence intervals for the derivatives with respect to fat and protein of each parent were quite wide.

Care should be taken in the ceteris paribus interpretation of the effect of PTA on the probability of being culled. Due to the high correlation between the PTA of a given parent, single PTA do not generally change in isolation. When evaluating the relative probability of two heifers being culled, the joint effect of changes in all of a parents' PTA should be considered. However, the summing of positive and negative effects tends to cancel the effects leaving little change in probability of being culled with different parents. This finding is consistent with the low pseudo R^2 of the models. The multicollinearity also widened the confidence intervals listed in Table 8.3.

Table 8.4 represents the predictive ability of the FE probit model. The model's performance is poor. The probit model did not predict the culling of any animals much less the 1,370 animals that were actually culled. The poor predictive ability of the FE probit model is consistent with the low pseudo R^2 reported in Table 8.2, and the small effect of reasonabley sized changes in the PTA on probability of being culled. Model performance is just as poor when the confidence intervals of the probabilities of being culled were considered. The 95% confidence intervals of only five animals included 0.5. Of these five animals, three were actually culled.

Use of the traditional 0.5 threshold in classifying predicted probabilities into predicted outcomes does affect the predictive ability of the models. The low pseudo R^2 suggests the models are explaining little of the variation in culling beyond the mean. Using a threshold of 0.5 when the mean level of culling of was 24.4% predisposed the models to low specificity, and thus predictive value. This is simply an application of the well known result that the prevalence affects the predictive ability of a model.

To further evaluate the poor performance of the model, alternative functional forms of the model were considered. First, the addition of quadratic terms of each element of $\mathbf{PTA}_{hs}$ did not alter the coefficient signs of the original variables, nor did the predicted culling outcome change for any animal. The hypothesis that the coefficients of the quadratic terms were jointly zero was not rejected at a probability of 0.38 by a likelihood ratio test with a chi squared statistic of 6.43 and 6 degrees of freedom. Second, G, the binary response function, in the population model of Equation (1) was specified as the asymmetrical log-log link function. Again, the coefficient signs of $\mathbf{PTA}_{hs}$, and the culling predictions were identical to the FE probit model.

Given the inability of the FE probit model to predict culling prior to second lactation, the predictive ability of parent PTA conditional on more information was explored. First predictions from the FE logit model were considered. Unlike the FE probit model which made predictions conditional on $\mathbf{PTA}_{hs}$ and the herd average of PTA, the FE logit predicts conditional on $\mathbf{PTA}_{hs}$ and the number of animals in the herd that were culled. Due to the conditioning on the latter the FE logit is not useful in a priori prediction of culling.

Table 8.5 presents the culling predictions for cohorts in which a single animal of the heifer cohort was culled prior to second lactation. The difficulty in calculating predicted probabilities where more than one animal is culled ~~is~~ presents another problem with using the FE logit model for out of sample predictions of which animals would subsequently be culled. Of the 174 cohorts that culled a single animal the FE logit model correctly predicted the number of animals culled in 35 cohorts. However, for 13 of those 35 cohorts the incorrect animal was predicted to have been culled. The poor predictive abilities of the models are again consistent with the low pseudo R^2 reported in Table 8.2, the small effect of reasonable changes in the PTA on probability of being culled, and the FE probit predictions.

It was considered that the poor predictive ability of the model may be due to the attempt to predict voluntary and involuntary culling. Specifically, since the producer presumably has little choice with regards to involuntary culls, or alternatively, the culling is not due to an animal's genetics but rather a random environmental shock, parent's PTA would be expected to predict voluntary but not involuntary culling. This ability to correctly classify culls conditional on $\mathbf{PTA}_{hs}$ and the fact that the animals had been culled was evaluated with a multinomial logit model and an ordered probit model. Pragmatically the multinomial logit model was rejected for the nonsensical results that the probability of being culled for low production increased with sire PTA milk, and that increasing dam PTA milk increased the probability of a heifer being culled prior to first lactation. If the herd effect was correlated with PTA the estimates from the multinomial logit model would be asymptotically biased.

Theoretically, the ordered FE probit by accounting for the herd effect and utilizing the information contained in the ordering of the outcomes, should result in performance superior to the multinomial specification. The results of the ordered FE probit model are presented in Table 8.6.

The coefficients from ordered models are very difficult to interpret (Greene, 1993). While in general, even the signs of the coefficient can not be interpreted, in the present case the coefficient signs do represent the direction a change in the respective independent will have on the probability of each type of cull. The signs of all the coefficients except for that associated with sire PTA milk are identical to those from the FE probit. As the standard errors have not been corrected for correlation among the residuals, the true *P* values will be larger than those reported. Table 8.7 represents the mean effect and 95% confidence intervals of a 100 kg change in PTA on the probability of being culled for each culling classification evaluated at the mean cohort size. Except for the effect of dam PTA protein, the effects from the ordered FE probit model were of the same sign and similar magnitude as those effects from the FE probit and logit models. The reason for the negative effect of dam protein is that negative herd mean dam protein effect overwhelms the positive individual dam protein. The standard errors of the derivatives are generally large relative to the mean effects, multicollinearity would increase the standard errors while leaving the mean effects asymptotically unbiased.

Table 8.8 represents the predicted culling classification from the ordered FE probit model versus the actual culling classification for only those animals which were actually culled. Similar to the dichotomous FE probit model predictions, the predicted outcomes performed poorly and fell overwhelmingly into the most prevalent classification which for the ordered FE probit model was culled for other reasons.

The results in Table 8.7 suggest that the direction of the effect of changing a given parent's PTA is consistent for each of the three culling classifications. In general the magnitude of that effect is lowest for production culls and highest for heifers culled prior to lactation, although it is doubtful that the differences among the classifications are practically significant. The relatively large standard errors indicate the differences among the classifications are not statistically different either. While it is plausible for genetics to have the greatest effect on heifer culling, one would have expected genetics to have greater effects on culling for production than other culling reasons. The results in Table 8.7 indicate that genetics play a greater role in culling of first lactation animals for what has traditionally been considered involuntary reasons and that genetics play a lesser role in culling for low production which has generally considered a voluntary culling decision. This result is difficult to accept. A more plausible conclusion is that, at a minimum, the classification of culls into voluntary and involuntary for lactating animals is inappropriate and rather these results are a function of modeling nonexistent classifications. The failure to correctly predict the culling classification supports this argument. Of course, the alternative explanation for the unexpected results with regard to Table 8.7 is classification error. But again this explanation by suggesting producers have such difficulty classifying animals into voluntary and involuntary classes further supports the lack of utility of this approach.

Upon reconsideration, the failure to find results consistent with the concept of voluntary and involuntary culling is not surprising. Every culling decision involves a voluntary choice as supported by Grohn et al. (1998) whose "research indicates dairy farmers consider many factors, including diseases, milk yield, conception status, parity, and stage of lactation, when deciding whether and when to cull a cow." The dairy industry must resolve this intuitively obvious statement by Grohn et al. with another equally generally accepted view that "Dairy cows may be culled for either involuntary reasons (i.e., death, acute disease, infertility) or voluntary reasons (i.e., low yield)" which is included in their introduction. Sudden death and sterility, but not infertility, are the only afflictions of a milk cow which result in a producer having no choice in the culling decision. This is not to suggest animal removal reasons have no place in herd management, but rather the concept of the voluntary nature of the removals has little place, particularly in culling and replacement strategies.

Moreover, the issue of voluntary versus involuntary culling is not merely a matter of schematics, but rather may have practical importance in improving producers' decisions. The concept of involuntary and voluntary culls may be the greatest impediment to producers realizing increased profit through the strategy of lower culling rates. For if producers and their advisors continue to believe they have no choice in many of their culling decisions, their opportunity to change these decisions and their culling rates is severely inhibited.

Table 8.9 contains the results of regressing culling on MEM using the FE probit and logit models. As expected in each model the coefficient associated with MEM is negative indicating higher production has a protective effect on the probability of being culled. The probability of being culled drops by 7.3% and 0.1% for every extra 1,000 kg. of MEM for the FE probit and FE logit models, respectively. In the FE probit model, first lactation milk production is explaining almost 15% of the variation in the probability of first lactation culling, and when conditioned on the number of heifers in the cohort that were culled during first lactation, MEM explained over 20% of the variation in culling in the FE logit model.

While first lactation production did not explain a large portion of the variation in first lactation milk production it's predictive ability, as presented in Tables 8.10 and 8.11, was fair and improved over the predictive ability of parents' genetics. Of those animals the model predicted would be culled 71.5% actually were, while 83.0% of the animals predicted to be kept were correct. The primary shortcoming of the FE probit model's predictive performance was predicting only 656 (5.4%) of the heifers would be culled. While 469 of these predictions were correct, the predictions underestimated the 20% culling that actually occurred in the population. Among the 238 cohorts with a single animal culled, the FE logit model correctly predicted a single animal was culled in 106 of these heifer cohorts. And for 73.6% of these 106 cohorts the FE logit model correctly predicted which animal was actually culled.

So despite a mean probability of being culled of 20%, using MEM the models still had reasonable predictive performance. This suggests the poor predictive ability of the models based on parents' genetics was not due to the low mean culling outcome of 0.24, but rather the poor model performance was due to the inability of parents' genetic estimates to explain the variation in culling.

Inability of parents' genetic estimates to predict culling, despite milk production affecting producers culling decisions can be explained based on Chapter 7 results. While milk production plays a role in producers' culling decisions, since from Chapter 7 it was revealed that parents' genetic estimates explained less than 7% of the variation in first lactation milk production, it is not surprising that genetics explaining virtually none of the culling decisions.

4. Conclusion

Despite evidence in the present data set that first lactation milk production plays a role in producer culling decisions, parents' PTA of milk, fat, and protein available following the daughters' births did not predict subsequent culling during rearing or first lactation. Even when conditioning on information that would not be available for culling predictions, parents' PTA were still unable to predict culling. Given that first lactation milk production explains less than 20% of the variation in culling, the poor predictive ability of parents' PTAs for subsequent culling are consistent with the Chapter 7 results that parents' genetic estimates explain less than 7% of the within herd variation in first lactation milk production. Furthermore, the failure to predict whether a culled animal was culled for voluntary or involuntary reasons, along with the unexpected effects that parents' genetics had on the probability of being culled for each of these classifications suggest voluntary and involuntary culls do not exist, and may be a barrier to management. Moreover, the consistency of the result across the unordered FE probit and logit models, as well as the ordered FE probit model suggests these results are not a function of the distributional and other assumptions made in each of the models. These results also suggest that Michigan producers are not using parent genetics in heifer culling decisions.

Acknowledgments

The authors thank Al Thelen, Terry Speer and Lyle Smith at Michigan DHI, as well as, Leigh Walton and H. D. Norman at the Animal Improvement Laboratory, USDA for assistance with data collection. Jeff Wooldridge's econometric comments were greatly appreciated. The financial support of the primary author by the Department of Agricultural Economics through the Animal Industry Initiative funds provided by the Agricultural Experiment Station at Michigan State University while this work was conducted is gratefully acknowledged.

References

Abrevaya, J. 1997. The equivalence of two estimators of the fixed-effects logit model. Economics Letters, 55:41-43.

Allaire, F. R. 1981. Economic consequences of replacing cows with genetically improved heifers. J. Dairy Sci. 64:1985-1995.

Allaire, F. R., and E. P. Cunningham. 1980. Culling on low milk yield and its economic consequences for the dairy herd. Livest. Prod. Sci. 7:349-359.

Cassell, B. G., B. B. Smith, and R. E. Pearson. 1993. Influence of herd-life opportunity and characteristics of cows and herds on different net income functions. J. Dairy Sci. 76:1182-1190.

Chamberlain, G. 1984. Ch. 22 Panel data. In: Z. Griliches and M. D. Intriligator (Editors), Handbook of Econometrics, Volume II. Elsevier, North-Holland, pp. 1247-1318.

Congleton, W. R. 1988. Dairy cow culling decisions. 2. Profitability and genetic trend in herds with culling on production versus income. J. Dairy Sci. 71:1905-1915.

Congleton, W. R., and L. W. King. 1985. Culling cows on predicted future profitability with a variable planning horizon. J. Dairy Sci. 68:2970-2983.

Connor, L. J., L. G. Hamm, S. B. Nott, D. Darling, W. Bickert, R. Mellenberger, H. A. Tucker, O. B. Hesterman, J. A. Partridge, and J. H. Kirk. 1989. Summary of the 1987 Michigan State University dairy farm survey. Research Report 498, Michigan Ag. Exper. Station, East Lansing.

Greene, W. H. 1993. Econometric Analysis, 2nd ed., Macmillan Publishing Company, New York.

Grohn, Y. T., S. W. Eicker, V. Ducrocq and J. A. Hertl. 1998. Effect of diseases on the culling of Holstein dairy cows in New York state. J. Dairy Sci. 81:966-978.

Hausman, J. A. 1978. Specification tests in econometrics. Econometrica 46:1251-1270.

Heckman, 1981. Ch. 4 The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: C. F. Manski and D. McFadden (Editors), Structural Analysis of Discrete Data with Econometric Applications, the MIT Press, Cambridge, Massachusetts, pp. 179-195.

Heinrichs, A. J., S. J. Wells, H. S. Hurd, G. W. Hill, and D. A. Dargatz. 1994. The national dairy herd evaluation project: A profile of heifer management practices in the United States. J. Dairy Sci. 77:1548.

Henderson, C. R. 1963. Selection index and expected genetic advance. Page 141 *in* Statistical Genetics and Plant Breeding ; a symposium and workshop. W. D. Hanson and H. F. Robinson, ed. Natl. Acad. Sci.-NRC, Washington, DC.

Hsiao, C. 1986. Analysis of Panel Data. Cambridge University Press, Cambridge.

Huirne, R., A. Dijkhuizen, R. King, and S. Harsh. 1993. Goal, critical success factors, and information systems needs on swine and dairy farms – Identified by workshops. Zeitschrift für agrainformatik 3:61-65.

Killen, L., and B. Kearney. 1978. Optimal dairy cow replacement policy. Ir. J. Agric. Econ. Rur. Sociol. 7:33-40.

Korver, S., and J. A. Renkema. 1979. Economic evaluation of replacement rates in dairy herds II. Selection of cows during the first lactation. Livest. Prod. Sci. 6:29-37.

Kuipers, A. 1980. Development and comparison of selection criteria for cows and bulls with a dairy herd simulation model. Ph.D. Thesis, Univ. Wisconsin, Madison.

McCullough, D. A., and M. DeLorenzo. 1996. Effects of price and management level on optimal replacement and insemination decisions. J. Dairy Sci. 79:242-253.

Michigan Dairy Herd Improvement Association. 1996. Michigan dairy production herd evaluation report.

Pearson, R. E., and A. E. Freeman. 1973. Effect of female culling and age distribution of the dairy herd on profitability. J. Dairy Sci. 56:1459-1471.

Quinton, V. M., and C. Smith. 1997. An empirical check on best linear unbiased prediction genetic evaluation using pig field recording data. Can. J. Anim. Sci. 77:211-216.

Rogers, G. W., J. A. M. van Arendonk, and B. T. McDaniel. 1988. Influence of production and prices on optimum culling rates and annualized net revenue. J. Dairy Sci. 71:3453-3462.

Stiratelli, R., N. Laird, and J. H. Ware. 1984. Random-effects models for serial observations with binary response. Biometrics 40:961-971.

Van Arendonk, J. A. M. 1991. Use of profit equations to determine relative economic value of dairy herd life and production from field data. J. Dairy Sci. 74:1101-1107.

Van Arendonk, J. A. M. and A. A. Dijkhuizen. 1985. Studies on the replacement policies in dairy cattle. III. Influence of variation in reproduction and production. Livest. Prod. Sci. 13:333-349.

Weigel, D. J., B. G. Cassel, I. Hoeschele, and R. E. Pearson. 1995. Multiple trait prediction of transmitting abilities for herd life and estimation of economic weights using relative net income adjusted for opportunity cost. J. Dairy Sci. 78:639-647.

Table 8.1.
Parent PTA of milk, fat and protein for 5,619 animals.

Variable[1]	$\bar{X}$	SD	Minima	Maxima
PTA Milk of Sire (100 kg)	8.43	2.38	-9.53	14.02
PTA Fat of Sire (100 kg)	0.26	0.09	-0.40	0.53
PTA Protein of Sire (100 kg)	0.23	0.07	-0.30	0.43
PTA Milk of Dam (100 kg)	3.66	2.71	-8.54	14.24
PTA Fat of Dam (100 kg)	0.11	0.10	-0.35	0.57
PTA Protein of Dam (100 kg)	0.10	0.08	-0.29	0.41

Table 8.2.
Coefficients and their standard errors from the FE probit and logit models of regressing culling on parent's PTA.

	FE Probit		FE Logit	
Variable	b	SE[1]	b	SE
Sire Milk	-0.001	0.012	-0.005	0.023
Sire Fat	0.221	0.270	0.427	0.515
Sire Protein	-0.799†	0.461	-1.478	0.905
Dam Milk	-0.056**	0.018	-0.106**	0.034
Dam Fat	0.690*	0.298	1.332*	0.634
Dam Protein	1.248†	0.684	2.327†	1.258
Herd Mean Sire Milk	-0.026	0.031		
Herd Mean Sire Fat	0.471	0.732		
Herd Mean Sire Protein	0.535	1.328		
Herd Mean Dam Milk	0.118*	0.057		
Herd Mean Dam Fat	0.696	0.970		
Herd Mean Dam Protein	-4.376†	2.313		
X^2	28.6**	12 df	18.6**	6 df
Pseudo R^2	0.0045		0.0048	
n	5,619		4,652	

[1] Robust to correlation among the residuals.

$^{\dagger}P \leq 0.10$; $^{*}P \leq 0.05$; $^{**}P \leq 0.01$.

Table 8.3.
Mean effect of a 100 kg change in PTA on probability of being culled, and associated 95% confidence intervals (CI).

	FE Probit		FE Logit	
Variable	Mean	95% CI	Mean	95% CI
Sire Milk	-0.002	(-0.009, 0.005)	-0.001	(-0.012, 0.010)
Sire Fat	0.090	(-0.064, 0.244)	0.105	(-0.146, 0.355)
Sire Protein	-0.227	(-0.493, 0.039)	-0.363	(-0.794, 0.069)
Dam Milk	-0.012	(-0.023, -0.001)	-0.026	(-0.042, -0.010)
Dam Fat	0.246	(0.063, 0.429)	0.327	(0.021, 0.633)
Dam Protein	0.200	(-0.217, 0.617)	0.571	(-0.034, 1.181)

Table 8.4.
Predictive ability of FE probit model.

	Actual		
Predicted	Culled	Kept	Total
Culled	0	0	0
Kept	1370	4249	5619
Total	1370	4249	5619

Table 8.5.
Predictive ability of FE logit model for the 174 heifer cohorts with a single animal culled.

Predicted Number of Animals Culled per Herd	Number of Herds	Percentage of Herds
0	137	78.7
1	35	20.1
2	2	1.1
Total	174	99.9

Table 8.6.
Coefficients and their standard errors from the ordered FE probit model of regressing culling classification on parent's PTA.

Variable	b[1]	SE
Sire Milk	0.001	0.012
Sire Fat	0.237	0.274
Sire Protein	-0.750	0.484
Dam Milk	-0.040*	0.018
Dam Fat	0.724*	0.341
Dam Protein	0.554	0.681
Herd Mean Sire Milk	-0.025	0.024
Herd Mean Sire Fat	0.672	0.558
Herd Mean Sire Protein	0.322	0.972
Herd Mean Dam Milk	0.111**	0.042
Herd Mean Dam Fat	0.462	0.786
Herd Mean Dam Protein	-3.998*	1.686
μ_1	0.673	0.111
μ_2	1.140	0.112
μ_3	1.325	0.112
X^2	24.4*	12 df
Pseudo R^2	0.0027	
n	5,619	

*$P \leq 0.05$; **$P \leq 0.01$. As the standard errors have not been corrected for correlation among the residuals the reported *P* values understate the true values.

Table 8.7.
Mean effect and standard errors (SE) of a 100 kg change in PTA on the probability of being culled for each culling classification.

	Culling Classification					
Variable	Other[1]		Production[2]		Heifer[3]	
	Mean	SE	Mean	SE	Mean	SE
Sire Milk	-0.0003	0.0012	-0.0001	0.0005	-0.0004	0.0018
Sire Fat	0.0367	0.0271	0.0132	0.0104	0.0469	0.0409
Sire Protein	-0.0783	0.0478	-0.0283	0.0182	-0.1000	0.0715
Dam Milk	-0.0027	0.0018	-0.0010	0.0007	-0.0035	0.0027
Dam Fat	0.0874	0.0342	0.0316	0.0131	0.1116	0.0528
Dam Protein	-0.0001	0.0935	-0.0001	0.0263	-0.0002	0.1005

[1] Culled during first lactation for reasons other than low production.
[2] Culled during first lactation for low production.
[3] Culled prior to first lactation.

Table 8.8.
Predictive ability of ordered FE probit model for those animals which were actually culled.

	Actual			
Predicted	Other[1]	Production[2]	Heifer[3]	Total
Other[1]	678	187	502	1367
Production[2]	0	0	0	0
Heifer[3]	0	2	1	3
Total	678	189	503	1370

[1] Culled during first lactation for reasons other than low production.
[2] Culled during first lactation for low production.
[3] Culled prior to first lactation.

Table 8.9.
Coefficients and their standard errors from the FE probit and logit models of regressing culling on MEM.

