



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

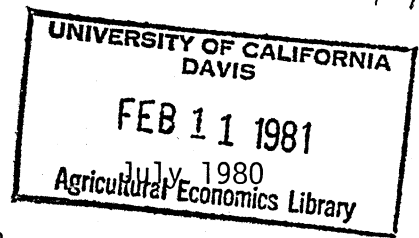
Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

LIVESTOCK

1980



The Economic Value of Genetic Information  
in Selecting Sires

Richard K. Perrin\*

Until quite recently, economists have not given much attention to what one might call the genetic decisions made by farmers, or to the decisions made by crop and livestock breeders on behalf of farmers. What I have in mind here is the choice of genetic stock to be used in final production of crops and livestock products, such as the choice of a variety or the choice of a sire. The purpose of this paper is to provide a brief review of the studies that economists have conducted in the breeding area, and then to focus on the economics of one rather narrow aspect of genetic choice, the value to a producer of genetic information about potential sires.

The Selection Decision

A universal element common to all plant and animal breeding is the selection choice. This paper is concerned with the final selection choice which is made for commercial production, for which the objective will be to maximize the economic value (expected profit or expected utility) of the individual plants or animals to be used in the final stage of production. In order to describe the nature of the choice set, it is useful to start with a simple additive genetic model. A version of the standard model which is readily understood by economists is one which specifies that the j-th observed (phenotypic) outcome, or "record", of the i-th genotype is determined as follows:

$$y_{ij} = x_{ij}b + g_i + e_{ij} \tag{1}$$

\*Professor of Economics, N. C. State University. I am indebted to Jerry Carlson and other colleagues for helpful comments on this paper.

*Presented at AAEE meetings, Urbana - July 1980.*

where  $y_{ij}$  is the measured outcome,<sup>1</sup>  $x_{ij}$  is a vector of controlled variables,  $g_i$  is the additive genetic contribution of the  $i$ -th genotype, and  $e_{ij}$  is a random "environmental" effect. The genetic effects  $g_i$ , which are not directly observable, are assumed to be fixed for a given genotype  $i$ , but randomly distributed across the genotypes comprising the population being considered, with mean 0 and variance  $\sigma_g$ .<sup>2</sup> The  $e_{ij}$  are assumed to be independently distributed, with mean 0 and variance  $\sigma_e$ .<sup>2</sup> The  $g_i$  and  $e_{ij}$  are assumed uncorrelated. This is equivalent to a random coefficients regression model, or more precisely a mixed model, since the  $b$  coefficients are assumed to be fixed. Much of the quantitative genetics literature is devoted to the problem of estimating the  $g_i$  parameters from the observations on animal  $i$  and/or his relatives. Henderson describes procedures for estimation of the model described in equation (1).

Given such a model, it is easy to describe the selection decision as a problem of choosing between alternative genotypes,  $i = 1, \dots, K$ , without having perfect information about the  $g_i$ , the additive effect of each alternative. Under this simple model, there are two sources of stochastic error associated with any one of the  $K$  prospects,<sup>2</sup> the first being errors in estimating  $g_i$  and the second being the environmental error term.

Consider the selection of a dairy sire and the selection of a wheat variety as two examples of a selection decision using this model. In choosing a wheat variety, the elements of the choice set consist of entire populations of individuals, for each of which the genotypic effect is fairly well established as a result of prior testing of the population, and for which any remaining genotypic variability will cancel out due to the large number of

individual plants being grown. Given the model specified above, the choice is fairly simple--for all the farmer has to do is to choose the variety with the largest genetic effect. As a practical matter, though, our model has assumed away the most important feature of the variety choice, which is the possibility of genotype-by-environment interactions. If such interactions exist, then the choice which will maximize returns in one environment may not maximize returns in another, and we would have to re-specify this simple model to take this aspect of decision-making into consideration.

