

# Exploring the Existence of Grader Bias in Beef Grading

## ABSTRACT

The United States Department of Agriculture (USDA) beef grading system plays an important role in marketing and promoting beef. USDA graders inspect beef carcasses and determine a quality grade within a few seconds. Although the graders are well-trained, the nature of this grading process may lead to grading errors. Significant differences in the USDA graders' 'called' and 'camera-graded' quality grades were observed as well as variations in quality grades across seasons and years. Under grid pricing, producers gained financially from grades called by USDA graders rather than grades measured by cameras.

*Key words:* Beef grading system, Grader bias, Quality grade, Marbling score.

**JEL Classifications:** Q13, Q18

## **I. Introduction**

The United States Department of Agriculture (USDA) beef carcass grading standards are comprised of USDA quality and yield grades, which are designed to assess the eating quality and the amount of lean edible meat from a carcass, respectively. Producers use these grades to roughly predict the market value of cattle they sell to packers and have a financial incentive to produce the best tasting and high-quality beef under the USDA grading system. Consumers make informed purchasing decisions using USDA quality grades and labels. In short, the system simplifies the marketing process and makes communication among producers, packers, and consumers easier (Field 2007).

The integrity of the beef grading system is assured by accurate and precise grading. In reality, however, graders employed by the USDA determine carcass grades by a brief visual inspection that takes only a few seconds. Although USDA graders are well-trained and independent of both producers and packers, the nature of the grading process could lead to grading errors. These errors could diminish the incentive to produce a higher-quality product (Chalfant *et al.* 2003). When quality grades called by USDA graders are lower than actual quality grades, cattle producers take a loss on transactions with packers. In the case of beef consumers, they have to pay more (or less) than the actual value of beef due to the grading errors. The grading errors in quality grades impede the communication among beef consumers, producers and packers. The influence of grading errors on the efficiency of the market and promotion of beef can be minimized, if the errors are not systematically biased across time and location (Hueth *et al.* 2007). Grading accuracy and consistency, thus, are crucial for improving producers', packers', and consumers' confidence in the efficiency of the beef marketing system.

In 2006, two camera-based grading systems were approved by the USDA in order to improve beef carcass grading accuracy and uniformity within the industry<sup>1</sup>. In August 2014, the USDA Agricultural Marketing Service (USDA-AMS) sought public input for possible revisions to the *U.S. Standards for Grades of Carcass Beef* to help adjust for recent improvements and trends in the raising and feeding of cattle. Although the USDA-AMS has been working on improving the accuracy of beef grading, there are relatively few studies that looked at the presence and sources of grader bias. Mafi *et al.* (2013) documented that cameras/instruments were more accurate and consistent than the USDA graders in assessing marbling score to determine quality grade. They also found that cameras/instruments reduced grader-to-grader and plant-to-plant variations. Hueth *et al.* (2007) used a behavior model and showed the existence of grader bias in assigning yield grade. They defined grading as biased when the distribution of the “true” (the grade that should be assigned according to the USDA standards) yield grade systemically differs from that of the ‘called’ (the grade actually assigned by a USDA grader) yield grade. To measure the divergence of two distributions, they estimated a mean and variance of the true yield index<sup>2</sup> and compared them with the sample mean and standard deviation of the index. They also estimated cutoff values for each yield grade to capture the USDA graders’ behavior.

The current study builds on the previous literature by looking for evidence of the existence and possible sources of grading errors using data from two large-scale Midwest packing plants from 2005 through 2008. The data on quality grade called by USDA graders (‘called’ quality grade) and ‘camera-graded’ quality grade of each carcass were provided along with year, month, and day of the week when cattle were processed.

The specific objectives of this study are threefold. First, we analyzed the difference between ‘called’ and ‘camera-graded’ quality grade. Then, using these given quality grades, we estimated the cutoff points for each quality grade (e.g. Choice or Select) and compared them with the USDA standards cutoff points for each grade. From the analysis, we expected to find possible sources of grading errors. One of the significant errors in ratings is known as "central tendency bias". The existence of central tendency bias may be shown in beef grading if USDA graders do not follow the USDA standards and have a tendency to call grades close to the mean and avoid calling extreme grades. Second, we further investigated the patterns of grading errors by conducting seasonal and annual analyses to extend the existing literature by estimating seasonal and annual cutoff points. Existing research documented seasonal changes in beef carcass characteristics (Gray et al. 2012), the number of cattle marketed, and consumer demand (McCully 2015). The patterns of estimated intervals for quality grades across seasons and years were compared with Choice-Select spread, physical characteristics of beef carcasses, and the number of slaughter cattle processed in order to help identify possible sources of grading errors. Finally, since the USDA intends to more widely utilize the camera grading system in the future, it is worth while to analyze and discuss the impact of potential changes on producers and packers. For this analysis, weekly weighted averages of premiums and discounts for each quality grade were collected from USDA-AMS 5-Area Weekly Direct Slaughter Reports. The premium and discount data, along with ‘called’ and ‘camera-graded’ quality grades, allow the measurement of the financial impact of fully utilizing the camera grading system on cattle producers and packers.

To our knowledge, this is the first study that quantifies the variations in beef quality grading by USDA graders and camera systems across seasons and years. In addition, we address the impact of increased utilization of cameras in grading on cattle producers and packers. These

analyses were possible, because the data used contain a much larger number of observations over the years than those in earlier studies (Hueth *et al.* 2007, Mafi *et al.* 2013).