	FE Probit		FE Logit	
Variable	b	SE[1]	b	SE
MEM	-0.032***	0.001	-0.060***	0.002
Herd Mean MEM	0.023***	0.002		
X^2	1023.8***	2 df	1798.46***	1 df
Pseudo R^2	0.148		0.209	
n	12,128		10,760	

[1]Robust to correlation among the residuals.
***$P \leq 0.001$.

Table 8.10.
Predictive ability of FE probit model of regressing culling on MEM.

	Actual		
Predicted	Culled	Kept	Total
Culled	469	187	656
Kept	1,952	9,520	11,472
Total	2,421	9,707	12,128

Table 8.11.
Culling predictive ability of FE logit model using MEM for the 238 herds with a single animal culled.

Predicted Number of Animals Culled per Herd	Number of Herds	Percentage of Herds
0	132	55.5
1	106	44.5
Total	238	100.0

Chapter 9

COMPLEX SOLUTION TO THE VALUE OF GENETIC INFORMATION IN PREPARTUM HOLSTEIN HEIFER SELECTION

Introduction

Genetics is a source of information that aids livestock producers in animal breeding and selection decisions. Like other information sources, the value of genetic information is difficult to discern. Previous studies on the value of genetic information have addressed the issue of how genetic information should be valued, but did not address the value of the genetic information *per se*. Hedonic pricing has been used to marginally value genetic traits of sires to form a pricing mechanism for semen (Schroeder et al. 1992; Richards and Jeffrey, 1996). At a more aggregate level, consideration has been given to assessing the proper economic weights of various genetic traits to maximize producer income (Harris and Freeman, 1993). But again the value of this genetic information was not directly addressed.

The value of information (Antonovitz and Roe, 1987; Nermuth, 1982) is a natural framework to assess the value of genetics in producer decision making. In this framework the information is represented by some arbitrarily accurate signal of an underlying random variable or event. Based on the signal, the information user makes a decision. The value of the information is represented by the increase in the decision maker's welfare by using the signal versus not using the signal in the decision. This framework suggests that the value of the information is dependent on the signal and the decision to be made based on the signal.

Genetic evaluations are signals of an animal's true genetic ability. Furthermore these signals are dynamic – as more information is collected on the animals and their relatives more informative signals are produced. The value of the genetic information is specific to the context of the decision, here we focus on the value of genetic information in selection of prepartum Holstein heifers.

Given a constant herd size, no seasonal effects on reproduction and production, and culled cows being sold for meat (not dairy) purposes, optimization (McCullough and DeLorenzo, 1996; Rogers et al., 1988; Congleton and King, 1985; Van Arendonk and Dijkhuizen, 1985; Killen and Kearney, 1978) and simulation (Allaire, 1981; Allaire and Cunningham, 1980) studies were unanimous in their conclusions that a culling rate of 20% to 30% optimizes producer profit. While other simulation studies (Congleton, 1988; Kuipers, 1980; Korver and Renkema, 1979; Pearson and Freeman, 1973), which considered subjectively selected culling rates or strategies for comparison, indicated that profit would be increased with culling rates below the average 37% culling rate[1] in the Michigan dairy industry in 1996 (Michigan Dairy Herd Improvement Association, 1996).

The results of these multiple studies are compelling because, not only did the various models employ different assumptions and analytical techniques, but as well, the models operated under economic conditions representative of a number of countries (i.e., the U.S.A, England, Ireland, Netherlands). In the models, the increased profit associated with use of the optimal culling rates varied from 2% to 40%.

---

[1] Culling rate, annually defined as, the number of animals leaving the herd due to sale or death divided by the average herd size.

Comparison of only the mean profit by the studies implies risk neutrality, whereas a majority of Michigan and Dutch dairy producers are risk adverse (Huirne et al., 1993). The risk preferences of decision makers can have significant effects on their choices, so the optimal rule assuming risk neutrality cannot necessarily be extrapolated to those who are risk averse.

At least under risk neutrality, one direct ramification of Michigan's excessive culling rate is that Michigan dairy producers do not need all potential replacement dairy heifers to enter the lactating herd, necessitating a method of heifer selection.

This need for heifer selection is in direct contrast to the present situation in which most producers rear and freshen nearly all of heifers produced on the farm (Connor et al., 1989; Heinrichs et al., 1994). Animals are then selected largely on the basis of the level of first lactation and expected future milk production. The simulation studies revealed that while this practice maximizes genetic progress and milk production, the economic performance is lower due to failure to consider the opportunity costs of higher culling rates.

The culling models can lend little insight into heifer selection. For while the lactating herd, which was the primary focus of these studies, was modeled with sophistication, the heifer modeling was simplistic. None of the culling studies dealt with the issue of when in the rearing period was the most profitable time to make the selection decision. Generally the models assumed that replacement heifers were available when required. This was operationalized in the studies through either the sale of all animals at birth and purchase of required replacements, or raising of all young stock with subsequent sale of the excess animals.

Regardless of the method of heifer procurement, most optimization studies assumed naïve expectations with respect to future heifer profitability, and as a result employed random replacement selection. Many simulation studies arbitrarily selected replacements based on their simulated estimated breeding value (EBV) of milk. The EBV milk, which represents the animal's estimated genetic ability to produce milk, is the average of the parents' potential transmitting abilities (PTAs) for milk. PTAs, which are estimated by the USDA, represent the genetic ability the parents are expected to transmit to their offspring.

While selection based on genetics has intuitive appeal and indeed theoretical justification (Henderson, 1963), empirical support for heifer selection on this basis is lacking.

The objective of this study was to determine the value of parents' PTAs of milk, fat and protein and their associated reliabilities in the prepartum selection of Holstein heifers, and assess whether this value was sufficient to prompt producers to select heifers on this basis. In order to accomplish the above objective it was necessary to derive an optimal heifer decision rule.

Theoretical Framework for the Value of Genetic Information in Prepartum Heifer Selection

The ultimate objective of heifer selection is to choose the most profitable animals.

Van Arendonk showed that lifetime profit corrected for the opportunity cost of postponed replacement is the correct method for ranking animals in terms of profitability. This measure of an animal's lifetime profit explicitly recognizes that, in retaining the animal in the herd, the average potential replacement's profit was foregone.

With perfect information, heifer selection would be a two step procedure for producers. Given that they desire to select *N* number of heifers, the heifers would be ranked by lifetime profit corrected for opportunity cost and the top *N* selected. The remaining heifers would then be culled. The second decision is when in the rearing period these remaining heifers should be sold. The profit maximizing producer would sell the animals in the period in which the gross sale price less the cumulative average variable rearing cost is the greatest. This can be mathematically represented as

(1)

$$\max_{s_{at}} \Pi_h^{perfect} = \sum_a \left[\left(1 - \sum_t s_{at}\right)\pi_a + \sum_t s_{at} r_t - \sum_t \left(1 - \sum_{j=1}^{t-1} s_{aj}\right) c_t \right], \quad a = 1, \ldots, A_h; t = 1, \ldots, T;$$

subject to

$$\sum_t s_{at} \in (0,1)$$
$$\sum_a s_{at} = n_h$$

Where

Π_h is the profit of herd h from heifer selection

$$s_{at} = \begin{cases} 1 & \text{if animal } a \text{ is not selected (i.e. is culled) in period } t \\ 0 & \text{otherwise} \end{cases}$$

π_a is the lifetime profit of animal a corrected for opportunity cost

r_t is the revenue from the sale of a heifer in period t

c_t is the average variable rearing cost of a heifer in period t

n_h is the number of heifers, in herd h, to be sold

A_h is, for herd h, the total number of heifers which form a cohort for selection purposes

T is the total number of periods in the rearing process.

The simplicity of the objective function (1) is apparent when animals to be culled and those not culled are separately considered. For that portion of the herd culled in period i the profit function simplifies to

$$\Pi_h^{culled} = \sum_a \left[r_i - \sum_{t=1}^{i} c_t \right], \quad a = 1,\ldots,A_h^{culled}; \quad t = 1,\ldots,T.$$

While for the portion of the heifer herd that enters the lactating herd the profit function is

$$\Pi_h^{selected} = \sum_a \left[\pi_a - \sum_{t=1}^{T} c_t \right], \quad a = 1,\ldots,A_h^{selected}; \quad t = 1,\ldots,T.$$

Unfortunately, corrected lifetime profit of animals is generally not available and furthermore, can not be determined with certainty prior to the culling of the animals. However, the literature has focused on indicators of lifetime profitability that are apparent early in the productive life of milk cows. No work has considered signals apparent prior to productive life.

The correlation between first lactation actual milk and lifetime profit corrected for opportunity cost has been reported at 0.54 (Cassell, Smith and Pearson, 1993). While Weigel et al. determined the correlation between corrected lifetime profit and first lactation mature equivalent milk production, which is actual milk production standardized for effect of age and season of calving, to be 0.51. For purposes of heifer selection, first lactation milk production may then be a suitable proxy for corrected lifetime profit. Then by choosing animals based on first lactation milk production producers have selected the correct animals.

However, this strategy of milking all heifers and heavily culling the herd to select the highest producing animals is costly. Given the large price differential between a heifer just prior to entering milk production and her value as a cull cow, the increased profit from even randomly selling some heifers prior to milking, and thereby lowering the culling rate, more than compensates for the decreased ability to identify and select the highest producing animals. In addition, with fewer heifers in the milking herd the proportion of older, more profitable cows in the herd rises.

Currently the information provided to Michigan producers upon which to base culling predictions is limited to the heifers' EBV of milk, fat and protein production. The accuracies of these signals, termed reliabilities, are also estimated by the USDA. Reliability is the squared correlation of an animal's estimated genetic value and its true genetic value. Thus reliabilities are the measures of the genetic signal's informativeness. Because EBV milk provide signals of the genetic merits for milk production, which should on average manifest as actual performance, EBV milk or alternatively parents' PTA milk, may have a role in heifer selection.

Working with swine, Quinton and Smith found the within herd rank correlation with progeny performance was higher for parents' genetic estimates than for parents' phenotypes. However they warn the benefits of genetic estimates over phenotype in selection may be less in practice than in theory. From Chapter 7 it was apparent heifers parents' PTAs explained less than 7% of the within herd variation in subsequent first lactation milk production, and Pearson rank correlations between the linear model predictions of first lactation milk production and actual production cluster around the value of one-third.

This result is not surprising for, due to Mendelian sampling and environmental variance, the reliability of a prepartum heifer's genetic estimated producing ability is only one-sixteenth of the parents' summed PTA reliabilities. At a maximum, a heifer's parents' PTAs will have reliabilities of 0.40 and 0.10 in predicting the daughter's genetic ability to produce milk, and actual subsequent milk production, respectively. So genetics can be considered a distorted or incomplete signal of lifetime profit corrected for the opportunity cost of postponed replacement.

Due to the positive correlation between PTA fat and protein with PTA milk, genetic selection index theory, as well as statistical theory, suggest the signal will be improved by incorporating information on these three PTAs of the parents. Moreover, by indicating the degree of distortion in the PTA signals, the reliabilities may themselves be valuable in heifer selection. Following the definition of Nermuth, the signal incorporating information on the parents' three PTAs and associated reliabilities must be more informative than a signal based solely on parents' PTAs of milk, given that the latter is a subset of the former. Furthermore, regardless of the source of information for the signal, the signal will improve in informativeness over the course of the rearing period of the heifers as the genetic estimates are periodically re-estimated as more information on the heifer's relatives are collected.

The application of capital asset replacement theory presents a caution, in addition to that provided by Quinton and Smith, in the utility of genetics in heifer selection. The study (Allaire, 1981) which based replacement selection on EBV had a lower optimal culling rate than all of the studies which randomly selected heifers (Allaire, and Cunningham, 1980; Congleton, and King, 1985; Killen, and Kearney, 1978; Rogers, van Arendonk, and McDaniel, 1988; Rogers, van Arendonk, McDaniel, 1988), except one (McCullough, and DeLorenzo, 1996). Standard capital asset replacement theory indicates if the profitability of the replacements selected based on genetics was higher than those selected randomly, the optimal length of the more profitable heifers' lives would decrease (Robison and Barry) which is equivalent to an increase in the culling rate of the simulation studies relative to those of the optimization studies. Then one possible explanation for the similarity between the most profitable culling rates of the optimization and simulation studies, is that the genetic information is not valuable in heifer selection.

Uncertainty in predicting a heifer's subsequent profitability can be incorporated into the initial objective function by noting that lifetime profit corrected for opportunity cost is a function of first lactation milk production (*L*) which in turn is a function of the heifer's estimated genetic information (G_{at}) which should increase in informativeness over time.

(2)

$$\max_{s_{at}} \Pi_h = \sum_a \left[\left(1 - \sum_t s_{at}\right) \pi_a\left(L_a\left(G_{at}\right)\right) + \sum_t s_{at} r_t - \sum_t \left(1 - \sum_{j=1}^{t-1} s_{aj}\right) c_t \right], \quad a = 1, \ldots, A_h; t = 1, \ldots, T$$

The profit associated with random selection can be represented as

(3)

$$\max_{s_t} \Pi_h^{random} = \sum_a \left[\left(1 - \sum_t s_t\right) E[\pi_a(L_a(G_a))] + \sum_t s_t r_t - \sum_t \left(1 - \sum_{j=1}^{t-1} s_j\right) c_t \right], \quad a = 1,\ldots,A_h; t = 1,\ldots,T$$

The lifetime profits corrected for opportunity costs is the expectation (E) of the heifer cohort, and the only selection decision is with respect to which period to sell the excess heifers.

The *ex ante* value of genetics in heifer selection can then be assessed by Π_h - Π_h^{random}. If genetics are of no value in heifer selection then $\Pi_h = \Pi_h^{random}$. The *ex post* value of genetic information in heifer selection is represented by $\Pi_h^{pefect} - \Pi_h^{random}$.

Theoretical Decision Rule for use of Genetic Information in Prepartum Holstein Heifer Selection

To assess the value of genetic information in prepartum Holstein heifer selection, a decision rule transforming the information into action by the decision maker is required. Given genetic theory's silence on how to optimally select heifers based on dynamic genetic information, a decision rule was developed.

Let the optimal decision rule for prepartum Holstein heifer selection be s_{at}(I, p) subject to

$$0 \leq p_t \leq 1$$

$$S = \prod_{t=1}^{T} p_t$$

Where

I is a partitioned matrix of dimension A_h by T, composed of elements $\mathbf{g}_{at}\mathbf{w}_t$, where $\mathbf{g}_{at}$ is a 1 x B vector, containing the B pieces of genetic information for animal a available at time t; $\mathbf{w}_t$ is a B x 1 vector of weights for the B pieces of genetic information in time period. $\mathbf{g}_{at}\mathbf{w}_t$ then is a composite signal of animal a's genetic information at time t.

p is a 1 x T vector whose elements, p_t, are the proportions of animals to be selected in period t.

S is the overall proportion of animals to be selected.

The decision rule begins with the first period (t=1) in which the heifers are ranked by $\mathbf{g}_{at}\mathbf{w}_t$ and the top $p_1 \cdot A_h$ are selected, and the remaining $(1-p_1) \cdot A_h$ are sold. The number of animals to be selected, $p_1 \cdot A_h$, is rounded to the nearest integer. The remaining animals are then re-ranked in the second period (t=2) by the second period composite signals, with $p_2 \cdot (p_1 \cdot A_h)$ selected and the remaining sold. This process continues to the T^{th} period, by which time only the desired overall percentage of heifers remain. These animals will then realize their first lactation and associated estimated profit. Heifer selection can then be considered in a dynamic value of information issue, weighing the value of the more informative signal versus the cost of its attainment.

Deriving and Testing the Optimal Decision Rule - Empirical Methods

With the herd, as opposed to the individual animal, being the unit of interest, and without knowledge of either the temporal relationship among the genetic information or between the information and lifetime profit, the dynamic decision rule of how and when to select heifers is not amenable to standard optimization techniques such as dynamic programming. Rather Modified Box-Complex, a derivative free non-linear search technique was employed. This method of sequential searching has been found useful in globally optimizing nonlinear multivariable objective functions subject to constraints (Harris, 1981).

Based on the description by Harris, Kuester and Mize's procedure for Box's original Complex algorithm was modified to include Nelder-Mead's flexible polyhedral procedures.

In exploring the effect of the input parameters on the objective value it was noted Modified Box-Complex was not consistently converging to a global optimum. Rather, over a set of 10 runs with a given set of parameters and 100 randomly generated vertices or vectors of search variables it was noted the final objective values varied by as much as 4.4%. The random creation of 1,000 initial vertices, of which only the 100 with the highest objective values were subsequently used in the algorithm improved the consistency of the results. Using this technique, for the same set of input parameters, the range of the objective values decreased 25% despite increasing the number of runs to 20. The search variables were limited to three decimal places for the 10 runs and then two decimal places for the runs of 20.

The data set consisted of Michigan Holstein heifers that were born between July 1, 1992 and December 31, 1992 and first calved between July 1, 1994 and December 31, 1994. Genetic information collected consisted of the heifers' parents' PTA for milk, fat and protein along with the milk-fat reliability and protein reliability from the 1993 to 1994 semi-annual sire and dam animal model evaluations calculated by the Animal Improvement Programs Laboratory, USDA (Beltsville, MD), the exception being the dam genetic estimates from July 1993 which were unavailable. Heifer birth dates, calving dates and first lactation mature equivalent milk production were collected from Michigan DHI.

In effect then the data sample consisted of, for each herd, a cohort of heifers that were born within a six month window of time and calved within a six month window of time, with the two windows being separated by two years. Herds whose heifer cohort contained less than 10 heifers were deleted, leaving a sample consisting of 1,982 heifers in 115 herds. The lifetime profit corrected for opportunity cost, associated with a given level of mature equivalent first lactation milk production, was deterministically estimated for each heifer based on information provided by Weigel et al. Generation of this profit measure likely misrepresented the true distribution of heifer profits in two ways. The use of a linear relationship between lifetime profit corrected for opportunity cost and first lactation milk production assumed a constant marginal profit for milk, and as a result over estimated the profit of high producing animals. Secondly, Weigel et al. generated the relationship based on nonoptimal culling practices, which corresponded to lower heifer profits. The resulting distribution of heifer profits would be narrower than the true distribution due to compression of the upper tail of the distribution.

Table 9.1 contains descriptive statistics for $\mathbf{g}_{a1}$, $\mathbf{g}_{a4}$, L_a, and π_a from the data sample. The milk PTAs are one to two orders of magnitude greater than the other genetic information. Relative to their mean, the reliabilities have less variation than the PTAs. The respective means of the genetic information rose consistently over the four periods. Within herd, π_a generally had a range of under $2,500.

Thirty-eight search variables then impacted the objective value: the ten weights for the genetic information in periods one, three and four, along with the five weights in period two, and the proportion of heifers to be selected in each of periods one through three.

The number of animals selected in the fourth and final period was equal to S, so the number sold in the final period was $p_1 * p_2 * p_3 * A_h - S * A_h$ if the result was positive. To increase efficiency of the algorithm, the implicit constraint on the number selected, was incorporated in the objective function as an artificial activity. Specifically, during the last of the four periods, if $S * A_h > p_1 * p_2 * p_3 * A_h$ heifers could be purchased for $1,000 and a $0 lifetime profit corrected for opportunity cost. To prevent difficulties with finding corner point solutions, p_1, p_2 and p_3, were permitted to be greater than one, however the number of animals available in periods prior to the fourth was constrained to be less than or equal to the number in the previous period.

The input parameters required for the heifer selection problem included S and the r_t. To make the transition from Michigan's 1996 average culling rate of 37% to a 30% culling rate suggests 80% of heifers need to be selected. In their calculation of 1994 π_a, Weigel et al. included the cost of rearing, so only the net revenue from the sale of heifers in each period was required for the empirical objective function. To generate the net revenues associated with heifer sale in a given period, it was assumed that fixed costs, such as building and machinery overhead, were sunk. Based on dairy heifer budgets for 1993 by Karszes, the total cost to raise a heifer in each period were calculated. The net return per period from sale of a heifer was then the 15% sunk portion of total cost to raise the animal in the respective period.

Given that the USDA genetic estimates are overwritten when re-estimated, an extended time frame would be required to collect multiple heifer cohorts for each herd in order to generate an optimal selection rule for each herd. Rather a cross-sectional approach to generate a single decision rule, which maximizes the summed profit of the individual herds was opted for. The advantage of this method, besides the decreased data requirements, is greater ease of decision rule implementation. However, the herd profits generated with the cross-section approach cannot be greater than those of the time-series approach. The new objective function associated with the cross-section approach can then be represented as

$$(4) \qquad \max_{s_{\alpha}} \sum_{h} \Omega_{h}, \quad h = 1, \ldots, H .$$

Where $\Omega_h = \Pi_h / A_h$.

The use of omega, the average herd profit per heifer, imposes equal weighting of each herd, regardless of its size, in the objective function.

Sample size calculations determined the requisite number of herds for one-tailed analysis of paired t tests to detect a $20 difference in profit per heifer between random selection and that associated with some selection rule which required information. Based on the calculations, to detect a $20 difference the 115 herds were randomly divided in two groups following stratification on herd size. For the average herd in Michigan, a $20 difference would result in roughly a $1,000 annual increase in herd profit which we felt would be sufficient to interest producers in collecting the necessary information and using the rule. The 58 herds used by Complex to derive the optimal decision rule, denoted as the rule deriving sample, were comprised of 1,034 animals and ranged in size from ten to one-hundred and thirty-two heifers, average of 17.8 and median of 14. The remaining 948 heifers in 57 herds of ten to seventy-three heifers, average of 16.3 and median of 14, termed the rule testing sample, was used to test the performance of the heifer selection rules.