In choosing a dairy sire, the farmer must make his choice on the basis of the predicted  $y$  (outcome) of a single offspring of a given cow, depending on whether that cow is mated with sire 1 or sire 2. The expected additive genetic effect of the offspring, according to the standard breeding assumptions will be the simple average of the sire's genetic effect and the dam's genetic effect, and since the dam's genetic effect is not subject to choice (in the sire selection problem), the decision boils down to a matter of choosing between an imperfectly known  $g_1$  versus an imperfectly known  $g_2$ . These genetic effects can be only imperfectly known, because they cannot be directly observed, partly because a bull doesn't exhibit such traits as milk production, but also because even his observable traits are the result of environmental effects and other fixed effects, as well as genetic effects. But one can observe the outcomes (milk production, for example) of some of the relatives of the bull, such as his dam or his daughters, and use this information to infer the genetic effect of the bull himself. This is the basic motivation of sire evaluation. The specific concern of this paper is the economic value of such information. Carlson has examined a closely related question, in finding that farmers seem

to be willing to pay more for sires whose genetic traits are better known. We will return to this issue after a brief review of other aspects of the selection decision which have been examined by economists.

#### Economic Considerations in the Selection Decision

In addition to the consideration of the value of genetic information just mentioned, three other aspects of the selection decision have received some attention by economists. The most venerable of these is the problem of choosing among desirable traits. In the model presented above we implied that there is just one dimension of the outcome which is of interest, that dimension which is measured by the variable  $y$ . But in fact, net profitability is the result of the outcome of a number of inherited physical traits such as yield, average daily gain, disease resistance, etc. If one knew the marginal value product of each trait, a logical procedure would be to calculate the total value of each potential sire or variety by multiplying the genetic contribution for each trait times the marginal value product of the trait, and then summing across all the traits of interest. This is exactly what the breeders have been doing, and the predicted total value is called the selection index value for that particular sire or variety.

Although breeders have been making selections on the basis of such selection indexes for forty years (since Smith and Hazel introduced the method), they have not been very comfortable with their ability to estimate marginal value products, or economic weights, as they call them, and various methods of avoiding the issue entirely have been suggested (Kempthorne and Nordskog, Pesek and Baker). Until recently, economists have not contributed much to the solution of the "economic weights" problem. Ladd and Gibson have recently used linear programming of representative firms to estimate the marginal value of increments

in various traits. Melton and Ladd have derived the VMP of various traits by first estimating production functions which have traits as arguments, and then examining the shadow price of traits by maximizing profits with traits constrained to specified levels. Merrel, Shumway, et al. avoided the problem of evaluation of the marginal value of alternative traits and simply budgeted out the net returns for each of 18 genetic prospects. They then select the prospect (breeding system, in this case) with the highest expected value.

A third aspect of the selection decision is the issue of the optimum level of adaptability and stability. This has been of primary concern in crop breeding, where environmental effects, as opposed to additive genetic effects, constitute the dominant component of variability in each prospect. Because of environmental variability across space, the crop breeder is often faced with the choice of breeding many varieties, each of which is very valuable in a narrow environmental "niche" but of low value elsewhere, versus breeding a few "adaptable" varieties which do not do so well in any particular niche but do moderately well in virtually all niches. A similar problem exists with respect to the stability of outcomes in the face of environmental variability across time at a particular location (due to weather variability, various levels of pest infestation, etc.). While the crop breeding literature on this choice problem seems to be substantial (see Neely for some references and comments), the only economic studies of which I am aware are those by Englander and Evenson, and Evenson and O'Toole.

A fourth aspect of the selection decision is that of choosing a selection differential. If for some reason you must select more than one sire or dam, you must decide how far down into the genetic rankings you should dip. This selection differential determines the expected gain in a given trait from one generation to another. The expected gain will be greatest of course if only the two

best selection elements from the current generation are mated. But in the case of cattle, which have very low reproductive rates, the only way the producer can maintain his herd size is to reduce the selection differential (keep more than just the one best hieffer), or to keep the existing cows to a greater age. Melton has recently examined the economics of this tradeoff in beef breeding.

### The Value of Information in Sire Selection

We return now to the first economic issue raised, that of the value of information which can be obtained about a sire. Assume that the value to a commercial producer of an individual animal can be described as the weighted average of a set of  $k$  performance traits (such as yield, average daily gain, etc. as previously mentioned), weighted by the marginal value product of each. The observed phenotypic value of an individual,  $V$ , can be described by extending the previous simple additive genetic model to the case of  $k$  traits as:

$$V = \underline{w}'\underline{y} = \underline{w}'\underline{g} + \underline{w}'\underline{e}, = v + \epsilon \quad (2)$$

where  $\underline{w}$  is a  $k \times 1$  vector of the marginal value products of the  $k$  performance traits,

$\underline{y}$  is a  $k \times 1$  vector of observed performance traits,

$\underline{g}$  is a  $k \times 1$  vector of additive genetic components of each trait,

$\underline{e}$  is a  $k \times 1$  vector of random environmental components of each trait,

$v$  is defined as the genetic component of value (a scalar), and

$\epsilon$  is defined as the error component of value (also a scalar).

