# Predicting Animals in Feedlot That Produce Discounted Carcasses

## A.S. Leaflet R2001

M. Ibarburu and J.D. Lawrence,
Iowa State University

### Introduction

U.S. beef producers are striving to improve the quality of their product. Most packers send economic signals to producers by paying quality grade and yield grade premiums based on the USDA standards in an effort to reward the higher quality beef producers and receive better quality animals in their slaughter plants. But the premiums and discounts for quality and yield grade vary between plants and grids. Therefore, identifying the characteristics of the lot before deciding which plant or grid system to sell their cattle is of special relevance for the producers.

Quality, yield, or weight discounts are less common, but large relative to the premiums and have a greater influence on the average prices received for cattle. Discounts are typically paid on carcasses with quality grade Select or lower, yield grade 4 or 5, and weights above 950 pounds or below 550 pounds. We will refer to these as *Discounted Carcasses* in this paper. While economically important, it is very difficult for producers to identify these animals before they are slaughtered. The focus of this analysis is to evaluate the ability to predict at different times in the feedlot the probability of producing a *Discounted Carcass* of each animal. The predictions are based on information observable at that point in time and we will consider two events: arrival and just before slaughter.

Predicting the animal carcass traits before slaughter permit the producers to sort their cattle into the grid with premiums and discounts that will favor each animal, i.e., higher quality cattle on grids with higher premiums and outlier cattle to grids with smaller discounts. This procedure could also be used to decide when to sell the animals by comparing marginal revenue to marginal cost of additional feeding. For example, consider a case where the producer estimates that his animals will not grade well if they are sold right now, but there is high chance of improving their grade by keeping them for four weeks. They could analyze if the increment in costs is paid by the lower quality discounts and the increase in pounds sold per animal.

Knowing the carcass performance of the animals in an earlier moment in the feedlot is much more useful, because this gives also the possibility of making management decisions according to the expectations they have on the animal's performance. Moreover, there are some management practices that could be made in order to decrease the probability of an animal to get a high discount at slaughter. They could also try to get rid of animals that are more likely to produce a *Discounted Carcass*.

Date: November 15, 2004.
The authors gratefully acknowledge Tri-County Steer Carcass Futurity and Darrell Busby for the use of their data

### Data

The data used for the analysis comes from 14,735 heifers and steers fed in 12 different Iowa feedlots and sold between April 2002 and June 2004. Of these animals, 28% qualified for a *Discounted Carcass*. For each observation (animal) there are variables measured at five different moments: Birth, Delivery (the time the animal enters in the feedlot), Start (just after few days of warm-up in the feedlot), Re-Implant (when they receive a second implant) and Slaughter (when they are sorted for slaughter). Feedlot performance (average daily gain, health treatment and estimated feed efficiency) and carcass data were also collected.

### Model

The dependent variable is binary; the carcass either receives a discount or it does not. Logit regression estimates the change in the probability of producing a *Discounted Carcass* in response to changes in the values of variables measured. Logit regressions were run to determine the variation in the probability of an animal to qualifying for *a 'Discounted Carcass'* as a function of the variability on the characteristics measured at each moment and on the variables that still could be changed by in-feedlot management.

The first step in determining which of the variables include in the regression is identifying:

- which variables values are known at each data collection moment
- which variables could be modified at each moment that could change the probability of getting a higher quality animal
- which variables could be predicted with some accuracy at each moment
- and which combination of variables will lead to multicollinearity[1] problems that would affect the results of the regression

[1] - Multicollinearity is a statistical problem that arises when at least one of the variables used is a linear, or near-linear, relationship of other variables used. In this case is very difficult to identify the effects of these variables that are so closely related.

For each of the regressions an estimate of the change in the probability of getting a *Discounted Carcass* as a function of one unit change of the independent variable (dP/dX) was calculated. The p-value[2], and the potential effect of each of the variables were also estimated.

The criteria used to determine the accuracy of the test included the goodness of fit in the entire set and the correctly predicted percentage on the two subsets of values of the dependent variable giving more weight to the percentage of predicted *Discounted Carcass* that actually were *Discounted Carcass*. Trying to not affect the percentage of predicted *non-Discounted Carcass* that actually were *non-Discounted Carcass*.