To evaluate the effect of the input parameters S and r_t on rule performance the three different scenarios listed in Table 9.2 were considered. In scenario 2, the effect of the percentage of heifers selected on the optimal rule was evaluated through specifying a 70% selection percentage. This latter selection rate represents the lower selection percentage required by producers with minimal heifer death loss or desiring a lower cow culling rate than 30%. While in scenario 3 the effect of the r_t were evaluated. Specifically, later sale of heifers incurred increasing losses and the absolute difference between r_1 and r_2, and r_2 and r_3 was less relative to the other scenarios. Modified Box-Complex was used to develop an optimal heifer selection rule for each scenario using the rule deriving sample of 58 herds. Complex was used to generate twenty solutions for each scenario. The optimal rule for each scenario, was the one of the twenty runs with the highest value of the objective function.

Using a simulation program and the rule testing sample of 57 herds, the Complex selection rule was compared with random selection and another selection rule under each scenario. The simulation program simply calculated the average profit per heifer for each herd given S, r_t, $\mathbf{w}_t$, $\mathbf{p}$ and the animal information. Members of the dairy industry felt random selection was naïve, and suggested a rule based on static genetic Selection Index theory which consisted of zero weights on all the genetic information except the parents' PTA milk which were equally weighted. This was the same rule used in many of the culling simulation studies. The signal based on PTA milk was less informative than the composite signal consisting of PTA milk, fat, protein, and associated reliabilities. For both the random and theoretical genetic rules, all selection was assumed to occur in the period during which the net revenue from heifer sales was the greatest, which was period four for scenarios 1 and 2, and period one for scenario 3. It was assumed that heifers were sold in the same periods that the decision that they would not be selected to remain in the heifer herd was made. In comparing Complex's rules with the theoretical genetic selection rule paired *t* tests were again employed, however, a zero difference was tested using a two-tailed distribution.

Results

Unless otherwise stated, all results presented are evaluating the performance of the rules on the 57 herd rule testing data which was not used in deriving Complex's optimal rules. The artificial activity of purchasing heifers did not enter the solution of any of the rules generated by Complex indicating the constraint on the number selected was met.

The rules with the highest objective value for each scenario developed by Complex using the rule deriving sample, along with the rule from selection index theory, are listed in Table 9.3.

In conducting the twenty runs of each scenario occasionally rules, but never the rule with the highest objective value, would generate the same value of the objective function but with different **p** and $\mathbf{w}_t$. Instances were found where a given herd's profit varied between the two rules, suggesting selection rules that make different selection decisions for a considerable portion of the herds can result in the same value of (4).

More interestingly, in scenario 1 a case also existed where the profit for each herd was identical between the rules suggesting more than a single vector of search variables existed which resulted in the same animal selection decisions, $\mathbf{w}_4$ for these rules are reported in Table 9.4 (each rule conducted all selection in the fourth period). A rule created by the average of the weights resulted in the same profit as the original rules for each herd but one which experienced a relative loss of $58. This suggested the search space between these two local optima was a shallow valley. The great similarity of the values of the elements of $\mathbf{w}_4$ which correspond to each parents' milk PTA, and to a lesser extent sire protein reliability and dam milk-fat reliability suggest these weights are driving the similar results.

Averaging the weights from the twenty scenario 1 runs, which varied in objective values from $9,926 to $9,653, and applying the rule to the rule deriving sample resulted in a value of an objective function of $9,191 which was above random selection's objective value of $8,008. This suggests the valley between the local optima could be quite deep, extending almost half way down to the floor of random selection. This low value of the average also suggests why Complex encountered difficulty optimizing this objective function, for this average, denoted as the centroid in Complex, is used extensively as a reference point for movement of the vertices.

Optimal selection of each herd would have resulted in an objective function of $15,519, however focusing on the final period data for even a single herd it was sometimes unclear how to derive a rule to optimally rank the animals. So this figure only represents the absolute maximum any optimal rule could ever achieve, and it is highly unlikely that any cross-section rule would approach this level of performance. Using the genetic information, Complex accounted for approximately 25% of the range in profit between perfect selection and random selection. This performance is consistent with the low positive rank correlations between predictions of first lactation production and actual production (presented in Chapter 7), and therefore lifetime profit corrected for the opportunity cost of postponed replacement. For a given herd in the rule deriving sample, Complex accounted for between -63% and 100% of the range in profit between perfect selection and random selection.

For scenario 1, the optimal Complex rule selected all heifers in the final period. Figure 9.1 presents a comparison of the performances of Complex's optimal rule derived under scenario 1 conditions and the theoretical genetic selection rule with random selection. Specifically, the difference in average heifer profit for each herd for each rule less the profit from random selection is presented.

Random selection was not dominated in the first nor second degree by either of the rules which utilized the information. While for the majority of herds random selection was inferior to at least one of the other two selection rules, random selection provided the highest profit of all three rules for ten of the 57 herds. Including ties, Complex's rule had the highest objective value for 39 herds under scenario 1, while for 33 herds the highest average profit per heifer was associated with the theoretical genetic selection rule.

Table 9.5 presents the means and sample variances of the increase in herd average profit per heifer for each of the two rules utilizing genetic information over random selection for the three scenarios. Using a one-tailed *t*-Test the hypothesis that Complex's rule was not \$20 more profitable than random selection was rejected at $P = 0.045$ (*t* statistic = 1.73). The mean profit difference of \$27.39 between selection based on genetic theory and random selection was not significantly greater than \$20 ($P = 0.12$, *t* statistic = 1.20).

Figure 9.2 presents the increase in average profit per herd through use of Complex's rule versus the rule based on genetic Selection Index theory. For the majority of herds the difference in profit is minimal, although a few herds would greatly benefit from use of one rule over the other. The less than \$4 difference in mean profit of the rules was not significantly different statistically ($P = 0.28$ two-tailed, *t* statistic = 1.09) nor pragmatically.

For scenario 2, Complex was used to generate a new optimal selection rule with the 58 herds in the rule deriving sample. Again, this rule conducted all animal selection in the fourth period.

First this rule was evaluated against the Complex rule generated under scenario 1 by specifying in the simulation of the scenario 1 rule on the 57 herd rule testing sample that 30% of the heifers would need to be sold in the final period rather than 20%. While a few herds in the rule testing sample would greatly profit over the use of one rule over the other, the average loss per heifer by using the Complex rule derived under scenario 1 when applied to scenario 2 was $2.46 per heifer as opposed to using the optimal rule for scenario 2. (Under scenario 2 with the rule deriving data sample the optimal rule derived under scenario 2 was $4.76 more profitable than the optimal rule derived under scenario 1.) This small difference was not statistically different than zero (P = 0.45 two-tailed, t statistic = 0.76) and was not felt to be sufficient to warrant the industry employing another rule for the different proportion to be selected.

The Complex optimal rule derived under scenario 1 was then tested against the theoretical genetic and random selection rules under the conditions of scenario 2. The comparison of the performances of Complex's optimal rule and the theoretical genetic selection rule with random selection under scenario 1 was substantially similar to that under scenario 1 presented in Figure 9.1. The distribution was slightly wider and flatter than that in Figure 9.1, with random selection being the most profitable method of heifer selection for 15 herds, 5 of which random selection was the most profitable method of selection under scenarios one and two. Neither Complex's rule nor the theoretical genetic selection rule exhibited first or second degree stochastic dominance over random selection. Including ties, Complex's rule and the selection rule based on genetic theory represented the maximum profit for 32 and 33 herds, respectively.

The $32.07 increase in mean profit associated with Complex's rule was not statistically $20 greater than random selection (P = 0.06, t statistic = 1.54). However, the mean herd \$35.21 increase in average profit per heifer as a result of selection based on genetic theory was statistically greater than \$20 ($P = 0.03$, t statistic = 1.96).

Again, while slightly flatter and wider, the histogram comparing the average profit per herd through use of Complex's rule versus the rule based on genetic theory was very similar to that of Figure 9.2. Similar to the results under scenario 1, the \$3.14 difference in profits associated with the two rules which make use of the information were not statistically ($P = 0.45$ two-tailed, t statistic = -0.76) nor practically different.

However, this result that the profit resulting from selection based on genetic theory was greater than the profit of even the rule derived by Complex under scenario 2 to maximize the objective function was unexpected. This result suggested the existence of a rule within Complex's search space superior to the rules Complex found. Applying the static theoretical genetic selection rule to the 58 herds which Complex used to derive its selection rules resulted in a herd average profit per heifer of \$166 versus \$177 for Complex's best rule under scenario 2. Applying Complex's optimal rule from scenario 1 to the conditions of scenario 2 and the rule deriving data sample resulted in simulated profits of \$172. As such, the suggestion that Complex did not find a global maximum can not be supported. Rather this unexpected result is due to sampling error among the choice variables, such that the apparently optimal w_4 and p from the rule deriving sample were not optimal for the rule testing sample.

The presence of sampling error then indicates the existence of sampling variation between the rule deriving and rule testing samples. However, it is unclear whether sampling variation is necessarily the sole or primary source of the sampling error. The creation of the two samples through random assignment of herds following stratification on herd size, and the reasonably large size of the data samples would decrease the probability of large sampling variation. It is possible that the level of sampling variation was not great, but the variation was magnified by over-fitting Complex to the rule deriving data resulting in the different optimal choice vectors for the two samples.

However it is not clear that over-fitting of the model would be expected to account for these results either. Over-fitting could be due to exploiting individual herd effects in the rule deriving sample or by over-parameterization across the entire rule deriving sample. With 57 or 58 randomly assigned herds in each sample and equal weighting of each herd in the objective function, any individual, or group of idiosyncratic herd effects would be expected to be countered by other herd effects in the sample and as a result it is not clear that herd effects could have been exploited in rule development. In terms of model parameterization, of the 38 search variables, only thirteen parameters, the ten of $\mathbf{w}_4$ and the three of $\mathbf{p}$, played any role in the scenario 1 and 2 results. Furthermore, within $\mathbf{w}_4$ the ten variables are not independent. While sire PTAs were expected to be uncorrelated with dam PTAs, within each parent the PTA for sire, fat and protein were highly correlated, and the reliabilities were also highly correlated, and as discussed previously, the reliabilities also had limited variability. Given the modest number of parameters, the high degree of correlation between many of them, and data sample sizes of nearly 1,000, it was not clear to what degree the model was over-parameterized. So it was unclear to what degree over-fitting of the optimal rule to either individual herd effects or the entire rule deriving data sample was possible.

By way of analogy, assume that interest lies in forecasting using either a simple rule or complicated rule, each of which are optimized through linear regression. Further assume that the simple rule is the globally optimal rule, and is within the parameter space of the complicated rule. It is very possible that the more complicated rule derived using the rule deriving sample data could exploit the idiosyncrasies contained in that data sample to produce a model with a higher R^2. Statistically, the more complicated rule cannot result in a lower R^2 (although the adjusted R^2 may in fact be lower).

However, when applied to the out of sample rule testing data the forecast ability of the more complicated rule will be worse as the idiosyncrasies of the rule deriving data sample will not be present in the out of sample data. In general, increases in the number of parameters in the complicated rule's parameter space increases the probability of over fitting the rule to the in sample data. Increasing the size of the samples would lower the probability of getting samples which differ in important ways, and would also minimize the effect of any one observation or cluster of observations on the deriving of the optimal weights for the complicated rule.

Under scenario 3 with the increasingly negative returns to heifer rearing, only one of the twenty rules generated by Complex completed all the necessary heifer selling in the first period, and this rule had the third lowest value of the objective function. The remaining rules generally selected 83 to 93% of the heifers from the rule deriving sample herds in the first period, the remaining majority of the selection occurred in the second period in 11 of the rules and in the third period in three rules. Commonly, a few herds sold animals in the fourth period. As the latter tended to occur primarily in the larger herds, the effect on the objective function would be small as the loss in profit was averaged over the entire herd before entering the objective function. The optimal rule selected 0.9, 0.91, 1.0 proportion of the heifers in periods one to three respectively, with a few herds selling heifers in the fourth period. Of the three scenarios, this one had the greatest range in objective values at 5.1%.

The variability of the periods in which selection occurred under the various r_t may reveal the factors that impacted the simple appearing trade-offs made in heifer selection. Within a herd, the range in π_a was generally between $1,000 to $2,500, so the profit from making a better choice in terms of which animals were selected could potentially more than compensate for conducting some selection in periods which were nonoptimal in terms of maximizing the herd profit from r_t. Under scenarios 1 and 2, with the improved information in the fourth period and the $44 cost to total herd profit for each animal selected in the third period all selection was conducted in the fourth period. However, under scenario 3, Complex was routinely willing to pay $20 and occasionally $50 to attain the improved genetic information and/or simply spread selection out over multiple periods to improve those selected. It is not clear which effect dominated.

Similar to the approach taken with scenario 2, the Complex rule from scenario 1 was adapted to operate in scenario 3 by forcing all selection to occur in the first period based on $\mathbf{w}_4$. The resulting rule, applied to the rule testing data, was $9.72 more profitable per heifer for the average herd than using the Complex rule developed under scenario 3 conditions. In contrast, simulating the profit from use of this adapted rule on the rule deriving sample data resulted in a loss of $11.80 as compared to the scenario 3 developed Complex rule. This again suggested the increased profit was the result of sampling error.

The modified rule from scenario 1 was compared against selection based on genetic theory and random selection. The distributions for scenario 3 corresponding to Figure 9.1 were again substantially similar. The selection rules based on the information were not first or second order stochastic dominant to random selection. The majority of herds profited from at least one selection rule based on the information versus random selection. Interestingly, of the 15 herds for whom random selection was the most profitable selection method under this scenario, for 5 of these herds random selection was the superior rule regardless of scenario. Under scenario 3, the rule generated by Complex accounted for the highest profit per heifer in 35 herds and while selection based on genetic theory resulted in the highest profit for 34 herds.

Complex's rule's $32.04 was statistically greater than $20 more profitable per heifer than random selection (P = 0.04, t statistic = 1.76), while the $25.09 difference between the theoretical genetic selection rule and random selection was not (P = 0.21, t statistic = 0.82).

The $6.95 difference in mean average profit per heifer between the two selection rules which used the information was not statistically different than zero (P = 0.12 two-tailed, t statistic = 1.58), and similar to previous scenarios a few herds would profit greatly from use of one rule over the other.

Summary and Conclusions

Modified Box-Complex had some difficulty in optimizing the heifer selection rule, which is not surprising given the 38 potentially relevant search variables and the discontinuous nature of the objective function. Even more troubling was the sampling error which was sufficiently large that the optimal decision rule derived under one sample was clearly non-optimal when applied to the other sample. It was not clear whether the sampling error was due solely to sampling variation, or whether the variation was magnified through over-fitting of the model to the rule deriving data sample. The sampling error not only makes interpretation of the profits resulting from rule use difficult, but also does not bode well for herd individualized rules. For even under the USDA's current quarterly schedule of genetic data estimation, a decade of prospective data collection could easily be necessary to amass sufficient information for developing and testing individual herd rules.

The suggestion that for some herds random heifer selection would consistently be more profitable than selection based on the genetic information warrants reconsideration. While a priori it was plausible that genetics may have been of little value in heifer selection, it is difficult to understand how, for a given herd, genetics could have negative value resulting in random selection of heifers consistently providing higher profits than selection based on genetics. It is possible that the single cohorts of heifers which represented these five herds experienced a random event which led to poor performance of the genetically superior heifers. In this situation, it is not surprising that selection based on genetics would consistently have resulted in lower profits for this cohort of heifers regardless of scenario. This suggests some of the differences between herds in terms of profit per heifer was likely due to systematic herd effects and some was due to random noise.

While the relative profit distributions evaluated were for a cross-section of herds, with each herd consisting of a single cohort of heifers, the profit distribution of a given herd for a number of cohorts would be of primary interest to dairy producers. The random noise experienced across the herds would also be present across cohorts of a given herd. However, the systematic herd effects can be decomposed into individual cohort effects and a herd management effect constant across cohorts. Then for a given herd while the herd management effect would affect the mean of the distribution of profits from repeated rule use on numerous cohorts it would not contribute to the width of the profit distribution resulting in a narrower distribution than displayed in Figures 9.1 and 9.2.

As mentioned in the introduction, the selection of replacements which are more profitable than the current defenders dictates the replacement rate should rise concomitantly. However, it is difficult to determine to what degree the optimal culling rate in the cow herd and therefore the selection rate in the heifer herd would change. Following the initial period of heifer selection the rate of heifer selection would enter a steady state. It is unclear what this steady state rate would be because it depends on the relative fecundity of the cows, which are no longer being culled, compared to the heifers which are now being sold. These are issues best addressed through linking the output of a heifer selection model into a stochastic dynamic programming model of the cow herd.

With the similar performance of the two rules, the rule suggested by genetic theory of selecting on EBV of milk is preferred over Complex's rule which is more complex and requires information not currently reported to producers. It can be concluded that while the composite signal based on parents' PTAs of milk, fat and protein and associated reliabilities was more informative than the signal based solely on parents' PTA of milk, the values of the two signals were pragmatically equivalent. This similar performance of the two rules which utilized the information suggests the value of the genetics information presides primarily in the parents milk PTA, while the other PTAs and reliabilities have relatively little value in prepartum Holstein heifer selection.

While it is likely that the solids PTAs are not valuable in heifer selection due to the high correlation with PTA milk, the limited value may also have sourced from limited variability in the data, this is likely more relevant for the reliabilities. Given that Complex's performance did not suffer when the decimal places for the weights was limited from three to two it is unlikely that the apparent unimportance of the genetic information other than PTA milk was due to scale problems whereby the one to two larger order of magnitude of PTA milk relative to the other genetic information simply overwhelmed the signal.

Complex's optimal selection rule from scenario 1 appeared robust to adaptation to the various other scenarios. For any of the three scenarios there appeared to be little systematic difference between the distribution of profits resulting from the rules derived by Complex and that suggested by genetic Selection Index theory. The robustness of these results suggests genetic information is valuable in heifer selection, and the *ex ante* value is estimated at greater than $20 per heifer realized over the course of the heifers' lives. The herd profit associated with a $20 per heifer increase in profit would make non-random heifer selection a potentially worthwhile endeavor for Michigan producers.

While profitable, selection based on EBV of milk is likely capturing only a portion of the profit potentially available through heifer selection. Recalling that lifetime profit corrected for the opportunity cost of postponed replacement was a deterministic function of first lactation milk production, this low capture ratio is then consistent with the results presented in Chapter 7 revealing a diffuse distribution of within herd rank correlations between parent PTA based predictions of first lactation milk production and actual production. To capture this elusive profit the dairy industry needs to continue to explore potential information sources other than EBV on which to base methods of heifer selection.

Acknowledgments

The authors thank Al Thelen, Terry Speer and Lyle Smith at Michigan DHI, as well as, Leigh Walton and H. D. Norman at the Animal Improvement Laboratory, USDA for assistance with data collection. Jason Shelagowski's assistance in FORTRAN programming is acknowledged. The financial support of the primary author by the Department of Agricultural Economics through the Animal Industry Initiative funds provided by the Agricultural Experiment Station at Michigan State University while this work was conducted is gratefully acknowledged.

References

Allaire, F. R. "Economic consequences of replacing cows with genetically improved heifers." *J. Dairy Sci.* 64(1981):1985-1995.

Allaire, F. R., and E. P. Cunningham. "Culling on low milk yield and its economic consequences for the dairy herd." *Livest. Prod. Sci.* 7:(1980):349-359.

Antonovitz, F., and T. Roe. "Chapter 9: Economic Efficiency and Market Information" in *Economic Efficiency in Agricultural and Food Marketing*, ed. R. L. Kilmer and W. J. Armbruster, Ames:Iowa State University Press, 1987.

Box, M. J. "A new method of constrained optimization and a comparison with other methods." *Comput. J.* 8(1965):42-52.

Cassell, B. G., B. B. Smith, and R. E. Pearson. "Influence of herd-life opportunity and characteristics of cows and herds on different net income functions." *J. Dairy Sci.* 76(1993):1182-1190.

Congleton, W. R. "Dairy cow culling decisions. 2. Profitability and genetic trend in herds with culling on production versus income." *J. Dairy Sci.* 71(1988):1905-1915.

Congleton, W. R., and L. W. King. "Culling cows on predicted future profitability with a variable planning horizon." *J. Dairy Sci.* 68(1985):2970-2983.

Connor, L. J., L. G. Hamm, S. B. Nott, D. Darling, W. Bickert, R. Mellenberger, H. A. Tucker, O. B. Hesterman, J. A. Partridge, and J. H. Kirk. "Summary of the 1987 Michigan State University dairy farm survey." *Research Report 498*, Michigan Ag. Exper. Station, East Lansing, 1989.

Harris, B. L., and A. E. Freeman. "Economic weights for milk yield traits and herd life under various economic conditions and production quotas." *J. Dairy Sci.* 76(1993):868-879.

Harris, T. R. "Appendix A: The Modified Box-Complex." *Analysis of Irrigation Scheduling for Grain Sorghum in the Oklahoma Panhandle.* Oklahoma State University Ph.D. Dissertation, 1981.

Heinrichs, A. J., S. J. Wells, H. S. Hurd, G. W. Hill, and D. A. Dargatz. "The national dairy herd evaluation project: A profile of heifer management practices in the United States." *J. Dairy Sci.* 77(1994):1548.