In extending model (1) the effect of controlled variables  $x$  is omitted because it simplifies the analysis and because these variables can reasonably be assumed to be fixed with respect to the comparison of sires.

In selecting a sire, a commercial producer is most interested in the sire's genetic value,  $v$ , since this is the component of value which will be transmitted to the offspring. This is an unobservable parameter. What might be observed as proxies of this value are the phenotypic value of the sire,  $V$ , a selection index constructed from his record,  $I$ , or the performance of some of the relatives of the sire. Any of these observations provide information about the genetic value of the sire, but even all of them together will not provide perfect information about  $v$ . This information will have value to the producer if he can expect to make a more profitable selection decision having obtained the information than he could without it.

Consider the case of a producer who has a choice between sire 1 and sire 2 and assume that (i) the payoff from choosing sire  $i$  is measured by  $1/2 v_i$  per offspring sired, (ii) the objective of the producer is to maximize expected payoff from the offspring (no risk aversion), and (iii) prior knowledge about the genetic values of the two sires is represented by the normal probability density functions  $f(v_i) = N(v_i | m_i, \sigma_i^2)$ , where  $m_i$  is the expected value of  $v_i$  and  $\sigma_i^2$  is the variance. Figure 1 presents such priors for two hypothetical sires. Since the expected contribution of each sire to the value of offspring is  $(1/2)m_i$ , these expected values are the best estimates the producer has as to the relative values of the two animals as sires. Given the above assumptions, clearly the best choice of the two shown in Figure 1 is sire 2, which has the highest expected genetic value. But the question here is, what is the value to the producer of being able to obtain more information about the genetic value of one of the sires, say sire 1. The value of this information is a separate issue from the value of the animal, and it arises only insofar as that information can be expected to change the decision that would have been made in the absence of the information.



If the producer acquired perfect information about  $v_1$ , then he would choose sire 1 if  $v_1$  exceeds  $\hat{m}_2$ , and would realize an expected payoff per offspring of  $(1/2)v_1$ . But if  $v_1$  is less than  $\hat{m}_2$ , he would stay with his prior choice of sire 2 with expected payoff per offspring of  $(1/2)\hat{m}_2$ . Hence the decision, and therefore the payoff, depends upon the value of  $v_1$  which will be revealed only when the information is received. Weighting these possible payoffs ( $\hat{m}_2$  or  $v_1 > \hat{m}_2$ ) by the prior density function on  $v_1$  yields the expected payoff prior to obtaining the information but given that the information is to be obtained. The difference between this expected payoff and the expected payoff which would be realized without information ( $\hat{m}_2$  in this case) is the expected value of perfect information (EVPI) about sire 1. In this case in which  $\hat{m}_2 > \hat{m}_1$ , EVPI is therefore

$$EVPI = \int_{-\infty}^{\hat{m}_2} \hat{m}_2 N(v_1 | \hat{m}_1, \sigma_1^2) dv_1 + \int_{\hat{m}_2}^{\infty} v_1 N(v_1 | \hat{m}_1, \sigma_1^2) dv_1 - \hat{m}_2, \quad (3a)$$

On the other hand if  $\hat{m}_1 > \hat{m}_2$ , then sire 1 would be the best selection in the absence of information, so that EVPI would in that case be

$$EVPI = \int_{-\infty}^{\hat{m}_2} \hat{m}_2 N(v_1 | \hat{m}_1, \sigma_1^2) dv_1 + \int_{\hat{m}_2}^{\infty} v_1 N(v_1 | \hat{m}_1, \sigma_1^2) dv_1 - \hat{m}_1. \quad (3b)$$

Expressions 3a and 3b can be re-written (see the appendix for a derivation) as:

$$EVPI = \sigma_1 [-z + zF_N(z) + f_n(z)] = \sigma_1 L_N(z), \quad (4)$$

$$\text{where } z = \frac{|\hat{m}_2 - \hat{m}_1|}{\sigma_1},$$

$f_N$ ,  $F_N$  are standard normal density and cumulative functions, respectively, and  $L_N(z)$  is the unit normal linear loss integral, tabulated in Winkler (Table 6).