A few points must be made about the terminology and assumptions used in this research. The term “grader bias” in this paper does not imply deception or dishonesty, but simply that the called quality grade is different from the USDA standards. In this paper, we assume that ‘camera-graded’ quality grade is not identical with the “true” USDA quality grade. Grades determined by cameras can be biased due to the initial settings, sensitivity, accuracy, and errors related to calibration of cameras (Moore *et al.* 2010, Mafi *et al.* 2013). Furthermore, part of our data was collected before the camera grading system was officially approved by the USDA in 2006. Given that there are some errors that can be associated with camera grading, the quality grade measured by the camera is unlikely to be identical with the “true” quality grade. These factors led us to develop a different behavior model from the model developed in Hueth *et al.* (2007).

## **II. Model**

There are eight USDA quality grades: Prime, Choice, Select, Standard, Commercial, Utility, Cutter, and Canner. The factors that are used to determine the quality grade are the degree of marbling and the maturity class, which are classified into 9 and 5 different levels<sup>3</sup>, respectively. The degree of marbling and the maturity class are combined to determine the final quality grade (Hale *et al.*, 2015). When slaughter cattle are processed before 42 months of age, their carcasses are categorized as Prime, Choice, Select, or Standard according to marbling score. If slaughter cattle are processed after 42 months of age, the carcasses are graded as Commercial, Utility, Cutter, or Canner. The USDA graders subjectively determine both maturity and marbling class

based on the descriptions and illustrations provided in the standards and their own practical work experiences.

Results of the 2005 National Beef Quality Audit determined that more than 97% of carcasses in U.S. fed beef plants were classified as A-maturity (9-30 months) (Garcia *et al.* 2008). Hence, in this study we assume that maturity class was A (9-30 months) or B (30-42 months). Given the maturity class, the primary determinant of quality grade will be the marbling score. The analysis of this study includes beef carcasses which are graded as Prime, Choice, Select, and Standard. Given this exclusion, a model that uses marbling score as the determinant of the quality grade is specified as follows.

Let  $MSI_k$  be the Marbling Score Interval for quality grade  $k$ . These intervals allow us to express quality grade in a functional form:

(1) Quality Grade =

$$\{k \mid \text{Marbling Score} \in MSI_k, k = \text{Prime, Choice, Select, Standard} \mid \text{Maturity} \leq 42 \text{ months}\}.$$

Let  $c_i$  be a ‘called’ quality grade,  $m_i$  be a ‘camera-graded’ quality grade, and  $t_i$  be a “true” quality grade for a carcass  $i$ . “True” quality grade is unobserved. Using these definitions, the ‘called’ and “true” quality grade can be expressed as follows:

$$(2) \quad c_i = m_i + u_i, \quad u_i \sim N(0, \sigma_u^2), \quad t_i = m_i + v_i, \quad v_i \sim N(0, \sigma_v^2),$$

where  $u_i$ , and  $v_i$  are error terms for ‘called’ and “true” quality grades, respectively. We assume that error terms are distributed normally with mean zero and standard deviations,  $\sigma_u$  and  $\sigma_v$ . This assumption allows the use of a likelihood function to estimate cutoff points and standard errors.

The USDA Standard Marbling Score Intervals ( $\widehat{MSI}_k$ ) for each quality grade are  $\widehat{MSI}_{\text{Prime}} = [8.0, +\infty)$ ,  $\widehat{MSI}_{\text{Choice}} = [5.0, 8.0)$ ,  $\widehat{MSI}_{\text{Select}} = [4.0, 5.0)$ , and  $\widehat{MSI}_{\text{Standard}} =$

$(-\infty, 4.0)$ . The  $\widetilde{\text{MSI}}_{\text{prime}}$  means that USDA graders should call Prime when an observed marbling score is greater than or equal to 8.0. Other quality grades should be called in a similar way, in that a grade is ‘called’ when marbling score falls within the indicated interval.

Since our data indicates that the ‘called’ quality grade is not identical with the ‘camera-graded’ quality grade, we presume that the USDA graders have their own marbling score intervals, which could be different from those of the USDA standards. Using this premise, the USDA graders Marbling Score Intervals ( $\widetilde{\text{MSI}}_k$ ) are defined by the following implicit cutoff points ( $C_k$ ,  $k = \text{Prime, Choice, Select, and Standard}$ )

$$\begin{aligned}\widetilde{\text{MSI}}_{\text{prime}} &= [C_{\text{prime}}, +\infty), \quad \widetilde{\text{MSI}}_{\text{choice}} = [C_{\text{choice}}, C_{\text{prime}}), \\ \widetilde{\text{MSI}}_{\text{select}} &= [C_{\text{select}}, C_{\text{choice}}), \quad \text{and} \quad \widetilde{\text{MSI}}_{\text{standard}} = (-\infty, C_{\text{select}}).\end{aligned}$$

If these implicit cutoff points are different from those of the USDA standards across time, then we can conclude that grader bias exists.

We assume that the ‘called’ quality grade and the probability of the ‘called’ quality grade being the “true” quality grade are independent. Then the likelihood function can be defined as follows:

$$\begin{aligned}(3) \quad L^i(c_i, m_i \mid \sigma_u, \sigma_v, C_{\text{prime}}, C_{\text{choice}}, C_{\text{select}}) \\ = \mathbf{I}(c_i = \text{standard}) \left\{ \Phi\left(\frac{C_{\text{select}} - m_i}{\sigma_u}\right) \times \Phi\left(\frac{4 - m_i}{\sigma_v}\right) \right\} \\ \times \mathbf{I}(c_i = \text{select}) \left\{ \left[ \Phi\left(\frac{C_{\text{choice}} - m_i}{\sigma_u}\right) - \Phi\left(\frac{C_{\text{select}} - m_i}{\sigma_u}\right) \right] \times \left[ \Phi\left(\frac{5 - m_i}{\sigma_v}\right) - \Phi\left(\frac{4 - m_i}{\sigma_v}\right) \right] \right\} \\ \times \mathbf{I}(c_i = \text{choice}) \left\{ \left[ \Phi\left(\frac{C_{\text{prime}} - m_i}{\sigma_u}\right) - \Phi\left(\frac{C_{\text{choice}} - m_i}{\sigma_u}\right) \right] \times \left[ \Phi\left(\frac{8 - m_i}{\sigma_v}\right) - \Phi\left(\frac{5 - m_i}{\sigma_v}\right) \right] \right\} \\ \times \mathbf{I}(c_i = \text{prime}) \left\{ \left[ 1 - \Phi\left(\frac{C_{\text{prime}} - m_i}{\sigma_u}\right) \right] \times \left[ 1 - \Phi\left(\frac{8 - m_i}{\sigma_v}\right) \right] \right\}\end{aligned}$$