Henderson, C. R. "Selection index and expected genetic advance." Page 141 in *Statistical Genetics and Plant Breeding ; a symposium and workshop.* W. D. Hanson and H. F. Robinson, ed. Natl. Acad. Sci.-NRC, Washington, DC, 1963.

Huirne, R., A. Dijkhuizen, R. King, and S. Harsh. "Goal, critical success factors, and information systems needs on swine and dairy farms – Identified by workshops." *Zeitschrift für agrainformatik* 3(1993):61-65.

Karszes, J. "Dairy replacement programs: costs and analysis Western New York, 1993." *1994 Heifer Management Symposium* Mimeograph Series Cornell Cooperative Extension,

Killen, L., and B. Kearney. "Optimal dairy cow replacement policy." *Ir. J. Agric. Econ.* Rur. Sociol. 7(1978):33-40.

Korver, S., and J. A. Renkema. "Economic evaluation of replacement rates in dairy herds II. Selection of cows during the first lactation." *Livest. Prod. Sci.* 6(1979):29-37.

Kuester, J.L., and J.H. Mize. *Optimization Techniques with Fortan.* pp 368-385. New York: McGRaw-Hill Book Co., 1973.

Kuipers, A. *Development and Comparison of Selection Criteria for Cows and Bulls with a Dairy Herd Simulation Model.* Univ. Wisconsin, Madison Ph.D. Thesis, 1980.

McCullough, D. A., and M. DeLorenzo. "Effects of price and management level on optimal replacement and insemination decisions." *J. Dairy Sci.* 79(1996):242-253.

Michigan Dairy Herd Improvement Association. *Michigan Dairy Production Herd Evaluation Report*, 1996.

Nelder, J.A., and R. Mead. "A simplex method for function minimization." *Comput. J.* 7(1964):308-313.

Nermuth, M. "Information Structures in Economics" in *Lecture Notes in Economics and Mathematical Systems*. Monograph 196, ed. M. Beckman and H. P. Kunzi, pp 12-44. Berlin:Springer-Verlag, 1982.

Pearson, R. E., and A. E. Freeman. "Effect of female culling and age distribution of the dairy herd on profitability." *J. Dairy Sci.* 56(1973):1459-1471.

Quinton, V. M., and C. Smith. "An empirical check on best linear unbiased prediction genetic evaluation using pig field recording data." *Can. J. Anim. Sci.* 77(1997):211-216.

Richards, T. J., and S. R. Jeffrey. "Establishing indices of genetic merit using hedonic pricing: AN application to dairy bulls in Alberta." *Can. J. Ag. Econ.* 44(1996) 251-264.

Robison, L. J., and P. J. Barry. "Chapter 21, Replacement Principles" in *Present Value Models and Investment Analysis*. 1st ed. The Academic Page, Northport, Alabama, 1996.

Rogers, G. W., J.A.M. van Arendonk, and B. T. McDaniel. "Influence of production and prices on optimum culling rates and annualized net revenue." *J. Dairy Sci.* 71(1988):3453-3462.

Schroeder, T. C., J. A. Espinosa, and B. K. Goodwin. "The value of genetic traits in purebred dairy bull services." *Rev. Ag. Econ.* 14(1992):215-226.

Van Arendonk, J. A. M. "Use of profit equations to determine relative economic value of dairy herd life and production from field data." *J. Dairy Sci.* 74(1991):1101-1107.

Van Arendonk, J. A. M. and A. A. Dijkhuizen. "Studies on the replacement policies in dairy cattle. III. Influence of variation in reproduction and production." *Livest. Prod. Sci.* 13(1985):333-349.

Weigel, D. J., B. G. Cassel, I. Hoeschele, and R. E. Pearson. "Multiple trait prediction of transmitting abilities for herd life and estimation of economic weights using relative net income adjusted for opportunity cost." *J. Dairy Sci.* 78(1995):639-647.

Table 9.1. Genetic information for the first and fourth periods, along with first lactation milk production and lifetime profit corrected for the opportunity cost of postponed replacement.

Variable[1]	$\bar{x}$	SD	Minima	Maxima
$\mathbf{g}_{a1}$				
PTA Milk of Sire, lb	1,894.2	470.1	-1086	3093
PTA Fat of Sire, lb	59.3	19.6	-28	110
PTA Protein of Sire, lb	51.0	14.2	-19	94
REL of Sire Milk-Fat, %	88.8	8.6	24	99
REL of Sire Protein, %	88.2	8.9	24	99
PTA Milk of Dam, lb	874.4	574.7	-1269	2575
PTA Fat of Dam, lb	25.6	21.5	-44	109
PTA Protein of Dam, lb	23.8	16.5	-39	82
REL of Dam Milk-Fat, %	47.9	6.2	12	74
REL of Dam Protein, %	47.7	6.2	12	74
$\mathbf{g}_{a4}$				
PTA Milk of Sire, lb	2,084.1	515.0	-1073	3347
PTA Fat of Sire, lb	64.0	21.0	-28	116
PTA Protein of Sire, lb	55.1	16.8	-19	94
REL of Sire Milk-Fat, %	95.3	7.1	28	99
REL of Sire Protein, %	95.1	7.3	28	99
PTA Milk of Dam, lb	928.2	647.6	-1240	2912
PTA Fat of Dam, lb	27.1	24.6	-57	117
PTA Protein of Dam, lb	25.4	18.8	-35	103
REL of Dam Milk-Fat, %	51.4	5.23	26	85
REL of Dam Protein, %	51.3	5.2	24	85
L_a, lb	25,995.6	5,293.1	6,746.0	41,612.0
π_a, $	142.0	602.9	-2,050.6	1,920.7

[1] Reliability (REL).

Table 9.2. Scenarios used to develop and test heifer selection rules.

Parameter[1]	Scenario 1	Scenario 2	Scenario 3
S	0.8	0.7	0.8
r_1	\$41	\$41	-\$10
r_2	\$75	\$75	-\$30
r_3	\$114	\$114	-\$60
r_4	\$158	\$158	-\$140

[1] S - Overall proportion of heifers to be selected in each herd.
r_1 - Net revenue from sale of heifer in period 1.
r_2 - Net revenue from sale of heifer in period 2.
r_3 - Net revenue from sale of heifer in period 3.
r_4 - Net revenue from sale of heifer in period 4.

Table 9.3. Data utilizing heifer selection rules.

Search Variable	Variable Number	Optimal Rules Generated by Complex			Rule based on Selection Index
		Scenario 1	Scenario 2	Scenario 3	
p_1	X(1) =	1.10	1.24	0.90	1.00[3]
p_2	X(2) =	1.20	1.22	0.91	1.00[3]
p_3[2]	X(3) =	1.05	1.18	1.14	1.00[3]
na[1]	X(4) =	-0.40	-1.16	0.16	0.00
na	X(5) =	0.13	0.97	0.19	0.00
na	X(6) =	-0.55	0.23	-0.16	0.00
na	X(7) =	-0.19	1.07	-0.28	0.00
na	X(8) =	-0.17	0.16	0.04	0.00
na	X(9) =	-10000.00	-10000.00	-10000.00	0.00
w_1					
sire milk	X(10) =	0.28	0.64	0.01	0.50
sire fat	X(11) =	0.58	-0.63	0.78	0.00
sire protein	X(12) =	0.42	0.68	-0.19	0.00
sire reliability milk-fat	X(13) =	0.16	-0.25	-0.21	0.00
sire reliability protein	X(14) =	0.08	-0.29	0.59	0.00
dam milk	X(15) =	-0.02	-0.61	0.21	0.50
dam fat	X(16) =	0.02	-0.42	0.62	0.00
dam protein	X(17) =	0.52	0.73	-0.69	0.00
dam reliability milk-fat	X(18) =	0.18	0.81	-0.23	0.00
dam reliability protein	X(19) =	-0.78	0.03	-0.43	0.00
w_2					
sire milk	X(20) =	-0.48	-0.18	0.36	0.50
sire fat	X(21) =	0.15	-0.71	-0.12	0.00
sire protein	X(22) =	0.47	-0.95	-0.28	0.00
sire reliability milk-fat	X(23) =	0.24	0.88	-0.62	0.00
sire reliability protein	X(24) =	0.29	0.76	-0.13	0.00
na	X(25) =	-0.50	0.47	-0.70	0.00
na	X(26) =	0.47	0.09	-0.18	0.00
na	X(27) =	-0.56	0.26	-0.05	0.00
na	X(28) =	-0.78	-0.73	0.18	0.00
na	X(29) =	0.52	0.76	-0.07	0.00
w_3					
sire milk	X(30) =	0.45	0.78	0.17	0.50
sire fat	X(31) =	0.07	0.69	0.34	0.00
sire protein	X(32) =	0.16	-0.84	-0.43	0.00
sire reliability milk-fat	X(33) =	0.33	0.45	0.68	0.00
sire reliability protein	X(34) =	-0.16	-0.40	0.58	0.00
dam milk	X(35) =	0.00	-0.48	0.12	0.50
dam fat	X(36) =	-0.64	0.89	-0.01	0.00
dam protein	X(37) =	-0.03	-0.92	-0.09	0.00
dam reliability milk-fat	X(38) =	-0.09	-0.93	0.08	0.00
dam reliability protein	X(39) =	0.74	0.09	0.03	0.00
w_4					
sire milk	X(40) =	0.07	0.30	0.46	0.50
sire fat	X(41) =	0.70	0.99	-0.16	0.00
sire protein	X(42) =	0.66	0.69	-0.23	0.00
sire reliability milk-fat	X(43) =	-0.19	0.86	0.18	0.00
sire reliability protein	X(44) =	-0.31	-0.96	-0.05	0.00
dam milk	X(45) =	0.11	0.33	0.29	0.50
dam fat	X(46) =	0.11	0.70	-0.17	0.00
dam protein	X(47) =	-0.11	-0.83	-0.14	0.00
dam reliability milk-fat	X(48) =	0.20	0.76	-0.15	0.00
dam reliability protein	X(49) =	0.26	-0.42	-0.27	0.00

[1] Not applicable.

[2] Of the heifers still present in the fourth period, the proportion selected is determined by the desired culling rate in the cow herd.

[3] For the rule based on Selection Index, all selection occurs in the period which generates the greatest return to heifer rearing. For the rule displayed this occurs in the fourth period and would be consistent with the conditions of Scenarios 1 or 2.

Table 9.4. Fourth period weights, and their average, for two scenario 1 rules with identical selection decisions for the 58 rule deriving herds.

w_4	Variable Number	Rule A	Rule B	Average
sire milk	X(40) =	0.69	0.70	0.70
sire fat	X(41) =	0.50	0.43	0.47
sire protein	X(42) =	0.52	0.43	0.48
sire reliability milk-fat	X(43) =	0.01	-0.06	-0.03
sire reliability protein	X(44) =	-0.24	-0.26	-0.25
dam milk	X(45) =	0.29	0.30	0.30
dam fat	X(46) =	-0.05	-0.03	-0.04
dam protein	X(47) =	0.41	0.23	0.32
dam reliability milk-fat	X(48) =	0.13	0.11	0.12
dam reliability protein	X(49) =	-0.14	0.25	0.06

Table 9.5. Means and variances of the increase in herd average heifer profit of the selection rules utilizing genetic infomation over random selection for the three scenarios.

Scenario	Profit from Heifer Sale				% Selected	Mean and Variance of Profit/Heifer from Rule vs Random Selection		
	Period					Rule		
	1	2	3	4		Complex	Genetic Theory	Difference
1	$45	$71	$114	$158	80%	$31*	$27	$4
						2443	2159	
2	$45	$71	$114	$158	70%	$32	$35*	-$3
						3497	3413	
3	-$10	-$30	-$60	-$140	80%	$32*	$25	$7
						2660	2184	

*Significantly different ($P \leq 0.05$) than $20, using one-tailed t-Test.

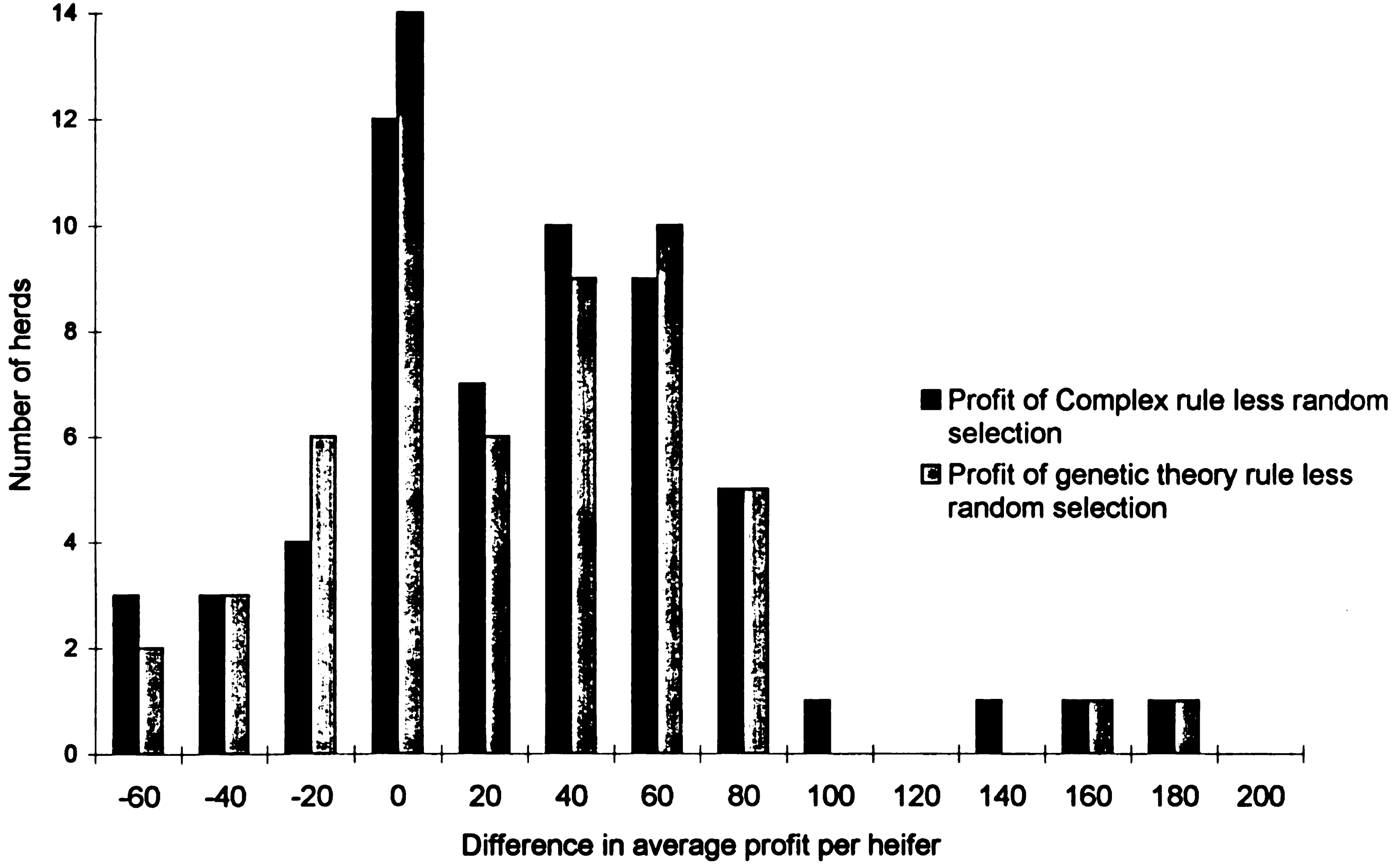

Figure 9.1. Average profit per heifer from Complex and theoretical genetic selection rules relative to random selection under scenario 1 for 57 herds.

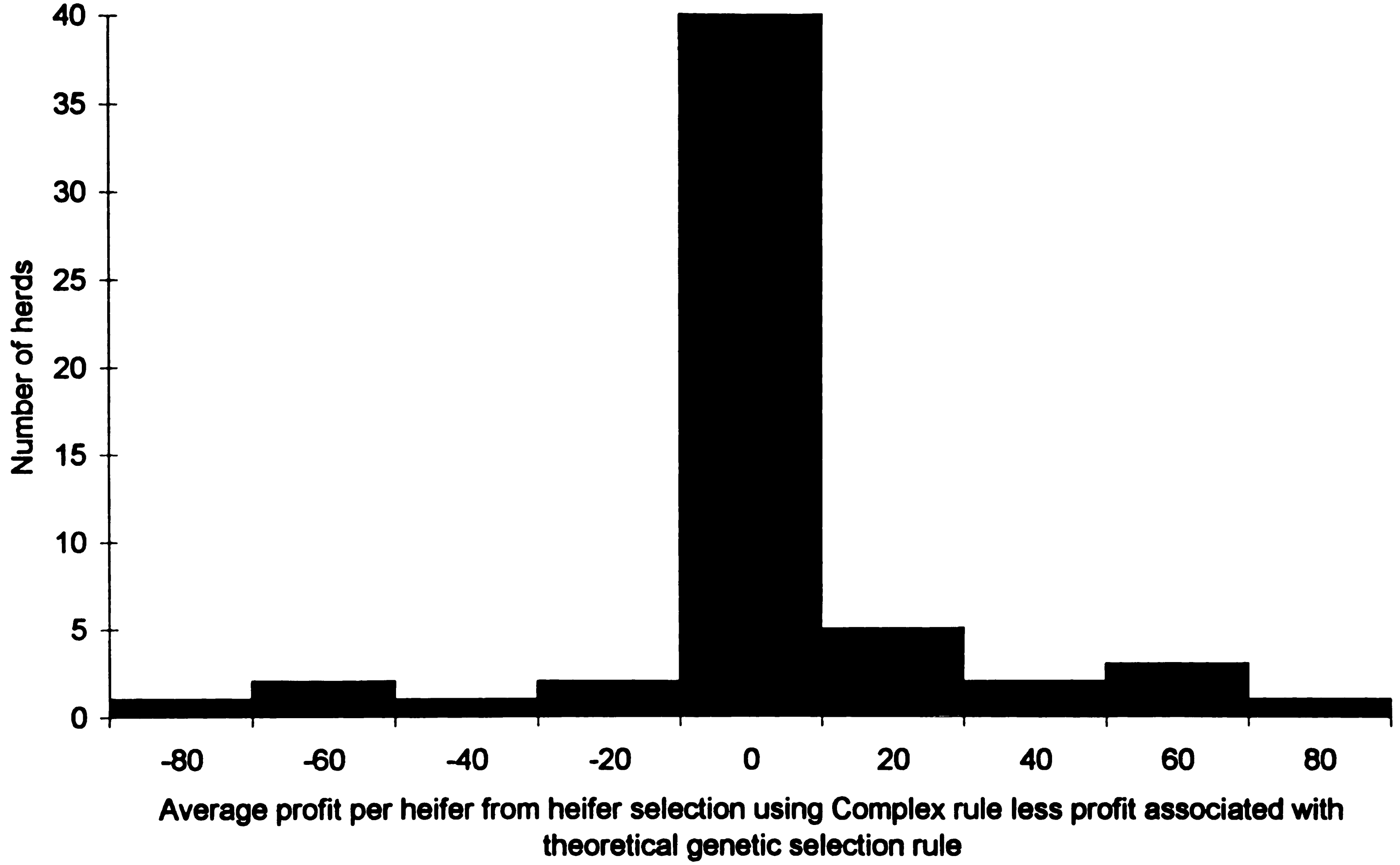

Figure 9.2. Histogram of difference in average profit per heifer between Complex and theoretical genetic selection rules under scenario 1 for 57 herds.

Chapter 10

CONCLUSIONS

In terms of dairy production, the literature revealed that assuming risk neutrality, a constant herd size, culls are sold for beef not dairy purposes, and no strong seasonal effects on milk production or reproduction, the optimal culling rate is likely between 20 and 30%. Prior to addressing the research conducted for this dissertation, it is worthwhile considering why Michigan apparently has a higher than optimal culling rate, for until producers recognize they may have an excessive culling rate no consideration will be given to heifer selection. While a violation of any of the four assumptions could support the Michigan DHI average culling rate of 37%, these four assumptions are not as restrictive as perhaps they initially appear.

While the objective functions of the stochastic dynamic programs and simulations implicitly assumed risk neutrality, the decision to lower the culling rate by selling prepartum heifers also has advantageous risk characteristics-- early positive cash flow and highly reversible in the short and long term. Turning to the second assumption, while for herds which are contracting in size, a higher culling rate could be accepted as the herd is liquidated, the majority of Michigan dairy herds which are not maintaining a constant herd size are likely expanding. The high average culling rate is more difficult to explain in these herds where, due to expansion, the facility is under capacity. As a result, the zero opportunity cost of postponed replacement should result in lower culling rates which would be expected to rise as livestock capacity is reached.

In terms of the third assumption, while some producers certainly can sell their culled first lactation heifers as dairy instead of beef, the logic of this is unclear. Assume that, despite having been culled (i.e., is inferior to the majority of non-culled animals in the seller's herd), the culled animal has higher milk production than the average animal in a potential buyer's herd.

It is possible that the better performance of the culled animal relative to the buyer's herd is due to genetics. With the heritability of milk being between 0.25 to 0.30, genetics account for less than one-third of the variation in milk production. However, from Chapter 7 it was evident that only 7% of the within herd variation in first lactation production was explained by estimates of parents' genetics. This poor performance of the parents' predicted transmitting ability (PTA) is due to the fact that the PTA are only estimates of the parents', and therefore their daughters', additive genetic values of milk production.