Since  $L_N$  is a decreasing function of  $z$ , it is clear that the value of information is the greatest when  $m_1 = m_2$ , and declines as these expected values diverge. This result is intuitively appealing since the farther apart are  $m_1$  and  $m_2$  (no matter which is the larger), the less likely it is that the information about  $v_1$  will affect the decision, and therefore the less is the expected gain from acquiring that information.

EVPI is the value of perfect information about  $v_1$ . For sample information about  $v_1$  which is less than perfect, the value (derived by Raiffa and Schlaifer, Chap. 11) is:

$$EVSI = \sigma_1^* L_N(z^*), \quad (5)$$

$$\text{where } \sigma_1^* = (\sigma^2 - \sigma^{*2})^{1/2}, \text{ and}$$

$$z^* = \frac{|m_2 - m_1|}{\sigma_1^*},$$

that is, use the same formula as for EVPI, but replace the prior variance on  $v_1$  with the reduction in variance due to the information about  $v_1$ . We are now ready to apply these results to a particular sire selection problem.

In order to estimate the value of various kinds of information about  $v_1$ , all that remains is to determine the reduction in the variance of the prior that is associated with each kind of information. The appendix to this paper derives the following results which are useful for that purpose. If the producer possesses no information specific to sire 1, then his prior on  $v_1$  is the distribution of genetic values (assumed normal) in the population, namely  $f(v_1) = N(v_1 | 0, \sigma_v^2)$  if value is expressed in departures from the mean. Then the posterior density function after observing  $V_1$ , the phenotypic value of the

sire is:

$$f''(v_1) = N[v_1 | \frac{\sigma_v^2}{\sigma_V^2} v_1, (1 - \frac{\sigma_v^2}{\sigma_V^2})\sigma_v^2], \quad (6)$$

where  $\sigma_v^2$  is the variance of phenotypic value in the population. The posterior density function after observing  $I_1$ , the Smith-Hazel selection index for sire 1 is:

$$f''(v_1) = N[v_1 | I, \sigma_v^2 - \sigma_I^2], \quad (7)$$

where  $\sigma_I^2$  is the variance of  $I$  in the population. The posterior density function after observing both  $I_1$  and  $I_2$ , the index value for a full sib, is

$$f''(v_1) = N[v_1 | C_1 I_1 + C_2 I_2, (C_3^{-1} + C_4^{-1})^{-1}] \quad (8)$$

where  $C_1 = [1 + C_3 C_4^{-1}]^{-1}$ ,

$$C_2 = [1 + C_4 C_3^{-1}]^{-1},$$

$$C_3 = 3\sigma_v^2 + \sigma_\epsilon^2,$$

$$C_4 = \sigma_v^2 - \sigma_I^2, \text{ and}$$

$\sigma_\epsilon^2$  = the variance of the random component of phenotypic value in the population.

Finally, of course, if perfect information about  $v_1$  were obtained, the posterior density function will be

$$f''(v_1) = N(v_1 | v_1, 0) \quad (9)$$

These results now permit us to examine the value of information that is generated by boar testing stations.

The Vaule of Boar Test Information

At a number of swine testing stations around the country, the growth characteristics of young boars as they grow to maturity is measured in terms of average daily gain (ADG--measured in pounds per day), feed efficiency (FE--measured in pounds of feed per pound of gain) and backfat thickness (BF--measured in inches). Bereskin (1977) estimated that at the margin, a one unit increase in ADG will result in time-related cost decreases of \$7.80 per pig, that a one unit increase in FE (pounds of feed per pound of gain) will increase costs by \$18 per pig, and an increase of one inch in backfat will reduce market value by \$7.00 per pig. The vector of economic weights is therefore  $\underline{w}' = (\$7.80, -\$18, -\$7)$ . (These estimates differ substantially from those of Ladd and Gibson). The boars to be tested for these performance traits are consigned to the testing stations by private breeders, and at the end of the growth period, the boars are sold (generally at auction) at a sale supervised by the testing station. Often producers submit two litter mate boars (full sibs) to be tested together in a pen . A sale brochure prepared at the end of the test provides the growth data and a selection index value for each of the animals.