where,  $\mathbf{I}(\cdot)$  is an indicator function, and  $\Phi(\cdot)$  is the cumulative density function of the standard normal distribution. The likelihood function is derived from the assumption that the USDA graders call quality grade to maximize the probability of calling the “true” quality grade by using their own implicit intervals. Since the “true” quality grade is unknown to USDA graders, they call quality grade using visual inspection and their own implicit cutoff points. A log transformation of the likelihood function was used in the maximum likelihood estimation process. The estimated cutoff points provide information about grading behavior of USDA graders in assigning quality grades.

### **III. Data**

The data used in the analysis provides information on ‘called’ and ‘camera-graded’ quality grades of beef carcasses from May 2005 to October 2008. Figure 1 presents the distribution of ‘called’ quality grade for the entire sample ( $n=134,451^4$ ) and shows that 94.4% of beef carcasses were graded Choice or Select. While the ‘called’ quality grade was available for the entire sample, the ‘camera-graded’ quality grade was only available for the subsample of the data ( $n=18,080$ ). Because the values for both ‘called’ and ‘camera-graded’ quality grades are required to estimate the implicit cutoff points, the subsample ( $n=18,080$ ) of the entire data ( $n=134,451$ ) was used in estimating the cutoff points and conducting premium-discount analysis.

In our data, ‘called’ marbling grades were reported as USDA quality grades (Prime, Choice, Select, or Standard) and ‘camera-graded’ marbling scores were reported as numeric values (e.g. 5.0 for Small) for some of the carcasses in our sample and as a degree of marbling (e.g. Small 20) for the remaining carcasses. To make the marbling measurements consistent across carcasses and to estimate the cutoff points, we converted each degree of marbling into a



numeric marbling score. Figure 2 shows the distribution of the numeric ('camera-graded') marbling scores. Each number on the horizontal axis of Figure 2 corresponds to a degree of marbling score.

As shown in Figures 1 and 3, the distributions of 'called' quality grade from the entire sample (n=134,451) and the subsample (n=18,080) used in the analysis were similar. Both distributions show that most carcasses were graded as Choice or Select. The distribution from the entire sample (subsample) shows that the USDA graders graded 67.3% (70.2%) and 27.1% (27.2%) of carcasses as Choice and Select, respectively. The National Summary of Meat Graded Reports (USDA 2015) announced by the USDA-AMS at the beginning of each year showed that most carcasses were graded either Choice or Select (Table 1). Although the distributions for 'called' and 'camera-graded' quality grades differ a bit, in percentage terms, from the national averages reported in Table 1, the shape of both 'called' and 'camera-graded' quality grades are similar to the national summary indicating that the sample data used in the analysis closely represents the national level data.

As shown in Figure 3, 70.2% (27.2%) of carcasses were graded as Choice (Select) by the USDA graders, whereas 51.1% (35.8%) of the carcasses were graded as Choice (Select) by cameras. This indicates that the USDA graders tend to call more Choice and less Select compared to cameras. Also, the two distributions show that the USDA graders were more generous in grading carcasses compared to cameras.

The distributions of 'camera-graded' quality grade given 'called' quality grade are shown in Figure 4. These conditional distributions allow us to analyze the differences between 'called' and 'camera-graded' quality grades. If there were no divergences between these two grades then all the carcasses called as Prime by the USDA graders should be graded as Prime by the cameras.

However, as shown in Figure 4, out of 395 beef carcasses that were graded as Prime by the USDA graders, only 144 carcasses (36.5%) were graded as Prime by the cameras and the remaining 251 carcasses (63.5%) were graded as Choice. Furthermore, from all the carcasses graded as Choice by the USDA graders, 33.4% were graded as Select by the cameras. In the case of Select, 42.6% were graded as Standard by the cameras. These conditional distributions suggest that noticeable differences exist between ‘called’ and ‘camera-graded’ quality grades, except for Standard, and that the cameras generally assigned lower quality grades than the USDA graders.

Figure 5 illustrates the distribution of ‘called’ quality grade given ‘camera-graded’ quality grade. Almost all beef carcasses graded as Choice by the cameras were graded as Choice by the USDA graders. A similar patterns was observed for Prime, where a majority of carcasses graded as Prime by the cameras were also graded as Prime by the USDA graders. However, in the case of all carcasses graded as Select by the cameras, 65.6% were graded as Choice by the USDA graders and 96.1% of all carcasses graded as Standard by the cameras were graded as Select by the USDA graders. The comparison of conditional distributions in Figure 5 indicates that the difference between ‘called’ and ‘camera-graded’ quality grades was smaller when the USDA graders assessed Choice grade carcasses, but this was not the case for the other quality grade carcasses. This smaller divergence for Choice quality grade could be explained by Piazza and Izard (2009)’s findings: the more humans are exposed to the number of objects or sequence, the more likely they accurately repeat the sequence. As shown in Figure 3, 70.2% of carcasses in our observations were graded as Choice by the USDA graders. This could indicate that the USDA graders were more accurate in assessing Choice grade carcasses due to repeated exposures to Choice grade carcasses.