**Results**

The results of the regression at slaughter and at start moment for *Discounted Carcass* are shown in the appendix 1. The prediction power of the logit model was not very good with this data set. The percentage of correct predictions of the model with the variables that are known just before slaughter was 76%. However if the model were used to predict *Discounted Carcass* in this dataset, the probability of an animal of getting a discount given that the model predicted it will get the discount would be only 61%. The probability of an animal of not getting a discount given that the model predicted it will not get the discount would be 78%. Because the model is not linear, the effects of each variable are calculated as deviations from the mean, keeping all the other characteristics equal to the average, so that the magnitude of their effects could not be added linearly.

**Table1: Percentage of correct predictions**

| | at Slaughter | at Start |
|---|---|---|
| %Total | 75.54 | 75.34 |
| % of predicted *Discounted* that actually were *Discounted* | 61.47 | 60.98 |
| % of predicted *non-Discounted* that actually were *non-Discounted* | 77.80 | 77.53 |

There are some results worth noting. The two variables that affect the most the chances of producing a *Discounted Carcass* are the season that animals entered in the feedlot and the genetic type. Within the season that animals entered in the feedlot, spring placement was the variable that increased the most the chances of producing a *Discounted Carcass*, followed by winter placements. Animals that entered in the feedlot in summer had a little more chances of producing a *Discounted Carcass* than animals that entered in the feedlot in fall. Animals that were at least 25% Indicus breeds and animals that were at least 75% Continental breeds had more chances of producing a *Discounted Carcass*. Animals that were at least 75% British breeds had less chances of producing a *Discounted Carcass*.

Following in importance are hide color, sex, health treatments. Black animals had less chances of producing a *Discounted Carcass* than non-black cattle. Steers had more chances of producing a *Discounted Carcass* than heifers. Animals that belonged to *groups* that received more preventive health treatments had less chances of producing a *Discounted Carcass*. On the other hand, less healthy animals that received more *individual* health treatments had more chances of producing a *Discounted Carcass*.

Bigger animals, those animals with a higher frame score at Start and/or animals with a higher weight/age at start had a higher chance of getting a *Discounted Carcass* than smaller ones. Animals that belonged to groups that gained faster during the test period had less chances of producing a *Discounted Carcass*. And animals with higher average daily gain than their group average had less chances of producing a *Discounted Carcass*. Cattle with a higher disposition score were more likely than calmer ones to get a *Discounted Carcass*. After correcting for other variables effects there were some statistical differences between feedlots in the probability of getting quality, yield or weight discounts.

A similar regression was run with the variables that could be observed at or near delivery to the feedlot (Start). Although the accuracy of the model dropped because producers have less information, this decrease was very little, and effects of the variables were similar to just before slaughter. Thus, some conclusions could be made at arrival about the direction and magnitude of the effects of variables on the probability of getting a *Discounted Carcass*. This would help producers in choosing the management of the pen that lead to decrease the percentage of *Discounted Carcasses*.

The goodness of fit of the model was calculated on a lot base to determine the variation in the percentage of correct predictions between groups.

There was little variation between groups in:

- The percentage of correct predictions over all the

2 - Each p value represents twice the probability that the true value of the effect has any value with sign opposite to that of the observed value.

observations.

- The percentage of predicted *non-Discounted Carcass* that actually were *non-Discounted Carcass*

However, there was large variation between groups in the percentage of predicted *Discounted Carcass* that actually were *Discounted Carcass*.

The goodness of fit was also calculated in a Feedlot base to see if there is much variation in the percentage of correct predictions between groups of animals. The variation between feedlots was very little and a lot less variation than the variation between groups in the percentage of correct predictions. So that the differences in the percentage of correct predictions could not be explained by differences between feedlots in the accuracy of measuring animal characteristics.