Suppose a producer has the choice of selecting sire 1, a boar which has been tested, versus sire 2, a sire selected at random from the same population from which sire 1 was obtained. What is the value to the producer of having access to the test information for sire 1, prior to making his choice?

In this particular example there is no visual evaluation of either animal, nor any prior information about relatives, so the priors on the values of the two sires are the same, and consist of the distribution of  $v$  in the

population. If we have defined  $v$  as the deviation of genetic value from average, then  $m_1' = m_2' = 0$  and  $\sigma_1'^2 = \sigma_2'^2 = \sigma_v^2$ . Equations 6-9 provide the formulas for determining the expected value of the information provided by these tests. Bereskin (1977) provides the genetic correlation data from which the following variances can be determined for the population of hogs tested at central testing stations:

$$\sigma_V^2 = 35.02$$

$$\sigma_v^2 = 13.61$$

$$\sigma_\epsilon^2 = 21.41$$

$$\sigma_I^2 = 6.19.$$

Utilizing these variances and equations 6-9, the density function parameters shown in Table 1 can be calculated.

If one merely observes the value of the sire ( $V_1$ ) in the test in terms of his average daily gain, feed efficiency and backfat thickness, all multiplied by the marginal value products and summed, the variance of the density function describing beliefs about  $v_1$  falls from 13.61 to 8.32. The expected value of this information for each offspring resulting from this sire choice is determined from (6) to be \$.46. That is, the expected returns from the pig increases by this amount because the producer can expect to make a better decision if he observes  $V_1$  before making the sire choice. If the boar is expected to produce 40 litters of eight pigs each, the total value of the information is then \$147. (The posterior expected genetic value of sire 1 himself is then  $.39V$ , as noted in column 1 of Table 2). To obtain a net value of information, one must subtract from the \$147 the cost of conducting the test, which in North Carolina is about \$90 per boar, perhaps \$50 or so more than the cost of raising the animal at the producer's farm.

The selection index for the sire is more highly correlated with his genetic value ( $\rho = .67$  vs.  $\rho = .62$  in this particular example) than is his phenotypic value. As a result, observation of the index reduces variance more and has a higher total value, \$159. If a full sib of sire 1 is also tested and both selection indexes are observed, the value of the test information increases to \$164. Finally, if it were somehow possible to obtain perfect information about the genetic value of sire 1, this information would be worth \$236. While this amount of information will never be realized, it might be approached with a large number of observations on relatives of the sire his sibs, (his parents, and his offspring). This latter situation is close to the results of the extensive testing of dairy sire progeny.

#### Summary

The contribution of economists to the understanding of selection decisions has not been very extensive. One potential contribution has been examined here, the value of information generated by sire evaluation programs. Specifically, the value of information generated by central boar testing stations has been examined, in the case of a producer choosing between a tested boar and an untested boar. This value is estimated at up to \$164 per boar, less the extra cost of raising the boar under test conditions, which is approximately \$50 at the North Carolina test station. These estimates, however, assume that the producer has no prior information about the genetic value of either sire being considered. If in fact the producer has prior knowledge about the value or performance of some of the relatives of either boar, then the value of the test information will be reduced in proportion to the amount of prior information which is available. The exact relationship between the number of observations on relatives and the value of the index for the sire himself has not been explored, and remains a topic for additional research.