The distributional analyses in this section were not enough to confirm the existence of grader bias caused by the USDA graders since ‘camera-graded’ quality grade can also be different from the “true” USDA standard quality grade due to calibration errors or initial camera settings. The implicit cutoff points for each quality grade, thus, were estimated to further analyze the existence of grader bias and explore possible sources of the bias.

## **IV. Results**

### **1. Subsample Analysis**

Cutoff points for each quality grade were estimated using equation (3) to identify the implicit USDA graders’ interval. The existence of grading errors can be checked by comparing the estimated and USDA Standard cutoff points. As shown in Table 2, the estimated cutoff point for Prime was 8.90, which was greater than the USDA standards cutoff point of 8.00. The estimated interval for Prime,  $[8.90, +\infty)$ , indicates that the USDA graders called Choice when the marbling score was greater than 8.00 and that the USDA graders have higher standards for Prime.

Table 2 also shows that the estimated cutoff point for Choice was 4.50, which was lower than the cutoff point of 5.00 for Choice defined by the USDA standards. This difference between two cutoff points indicates that the USDA graders called Choice instead of Select when marbling score was less than 5.00<sup>5</sup>. The estimated cutoff points also identify the estimated implicit interval for Choice as  $[4.50, 8.90)$ . This interval is much wider than the one from the USDA standards for Choice,  $[5.00, 8.00)$ , indicating that the USDA graders had a tendency to call more Choice.

The estimated cutoff point for Select was 3.18. This value is smaller than 4.00, the value from the USDA standards for Select. Using the estimated cutoff points, the estimated intervals for Select and Standard quality grades were identified as  $[3.18, 4.50)$  and  $(-\infty, 3.18)$ , respectively.

These intervals indicate that USDA graders called Select when marbling score was less than the USDA standards cutoff point of 4.00 for Select, again indicating that the USDA graders were generous in grading beef carcasses with less marbling.

Potential sources of grader bias could be identified by comparing the estimated and USDA standards intervals across quality grades. While the estimated intervals for Prime and Standard were narrower than the USDA standards intervals, the estimated intervals for Choice and Select were wider than the USDA standards intervals. This nonconformity can be explained by a central tendency bias. This bias was mostly researched by educationalists. Saal *et al.* (1980) define this bias as a rater's (grader's) property or tendency to restrict a range of scores around a mean and to avoid awarding extreme scores. Existing studies in the area (Engelhard 1994, Myford *et al.* 2009, Leckie *et al.* 2011) found that there is a central tendency to a rater's scoring. Beef grading behavior is very similar to scoring behavior in schools. Both USDA graders and raters, although well trained, are human beings and evaluate subjects based on their subjective observations with given grading standards. These similarities have led us to consider the central tendency bias as the potential source of grader bias in beef carcass grading.

The narrow estimated intervals for Prime and Standard quality grades mean that the USDA graders tend to avoid calling extreme grades. The wider intervals for Choice and Select indicate that graders preferred to call the quality grade around the mean marbling score of 5.10 for our sample (Table 3). These results indicate that USDA graders tend to call central grades and avoid calling extreme grades, i.e., Prime and Standard. These grading patterns are evidence of the central tendency bias in beef carcass grading.

A reason for the central tendency bias in beef carcass grading may be found in the economic impact of quality grade to producers and packers. Producers can receive a premium or

discount based on the quality grade of a beef carcass, if slaughter cattle were sold or priced based on their eventual grade. As shown in Table 4, Choice grade carcasses do not receive any premium or discount when priced based on a grid pricing system<sup>6</sup>. Under grid pricing, calling Choice is a way to make a smaller impact on the financial rewards/losses of producers and packers. Moreover, calling central grades, especially Choice, may be a way to avoid complaints from producers and packers. If USDA graders call extreme grades (Prime and Standard) more frequently, the probability of receiving complaints and re-grading requests could be higher. Because USDA graders are independent from producers and packers, they may have no intention of affecting the profit margin of both producers and packers through their grading. According to Hueth et al. (2007), packing plants hire a “tagger” who identifies grader miscalls and requests re-grading. With the presence of a tagger, the USDA graders could become more generous in grading and have a tendency to call the central grades (Choice and Select) more often to avoid re-grading requests.

## **2. Seasonal and Annual Analyses**

Dynamics in beef carcass grading were analyzed by estimating seasonal and annual cutoff points and comparing them with the USDA standards. The results reported in Table 2 show that the estimated cutoff points for Prime varied significantly by season. The estimated cutoff point for Prime in the summer was 7.97, which was close to the USDA standards cutoff point for Prime. With respect to other seasons, the estimated cutoff points for Prime were noticeably higher than the USDA standards, indicating that during those seasons the USDA graders were much stricter in grading high quality beef carcasses compared to summer. The estimated interval for Choice in the summer, [4.32, 7.97), was narrower than those for other seasons, however the cutoff point of

4.32 was smaller compared to the USDA standards of 5.00, indicating that USDA graders were more generous and graded Select carcasses as Choice. The estimated interval for Select was the widest in fall and narrowest in winter. These seasonal differences in the estimated intervals and cutoff points can be caused by many factors such as seasonality in the Choice-Select spread, the volume of carcasses processed, the physical characteristics of beef carcass, and many other factors.

The Choice-Select spread, which is defined as the difference between the Choice and Select wholesale boxed-beef values, is used as an indicator of demand for high quality beef in the industry (McCully 2015). For example, when the Choice-Select spread reaches a high level ( $> \$8/\text{cwt}$ ), the industry assumes strong demand for high marbled beef, such as Choice, and when the spread is low ( $< \$3/\text{cwt}$ ), the industry assumes weak demand for Choice (McCully 2015). In this paper, the Choice-Select spread data were collected from USDA-AMS 5-Area Weekly Direct Slaughter Reports for the period covered in our data and were summarized in Figure 6. As illustrated in Figure 6, the average Choice-Select spread peaked during the cookout month, May, and during the holiday months, November through January, indicating a high demand for Choice beef during these months. The spread decreased significantly after May and the holidays indicating the lower demand for Choice beef. Especially, the spread in summer (June - August) was relatively lower than those for other seasons. It is also true that beef supplies do tend to increase in summer. When comparing the patterns of the Choice-Select spread with those of the estimated intervals, we can argue that the low Choice-Select spread (low demand for Choice beef) influenced the narrow interval for Choice (calling less Choice) in summer compared to other seasons. The similarity in two patterns suggests that the demand for specific quality grade beef possibly influences the grading behavior.