*The cost of making prediction errors*

The value of such a prediction model is whether it can increase producer profitability. That is, how much money is lost by not sorting the animals to the most profitable grid? Suppose there are only two options where to sell the animals. The first option is a grid that pays higher premiums for animals that are quality grade choice or better but it has higher discounts for low quality, high yield grades, or heavy or light animals. The second choice is a grid that pays lower premiums but also has lower discounts (for the example we can use the extreme case of a flat price, the base price for the market).

If the producer identifies that the animal's probability of getting a *Discounted Carcass* is high, he will choose to sell to the second grid (a flat price), and if he determines that there are low probabilities of getting a *Discounted Carcass* he would choose to sell on the first grid. Therefore there are 2 different possible errors:

- Estimating that the animal has low probability of getting a discount when its probability of getting a discount is high. Underestimating the potential discounts. Type I error.

- Estimating that the animal has high probability of getting a discount when its probability of getting a discount is low. Missing out on premiums. Type II error.

For the case of a type I error the cost of not sorting correctly for each animal *(i)* is:

$$CostI(i)= Grid1Prem(i)-Grid2Prem(i)$$

where *Grid1Prem(i)* is the premium the animal could get if it sold to the grid 1 (the one with higher premiums and discounts). And *Grid2Prem(i)* is the premium the animal could get if it sold to the grid 2 (the one with lower premiums and discounts). For the case of a type II error the cost of not sorting correctly for each animal *(i)* is:

$$CostII(i)= Grid2Prem(i)-Grid1Prem(i)$$

The total Type I error cost is the sum of Type I error cost for each animal is:

$$TotalCostI = \sum_{i=1}^{n} CostI(i) = \sum_{i=1}^{n} Grid1Prem(i) - Grid2Prem(i)$$

The total Type II error cost is the sum of Type II error cost for each animal is:

$$TotalCostII = \sum_{i=1}^{n} CostII(i) = \sum_{i=1}^{n} Grid2Prem(i) - Grid1Prem(i)$$

And the total cost of not being able to sort the animals correctly is the sum of the total Type I error cost and total Type II error cost.

*TotalErrorsCost= TotalCostI + TotalCostII*

The average cost ($/head) of predicting that the animals don't qualify for a discount given that they do qualify is *TotalCostI* divided by the number of animals mistakenly predicted as non-Discounted. The average cost of predicting that the animals don't qualify for a discount given that they do qualify was $59.10/head, and it is equal to $8.25 cwt for those animals mistakenly predicted as non-Discounted. The average cost ($/head) of predicting that the animals do qualify for a discount given that they don't qualify is *TotalCostII* divided by the number of animals mistakenly predicted as Discounted. The average cost of predicting that the animals qualify for a discount given that they don't qualify was $21.10/head, and it is equal to $2.85/cwt. for those animals mistakenly predicted as Discounted. Considering the entire dataset as one lot, the average cost of the errors (Type I errors + Type II errors) in the entire set was $12.43/animal, or $1.73 /cwt.

A typical case is selling all the animals to the same grid. Therefore a study of how much money is lost if all the animals in the dataset were sold to grid 1 or grid 2 was done following the same procedure. If all the animals in the entire dataset were sold under grid 2 (a flat price) regardless of predicted Discounted, the average cost of the errors in the entire set would be $15.08/animal, or $2.09/cwt. Similarly, if all the animals in the entire dataset were sold under grid 1 regardless of predicted Discounted, the average cost of the errors in the entire set would be $15.42/animal, or $2.14/cwt.

The cost of the low percentage of correct prediction from using the model has only $0.36/cwt (or $2.60/head) advantage with respect to selling all the animals in a flat price and $0.41/cwt (or $2.95/head) of advantage with respect to selling all the animals in the grid if it were possible to sort the animals individually to the different markets.

## Conclusions

Although the prediction power of the model is low, little of it is lost by measuring the characteristics just after the animals entered in the feedlot instead of just before slaughter, so that there is some room for management practices that could influence the outcome during the feeding period. For example, cattle that have a lower probability of grading Choice could be sold at a lighter weight rather than feeding to an marbling endpoint that is not likely to occur.

Based on the prediction equation the probability of producing a *Discounted Carcass* is reduced if the producer chooses black, healthy, tame, smaller, British (with no more than 25% of indicus genes) heifers placed in fall and winter. The odds of avoiding discounts are improved with good average daily gain and if the cattle receive better health treatment.