References

- Bereskin, Ben. Evaluating and Using Performance Records of Boars at Central Testing Stations. ARS-NE-82, U.S. Dept. of Agr. May 1977.
- Carlson, Gerald A. "Variability and market indicators of breeding values," In Applications of Economic in Plant and Animal Breeding, Staff Paper No. 98, Department of Economics, Iowa State U., Oct. 1979.
- Englander, A. Steven and Robert E. Evenson. "Stability, Adaptability and Targeting in Crop Breeding Programs," in Applications of Economics in Plant and Animal Breeding, Staff Paper No. 98, Department of Economics, Iowa State U., Oct. 1979.
- Evenson, R. E. and J. C. O'Toole. "Risk and uncertainty as factors in crop improvement research," in Roumasset, et al. (eds), Risk and Uncertainty in Agricultural Development, Agricultural Development Council, N.Y., 1979.
- Falconer, D. S. Introduction to Quantitative Genetics. Ronald Press, N.Y. 1960.
- Hazel, L. N., "The Genetic basis for constructing selection indexes," Genetics 28: 476-490. 1943.
- Henderson, C. R. "Prediction of future records," in Pollak, et al. (eds.) Proceedings of the International Conference on Quantitative Genetics, Iowa State University Press, Ames, 1977.
- Kempthorne, Oscar. An Introduction to Genetic Statistics. Iowa State University Press, Ames, Iowa. 1963.
- Kempthorne, O. and A. W. Nordskog. "Restricted selection indexes" Biometrics 15:10-19, 1959.
- Ladd, George W. and Graig Gibson, "Microeconomics of technical change: What's a better animal worth?" Am. Jl. Agr. Econ. 60 (May 1978): 236-240.
- Merrell, R. A., C. R. Shumway, J. O. Sanders and T. D. Tanksley, Jr. "Economics of swine crossbreeding systems," So. Jl. Agr. Econ. 11(1979): 89-96 Dec. 1979.
- Melton, Bryan. "Economics of beef cow culling and replacement decision under genetic progress," (Unpublished paper, 1979).
- Melton, Bryan and George Ladd. "Economic value of genetic shifts in the production function," in Applications of Economics in Plant and Animal Breeding, Staff Paper No. 98, Department of Economics, Iowa State U., Oct. 1979.
- Neeley, Doug, Wayne McProud and John Yohe. "Diversity by breeding: genetic variability in the farmer's field," in Roumasset, et al (eds.) Risk and Uncertainty in Agricultural Development, Agricultural Development Council, N. Y., 1979.

Pesek, J. and R. J. Baker. "Application of index selection to the improvement of self-pollinated species," Can. J. Plant Sci. 50:257.276, May 1970.

Raiffa, Howard and Robert Schlaifer. Applied Statistical Decision Theory, MIT Press, Cambridge, 1961.

Smith, H. Fairfield. "A discriminant function for plant selection, Ann. Eugen. (Lond.) 7:240-250. 1936.

Winkler, Robert L. Introduction to Bayesian Inference and Decision. Holt, Rinehart and Winston, New York, 1972.



### Abstract

Contributions by economists to the understanding of the genetic selection choice are reviewed. Concepts for determining the expected value of sire test information are then developed and applied to central boar testing stations. For a producer choosing between a tested and an untested boar, the test information is estimated to have a value of about \$160, less testing costs.

### Footnotes

<sup>1</sup> $y_{ij}$  can be any measured trait or index of traits, measured in physical units or money value.

<sup>2</sup>As used in this paper, "prospect" refers to a choice alternative that has a probability distribution of outcomes.

Table 1. Density function parameters and the value of information for different amounts of information about the genetic value of boars.

<u>Information available</u>	<u>Subjective pdf for unknown <math>v_1</math></u>		<u>Value of information</u>	
	<u>mean</u>	<u>variance</u>	<u>per offspring</u>	<u>total</u>
None	0	13.61	0	0
Observe sire's phenotypic value, $V_1$	$.39 V_1$	8.32	\$.460	\$147
Observe sire's selection index, $I_1$	$I_1$	7.42	\$.498	\$159
Observe selection index for sire and sib	$.95 + .05 I_s$	7.03	\$.513	\$164
Perfect information	$v_1$	0	\$.738	\$236

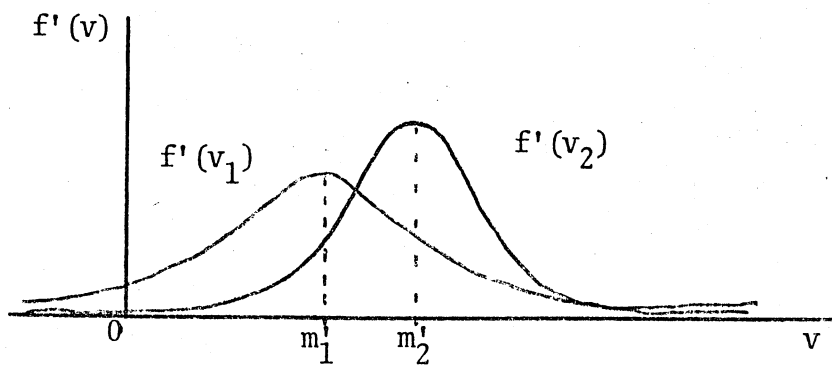


Figure 1. Probability density functions describing prior beliefs about the genetic value of two sires.