As reported in Table 3, the majority of cattle in our sample were processed in spring and summer seasons, 43.1% and 45.1%, respectively, and in 2006 and 2007, indicating that for our sample the volume of slaughter cattle fluctuated greatly by season and year. The seasonality in number of slaughter cattle processed can be explained by the fact that the majority of calves are born in the spring months, weaned in fall, and either backgrounded or placed on feed during October and November. The majority of these cattle are marketed and slaughtered during the summer months or later of the following year. These trends in seasons and years from our findings are consistent with the national averages reported in the monthly *Cattle on Feed* report provided by USDA NASS (NASS, USDA 2015) and summarized in Figure 7 for the time period covered in our data. Marketing of cattle tends to be highest in May through August of every year, which covers the last month of spring and all the months of summer. During the spring and summer months, the busy time of the year, USDA graders were more generous and were more likely to call Choice when the actual quality grade was Select. The estimated cutoff points for Choice in fall and winter were 4.56 and 4.81, respectively, and were closer to 5.00, the USDA standards cutoff for Choice, compared to spring and summer. Seasonal variations in the number of slaughter cattle processed at the packing plants can influence the grades called by USDA graders. Graders had a tendency to call more central grades during the busy seasons of the year which may be associated with taking shorter breaks, working longer hours, and/or using more temporary help.

We observed seasonal and yearly variations in carcass characteristics such as marbling score (Table 3), ribeye area, fat thickness, and hot carcass weight (Figure 8) in our data. High grain and oilseed prices between 2006 and 2008 increased the cost of production for beef cattle producers. Beef cattle producers can respond to high feed ingredient costs by adjusting the types

and amount of ingredients in feed rations, and the length of time spent in the feedlot, which in return can affect the quality grade of slaughter cattle. Other factors, such as age at slaughter and the type of breed, can explain variations in carcass characteristics by season and over time (Gray *et al.* 2012). These seasonal variations in carcass characteristics can influence the graders' judgement and serve as one of the potential sources of grader bias in beef carcass grading.

Figure 6 shows that average Choice-Select spread in 2008 (5.31\$/cwt) was less than those in 2005, 2006, and 2007 (9.33\$/cwt, 13.81\$/cwt, and 9.73\$/cwt, respectively) indicating lower demand for higher quality beef in 2008, the period that overlaps with global financial crisis. During the economic recession, the demand for Choice beef declined as shown by the decrease in the Choice-Select spread (Figure 6). Changes in demand may influence USDA graders and lead to calling less Choice. The entire sample, thus, is separated into two subsamples (before and during the crisis) to analyze the potential impact of the economic recession on grading behavior.

As shown in Table 2, the interval for Choice during the crisis, [5.04, 8.47], was significantly narrower than the one before the crisis, [4.20, 8.28). Both the lower and upper cutoff points during the crisis were significantly higher than the cutoff points before the crisis. The estimation results also show that the estimated cutoff points for each quality grade, 3.93, 5.04, and 8.47, were close to the USDA standards, 4.00, 5.00, and 8.00, after the crisis broke out. These results indicate that the USDA graders were more precise and careful when grading. Their possible awareness of higher demand for cheaper beef cuts during the recession might have influenced their grading. It is possible that USDA graders were trying to avoid grading errors to prevent giving financial advantages/disadvantages to either producers or packers.

### **3. Premiums and Discounts Analysis<sup>7</sup>**



The trend in premiums and discounts for each quality grade during our sample period is illustrated in Figure 9. The data were collected from USDA-AMS 5-Area Weekly Direct Slaughter Reports for the period covered in our data. The premiums and discounts for Choice are zero, since it serves as a base quality grade from which premiums and discounts are added/subtracted for Prime, Select, and Standard<sup>8</sup>. In 2008, premiums for Prime decreased, and discounts for Select and Standard also decreased (Figure 9). This means that the premium-discount spread between Prime and Select as well as Prime and Standard became narrower. Because the change in premiums and discounts relates to consumer preferences and because packers send signals to producers about the quality of beef demanded through premiums and discounts<sup>9</sup>, the narrow spread in 2008 implies that consumers preferred less expensive beef instead of high quality beef as their income declined.

Our quality grade data includes weights of each beef carcass. Using weekly weighted averages of premiums and discounts provided by the USDA and ‘camera-graded’ and ‘called’ quality grades along with the weight of each carcass from our dataset, we were able to calculate the premiums and discounts of ‘camera-graded’ and ‘called’ quality grade for each carcass. The difference in ‘camera-graded’ and ‘called’ quality grade premiums and discounts indicates how much producers or packers would have financially gained or lost if USDA graders were to be replaced by cameras during our sample period. By measuring this difference, we can forecast how replacing human graders with cameras may influence the future earnings of producers and packers. For example, if ‘camera-graded’ quality grade discounts were greater than ‘called’ quality grade discounts, the difference of the discounts provides the amount of money that producers or packers may lose if USDA graders were replaced by cameras. Although the amount of money that producers or packers could have lost is not identical with what they will lose in the

future, we could roughly estimate the financial impact of the replacement on producers and packers. However, in the analysis we do not account for the dynamics of the market. If the volume of beef is changed by the full adoption of cameras, then the premium or discount of beef carcass may be altered. Because we do not account for this change in the analysis, the findings of this section need to be interpreted with caution.