**Appendix 1: Results of the Logit Regression at Slaughter and at Start**
**Dependent Variable: Discounted Carcass**

| **Moment** | **Slaughter** | **Slaughter** | **Slaughter** | **Start** | **Start** | **Start** |
|---|---|---|---|---|---|---|
| **Variable** | **dP/dX** | **P_value** | **PotEffect** | **dP/dX** | **P_value** | **PotEffect** |
| Intercept | -0.063 | 0.615 | 0.000 | -0.358 | 0.000 | 0.000 |
| BlackHide (1 if balck, 0 otherwise) | -0.085 | 0.000 | -0.085 | -0.083 | 0.000 | -0.083 |
| Male (1 if male, 0 if female) | 0.089 | 0.000 | 0.089 | 0.043 | 0.001 | 0.043 |
| More than 75% British | -0.081 | 0.000 | -0.081 | -0.074 | 0.000 | -0.074 |
| More than 75% Continental | 0.051 | 0.004 | 0.051 | 0.051 | 0.002 | 0.051 |
| More than 25% Indicus | 0.123 | 0.000 | 0.123 | 0.129 | 0.000 | 0.129 |
| Delivered in spring | 0.257 | 0.000 | 0.257 | 0.235 | 0.000 | 0.235 |
| Delivered in summer | 0.035 | 0.016 | 0.035 | 0.027 | 0.035 | 0.027 |
| Delivered in winter | 0.090 | 0.012 | 0.090 | 0.126 | 0.000 | 0.126 |
| Feedlot 1 | 0.042 | 0.118 | 0.042 | 0.018 | 0.450 | 0.018 |
| Feedlot 2 | -0.011 | 0.599 | -0.011 | 0.008 | 0.676 | 0.008 |
| Feedlot 3 | 0.034 | 0.110 | 0.034 | 0.012 | 0.501 | 0.012 |
| Feedlot 4 | 0.013 | 0.564 | 0.013 | 0.055 | 0.003 | 0.055 |
| Feedlot 5 | 0.074 | 0.001 | 0.074 | 0.075 | 0.000 | 0.075 |
| Feedlot 6 | 0.036 | 0.116 | 0.036 | 0.042 | 0.039 | 0.042 |
| Feedlot 7 | -0.094 | 0.028 | -0.094 | -0.090 | 0.023 | -0.090 |
| Feedlot 8 | -0.079 | 0.072 | -0.079 | -0.110 | 0.006 | -0.110 |
| Feedlot 9 | 0.055 | 0.018 | 0.055 | 0.038 | 0.069 | 0.038 |
| Number of individual health treatments | 0.040 | 0.000 | 0.081 | 0.041 | 0.000 | 0.081 |
| Group health treatments | -0.008 | 0.001 | -0.082 | | | |
| Average Disposition Score (from 1 if tame to 6 if wild)* | 0.030 | 0.000 | 0.053 | 0.018 | 0.003 | 0.036 |
| Frame Score at delivery | 0.000 | 0.003 | 0.060 | 0.000 | 0.017 | 0.044 |
| Muscle Score at delivery | 0.000 | 0.164 | 0.018 | 0.000 | 0.223 | 0.014 |
| Body Condition Score at delivery | -0.013 | 0.292 | -0.020 | -0.007 | 0.552 | -0.010 |
| Weight/Age at Start | 0.054 | 0.000 | 0.081 | 0.052 | 0.000 | 0.078 |
| Home weight at delivery | 0.000 | 0.151 | 0.040 | 0.000 | 0.506 | 0.014 |
| (Self ADG between Start and Slaughter/ Group average ADG between Start and Slaughter)*100 | -0.001 | 0.013 | -0.045 | | | |
| Group average ADG between Start and Slaughter | -0.075 | 0.001 | -0.076 | | | |
| Lbs of feed/ Lbs of weight gained | -0.003 | 0.735 | -0.008 | | | |

* Is Disposition Score at Start for Start moment instead of the average because the average is not available at Start