Once she has produced milk, and thereby contributes to her own PTA, the accuracy of the culled animal's estimated genetic ability to milk will likely improve to explaining over 10% of the variation in milk production. So while it is possible that the cull may be genetically superior, given the large estimation error in the cull's estimated genetic value it is difficult to tell how much of her performance is due to genetics and how much is due to environment.

The amount of variation explained by genetic estimates is small compared to the 40% of variation in first lactation milk production that herd, conditional on parent PTA, explained in the Chapter 7 analysis. So unless a large part of this herd effect is permanent (i.e., is a permanent environmental effect) it is highly likely that the culled animal would rank similarly in the buyer's herd as in the seller's herd, namely at the bottom.

Furthermore, initially it would seem that having been identified as a cull, the animal's value would be less than one not so identified. However, given the altered probability distribution of the animal's potential returns following calving, as no uncertainty exists as to whether the animal will survive calving and come to milk, it is possible the animal may be worth the same or more.

Turning to the last assumption of no strong seasonal effects, it is accepted that compared to areas such as Florida and Texas, the level of heat stress encountered by Michigan cattle can be effectively managed so it does not have prolonged effects on reproduction or milk production.

With this literature on optimal culling rates being applicable to much of the North American dairy industry, it is certainly disconcerting that this wealth of information, concentrated in the Journal of Dairy Science and Livestock Production Science, could have been overlooked by the academics of the dairy industry. My only explanation for this is agricultural economists tend not to subscribe to such applied journals, while the biologically based veterinarians and animal scientists are not versatile in the techniques used in these studies. The other potential impediment to embracement and dissemination of these studies is that the results may not have coincided with some of the scientists' assumptions and beliefs that maximizing the inflow of heifers, which represent the best genetics and thus milk production potential of the herd, would maximize profit.

Of course, as discussed in Chapter 3, the other reason the industry may not have recognized the nonoptimality of its culling rates is simply failure to recognize that the culling rates in the models and those reported by many dairy herd improvement associations are defined and calculated differently.

The suggestion, based on this information of optimal culling rates, for producers and their advisors to consider lowering their culling rates is generally not embraced, at least not initially. My explanation for this resistance is that culling and culling rates are not largely an active decision. Rather rules of thumb have been developed to guide under what conditions animals should be culled. These culls are then usually designated as involuntary and account for roughly a 20% culling rate. According to another rule of thumb the producer then culls an extra 10 to 15% for voluntary purposes. Some of the culling studies (see Chapter 3, especially Table 3.1) suggest these rules of thumb for involuntary culling are wrong, resulting in the culling of animals that should not have been. More fundamentally, as discussed, the category of involuntary culls is so small as to be negligible. Rather producers must focus on culling the least profitable animals and not worry about classifying the culled animals into categories of voluntary and involuntary culls. Results from Chapter 8 support the finding that there is little difference in the voluntary nature of culls classified as voluntary or involuntary. As long as producers and their advisors are convinced that they have no choice in the culling decisions of the majority of their cows they will be ill prepared to critically evaluate their culling decisions. Certainly, it is worthwhile for producers to monitor why cows are being culled but for the purposes of monitoring management and not to aid in culling decision making.

A counter argument to the above explanation is that producers and their advisors are making active decisions concerning culling, and based on the poor milk production of the current assets and/or the low probability of the current animals' reproductive success the producer is making the optimal decision to maximize the inflow of heifers. While this argument is logical in the short-term, the fact that this actively assessed culling rate always exactly equals the number of available heifers leaves this argument ringing hollow. Furthermore, if heifers were sold or purchased this was likely the result of an active decision with regard to individual heifers and not reflective of an active decision at the herd level.

It is quite possible, that heifer pressure is driving the high culling rate. It may be that the culling rules of thumb were developed in response to the need to cull sufficient cows to make room for the first lactation animals. A prescription to lower the culling rate then simply starts with determining a target culling rate. Keep a sufficient number of heifers to meet this target rate and then rank the cows in terms of the order that they would be culled. Starting at the bottom of the list with the cows the producer would most prefer to cull, cows are designated to be culled until the proper number to meet the target rate has been reached. The optimization and simulation studies on culling rates suggest 5% of the first calf heifers should be culled for low production.

We then arrive at the issue of how to select those heifers. Certainly random selection is one option. The derivation of a heifer selection method which can select the better heifers would result in more than the 5 to 10% increase in profit for the average Michigan producer through use of a 20 to 30% culling rate and random heifer selection.

First, I define better as higher lifetime profit corrected for the opportunity cost of postponed replacement. Of the production traits apparent early in life, first lactation milk production has the highest correlation with this profit number. Based on genetic Selection Index theory, heifers' parents' PTA should be unbiased predictors of daughters' subsequent first lactation milk production. Therefore, the parents' PTA should also be useful in predicting heifers' lifetime profit corrected for the opportunity cost of postponed replacement.

However, there is little in the genetics literature to suggest how well parents' PTA actually predict daughters' subsequent milk production. Based on a sample of 5,123 Michigan heifers, parents' PTA available just prior to the heifers' calvings explained less than 7% of the within herd variation in first lactation mature equivalent milk production. Despite geneticists consideration of herd as a fixed effect, my Chapter 7 analysis revealed that it can be treated as random resulting in greater efficiency of the estimates.

To be useful in heifer selection it is the rankings of heifers based on predicted first lactation production that is of importance, not how much of the variation is explained. That PTA may be useful in heifer selection was suggested by the results that the within herd rank correlations between predicted and actual first lactation mature equivalent were positive for 65% of the herds, and clustered in the area of 0.3. The omission of heifers which were culled prior to first lactation, did not appear to cause selection bias in the sample data suggesting that these results are not limited only to the sample used in the statistical analyses.

Given some evidence that parents' PTA may be useful in predicting milk production and therefore lifetime profit corrected for the opportunity cost of postponed replacement, I next turned to the evaluation of an optimal heifer selection rule.

First, in Chapter 8 I considered whether parents' PTA from the period in which a cohort of 5,619 heifers were born could predict subsequent heifer cullings during rearing or first lactation. The logic behind such an approach is that producers are currently likely making optimal decisions with regard to which heifers to cull and which to keep in the herd. However, by waiting to freshen the heifers to make this decision they are in essence paying too high a price to make this decision. Rather the culling literature reviewed in Chapter 3 suggests making the heifer selection decisions prior to calving would be more profitable even after taking into account the cost of decreased ability to pick the most profitable heifers. Then if genetics could be used to identify the heifers which subsequently would be culled, an optimal heifer selection rule would largely be accomplished.

Regardless of the model specification, including the distributional assumption for the probability of being culled, PTA of milk, fat, and protein for each heifer's parents did not predict subsequent culling during rearing or first lactation. This despite evidence that first lactation milk production did influence producers' culling decisions. This inability of genetic estimates to predict culling outcomes surprisingly also indicated producers were not using genetics in their culling decisions. The culling reasons were classified into the broad categories of supposed voluntary and involuntary, and then the ability of genetics to predict culling in each of these classifications, as well as, the effect of a change in PTA on the propensity of a heifer being culled were evaluated. Based on these analyses it was concluded that there was no difference in the voluntary nature of the culling decisions in the two classifications, and thus no evidence of involuntary or voluntary culls.

Given the failure of regression techniques to predict which heifers would be culled. I turned to Modified Box-Complex to solve the issue of how to select heifers to maximize profit and assess the expected increase in profit versus random selection. The objective of the heifer selection rule was to maximize herd profit, given the proportion of heifers to be selected. Herd profit then consisted of the returns from those heifers that were sold and the lifetime profit corrected for the opportunity cost of postponed replacement of those that were selected. Due to data limitations, herd profit was summed over a number of herds, resulting in a single rule for all farms. Herd profit was on a per heifer basis so each herd, regardless of herd size, had equal weighting in the objective function.

While Complex was used to generate an optimal rule for each of three scenarios, it was found the performance of the rules varied little when adapted to a particular scenario, suggesting robustness of the rule. As a result, the performance of a single Complex rule was tested against random selection and selection based on estimate breeding value (EBV) of milk under three scenarios. EBV milk is the average of the parents' PTA of milk. Generally selection was assumed to occur at the most profitable time in terms of returns from those sold.

Regardless of scenario, Complex's rule and the EBV of milk selection rule performed very comparably, suggesting parents' PTA fat, protein and the associated reliabilities provide little value in heifer selection. Given the decreased information requirements for heifer selection on EBV of milk versus Complex's rule, EBV of milk is the recommended method of heifer selection.

Generally, for every heifer in the cohort, selection based on EBV of milk was $20 more profitable than random selection. While the performance of the rule would vary by cohort, on average a producer with a herd of 100 cows would benefit by roughly $1,000 per annum. This increase in profit, due solely to heifer selection, would be realized over the course of those selected heifers' lives. Other profit would also accrue if, in addition to the heifer selection, movement was made to a more optimal culling rate among the cow herd.

The derivation and testing of the heifer selection rules on different data sets also provided some insight into the presence and potential causes of sampling error. The 115 herds were randomly assigned, following stratification on herd size, to one of the two data sets resulting in nearly 1,000 animals in each data set. Despite the size of these randomly created data sets, the optimal rule derived under one sample was non-optimal when applied to the other data sample. This presence of sampling error indicated the existence of sampling variation between the rule deriving and rule testing data samples. It was not clear whether the sampling error was due solely to sampling variation, or whether the variation was magnified through over-fitting of the model to the data sample used to derive the selection rule.

The sampling error not only makes interpretation of the profits resulting from rule use difficult, but also does not bode well for herd individualized rules. For even under the USDA's current quarterly schedule of genetic data estimation, a decade of prospective data collection could easily be necessary to amass sufficient data for developing and testing individual herd rules. The possibility of the sampling error being due to sampling variation serves as a caution to anyone basing conclusions on data analysis of single medium size or small data sets, especially in the absence of a rigorous theoretical basis to aid assessment of the results.

If parents' PTA and reliabilities were the only information available to select heifers, Complex revealed that EBV milk represented a nearly optimal method of using this genetic information in heifer selection. While heifer selection based on EBV milk is certainly worthwhile, evidence was presented that this method of selection is capturing only a fraction of the profit potentially available though heifer selection. Given that the lifetime profit corrected for the opportunity cost of postponed replacement was deterministically estimated from first lactation production, this then suggests EBV are not optimally ranking the heifers in terms of milk production which is consistent with the results presented in Chapter 7. More accurate methods of identifying a heifer's lifetime profit corrected for the opportunity cost of postponed replacement, or alternatively her subsequent milk production would enable producers to profit more from heifer selection. Serum β-lactoglobulin, discussed in Chapters 4 and 7, and genetic molecular markers, mentioned in Chapters 5 and 7, each have potential as more accurate information on which to base selection. Whereas the molecular markers could likely be identified at birth or earlier, the serum protein has the interesting dynamic component of not being available until mid-pregnancy.

The present work does not address the issue of the optimal heifer selection rate, rather it is taken as given. This rate is impacted by two factors: the desired optimal culling rate in the cow herd, and the second is the number of heifers available. As the profitability of the heifers rise, due to more accurate selection, the culling rate in the cow herd should also rise, but it is not clear to what degree. The answer to the optimal cow culling rate is then best answered through integrating a heifer selection model into the dynamic optimization models discussed in Chapter 3.

The percentage of heifers selected will also be a function of the number available. Clearly, once selection has occurred in a cohort of heifers, those heifers who were sold prior to calving will not have contributed a calf to the farm so the number of heifers will drop. But this will be compensated, to some degree, by calvings of those cows which otherwise would have been culled to make room for the heifers. This issue would most easily be addressed through a simulation model of a cow herd which could track animal numbers.

Selection based on EBV milk also has future potential applications when sexed semen becomes commercially available. By largely removing the uncertainty of which pregnancy will result in a heifer, the selection decision turns to selecting the parents of the heifer.

Aside from the issue of heifer selection, a number of interesting aberrations of animal valuation have become apparent. As discussed earlier, this has included producers being able to sell cull first calf heifers at the same price as springing heifers. Another example is the ability of producers to buy cull cows and then proceed to milk these cows and then eventually re-sell these cows as culls. While the latter example can be explained by producers exploiting the dairy industry's nonoptimal culling strategy it is nonetheless interesting that the phenomenon persists. These observations indicate the need for further work on the valuation of cattle. Real options, a method which unlike traditional net present value analysis, explicitly accounts for the ability to sell the asset throughout its life may be a fruitful extension of the current cattle valuation work.

In summary, it is my hope that this research has provided two fundamental pieces of information to the Michigan dairy industry. Firstly, there is an optimal culling rate of 20% to 30%, and I expect the upper half of this range would be an appropriate starting target for many herds. Where culling rate is defined as the number of animals sold in a year divided by the average herd inventory. This range of optimal culling rates assumes a constant herd size, so for a herd undergoing expansion an even lower rate is dictated. Studies suggest moving from an non-optimal to optimal culling rate will likely improve profit by 2% - 40%. My feeling is that for most Michigan herds 5% to 10% is the more likely return associated with moving from, say a 40% culling rate to a 30% culling rate.

Secondly, as a result of the lower culling rate, producers will have excessive heifers. This dissertation provides empirical support for those suggesting selection based on EBV milk. The prescription for heifer selection is then to, based on a heifer enterprise budget, determine the raising period in which the maximum return from selling heifers is achieved. Then sell the heifers with the lowest EBV of milk. I expect this to return $20 more per heifer in the group than picking them randomly. This return will be generated over the course of the lifetimes of the animals in the cohort.

If the history of the culling literature is any indication, these results will need an active extension effort to disseminate the information to producers and their advisors. This dissertation can serve as the basis for an intuitive understanding of the important issues of cow culling and heifer selection. But I would suggest an appropriate place to begin any extension effort directed to dairy producers, their advisors, or extension educators would be with the following two questions:

"What is your current culling rate?"

"Why?"

APPENDICES

APPENDIX A

FIXED EFFECT LOGIT (CONDITIONAL LOGISTIC) PROOF FOR A CLUSTER OF 3 WITH ONE SUCCESS

The log-likelihood function for clusters of 3 (i.e., $t = 1, \ldots, T$, and $T = 3$), conditioning on $\sum_t y_{it}$, for $i = 1, \ldots, N$ (where N is assumed to approach infinity) is

$$L(\beta) = \sum_{i=1}^{N} \ell_i(\beta)$$

$$\text{where } \ell_i = \ln\left[\exp\left(\beta\sum_{t=1}^{3} x_{it}y_{it}\right) \Big/ \sum_{d\in K_i}\exp\left(\beta\sum_{t=1}^{3} x_{it}d_t\right)\right].$$

$$K_i = \left\{\mathbf{d} = (d_1, d_2, d_3) \,\middle|\, d_t \in \{0,1\} \text{ and } \sum_{t=1}^{3} d_t = \sum_{t=1}^{3} y_{it}\right\}$$

Preliminaries:

$$\text{let } P(y = 1|X) = \Lambda(X\beta) = \frac{\exp(X\beta)}{1+\exp(X\beta)}$$

a) if A, B, and C are independent then $P(ABC) = P(A)\cdot P(B)\cdot P(C)$

b) $P(ABC|D) = \dfrac{P(ABC)}{P(D)}$

c) $1 - \Lambda(X\beta) = 1 - \left[\dfrac{\exp(X\beta)}{1+\exp(X\beta)}\right] = \dfrac{1+\exp(X\beta)-\exp(X\beta)}{1+\exp(X\beta)} = \dfrac{1}{1+\exp(X\beta)}$

d) $\dfrac{e^x}{e^y} = e^{x-y}$

Assume one success in the cluster of three (i.e., $\sum_{t=1}^{3} y_{it} = 1$), as a result there are three possible alternative outcomes. Notice that if the value of the dependent variables are invariant within a cluster, the value of y_i is completely determined. As a result, these observations are dropped from the analysis as they contribute no information to the log-likelihood function. In general the number of potential outcomes (combinations) can be calculated from

$$\frac{T!}{\left(\sum_t y_{it}\right)!\left(T-\sum_t y_{it}\right)!}.$$

In step one, of the following proof (a) and (b) are used. An assumption of fixed effect logit is that y_{i1}, y_{i2}, y_{i3} are independent conditional on x_i, c_i, and $\Sigma_t y_{it}$, as a result, (a) is assumed to be true. In step two the probabilities are reported in the logistic distribution. Notice that in step two the denominator is the sum of the numerators since, by definition, $\sum_{t=1}^{3} y_{it} = 1$. In step three the exponential formulation of the logit model is explicitly stated, and (c) is applied. (Steps two and three have been omitted for the second and third joint conditional probabilities.) Step four results from arithmetic simplification of the expression in step three, notice the denominators of each fraction are the same and, as a result, cancel. The final step involves dividing both the numerator and denominator by the independent variables of one of the observations raised to the power of e. This is equivalent to dividing by one. In this case, those of the third observation (i.e., $x_{i3} + c_i$) have arbitrarily been chosen. As a result of the logistic function, namely the property noted in (d), the unobserved effect falls out, leaving differences of the observed independent variables. Notice that, within cluster, the independent variables are being differenced between observations. As a result, explanatory variables which do not vary among observations within a cluster do not contribute information to the assessment of β for that variable (i.e., observations that are concordant with respect to a epidemiological risk factor are dropped from the calculation of β for that risk factor).

$$P\left(y_{i1}=1,y_{i2}=0,y_{i3}=0\,\middle|\,x_i,c_i,\sum_{t=1}^{3}y_{it}=1\right)=\frac{P\left(y_{i1}=1\middle|x_i,c_i\right)\cdot P\left(y_{i2}=0\middle|x_i,c_i\right)\cdot P\left(y_{i3}=0\middle|x_i,c_i\right)}{P\left(y_i=1{,}0{,}0\middle|x_i,c_i\right)+P\left(y_i=0{,}1{,}0\middle|x_i,c_i\right)+P\left(y_i=0{,}0{,}1\middle|x_i,c_i\right)}$$

$$=\frac{\Lambda\left(x_{i1}\beta+c_i\right)\cdot\left[1-\Lambda\left(x_{i2}\beta+c_i\right)\right]\cdot\left[1-\Lambda\left(x_{i3}\beta+c_i\right)\right]}{\Lambda\left(x_{i1}\beta+c_i\right)\cdot\left[1-\Lambda\left(x_{i2}\beta+c_i\right)\right]\cdot\left[1-\Lambda\left(x_{i3}\beta+c_i\right)\right]+\left[1-\Lambda\left(x_{i1}\beta+c_i\right)\right]\cdot\Lambda\left(x_{i2}\beta+c_i\right)\cdot\left[1-\Lambda\left(x_{i3}\beta+c_i\right)\right]+\left[1-\Lambda\left(x_{i1}\beta+c_i\right)\right]\cdot\left[1-\Lambda\left(x_{i2}\beta+c_i\right)\right]\cdot\Lambda\left(x_{i3}\beta+c_i\right)}$$

$$=\frac{\exp(x_{i1}\beta+c_i)\Big/\left[1+\exp(x_{i1}\beta+c_i)\right]\cdot\left[1+\exp(x_{i2}\beta+c_i)\right]\cdot\left[1+\exp(x_{i3}\beta+c_i)\right]}{\dfrac{\exp(x_{i1}\beta+c_i)+\exp(x_{i2}\beta+c_i)+\exp(x_{i3}\beta+c_i)}{\left[1+\exp(x_{i1}\beta+c_i)\right]\cdot\left[1+\exp(x_{i2}\beta+c_i)\right]\cdot\left[1+\exp(x_{i3}\beta+c_i)\right]}}$$

$$=\frac{\exp(x_{i1}\beta+c_i)}{\exp(x_{i1}\beta+c_i)\cdot\exp(x_{i2}\beta+c_i)\cdot\exp(x_{i3}\beta+c_i)}$$

$$=\frac{\exp\left[\beta(x_{i1}-x_{i3})\right]}{1+\exp\left[\beta(x_{i1}-x_{i3})\right]\cdot\exp\left[\beta(x_{i2}-x_{i3})\right]}$$

$$P\left(y_{i1}=0,y_{i2}=1,y_{i3}=0\,\middle|\,x_i,c_i,\sum_{t=1}^{3}y_{it}=1\right)=\frac{P\left(y_{i1}=0\middle|x_i,c_i\right)\cdot P\left(y_{i2}=1\middle|x_i,c_i\right)\cdot P\left(y_{i3}=0\middle|x_i,c_i\right)}{P\left(y_i=1{,}0{,}0\middle|x_i,c_i\right)+P\left(y_i=0{,}1{,}0\middle|x_i,c_i\right)+P\left(y_i=0{,}0{,}1\middle|x_i,c_i\right)}$$

$$=\frac{\exp(x_{i2}\beta+c_i)}{\exp(x_{i1}\beta+c_i)\cdot\exp(x_{i2}\beta+c_i)\cdot\exp(x_{i3}\beta+c_i)}$$

$$=\frac{\exp\left[\beta(x_{i2}-x_{i3})\right]}{1+\exp\left[\beta(x_{i1}-x_{i3})\right]\cdot\exp\left[\beta(x_{i2}-x_{i3})\right]}$$

$$P\left(y_{i1}=0, y_{i2}=0, y_{i3}=1 \middle| x_i, c_i, \sum_{t=1}^{3} y_{it}=1\right)=\frac{P\left(y_{i1}=0 \middle| x_i, c_i\right)\cdot P\left(y_{i2}=0 \middle| x_i, c_i\right)\cdot P\left(y_{i3}=1 \middle| x_i, c_i\right)}{P\left(y_i=1,0,0 \middle| x_i, c_i\right)+P\left(y_i=0,1,0 \middle| x_i, c_i\right)+P\left(y_i=0,0,1 \middle| x_i, c_i\right)}$$

$$=\frac{\exp(x_{i3}\beta+c_i)}{\exp(x_{i1}\beta+c_i)\cdot\exp(x_{i2}\beta+c_i)\cdot\exp(x_{i3}\beta+c_i)}$$

$$=\frac{1}{1+\exp\left[\beta(x_{i1}-x_{i3})\right]\cdot\exp\left[\beta(x_{i2}-x_{i3})\right]}$$