Cattle are marketed mainly by three pricing methods: 1) live weight pricing, 2) dressed weight pricing, or 3) grid pricing (Schroeder *et al.*, 1998). When slaughter cattle are priced on a live or dressed weight basis, packers and producers negotiate prices based on the expected value of the cattle. The expected value is determined by expected quality and yield grade, weight premiums and discounts, by-products, slaughter costs (sellers generally pay transportation on dressed cattle sales), and the packer's profit. Since packers pay before cattle are graded by the USDA graders, packers can have financial gains if beef carcasses are graded at a higher quality grade than their expected value, and vice versa. Hence, under live and dressed weight pricing methods, only packers' earnings are influenced by the 'called' quality grade. When slaughter cattle are marketed based on yield and quality grade, i.e. grid pricing, price is based on the 'called' grade of each animal. Under grid pricing, the quality grade and yield grade influence producers' earning unlike live and dressed weight pricing. Therefore, under grid pricing, producers will lose financially when USDA graders call a lower quality grade than the true grade. In the case of the live and dressed weight pricing, packers will lose when USDA graders call a lower quality grade than the expected value for which they paid. Information on pricing method used for each carcass and expected value of cattle was unavailable to us, so we were not able to calculate the amount of money that each producer and each packer would gain or lose under different pricing methods. We, however, were able to calculate the combined financial

gains/losses of producers and packers after replacing human graders with cameras by calculating the difference between ‘called’ and ‘camera-graded’ quality grade premiums and discounts. The expected value of the cattle did not affect the calculated difference, since the expected values of ‘camera-graded’ and ‘called’ quality grades were identical for each cattle carcass and cancel out when the difference is calculated.

The differences reported in Table 5 were calculated by subtracting the sum of ‘called’ quality grade premiums and discounts from the sum of ‘camera-graded’ quality grade premiums and discounts. The average difference in value of  $-\$3.00/\text{cwt}$  is the amount of money producers and packers would have jointly lost on average per hundredweight of carcass if a camera grading system would have been used instead of USDA graders during our sample period.

Traditionally, live weight pricing was very popular. However, over the past two decades dressed weight pricing and grid pricing methods became increasingly popular. According to the USDA report, over 50% of cattle sold during the period covered in our data were sold on grid pricing. Specifically, 56.3% (in 2005), 53.3% (in 2006), 57.2% (in 2007), and 62.3% (in 2008) of cattle were sold based on grid pricing (USDA 2014). To calculate and interpret the change in the earnings of producers and packers, respectively, we assume that the proportion of the grid pricing in our sample is similar to the national level. Hence, the combined difference of  $-\$53,981$ , as reported in Table 5, can be separated into producers’ and packers’ differences,  $-\$29,825$  and  $-\$24,156$ , respectively. The difference of  $-\$29,825$  for producers implies that producers will lose financially when the number of USDA graders is reduced through increased use of the camera grading system under grid pricing. The difference for packers ( $-\$24,156$ ) implies that under dressed weight pricing, packers gain from grades called by USDA graders instead of camera grades. Here we are only considering transactions between producers and packers. In reality, the

process is more complex and depends on how packers profitably market high and low quality carcasses in the wholesale market. The discount for low quality carcass can be high if packers have difficulty profitably marketing low quality beef. At the same time, packers may pay high premiums for high quality carcasses if there is a demand for high quality beef. Our results from Table 5 imply that on average packers were penalized more for low quality carcasses. The discounts for 'camera-graded' and 'called' grades for Standard, on average, were \$15.6/cwt and \$15.7/cwt, respectively, while the premiums for 'camera-graded' and 'called' grades for Prime were \$4.2/cwt and \$11.7/cwt, respectively. Both 'called' and 'camera-graded' quality grade discounts for Standard were very similar to the national averages reported in Table 1. Whereas, 'called' and 'camera-graded' grade premiums for Prime were well below the national average.

Our results in this section are consistent with our findings in previous sections that the USDA graders were more generous in grading than the cameras. Since the USDA is working on reducing human graders, this might imply that producers and packers will lose financially if more cameras are used in grading.

Table 5 also shows that the difference in premiums and discounts has noticeably decreased after 2007. This result is consistent with our findings of annual data analysis. We found that, after the financial crisis started, USDA graders became much more precise and stricter in grading, and, at the same time, as illustrated in Figure 9, both premiums and discounts decreased. These changes could be one of the reasons why the difference decreased after 2007.

## **V. Conclusion**

The role of USDA graders is crucial in cattle and beef markets. Although USDA graders are well trained, a subjective determination of quality grades could cause grading errors. This study uses a

unique data set and provides a comprehensive analysis of existence and possible sources of grader bias in assigning quality grades to beef carcasses and adds to the existing body of research that has addressed this issue.

The analyses in this paper used data from two large-scale Midwest packing plants. The data includes ‘called’ and ‘camera-graded’ quality grades for each beef carcass from May 2005 to October 2008. We also used the USDA reported weekly weighted averages of premiums and discounts for each quality grade along with ‘called’ and ‘camera-graded’ quality grades to estimate the financial impact of the reduced use of USDA graders and adoption of a camera grading system on beef cattle producers and packers.

The results of the interval estimation analysis indicate that USDA graders’ called grades were noticeably different from those measured by the camera grading system. The analyses suggest that seasonality in Choice-Select spread, consumer demand, number of carcass processed, and carcass characteristics can influence grading behavior of human graders. We also observed a central tendency bias in the grading behavior of USDA graders.

Our results have important implications for the current debate surrounding the widespread adoption of camera grading systems at packing plants. After verifying the existence of systematic grader bias across time, we investigated the possible impact of using camera grading methods instead of USDA graders on the economic gains/losses of producers and packers. When grading errors are systematically biased, the reduction of USDA graders’ utilization can influence the financial rewards of producers and packers. The results of the premiums and discounts analysis support the findings of the interval estimation analysis and show that combined earnings of producers and packers will decline when more camera grading is utilized

in the beef grading system. Under grid pricing, producers will lose financially if camera grading is used instead of the USDA graders.

There are a number of limitations to the present work. First, in this paper we used data from 2005 through 2008. Conducting the analysis using newer data that was collected using more recent computerized technology in grading beef carcasses would provide more up to date information on beef grading and presence of grading errors. Second, we focused on investigating the financial impact of the USDA graders' replacement with cameras on packers and producers. The calculations were done without accounting for market response to changes in relative shares of different grades, hence the results provided in this paper need to be interpreted with caution. Third, it is also important to examine the welfare impact of the policy change on consumers. According to our results, we expect that beef prices will change when USDA graders are replaced by cameras. This price change will influence consumers' welfare in one way or another. Due to the lack of price information, it was not feasible to inspect this impact in this study. Hence, future research that focuses on using more recent data and more nationally representative samples in comparing 'called' and 'camera-graded' beef carcass grades is needed. Nonetheless, the findings of this study are relevant to a variety of policy questions.

## Footnotes

<sup>1</sup> Nine packing plants use these instruments to assist in grading operations for approximately 40 percent of the beef carcasses graded each day by USDA (USDA 2013).

<sup>2</sup> To define its yield grade standard, the USDA uses the following equation:

$$\text{yield index} = 2.50 + 2.5 \times \text{fat thickness} + 0.20 \times \text{kph} + 0.0038 \times \text{weight} - 0.32 \times \text{ribeye area}$$

where kph refers to kidney, pelvic, and heart fat.

<sup>3</sup> Degree of Marbling is segmented into abundant, moderately abundant, slightly abundant, moderate, modest, small, slight, traces, and practically devoid. Maturity classes are classified into A (9-30 month), B (30-42 month), C (42-72 month), D (72-96 month), and E (>96 month)

<sup>4</sup> The total number of observations in our data does not necessarily reflect all the cattle processed at the packing plants.

<sup>5</sup> If USDA graders follow the USDA standards, they should call Select when marbling score is greater than or equal to 4 and less than 5.

<sup>6</sup> There are three cattle pricing methods: live weight pricing, dressed weight pricing, and grid pricing. While the price of carcass is determined by 'called' yield and quality grade under grid pricing, the price is determined based on the expected value under live and dressed weight pricing.

<sup>7</sup> Financial terms (loss/gain) in this analysis are used to express the amount of money that producers or packers would have earned if USDA human graders had been replaced by camera-grading system during the research period and do not have any normative meanings.

<sup>8</sup> Choice as the par value without premium or discount represents all Choices and does not account for high Choice which may have a premium in some grid pricing scales.

<sup>9</sup> If there is a market for high quality beef, then packers penalize more heavily the low quality beef carcasses, while the premiums for high quality beef increase. However, when there is high demand for beef in general, then packers do not consider beef quality and decrease (increase) premiums (discounts).



## References

- Chalfant, J.A., and R.I. Sexton. "Marketing orders, grading errors, and price discrimination." *American Journal of Agricultural Economics* 84(February 2003):53-66.
- Engelhard, G. "Examining rater errors in the assessment of written composition with a many-faceted Rasch model." *Journal of Educational Measurement* 31(Summer 1994):93-112.
- Field, T.G. *Beef Production Management Decisions*, Fifth ed. New Jersey: Prentice Hall, 2007.
- Garcia, L.G., K. L. Nicholson, T. W. Hoffman, T. E. Lawrence, D. S. Hale, D. B. Griffin, J. W. Savell, D. L. VanOverbeke, J. B. Morgan, K. E. Belk, T. G. Field, J. A. Scanga, J. D. Tatum, and G. C. Smith. "National Beef Quality Audit-2005: Survey of targeted cattle and carcass characteristics related to quality, quantity, and value of fed steers and heifers." *Journal of Animal Science* 86(December 2008):3533-3543.
- Gray, G. D., M. C. Moore, D. S. Hale, C. R. Kerth, D. B. Griffin, J. W. Savell, C. R. Raines, T. E. Lawrence, K. E. Belk, D. R. Woerner, J. D. Tatum, D. L. VanOverbeke, G. G. Mafi, R. J. Delmore Jr., S. D. Shackelford, D. A. King, and T. L. Wheeler, L. R. Meadows, and M. E. O'Connor. "National Beef Quality Audit-2011: Survey of instrument grading assessments of beef carcass characteristics." *Journal of Animal Science* 90(December 2012):5152-5158.
- Hale, D.S., K.J. Goodson, and J.W. Savell. "USDA Beef Quality and Yield Grades." Internet site: <http://meat.tamu.edu/beefgrading/> (Accessed January 2, 2015).
- Hueth, B., P. Marcoul, and J. Lawrence. "Grader bias in cattle markets? Evidence from Iowa." *American Journal of Agricultural Economics* 89(November 2007):890-903.
- Leckie, G., and H. Goldstein. "Understanding uncertainty in school league tables." *Fiscal Studies* 32(June 2011):207-224.
- Mafi, G., B. Harsh, and J. Scanga. *Review of Instrument Augmented Assessment of USDA Beef*

*Carcass Quality Grades*. American Meat Science Association, 2013.

McCully, M. A. "Trends in the Choice-Select spread and implications to cattle producers."

Internet site: <http://www.cabpartners.com/articles/news/217/ChoiceSelectWhitePaper.pdf>.

(Accessed June 9, 2015).