APPENDIX B

MODIFIED BOX-COMPLEX FORTRAN CODE FOR THE HEIFER SELECTION PROBLEM

```
!------------------------------------------------------------------
      program complex
!
! This code solves the optimal heifer selection problem of assessing
!  when in the 4 periods prior to calving heifers should be selected
!  and how the genetic info in each period should be weighted to
create
!  an index on which the heifers are ranked for subsequent selection.
! X1 to X3 are the percentage of animals selected in periods 1 to 3.
! The number selected in the 4th period is the difference between
!  Selectp the constraint of the percentage of heifers to be selected
!  and Select3.
! The implicit constraint of meeting Selectp has been altered by
adding
!  the artificial activity of buying heifers in the 4th period for the
!  amount of Buy to make up for any short fall in heifers.
! The Select variables are constrained to less than 1.25 to prevent
!  difficulty finding an otherwise corner point solution of 1.0
! X10 to X19 are the weights for the genetic data in the 1st period.
! X20 to X29 are the weights for the 2nd period etc.
! Order of the genetic data: sire followed by dam, PTA milk, fat,
!  protein, milk-fat reliability, protein reliability.
! The objective value is the herd profit divided by the number of
!  heifers in the herd and then summed over all herds. Herd profit
!  is a function of R1,R2,R3,R4 (the net returns in each of the
!  respective rearing periods from sale of a heifer) and Lactprof
!  (the lifetime profit for the heifer corrected for the opportunity
!   cost of postponed replacement, this was generated based on 1st
lact
!   ME). Any bought heifers have a Lactprof=0.
! Upto 5000 original vertices can be created. Funcl, Selectl and Xl
!  hold these original large(l) number of vertices, the Xs of the K
!  vertices with the highest objective values are transferred to the X
!  array and the standard Modified Box-Complex algorithm begins.
! Rho determines how many of the potentially 5000 vertices are
generated.
! Prec controls precision of the Xs (ie if Prec=1000, the Xs will be
!  3 decimal places, if =100 then 2 decimal places etc to if =1 then
!  the Xs will be 0 or 1.
! For definition of the standard variables used in Box-Complex see
!  Kuester and Mize Optimization Techniques with Fortran (1973) pp
!  368-385. See Harris, Thomas R.("Appendix A: The Modified Box-
Complex"
!  Analysis of Irrigation Scheduling for Grain Sorghum in the
!  Oklahoma Panhandle, Oklahoma State University Ph.D. Dissertation,
!  1981) for application of the Modified operations of Nelder, J.A.,
!  and R. Mead ("A simplex method for function minimization." Comput.
!  J. 7(1964):308-313) to Box-Complex.
!------------------------------------------------------------------
      Integer*4 N, M, K, Itmax, Gamma, Beta, Herdnum, Xl(5000,52)
        Integer*4 It, Iev2, Iprint, R(1:4), Prec, Buy, H(52), Rho
        Integer*4 X(104,52), F(104), Select(1:104,1:5,1:100), G(52)
        Integer*4 Ievr, Ieve, Ievc1, Ievc2, Rankl(5000), Fl(5000)
        Integer*4 Selectl(1:5000,1:5,1:100), II, JJ, KK
        Dimension Xc(49)
      Real Alpha, Delta, Selectp, Psi, Theta, Rnum(52)
        Character*3 Endl
      Character*5 Again

        OPEN (UNIT=33, ACCESS='SEQUENTIAL',FILE='c:\misc\result.dat',
     +  FORM='FORMATTED', STATUS='OLD')

      Again = '     '
      Delta = 0.0
      N = 49
        M = 49
        K = 100
      Endl = '   '
        KK = 0
        Ievr = 101
        Ieve = 102
        Ievc1 = 103
        Ievc2 = 104

5     call Input(N,M,K,Itmax,Ic,Iprint,Alpha,Beta,Gamma,Xl,R,
     +  Again,KK,Prec,Buy,Herdnum,Selectp,Psi,Theta,Rho)

      ! Zero the Select array
          II = 0
          JJ = 0
          KK = 0
      Do While ( II .LT. 5 )
          JJ = 0
          II = II + 1
          Do While ( JJ .LT. 104 )
            JJ = JJ + 1
            KK = 0
          Do While ( KK .LT. 100 )
              KK = KK + 1
            Select(JJ,II,KK) = 0
            End Do
        End Do
      End Do
```

```
!     zero the X array
          II = 0
          JJ = 0
      Do While ( II .LT. 52 )
        II = II + 1
          JJ = 0
          Do While ( JJ .LT. 104 )
            JJ = JJ + 1
         X(JJ,II) = 0
        End Do
      End Do

!     Zero the Fl array
          II = 0
      Do While ( II .LT. 5000 )
          II = II + 1
          Fl(II) = 0
      End Do

!     Zero the F array
          II = 0
      Do While ( II .LT. 104 )
          II = II + 1
          F(II) = 0
      End Do

!     Zero the Rankl array
          II = 0
      Do While ( II .LT. 5000 )
          II = II + 1
          Rankl(II) = 0
      End Do

        Write(*,10)
10    Format(18X,'Complex procedure of box',//,18X,'Parameters')
      Write(*,11) N, M, K, Itmax, Ic
11    Format(2X,'N=',I2,3X,'M=',I2,3X,'K=',I3,2X,'Itmax=',
     + I4,2X,'Ic=',I2)
      Write(*,12) Alpha, Beta, Gamma, Delta
12    Format(2X,'Alpha=',F5.2,5X,'Beta=',I8,3X,'Gamma=',
     + I2,3X,'Delta=',F6.2)

      pause

        I = 0
      Do While ( I .LT. Rho )
          I = I + 1
          Call Funcl(N,M,K,Xl,Fl,I,Selectl,R,Prec,Buy,Again,Endl,
     +        Herdnum,KK,Selectp)
      End Do

      ! Rank Fl and store rankings in Rankl
          II = 0
          Do While (II .LT. Rho)
          II = II + 1
             Rankl(II) = 1
             JJ = 0
            Do While ( JJ .LT. Rho )
            JJ = JJ + 1
              If ( Fl(JJ) .GT. Fl(II) ) Then
                          Rankl(II) = Rankl(II) + 1
              End If
          End Do
        End Do

      ! Break any ties in Fl rankings
          II = 0
          Do While ( II .LT. Rho )
            II = II + 1
            JJ = II
           Do While ( JJ .LT. Rho )
             JJ = JJ + 1
           If ( Rankl(II) .EQ. Rankl(JJ) ) Then
             Rankl(JJ) = Rankl(JJ) + 1
           End If
          End Do
          End Do

!     Copy the X for the 100 vertices with the highest objective
!     value to the X array from the Xl array
        II = 0
      Do While ( II .LT. Rho )
        II = II + 1
          If ( Rankl(II) .LE. K ) Then
            JJ = 0
          Do While ( JJ .LT. 52 )
            JJ = JJ + 1
            X(Rankl(II),JJ) = Xl(II,JJ)
        End Do
       End If
        End Do

50    Call Consx(N,M,K,Itmax,Alpha,Beta,Gamma,Delta,X,F,It,Iev2,
     +  G,H,Xc,Iprint,R,Again,Select,Prec,Buy,Endl,Herdnum,Selectp,
```

```
     +  Psi, Theta)

      If ( Again .EQ. 'again' ) then
          goto 30
        End If

        If(It-Itmax) 22,22,30
30    Write(33,20) Itmax
20    Format(///,2X,'The Number of Iterations has exceeded',I4,10X,
     + 'Program Terminated')
22    Write(33,15) F(Iev2)
15    Format(//,2X,'Final Value of the function = ',I12)
      Write(33,16)
16    Format(/,2X,'Final X Values')

      Write(33,17) Prec, Alpha, Beta, Gamma, Psi, Theta, Buy, It
17    Format(/,1X,'Precision Variable = ',I8,'     Alpha = ',F6.3,
     +'   Beta = ',I6,/,1X,'Gamma = ',I7,'    Psi = ',F6.3,
     +'   Theta = ',F6.3,'   Buy = ',I6,/,'Iteration number = ',I3 )
      Write(33,21) Selectp, R(1), R(2), R(3), R(4)
21    Format(/,1X,'Selectp=',F6.4,1X,'R(1)=',I5,1X,'R(2)=',
     +  I5,1X,'R(3)=',I5,1X,'R(4)=',I5)

        II = 0
      Do While ( II .LT. Herdnum )
          II = II + 1
      Write(33,18) Select(Iev2,1,II),Select(Iev2,2,II),
     + Select(Iev2,3,II),Select(Iev2,4,II), Select(Iev2,5,II)
18    Format(/,2X,'Final Select Values = ', 5(I6,3X) )
      End Do

        II = 0
      Do While ( II .LT. 52 )
        II = II + 1
          Rnum(II) = Real(X(Iev2,II))/Real(Prec)
        End Do

      Do JJ = 1,M
                  Write(33,19) JJ, Rnum(JJ)
19      Format(/,2x,'X(',I2,') = ',F15.6)
      End Do

      Print*, ' '
        Print*, ' '
      Print*, 'program has finished processing for this group. '

999   Stop
```

```
End
```

```
!-------------------------------------------------------------
        subroutine Input(N,M,K,Itmax,Ic,Iprint,Alpha,Beta,Gamma,Xl,
     +R,Again,KK,Prec,Buy,Herdnum,Selectp,Psi,Theta,Rho)
!
!
!-------------------------------------------------------------
        Integer*4 N, M, K, Itmax, Ic, Iprint, Beta, Gamma, Xl(5000,52)
      Integer*4 R(1:4), KK, Prec, Buy, Herdnum, Herd1, Herd2
      Integer*4 Seed, Ihr, Imin, Isec, I100th, I, J, Rho
        Character*5 Again
        Character*8 Infile
        Character*12 Tempname
        Character*20 Filename
      Real Alpha, Renout, Selectp,Psi,Theta
      Logical Exists
      External Gettim

      Rho = 1000
      Herdnum = 0
      KK = KK + 1

      If ( KK .EQ. 1 ) Then
       Infile = 'c:\misc\'
!        Print*, 'Enter a File Name for the data file.'
!        Read(*,18) Tempname
!18     Format(A12)

       Tempname = 'insample.dat'
         Filename = Infile // Tempname
         Inquire(FILE=Filename,EXIST=Exists)


          If ( Exists ) Then
           OPEN
(UNIT=32,FILE=Filename,ACCESS='SEQUENTIAL',STATUS='OLD',
     +             FORM='FORMATTED',IOSTAT=IERROR)
            Iout = 32
            Print*, Filename
            Print*, 'Was Found '
          End If

         Do While( .NOT. Exists )
           Print*, 'File does not exist try another name.'
             Read(*,1) Tempname
1          Format(A12)
             Filename = Infile // Tempname
           Inquire(FILE=Filename,EXIST=Exists)

           If ( Exists ) Then
           OPEN
(UNIT=32,FILE=Filename,ACCESS='SEQUENTIAL',STATUS='OLD',
     +            FORM='FORMATTED',IOSTAT=IERROR)
            Iout = 32
            Print*, Filename
            Print*, 'Was Found '
            End If
         End Do
      End If

!        Itmax = 100
      Print*, 'Enter a Maximum number of Iterations'
        Read(*,2) Itmax
2     Format(I5)
      ! Unconstrained problem through addition of artificial activity
of
        ! buying animals in forth period.
        Ic = 0
        Iprint = 1
      Templact = 0

          I = 0
      Do While ( I .LT. 4 )
          I = I + 1
          R(I) = 0
      End Do

!     zero the Xl array
          I = 0
          J = 0
      Do While ( I .LT. 52 )
        I = I + 1
          J = 0
          Do While ( J .LT. 5000 )
            J = J + 1
          Xl(J,I) = 0
        End Do
      End Do

!      Print*, 'Enter a Alpha '
        !Read(*,3) Alpha
!3     Format(E12.4)
      Alpha = 2.0
```

```
!      Print*, 'Enter a Beta '
!        Read(*,4) Beta
!4     Format(I8)
      Beta = 1
!      Print*, 'Enter a Gamma '
!        Read(*,5) Gamma
!5     Format(I5)
      Gamma = 2
      Print*, 'Enter Select Percentage '
        Read(*,6) Selectp
6     Format(F6.4)
!      Selectp = .8
!      Print*, 'Enter a Psi '
!        Read(*,8) Psi
!8     Format(E7.4)
      Psi = 2.5
!      Print*, 'Enter a Theta '
!        Read(*,10) Theta
!10    Format(E7.4)
      Theta = .6
        Print*, 'Enter a R1 '
        Read(*,11) R(1)
11    Format(I5)
      Print*, 'Enter a R2 '
        Read(*,12) R(2)
12    Format(I5)
      Print*, 'Enter a R3 '
        Read(*,13) R(3)
13    Format(I5)
      Print*, 'Enter a R4 '
        Read(*,14) R(4)
14    Format(I5)
!      R(1) = 41
!        R(2) = 75
!        R(3) = 114
!        R(4) = 158

      Print*, 'Enter a precision variable '
        Read(*,15) Prec
15    Format(I6)

      Print*, 'Enter the cost to buy an animal in the fourth period. '
        Read(*,16) Buy
16    Format(I5)

      ! Initial Values
      Xl(1,1) = 1*Prec
```

```
        Xl(1,2) = 1*Prec
        Xl(1,3) = 1*Prec

!     generate Xl1 to Xl3 for other than 1st vertex,
!     Xl1 to Xl3 must be between 0 and 1
        I = 1
        J = 0
      Do While ( I .LT. Rho )
        I = I + 1
        J = 0
        Do While ( J .LT. 3 )
          J = J + 1
         Call GETTIM(Ihr,Imin,Isec,I100th)
         Seed = I100th
         Call RANDN(Seed, Ranout)
         Xl(I,J) = Anint(Ranout*Real(Prec))
           Xl(I,J) =
Anint(((Real(Xl(I,J))*.5)/Real(Prec)+.75)*Real(Prec))
        End Do
      End Do

!     Generate Xl4 to Xl49 for all vertices
        I = 0
      Do While ( I .LT. Rho )
        I = I + 1
        J = 3
        Do While ( J .LT. N )
          J = J + 1
        Call GETTIM(Ihr,Imin,Isec,I100th)
        Seed = I100th
        Call RANDN(Seed, Ranout)
!     Xl4 through Xl49 must be between -1 and 1
          Ranout = (Ranout*2)+(-1.0)
          Xl(I,J) = Anint(Ranout*Real(Prec))
        End Do
      End Do

!     Count the number of herds
        I = 1
20      J = 1
30      Read(Iout,40,End=100) Herd1
40      Format(I9)
        If ( J .EQ. 1 ) Then
          J = 2
          Herd2 = Herd1
        End If
        If ( Herd1 .NE. Herd2 ) Then
```

```
        I = I + 1
        Herd2 = Herd1
      End If
    Goto 30
100 Herdnum = I
    Rewind Iout

    Print*, ' '
    Print*, ' '
      Print*, 'There are', Herdnum, '   Herds. '
    Print*, ' '
    Print*, ' '


    Return
      End
```

```
C*********************************************************************
*****
          SUBROUTINE RANDN(NSEED,RANNUM)
C    Subroutine: RANDN
C    Procedure to produce a uniform random number from 0.0 to 1.0

C-----Start of Variables Passed and Returned--------------------------
-----
      INTEGER*4 NSEED
      REAL RANNUM
C     Where:
C       NSEED =    Starting seed value (0 = use defalt seed,
C                   Pos no = use this number to start the seed
C       RANNUM  =  Random number  returned (0.0 to 1.0)
C-----End of Variables Passed and Returned----------------------------
-----

C-----Start of Shared Common Blocks-----------------------------------
-----
C-----End of Shared Common Blocks-------------------------------------
-----

C-----Start of Local Variable Definition------------------------------
-----
      INTEGER*2 I, ICTL, TEMP
      REAL SEEDI(4), K(4)

      DATA SEEDI/2510.,7692.,2456.,3765./
      DATA ICTL/0/
C-----End of Local Variable Definition--------------------------------
-----


C**** Set the random number seed based on option selected
C---> Reset the seed back to the internal base seed for this
simulation
      IF (NSEED .LT. 0) THEN
        DO 100 I=1,4,1
  100   K(I) = SEEDI(I)
        NSEED = 0
        ICTL = 1
      ENDIF

C---> Set the base seed if not set earlier
      IF (ICTL .EQ. 0) THEN
        DO 110 I=1,4,1
```

```
110   K(I) = SEEDI(I) + NSEED
        ICTL = 1
      ENDIF

C**** Calculate the random number
      K(4) = 3 * K(4) + K(2)
      K(3) = 3 * K(3) + K(1)
      K(2) = 3 * K(2)
      K(1) = 3 * K(1)
      TEMP = K(1) / 1000
      K(1) = K(1) - TEMP * 1000
      K(2) = K(2) + TEMP
      TEMP = K(2) / 100
      K(2) = K(2) - 100 * TEMP
      K(3) = K(3) + TEMP
      TEMP = K(3) / 1000
      K(3) = K(3) - TEMP * 1000
      K(4) = K(4) + TEMP
      TEMP = K(4) / 100
      K(4) = K(4) - 100 * TEMP
      RANNUM = (((REAL(K(1))*.001+REAL(K(2)))*.01+REAL(K(3)))
     +                 *.001+REAL(K(4)))*.01

      RETURN
      END
```

```
!------------------------------------------------------------
      subroutine  Funcl(N,M,K,Xl,Fl,I,Selectl,R,Prec,Buy,Again,
     +  Endl,Herdnum,KK,Selectp)
!
!
!------------------------------------------------------------
      Integer*4 N,M,K,Xl(5000,52),Fl(5000),I,Selectl(1:5000,1:5,1:100)
      Integer*4 R(1:4), Prec, Buy, Herdnum, Herd, KK, ZZ, H, Temp1
        Integer*4
Save1(1:150,0:40),Index1(1:150,1:40),Lactprofit(1:150)
        Integer*4 Rank(1:150,1:40), Data1(1:150,1:49), Temp(1:41)
        Integer*4 II, JJ, Selectn, Count, Z, P, Sum1, Sum2, Herd1,
Herd2
        Real Selectp
        Character*5 Again
        Character*3 Endl

      ! Zero the Selectl array
          II = 0
          JJ = 0
          KK = 0
      Do While ( II .LT. 5 )
          JJ = 0
          II = II + 1
          Do While ( JJ .LT. 100 )
            JJ = JJ + 1
            KK = 0
         Do While ( KK .LT. 5000 )
              KK = KK + 1
           Selectl(KK,II,JJ) = 0
            End Do
        End Do
      End Do

        Temp1 = 0
        Herd = 0
      Do While ( Herd .LT. Herdnum )
          Herd = Herd + 1

      Sum1 = 0
        Sum2 = 0

        ! Optimal heifer selection problem


      ! Zero save array
          II = 0
```

```
   JJ = 0
  Do While ( II .LT. 150 )
   II = II + 1
   JJ = 0
   Do While ( JJ .LT. 41 )
    Save1(II,JJ) = 0
    JJ = JJ + 1
   End Do
 End Do

! Zero index array
  II = 0
  Do While ( II .LT. 150 )
   II = II + 1
   JJ = 0
   Do While ( JJ .LT. 40 )
    JJ = JJ + 1
    Index1(II,JJ) = 0
   End Do
 End Do

! Zero the Rank array
   II = 0
  Do While ( II .LT. 150 )
   II = II + 1
   JJ = 0
   Do While ( JJ .LT. 40 )
   JJ = JJ + 1
    Rank(II,JJ) = 0
   End Do
 End Do

! Zero out the temporary array for the data file
  II = 0
Do While ( II .LT. 41 )
  II = II + 1
  Temp(II) = 0
End Do

! Zero the Lactation Profit array
 II = 0
 Do While ( II .LT. 150)
  II = II + 1
  Lactprofit(II) = 0
 End Do

! Zero the save1 array
```

```
        II = 0
        JJ = 0
      Do While ( II .LT. 41 )
        JJ = 0
        Do While ( JJ .LT. 150 )
         JJ = JJ + 1
        Save1(JJ,II) = 0
       End Do
        II = II + 1
      End Do

      ! Zero the data1 array
        II = 0
        JJ = 0
      Do While ( II .LT. 49 )
        JJ = 0
        II = II + 1
        Do While ( JJ .LT. 150 )
         JJ = JJ + 1
        Data1(JJ,II) = 0
       End Do
      End Do


      ! Read in Data File for a single herd
       II = 0
       Do While ( Endl .NE. 'end' )

       Read(32,40,End=100) Herd1, (Temp(KK), KK=1,41), ZZ
40     Format(I9,1X,41(I10,1X),I10)

        II = II + 1

       If ( II .EQ. 1 ) Then
         Herd2 = Herd1
        End If
        If ( Herd2 .EQ. Herd1 ) Then
      ! Load Lactation Profit Array
        Lactprofit(II) = Temp(41)
         JJ = 0
        Do While( JJ .LT. N )
          JJ = JJ + 1
      ! Load genetic data into data1 array
          Data1(II,JJ) = Temp(JJ)
        End Do
          JJ = 0
        Do While( JJ .LT. 41 )
```

```
            JJ = JJ + 1
            Temp(JJ) = 0
          End Do
        Else
          Backspace 32
            II = II - 1
          Goto 101
          End If

50      End Do


100   Rewind 32

      ! Count = number of heifers
101   Count = II
        H = Count

      Selectn = Anint(Selectp * Real(Count))


      ! Initialize save1 array: pt at 1 in the 0 spot up to Count
        II = 0
      Do While ( II .LT. Count )
        II = II + 1
          Save1(II,0) = 1
      End Do

      ! Calculate number selected in periods 1 to 3
      Selectl(I,1,Herd) =
anint((Real(Xl(I,1))*Real(Count))/Real(Prec))
        If (Selectl(I,1,Herd) .GT. Count) Then
       Selectl(I,1,Herd) = Count
        End If
         Selectl(I,2,Herd) =
anint((Real(Xl(I,2))*Real(Selectl(I,1,Herd)))
     + /Real(Prec))
        If (Selectl(I,2,Herd) .GT. Selectl(I,1,Herd)) Then
        Selectl(I,2,Herd) = Selectl(I,1,Herd)
        End If
         Selectl(I,3,Herd) =
anint((Real(Xl(I,3))*Real(Selectl(I,2,Herd)))
     + /Real(Prec))
        If (Selectl(I,3,Herd) .GT. Selectl(I,2,Herd)) Then
         Selectl(I,3,Herd) = Selectl(I,2,Herd)
        End If
        ! Number selected in period 4 must meet constraint
```

```
         ! on number to be selected.
         Selectl(I,4,Herd) = Selectn
         ! Selectl(I,5,Herd) if negative is the number of animals
purchased
         !  in the fourth period.
      Selectl(I,5,Herd) = Selectl(I,3,Herd) - Selectl(I,4,Herd)


      Do 10 P = 10,40,10

      ! If Present Calculate Index
          II = 0
        Do While (II .LT. H)
        II = II + 1
          If ( Save1(II,(P-10)) .EQ. 1 ) Then
            JJ = P
              Do While ( JJ .LE. (P+9) )
!     Correct Data1 array column by 9 because genetic
!     for first period is in column 1 not ten
                        Index1(II,P) = Index1(II,P) + (
Xl(I,JJ)*Data1(II,(JJ-9)))
                        JJ = JJ + 1
          End Do
        End If
      End Do

      ! Rank Heifers if saved in the previous period
        II = 0
        Do While (II .LT. H)
        II = II + 1
         If ( Save1(II,P-10) .EQ. 1 ) Then
          Rank(II,P) = 1
          JJ = 0
         Do While ( JJ .LT. H )
          JJ = JJ + 1
           If ( Save1(JJ,P-10) .EQ. 1 ) Then
             If ( Index1(JJ,P) .GT. Index1(II,P) ) Then
                        Rank(II,P) = Rank(II,P) + 1
             End If
          End If
        End Do
         End If
      End Do

      ! Break any ties in heifer rankings
        II = 0
        Do While ( II .LT. H )
```

```
           II = II + 1
           JJ = II
          Do While ( JJ .LT. H )
            JJ = JJ + 1
          If ( Save1(II,P-10) .EQ. 1 ) Then
           If ( Save1(JJ,P-10) .EQ. 1 ) Then
              If ( Rank(II,P) .EQ. Rank(JJ,P) ) Then
                Rank(JJ,P) = Rank(JJ,P) + 1
              End If
              End If
          End If
         End Do
         End Do


    ! Sell Heifers with rank greater then Selectl(h,k,herd)
         Z = int(P/10)

       II = 0
       Do While (II .LT. H)
       II = II + 1
         If ( Save1(II,P-10) .EQ. 1 ) Then
          If ( Rank(II,P) .LE. Selectl(I,Z,Herd) ) Then
            Save1(II,P) = 1
          End If
       End If
     End Do

10    Continue

      II = 0
    Do While ( II .LT. H )
        II = II + 1
        Sum1 = Sum1 + Save1(II,40)*Lactprofit(II)
    End Do

    ! Objective value is average herd profit per animal in herd.
     Fl(I) = (R(1)*(H-Selectl(I,1,Herd)) +
   + R(2)*(Selectl(I,1,Herd)-Selectl(I,2,Herd)) +
   + R(3)*(Selectl(I,2,Herd)-Selectl(I,3,Herd)) +
   +  Sum1)

    If ( Selectl(I,3,Herd) .GE. Selectl(I,4,Herd) ) Then
        Fl(I) = Anint((Real(Fl(I)) +
Real(R(4))*(Real(Selectl(I,3,Herd))
   +  -Real(Selectl(I,4,Herd))))/ Real(Count))
    Else
```

```
          Fl(I) = Anint((Real(Fl(I)) +
Real(Buy)*(Real(Selectl(I,3,Herd))
     +  -Real(Selectl(I,4,Herd))))/ Real(Count))
        End If

      Temp1 = Temp1 + Fl(I)

      End Do

      Fl(I) = Temp1
        Temp1 = 0

        Return
        End
```

```
!-----------------------------------------------------------
      subroutine Consx(N,M,K,Itmax,Alpha,Beta,Gamma,Delta,X,F,It,
      +
Iev2,G,H,Xc,Iprint,R,Again,Select,Prec,Buy,Endl,Herdnum,Selectp,
     + Psi,Theta)
!
!
!-----------------------------------------------------------
      Integer*4 N, M, K, Itmax, Gamma, Beta, X(104,52), F(104)
        Integer*4 It, G(52), H(52), Iprint, R(1:4)
        Integer*4 Select(1:104,1:5,1:100), Prec, Buy, Herdnum
        Integer*4 J, JJ, KK, Fhigh
        Real Alpha, Delta, Selectp
        Dimension Xc(49)
        Character*3 Endl
        Character*5 Again
        Character*9 Exp, Con1, Con2, Red
        Character*10 Ref
        Real Psi, Theta

      Ievr = 101
      Ieve = 102
      Ievc1 = 103
      Ievc2 = 104
      Iev2 = 0

!     Check if implicit constraints present
!     Zero all X vertices other then the first
        It = 1
        Kode = 0
        If(M-N) 20,20,10
10    Kode = 1
20    Continue

        ! Check complex points against constraints
        Do 65 II = 2,K
          I = II
          K1 = II
          Call Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)

        If ( Again .EQ. 'again' ) then
            goto 240
          End If
65    Continue

      K1 = K
      Do 70 I = 1,K
          Call Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,Endl,
     +         Herdnum,KK,Selectp)
70    Continue

        Kount = 1
        Ia = 0

        ! Find point with lowest function value
        If (Iprint) 72, 80, 72
72    Write(*,21)
21    Format(/,2X,'Values of the function')
      Write(*,22) (J, F(J), J = 1,K)
22    Format(3('F(',I3,') = ',I10,4X))
80    Iev1 = 1
      Ref = '          '
        Exp = '         '
        Con2 = '         '
      Con1 = '         '
        Red = '         '

      Do 100 Icm = 2,K
          If (F(Iev1) - F(Icm)) 100, 100, 90
90        Iev1 = Icm
100   Continue

      ! Find point with highest function value
        Iev2 = 1
        Do 120 Icm = 2,K
          If (F(Iev2) - F(Icm)) 110, 110, 120
110       Iev2 = Icm
120   Continue

      If ( It .EQ. 1 ) Then
      Write(33,15) F(Iev2)
15    Format(///,2X,'Highest function value of the original vertices =
'
     + ,I10)
        Fhigh = F(Iev2)
      End If

      If ( F(Iev2) .GT. Fhigh ) Then
      Write(33,16) F(Iev2), It
16    Format(///,2X,'New highest function value and iteration = '
     + ,I10, I4)
        Fhigh = F(Iev2)
      End If
```

```
      ! Check convergence criteria
        If (F(Iev2)-(F(Iev1)+Beta)) 140, 130, 130
130   Kount = 1
      Goto 150

140   Kount = Kount + 1
      If (Kount-Gamma) 150, 240, 240

        ! Replace point with lowest function value
150   Call Centr(N,Iev1,Xc,X,K1)

      Do 160 JJ = 1,N
160       X(Ievr,JJ) = Anint((1.0+Alpha)*Xc(JJ)-
Alpha*Real(X(Iev1,JJ)))
        I = Ievr

        Call Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)

        If ( Again .EQ. 'again' ) then
           goto 240
          End If
          Call Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,Endl,
     +         Herdnum,KK,Selectp)

      Ref = 'reflection'

        ! If reflected point has the highest function value then expand
it,
        ! if the lowest function value then contract2 it, otherwise
contract1 it
        If (F(Iev2)-F(Ievr)) 310, 310, 300
300     If (F(Iev1)-F(Ievr)) 350, 350, 330

       ! Expansion of reflection point, where Psi is a positive
constant
310      Do 311 JJ = 1,N
311       X(Ieve,JJ) = Anint((1.0-
Psi)*(Xc(JJ))+Psi*(Real(X(Ievr,JJ))))
        I = Ieve

          Call Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)

         If ( Again .EQ. 'again' ) then
          goto 240
         End If
          Call Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,Endl,
     +        Herdnum,KK,Selectp)
```

```
      Exp = 'expansion'

       ! Replace point with the lowest function value (Iev1) with the
vertex,
       ! expansion or reflected, with the greater function value
         If (F(Ieve)-F(Ievr)) 315, 320, 320
315      Do 316 JJ = 1,N
316              X(Iev1,JJ) = X(Ievr,JJ)
         I = Iev1
          Goto 220
320      Do 321 JJ = 1,N
321              X(Iev1,JJ) = X(Ieve,JJ)
         I = Iev1
          Goto 220

      ! Contract2 of lowest function value point where reflected point
      ! had the lowest function value of all points

      ! 0<Theta<1
330    Do 331 JJ = 1,N
331       X(Ievc2,JJ) = Anint((1.0-Theta)*(Xc(JJ))+Theta*
     +                  (Real(X(Iev1,JJ))))
            I = Ievc2

         Call Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)

        If ( Again .EQ. 'again' ) then
           goto 240
         End If
         Call Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,Endl,
     +       Herdnum,KK,Selectp)

      Con2 = 'contract2'

       ! If contract2 point has function value > function value of
lowest
       ! point then check convergence criteria, otherwise reduction it

        If (F(Iev1)-F(Ievc2)) 340, 370, 370
340     Do 341 JJ = 1,N
341             X(Iev1,JJ) = X(Ievc2,JJ)
         I = Iev1
          Goto 220

      ! Contract1 of reflected point which doesn't have the lowest nor
highest function value
```

```
      ! 0<Theta<1
350   Do 351 JJ = 1,N
351       X(Ievc1,JJ) = Anint((1.0-Theta)*(Xc(JJ))+Theta*
     +                     (Real(X(Ievr,JJ))))
        I = Ievc1

          Call Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)

        If ( Again .EQ. 'again' ) then
            goto 240
          End If
          Call Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,Endl,
     +         Herdnum,KK,Selectp)

      Con1 = 'contract1'

      ! If contract1 point has highest function value replace Iev1
with it
      ! and check convergence criteria,
      ! if it has the lowest function value then reduction it,
otherwise
      ! replace Iev1 select reflected or contracted1 point whichever
has the
      ! highest function value and check convergence criteria

        If (F(Iev2)-F(Ievc1)) 360, 360, 362
360       Do 361 JJ = 1,N
361               X(Iev1,JJ) = X(Ievc1,JJ)
          I = Iev1
            Goto 220

362     If (F(Iev1)-F(Ievc1)) 363, 370, 370
363     If (F(Ievc1)-F(Ievr)) 364, 364, 360
364       Do 365 JJ = 1,N
365               X(Iev1,JJ) = X(Ievr,JJ)
          I = Iev1
            Goto 220


      ! Reduction of contracted points, 1 or 2, which continue to have
the
      ! lowest function value

370     JJ = 0
        II = 0
        Do While ( JJ .LT. N )
          JJ = JJ + 1
          II = 0
          Do While ( II .LT. K )
            II = II + 1
            If ( Abs(X(Iev2,JJ)-X(II,JJ)) .EQ. 1 ) Then
              X(II,JJ) = X(Iev2,JJ)
            Else
            X(II,JJ) = Anint((Real(X(Iev2,JJ))+Real(X(II,JJ)))/2)
            End If
        End Do
        End Do

      Do 374 I = 1,K
          Call Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)
374   Continue

        If ( Again .EQ. 'again' ) then
            goto 240
          End If

      Do 375 I = 1,K
          Call Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,Endl,
     +         Herdnum,KK,Selectp)
375   Continue

      Red = 'reduction'

       Goto 229

220   Continue

          Call Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)

        If ( Again .EQ. 'again' ) then
            goto 240
          End If

          Call Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,Endl,
     +         Herdnum,KK,Selectp)

229   If (Iprint) 230, 228, 230

230   Write(*,24)
24    Format(/,2X,'Coordinates of corrected point')
      Write(*,19) (Iev1, Jc, X(Iev1,Jc), Jc=1,N)
19    Format(3('X(',I3,',',I2,') = ',I8,4X))
      Print*, ' '
```

```
      Write(*,22) (I, F(I), I=1,K)
        Write(*,25)
25    Format(/,2X,'Coordinates of the centroid')
      Write(*,26) (Jc, Xc(Jc), Jc=1,N)
26    Format(3('Xc(C,',I2,') = ',F14.5,2X))
      Write(*,23) It
23    Format(/,2X,'Iteration number',I5)

      Print*, ' '
      If ( Ref .EQ. 'reflection' ) Then
        Print*, 'Reflection'
      End If
      If ( Exp .EQ. 'expansion' ) Then
        Print*, 'Expansion'
      End If
      If ( Con2 .EQ. 'contract2' ) Then
        Print*, 'Contraction 2'
      End If
      If ( Con1 .EQ. 'contract1' ) Then
        Print*, 'Contraction 1'
      End If
      If ( Red .EQ. 'reduction' ) Then
        Print*, 'Reduction'
      End If
        Print*, ' '

228   It = It + 1
      If (It-Itmax) 80, 80, 240
240   Return
      End
```

```
!------------------------------------------------------------
      subroutine Check(N,M,K,X,G,H,I,Kode,Xc,Delta,K1,Again,Prec)
!
!
!------------------------------------------------------------
      Integer*4 N, M, K, X(104,52), G(52), H(52), I
        Integer*4 K1, Prec
        Dimension Xc(49)
      Character*5 Again
      Real Delta
      Logical Kode, Kt

      A = 0
10    Kt = 0
      Call Const(G,H,Prec)

      ! Check against explicit constraints
        Do 50 J = 1,N
          If (X(I,J)-G(J)) 20,30,30
20      X(I,J) = G(J) + Int(Delta)
        Goto 50
30      If (H(J)-X(I,J)) 40,50,50
40      X(I,J) = H(J) - Int(Delta)
50    Continue

      A = A + 1
        If ( A .GT. 500 ) Then
          Print*, 'Error Occured -- not satisfying explicit
constraints '
          Print*, 'Please use a slightly different Alpha, Gamma, or
Delta'
          Print*, ' '
          Print*, ' '
        Again = 'again'
          goto 110
      End If

      If (Kode) 110,110,60

      ! Check against the implicit constraints
60    NN = N + 1

      Do 100 J = NN,M
          Call Const(G,H,Prec)

          If (X(I,J)-G(J)) 80,70,70
70        If (H(J)-X(I,J)) 80,100,100
```

```
80     levl = I
       Kt = 1

         Call Centr(N,levl,Xc,X,K1)

       Do JJ = 1,3
          X(I,JJ) = Anint((Real(X(I,JJ)) + Xc(JJ))/2)
       End Do
100   Continue

      If (Kt) 110,110,10
110   Return
      End
```

```
!-----------------------------------------------------------
      Subroutine Const(G,H,Prec)
!
!
!-----------------------------------------------------------
      Integer*4 G(52), H(52), Prec

        ! Optimal heifer selection problem
      G(1) = Anint(.65*Real(Prec))
        H(1) = Anint(1.25*Real(Prec))
      G(2) = Anint(.65*Real(Prec))
        H(2) = Anint(1.25*Real(Prec))
      G(3) = Anint(.65*Real(Prec))
        H(3) = Anint(1.25*Real(Prec))
      G(4) = -1000000
        H(4) = 1000000
      G(5) = -1000000
        H(5) = 1000000
      G(6) = -1000000
        H(6) = 1000000
      G(7) = -1000000
        H(7) = 1000000
      G(8) = -1000000
        H(8) = 1000000
      G(9) = -1000000
        H(9) = 1000000
        H(9) = -1000000
        H(10) = -1*Prec
        H(10) = 1*Prec
      G(11) = -1*Prec
        H(11) = 1*Prec
      G(12) = -1*Prec
        H(12) = 1*Prec
      G(13) = -1*Prec
        H(13) = 1*Prec
      G(14) = -1*Prec
        H(14) = 1*Prec
      G(15) = -1*Prec
        H(15) = 1*Prec
      G(16) = -1*Prec
        H(16) = 1*Prec
      G(17) = -1*Prec
        H(17) = 1*Prec
      G(18) = -1*Prec
        H(18) = 1*Prec
      G(19) = -1*Prec
        H(19) = 1*Prec
```

```
G(20) = -1*Prec
  H(20) = 1*Prec
G(21) = -1*Prec
  H(21) = 1*Prec
G(22) = -1*Prec
  H(22) = 1*Prec
G(23) = -1*Prec
  H(23) = 1*Prec
G(24) = -1*Prec
  H(24) = 1*Prec
G(25) = -1000000
  H(25) = 1000000
G(26) = -1000000
  H(26) = 1000000
G(27) = -1000000
  H(27) = 1000000
G(28) = -1000000
  H(28) = 1000000
G(29) = -1000000
  H(29) = 1000000
H(30) = -1*Prec
  H(30) = 1*Prec
G(31) = -1*Prec
  H(31) = 1*Prec
G(32) = -1*Prec
  H(32) = 1*Prec
G(33) = -1*Prec
  H(33) = 1*Prec
G(34) = -1*Prec
  H(34) = 1*Prec
G(35) = -1*Prec
  H(35) = 1*Prec
G(36) = -1*Prec
  H(36) = 1*Prec
G(37) = -1*Prec
  H(37) = 1*Prec
G(38) = -1*Prec
  H(38) = 1*Prec
G(39) = -1*Prec
  H(39) = 1*Prec
G(40) = -1*Prec
  H(40) = 1*Prec
G(41) = -1*Prec
  H(41) = 1*Prec
G(42) = -1*Prec
  H(42) = 1*Prec
G(43) = -1*Prec
  H(43) = 1*Prec
G(44) = -1*Prec
  H(44) = 1*Prec
G(45) = -1*Prec
  H(45) = 1*Prec
G(46) = -1*Prec
  H(46) = 1*Prec
G(47) = -1*Prec
  H(47) = 1*Prec
G(48) = -1*Prec
  H(48) = 1*Prec
G(49) = -1*Prec
  H(49) = 1*Prec


  Return
  End
```

```
!------------------------------------------------------------
      subroutine  Func(N,M,K,X,F,I,Select,R,Prec,Buy,Again,
     +  Endl,Herdnum,KK,Selectp)
!
!
!------------------------------------------------------------
      Integer*4 N,M,K, X(104,52), F(104), I, Select(1:104,1:5,1:100)
      Integer*4 R(1:4), Prec, Buy, Herdnum, Herd, KK, ZZ, H, Temp1
       Integer*4
Save1(1:150,0:40),Index1(1:150,1:40),Lactprofit(1:150)
       Integer*4 Rank(1:150,1:40), Data1(1:150,1:49), Temp(1:41)
       Integer*4 II, JJ, Selectn, Count, Z, P, Sum1, Sum2, Herd1,
Herd2
       Real Selectp
       Character*5 Again
       Character*3 Endl

       Temp1 = 0
       Herd = 0
      Do While ( Herd .LT. Herdnum )
        Herd = Herd + 1

      Sum1 = 0
       Sum2 = 0

       ! Optimal heifer selection problem


      ! Zero save array
        II = 0
        Do While ( II .LT. 150 )
         II = II + 1
         JJ = 0
         Do While ( JJ .LT. 41 )
          Save1(II,JJ) = 0
          JJ = JJ + 1
         End Do
       End Do

      ! Zero index array
        II = 0
        Do While ( II .LT. 150 )
         II = II + 1
         JJ = 0
         Do While ( JJ .LT. 40 )
          JJ = JJ + 1
          Index1(II,JJ) = 0
   End Do
 End Do

! Zero the Rank array
   II = 0
  Do While ( II .LT. 150 )
   II = II + 1
   JJ = 0
   Do While ( JJ .LT. 40 )
   JJ = JJ + 1
    Rank(II,JJ) = 0
   End Do
 End Do

! Zero out the temporary array for the data file
  II = 0
Do While ( II .LT. 41 )
  II = II + 1
  Temp(II) = 0
End Do

! Zero the Lactation Profit array
 II = 0
 Do While ( II .LT. 150)
  II = II + 1
  Lactprofit(II) = 0
 End Do

! Zero the save1 array
  II = 0
  JJ = 0
Do While ( II .LT. 41 )
  JJ = 0
  Do While ( JJ .LT. 150 )
   JJ = JJ + 1
  Save1(JJ,II) = 0
 End Do
  II = II + 1
End Do

! Zero the data1 array
  II = 0
  JJ = 0
Do While ( II .LT. 49 )
  JJ = 0
  II = II + 1
  Do While ( JJ .LT. 150 )
```

```
           JJ = JJ + 1
         Data1(JJ,II) = 0
       End Do
     End Do

     ! Read in Data File for a single herd
       II = 0
       Do While ( Endl .NE. 'end' )

       Read(32,40,End=100) Herd1, (Temp(KK), KK=1,41), ZZ
40     Format(I9,1X,41(I10,1X),I10)

         II = II + 1

       If ( II .EQ. 1 ) Then
           Herd2 = Herd1
         End If
         If ( Herd2 .EQ. Herd1 ) Then
     ! Load Lactation Profit Array
         Lactprofit(II) = Temp(41)
           JJ = 0
         Do While( JJ .LT. N )
             JJ = JJ + 1
     ! Load genetic data into data1 array
             Data1(II,JJ) = Temp(JJ)
         End Do
             JJ = 0
         Do While( JJ .LT. 41 )
             JJ = JJ + 1
             Temp(JJ) = 0
         End Do
       Else
         Backspace 32
           II = II - 1
         Goto 101
         End If

50      End Do


100  Rewind 32

     ! Count = number of heifers
101  Count = II
       N = Count

     Selectn = Anint(Selectp * Real(Count))
```

```
      ! Initialize save1 array: pt at 1 in the 0 spot up to Count
        II = 0
      Do While ( II .LT. Count )
        II = II + 1
          Save1(II,0) = 1
      End Do

      ! Calculate number selected in periods 1 to 3
      Select(I,1,Herd) = anint((Real(X(I,1))*Real(Count))/Real(Prec))
        If (Select(I,1,Herd) .GT. Count) Then
       Select(I,1,Herd) = Count
        End If
         Select(I,2,Herd) =
anint((Real(X(I,2))*Real(Select(I,1,Herd)))
     + /Real(Prec))
        If (Select(I,2,Herd) .GT. Select(I,1,Herd)) Then
       Select(I,2,Herd) = Select(I,1,Herd)
        End If
         Select(I,3,Herd) =
anint((Real(X(I,3))*Real(Select(I,2,Herd)))
     + /Real(Prec))
        If (Select(I,3,Herd) .GT. Select(I,2,Herd)) Then
         Select(I,3,Herd) = Select(I,2,Herd)
        End If
        ! Number selected in period 4 must meet constraint
        ! on number to be selected.
        Select(I,4,Herd) = Selectn
        ! Select(I,5,Herd) if negative is the number of animals
purchased
        !  in the fourth period.
      Select(I,5,Herd) = Select(I,3,Herd) - Select(I,4,Herd)


      Do 10 P = 10,40,10

      ! If Present Calculate Index
            II = 0
          Do While (II .LT. N)
          II = II + 1
            If ( Save1(II,(P-10)) .EQ. 1 ) Then
                JJ = P
                  Do While ( JJ .LE. (P+9) )
!     Correct Data1 array column by 9 because genetic
!     for first period is in column 1 not ten
```

```
                        Index1(II,P) = Index1(II,P) + (
X(I,JJ)*Data1(II,(JJ-9)) )
                        JJ = JJ + 1
           End Do
         End If
       End Do

    ! Rank Heifers if saved in the previous period
        II = 0
        Do While (II .LT. H)
        II = II + 1
          If ( Save1(II,P-10) .EQ. 1 ) Then
            Rank(II,P) = 1
            JJ = 0
           Do While ( JJ .LT. H )
             JJ = JJ + 1
              If ( Save1(JJ,P-10) .EQ. 1 ) Then
                If ( Index1(JJ,P) .GT. Index1(II,P) ) Then
                          Rank(II,P) = Rank(II,P) + 1
                End If
           End If
        End Do
          End If
       End Do

    ! Break any ties in heifer rankings
        II = 0
        Do While ( II .LT. H )
          II = II + 1
          JJ = II
         Do While ( JJ .LT. H )
           JJ = JJ + 1
         If ( Save1(II,P-10) .EQ. 1 ) Then
          If ( Save1(JJ,P-10) .EQ. 1 ) Then
             If ( Rank(II,P) .EQ. Rank(JJ,P) ) Then
               Rank(JJ,P) = Rank(JJ,P) + 1
             End If
             End If
         End If
        End Do
        End Do


    ! Sell Heifers with rank greater then select(h,k,herd)
         Z = int(P/10)

       II = 0
```

```
          Do While (II .LT. H)
          II = II + 1
            If ( Save1(II,P-10) .EQ. 1 ) Then
             If ( Rank(II,P) .LE. Select(I,Z,Herd) ) Then
               Save1(II,P) = 1
             End If
          End If
        End Do

10    Continue

         II = 0
       Do While ( II .LT. H )
           II = II + 1
           Sum1 = Sum1 + Save1(II,40)*Lactprofit(II)
       End Do

       ! Objective value is average herd profit per animal in herd.
        F(I) = (R(1)*(H-Select(I,1,Herd)) +
      + R(2)*(Select(I,1,Herd)-Select(I,2,Herd)) +
      + R(3)*(Select(I,2,Herd)-Select(I,3,Herd)) +
      +  Sum1)

       If ( Select(I,3,Herd) .GE. Select(I,4,Herd) ) Then
           F(I) = Anint((Real(F(I)) +
Real(R(4))*(Real(Select(I,3,Herd))
      +  -Real(Select(I,4,Herd))))/ Real(Count))
!          F(I) = (F(I) + R(4)*(Select(I,3,Herd)-Select(I,4,Herd)))/
Count
       Else
           F(I) = Anint((Real(F(I)) + Real(Buy)*(Real(Select(I,3,Herd))
      +  -Real(Select(I,4,Herd))))/ Real(Count))
!          F(I) = (F(I) + Buy*(Select(I,3,Herd)-Select(I,4,Herd)))/
Count
         End If

       Temp1 = Temp1 + F(I)

       End Do

       F(I) = Temp1
         Temp1 = 0

         Return
         End
```

```
!------------------------------------------------------------
      Subroutine Centr(N,Iev1,Xc,X,K1)
!
!
!------------------------------------------------------------
      Integer*4 N,Iev1,X(104,52),K1,Rk,II,JJ
        Dimension Xc(49)

        Do JJ = 1,N
          Xc(JJ) = 0.0
          Do II = 1,K1
            Xc(JJ) = Xc(JJ) + Real(X(II,JJ))
          End Do
        Rk = K1
          Xc(JJ) = (Xc(JJ)-Real(X(Iev1,JJ)))/(Real(Rk)-1.0)
      End Do

        Return
        End
```

APPENDIX C

FORTRAN CODE TO SIMULATE HERD PROFITS GIVEN A HEIFER SELECTION RULE

```
!-----------------------------------------------------------------
-
     program complex
!
! This program simulates the profit associated with a given rule for
!  heifer selection. It reads in the Xs and the data file containing
!  the genetic info and Lactprof and calculates the herd profit
divided
!  by the number of heifers for each herd in the input file.
!-----------------------------------------------------------------
-
      Integer*4 N, M, K, Herdnum, II
        Integer*4 R(1:4), Buy, Herdno(100,2)
        Integer*4 F(100)
      Real Selectp, X(52), Select(1:5,1:100)
        Character*8 Infile
        Character*12 Tempname
        Character*20 Filename
      Logical Exists

        OPEN (UNIT=33, ACCESS='SEQUENTIAL',FILE='c:\misc\result1.dat',
     +  FORM='FORMATTED', STATUS='OLD')



       Infile = 'c:\misc\'
         Print*, 'Enter a File Name that contains the Xs.'
         Read(*,2) Tempname
2      Format(A12)

         Filename = Infile // Tempname
         Inquire(FILE=Filename,EXIST=Exists)


          If ( Exists ) Then
           OPEN
(UNIT=34,FILE=Filename,ACCESS='SEQUENTIAL',STATUS='OLD',
     +            FORM='FORMATTED',IOSTAT=IERROR)
            Print*, Filename
            Print*, 'Was Found '
          End If

        Do While( .NOT. Exists )
          Print*, 'File does not exist try another name.'
            Read(*,1) Tempname
1           Format(A12)
```

```
            Filename = Infile // Tempname
           Inquire(FILE=Filename,EXIST=Exists)

           If ( Exists ) Then
            OPEN
(UNIT=34,FILE=Filename,ACCESS='SEQUENTIAL',STATUS='OLD',
     +            FORM='FORMATTED',IOSTAT=IERROR)
             Print*, Filename
             Print*, 'Was Found '
             End If
         End Do

      N = 49
        M = 49
        K = 100

5     call Input(N,M,K,X,R,Buy,Herdnum,Selectp)

      ! Zero the Select array
          I = 0
          II = 0
      Do While ( I .LT. 5 )
          I = I + 1
            II = 0
          Do While ( II .LT. 100 )
              II = II + 1
             Select(I,II) = 0
             End Do
      End Do

        Write(*,10)
10    Format(18X,'Complex procedure of box',//,18X,'Parameters')
      Write(*,11) N, M, K
11    Format(2X,'N=',I2,3X,'M=',I2,3X,'K=',I3)

      pause

      ! Zero the F(I) array
          II = 0
        Do While ( II .LT. 100 )
          II = II + 1
          F(II) = 0
        End Do


      Call Func(N,M,K,X,F,I,Select,R,Buy,Selectp,Herdno,Herdnum)
```

```
      Write(33,14) Selectp, Buy
14    Format(1X,'Selectp = ',F6.4,1X,'Buy = ',I5,/)

      Write(33,15) R(1), R(2), R(3), R(4)
15    Format(1X,'R(1) = ',I5,1X,'R(2) = ',I5,1X,'R(3) = ',I5,1X,
     +  'R(4) = ',I5,/)

        Write(33,16) Herdno(II,1)
16      Format(' Herdno ',2X,'F(',I6,')',2X,'Count',2X,'Select1',
     +   2X,'Select2',2X,'Select3',2X,'Select4',2X,'Select5' )

        II = 0
      Do While ( II .LT. Herdnum )
          II = II + 1
        Write(33,18) Herdno(II,1),F(II),Herdno(II,2),Select(1,II),
     +   Select(2,II),Select(3,II),Select(4,II), Select(5,II)
18      Format(1X,I6,5X,I6,3X,I6,3X,5(F6.1,3X) )
      End Do

      Do J = 1,M
          Write(33,20) J, X(J)
20      Format(/,2x,'X(',I2,') = ',F15.6)
      End Do

      Print*, ' '
        Print*, ' '
      Print*, 'program has finished processing for this group. '
      pause

        Goto 999

999   Stop
      End
```

```
!------------------------------------------------------------------
       subroutine Input(N,M,K,X,R,Buy,Herdnum,Selectp)
!
!
!------------------------------------------------------------------
       Integer*4 N, M, K, I, J
     Integer*4 R(1:4), Buy, Herdnum, Herd1, Herd2
       Character*3 Value
       Character*8 Infile
       Character*10 Value1
       Character*12 Tempname
       Character*20 Filename
     Real Ranout, Selectp, X(52)
     Logical Exists

     Herdnum = 0

      Infile = 'c:\misc\'
        Print*, 'Enter a File Name for the data file.'
        Read(*,18) Tempname
18    Format(A12)

        Filename = Infile // Tempname
        Inquire(FILE=Filename,EXIST=Exists)


         If ( Exists ) Then
          OPEN
(UNIT=32,FILE=Filename,ACCESS='SEQUENTIAL',STATUS='OLD',
     +             FORM='FORMATTED',IOSTAT=IERROR)
           Iout = 32
           Print*, Filename
           Print*, 'Was Found '
         End If

        Do While( .NOT. Exists )
          Print*, 'File does not exist try another name.'
            Read(*,1) Tempname
1         Format(A12)

            Filename = Infile // Tempname
          Inquire(FILE=Filename,EXIST=Exists)

         If ( Exists ) Then
          OPEN
(UNIT=32,FILE=Filename,ACCESS='SEQUENTIAL',STATUS='OLD',
     +             FORM='FORMATTED',IOSTAT=IERROR)
```

```
            Iout = 32
            Print*, Filename
            Print*, 'Was Found '
             End If
          End Do

       ! Unconstrained problem through addition of artificial activity
of
         ! buying animals in forth period.

          I = 0
       Do While ( I .LT. 4 )
          I = I + 1
          R(I) = 0
       End Do

!      zero the X array
          I = 0
       Do While ( I .LT. 52 )
         I = I + 1
         X(I) = 0.0
       End Do

       Print*, 'Enter Select Percentage '
         Read(*,6) Selectp
6      Format(F6.4)
!       Selectp = .8

         Print*, 'Enter a R1 '
         Read(*,11) R(1)
11     Format(I5)
       Print*, 'Enter a R2 '
         Read(*,12) R(2)
12     Format(I5)
       Print*, 'Enter a R3 '
         Read(*,13) R(3)
13     Format(I5)
       Print*, 'Enter a R4 '
         Read(*,14) R(4)
14     Format(I5)
!       R(1) = 41
!        R(2) = 75
!        R(3) = 114
!        R(4) = 158

       Print*, 'Enter the cost to buy an animal in the fourth period. '
         Read(*,16) Buy
```

```
16     Format(I5)

         I = 1
!      Count the number of herds
20       J = 1
30       Read(Iout,40,End=100) Herd1
40       Format(I9)
         If ( J .EQ. 1 ) Then
             J = 2
             Herd2 = Herd1
           End If
         If ( Herd1 .NE. Herd2 ) Then
             I = I + 1
             Herd2 = Herd1
           End If
       Goto 30
100    Herdnum = I
       Rewind Iout

       Print*, ' '
       Print*, ' '
         Print*, 'There are', Herdnum, '   Herds. '
       Print*, ' '
       Print*, ' '

       ! Read in X(I) from the result.dat file
       Do While ( Value .NE. '  X' )
           Read(34,50) Value
50       Format(A3)
       End Do

       Backspace 34

       I = 0
       Do While ( I .LT. 52 )
           I = I + 1
           Read(34,60,End=101) Value1, X(I)
60       Format(A10,2X,F15.6)
         Read(34,70,End=101) Value
70         Format(A3)
       End Do

101    Return
         End
```

```
!--------------------------------------------------------------
      subroutine
Func(N,M,K,X,F,I,Select,R,Buy,Selectp,Herdno,Herdnum)
!
!
!--------------------------------------------------------------
      Integer*4 N,M,K, F(100), I
      Integer*4 R(1:4), Buy, ZZ, Herdno(100,2), Herdnum
        Integer*4 Save1(1:150,0:40),Lactprofit(1:150)
        Integer*4 Rank(1:150,1:40), Data1(1:150,1:49), Temp(1:41)
        Integer*4 II, JJ, KK, LL, Selectn, Count, Z, P, Sum1, Herd1,
Herd2
        Real Selectp, X(52), Index1(1:150,1:40), Select(1:5,1:100)
      Character*3 Endl

      LL = 0
      I = 0

      Do While ( I .LT. Herdnum )
          I = I + 1

      F(I) = 0
          LL = LL + 1

      Sum1 = 0

        ! Optimal heifer selection problem
        II = 0
        JJ = 0

      ! Zero save array
          II = 0
          Do While ( II .LT. 150 )
            II = II + 1
            JJ = 0
            Do While ( JJ .LT. 41 )
              JJ = JJ + 1
              Save1(II,JJ) = 0
            End Do
        End Do

      ! Zero index array
          II = 0
          Do While ( II .LT. 150 )
            II = II + 1
            JJ = 0
            Do While ( JJ .LT. 40 )
        JJ = JJ + 1
        Index1(II,JJ) = 0.0
      End Do
  End Do

! Zero the Rank array
      II = 0
    Do While ( II .LT. 150 )
      II = II + 1
      JJ = 0
      Do While ( JJ .LT. 40 )
        JJ = JJ + 1
        Rank(II,JJ) = 0
      End Do
  End Do

! Zero out the temporary array for the data file
    II = 0
Do While ( II .LT. 41 )
    II = II + 1
    Temp(II) = 0
End Do

! Zero the Lactation Profit array
  II = 0
  Do While ( II .LT. 150)
    II = II + 1
    Lactprofit(II) = 0
  End Do

! Zero the save1 array
    II = 0
    JJ = 0
Do While ( II .LT. 41 )
    JJ = 0
    Do While ( JJ .LT. 150 )
      JJ = JJ + 1
    Save1(JJ,II) = 0
  End Do
    II = II + 1
End Do

! Zero the data1 array
    II = 0
    JJ = 0
Do While ( II .LT. 49 )
    JJ = 0
```

```
          II = II + 1
          Do While ( JJ .LT. 150 )
            JJ = JJ + 1
          Data1(JJ,II) = 0
         End Do
        End Do

        ! Read in Data File for a single herd
         II = 0
         Do While ( Endl .NE. 'end' )

         Read(32,40,End=100) Herd1, (Temp(KK), KK=1,41), ZZ
40       Format(I9,1X,41(I10,1X),I10)

          II = II + 1

         If ( II .EQ. 1 ) Then
            Herd2 = Herd1
          End If
          If ( Herd2 .EQ. Herd1 ) Then
        ! Load Lactation Profit Array
          Lactprofit(II) = Temp(41)
           JJ = 0
          Do While( JJ .LT. N )
              JJ = JJ + 1
        ! Load genetic data into data1 array
              Data1(II,JJ) = Temp(JJ)
          End Do
              JJ = 0
          Do While( JJ .LT. 41 )
              JJ = JJ + 1
              Temp(JJ) = 0
          End Do
         Else
          Backspace 32
            II = II - 1
          Goto 101
          End If

50       End Do


100     Rewind 32

        ! Count = number of heifers
101     Count = II
         N = Count
```

```
         Herdno(LL,1) = Herd2
           Herdno(LL,2) = Count

       Selectn = Anint(Selectp * Real(Count))

       ! Initialize save1 array: pt at 1 in the 0 spot up to Count
         II = 0
       Do While ( II .LT. Count )
         II = II + 1
           Save1(II,0) = 1
       End Do


       ! Calculate number selected in periods 1 to 3
        Select(1,I) = Anint(Real(X(1))*Real(Count)+0.000001)
         If (Select(1,I) .GT. Count) Then
        Select(1,I) = Count
         End If
          Select(2,I) = Anint(Real(X(2))*Real(Select(1,I))+0.000001)
         If (Select(2,I) .GT. Select(1,I)) Then
         Select(2,I) = Select(1,I)
         End If
          Select(3,I) = Anint(Real(X(3))*Real(Select(2,I))+0.000001)

         If (Select(3,I) .GT. Select(2,I)) Then
          Select(3,I) = Select(2,I)
         End If
         ! Number selected in period 4 must meet constraint
         ! on number to be selected.
         Select(4,I) = Selectn
         ! Select(I,5,Herd) if negative is the number of animals
purchased
         !  in the fourth period.
       Select(5,I) = Select(3,I) - Select(4,I)


       Do 10 P = 10,40,10

       ! If Present Calculate Index
             II = 0
           Do While (II .LT. N)
           II = II + 1
             If ( Save1(II,(P-10)) .EQ. 1 ) Then
                 JJ = P
                   Do While ( JJ .LE. (P+9) )
!      Correct Data1 array column by 9 because genetic
```

```
!     for first period is in column 1 not ten
                        Index1(II,P) = Index1(II,P) +
     +                        ( X(JJ)*Real(Data1(II,(JJ-9))) )
                        JJ = JJ + 1
          End Do
        End If
      End Do

      ! Rank Heifers if saved in the previous period
        II = 0
        Do While (II .LT. H)
        II = II + 1
         If ( Save1(II,P-10) .EQ. 1 ) Then
           Rank(II,P) = 1
           JJ = 0
          Do While ( JJ .LT. H )
            JJ = JJ + 1
            If ( Save1(JJ,P-10) .EQ. 1 ) Then
              If ( Index1(JJ,P) .GT. Index1(II,P) ) Then
                        Rank(II,P) = Rank(II,P) + 1
              End If
          End If
        End Do
         End If
      End Do

      ! Break any ties in heifer rankings
        II = 0
        Do While ( II .LT. H )
          II = II + 1
          JJ = II
         Do While ( JJ .LT. H )
           JJ = JJ + 1
         If ( Save1(II,P-10) .EQ. 1 ) Then
          If ( Save1(JJ,P-10) .EQ. 1 ) Then
            If ( Rank(II,P) .EQ. Rank(JJ,P) ) Then
              Rank(JJ,P) = Rank(JJ,P) + 1
            End If
            End If
         End If
        End Do
        End Do


      ! Sell Heifers with rank greater then select(h,k,herd)
          Z = int(P/10)
```

```
         II = 0
         Do While (II .LT. H)
         II = II + 1
           If ( Save1(II,P-10) .EQ. 1 ) Then
            If ( Rank(II,P) .LE. Select(Z,I) ) Then
              Save1(II,P) = 1
            End If
         End If
       End Do

10     Continue

         II = 0
       Do While ( II .LT. H )
           II = II + 1
           Sum1 = Sum1 + Save1(II,40)*Lactprofit(II)
       End Do

       ! Objective value is average herd profit per andimal in herd.
        F(I) = (R(1)*(H-Select(1,I)) +
     + R(2)*(Select(1,I)-Select(2,I)) +
     + R(3)*(Select(2,I)-Select(3,I)) +
     +  Sum1)

       If ( Select(3,I) .GE. Select(4,I) ) Then
           F(I) = Anint((Real(F(I)) + Real(R(4))*(Real(Select(3,I))
     +  -Real(Select(4,I))))/ Real(Count))
       Else
           F(I) = Anint((Real(F(I)) + Real(Buy)*(Real(Select(3,I))
     +  -Real(Select(4,I))))/ Real(Count))
         End If

       End Do

         Return
         End
```