Moore, C. B., P. D. Bass, M. D. Green, P. L. Chapman, M. E. O'Connor, L. D. Yates, J. A.

Scanga, J. D. Tatum, G. C. Smith, and K. E. Belk. "Establishing an appropriate mode of comparison for measuring the performance of marbling score output from video image analysis beef carcass grading systems." *Journal of Animal Science* 88(July 2010):2464–2475.

Myford, C.M., and E.W. Wolfe. "Monitoring rater performance over time: A framework for

detecting differential accuracy and differential scale category use." *Journal of Educational Measurement* 46(Winter 2009):371–389.

Piazza, M., and V. Izard. "How humans count: Numerosity and the parietal cortex."

*Neuroscientist* 15(June 2009):261-273.

Saal, F.E., Downey, R.G., and Lahey, M.A. "Rating the ratings: Assessing the psychometric

quality of rating data." *Psychological bulletin* 88(September 1980):413-28.

Schroeder, T.C., Davis, E.E. *Fed cattle Grid Pricing*. Texas Agricultural Extension Service, The

Texas A&M University System, 1998.

U.S. Department of Agriculture. *FSIS' and AMS' Field-Level Workforce Challenges*.

Washington, DC: U.S. Department of Agriculture, Audit Report 50601-0002-31, 2013.

U.S. Department of Agriculture. *2013 Packers and Stockyards Annual Report*. Washington, DC:

U.S. Department of Agriculture, Grain Inspection, Packers and Stockyards

Administration, 2014.

U.S. Department of Agriculture. *2015 National Summary of Meats Graded Reports – Fiscal Years (Beef)*. Internet cite: <https://catalog.data.gov/dataset/national-summary-of-meats-graded-historical-grading-volumes-beef> (Accessed October, 2015).

## Tables

**Table 1.** National Summary of Meat Graded (Million Pounds, percent of total graded in parentheses)

	2005	2006	2007	2008	Total
Prime	602 (3.1)	577 (2.9)	525 (2.6)	595 (2.9)	2,298 (2.9)
Choice	11,133 (57.3)	11,367 (56.2)	11,655 (58.0)	12,459 (61.0)	46,614 (58.1)
Select	7,679 (39.5)	8,279 (40.9)	7,872 (39.1)	7,312 (35.8)	31,142 (38.8)
Standard	29 (0.1)	6 (0.0)	56 (0.3)	69 (0.3)	161 (0.2)
Total	19,441 (100)	20,229 (100)	20,109 (100)	20,435 (100)	80,214 (100)

Source: AMS/USDA (<http://www.ams.usda.gov/reports/meat-grading>)

**Table 2.** Estimates of Standard Errors ( $\sigma_u, \sigma_v$ ) and Cutoff Values ( $C_k$ )

Period	$\sigma_u$	$\sigma_v$	$C_{Select}$	$C_{Choice}$	$C_{Prime}$	ln L
USDA Standards			4.00	5.00	8.00	
Whole Sample Analysis	0.89 (0.04)	0.99 (0.01)	3.18 (0.01)	4.50 (0.01)	8.90 (0.15)	-23,881.3
<i>Seasonal Analysis</i>						
Spring	0.89 (0.08)	0.96 (0.01)	3.13 (0.02)	4.55 (0.02)	8.59 (0.27)	-10,340.4
Summer	0.72 (0.02)	0.86 (0.01)	2.24 (0.04)	4.32 (0.02)	7.97 (0.06)	-9,669.4
Fall	1.12 (0.06)	0.82 (0.02)	1.35 (0.24)	4.56 (0.05)	8.54 (0.15)	-1,938.0
Winter	0.66 (0.09)	0.63 (0.06)	3.64 (0.06)	4.81 (0.04)	8.79 (0.21)	-536.3
<i>Annual Analysis</i>						
2005	0.85 (0.05)	1.12 (0.01)	1.94 (0.09)	3.87 (0.02)	9.02 (0.11)	-3,419.7
2006	0.93 (0.12)	1.05 (0.01)	2.28 (0.14)	3.99 (0.05)	9.21 (0.41)	-8,911.7
2007	0.65 (0.05)	0.58 (0.02)	3.54 (0.04)	4.88 (0.01)	8.54 (0.18)	-5,774.1
2008	0.67 (0.04)	0.60 (0.03)	3.80 (0.04)	4.97 (0.02)	8.53 (0.09)	-2,656.6
Before (May 2005 – Jul 2007 )	0.82 (0.04)	0.94 (0.01)	1.87 (0.09)	4.20 (0.02)	8.28 (0.12)	-18,002.3
During (Aug 2007 –Oct 2008)	0.66 (0.04)	0.61 (0.03)	3.93 (0.02)	5.04 (0.02)	8.47 (0.08)	-3,396.3

Note: Standard errors in parentheses; All the estimated parameters were significant at 1% level.

**Table 3.** Summary Statistics of Marbling Score

Period	Num. of obs.	Mean	Std. Dev.
Whole Sample Period	18,080	5.10	1.04
<i>Seasonal Analysis</i>			
Spring (Mar - May)	7,785	5.01	1.00
Summer (Jun - Aug)	8,160	5.11	1.05
Fall (Sep - Nov)	1,485	5.34	1.03
Winter (Dec - Feb)	650	5.67	1.08
<i>Annual Analysis</i>			
2005 (May-Dec)	2,408	4.53	1.00
2006 (Jan-Dec)	6,304	4.79	1.01
2007 (Jan-Dec)	6,140	5.30	0.87
2008 (Jan-Oct)	3,228	5.77	0.95

**Table 4.** USDA Reported Average Premiums and Discounts, (May 2005 – October 2008, \$/cwt)

---

	<u>Prime</u>	<u>Choice</u>	<u>Select</u>	<u>Standard</u>
Premiums and Discounts	15.43	0.00	-9.78	-15.79

---

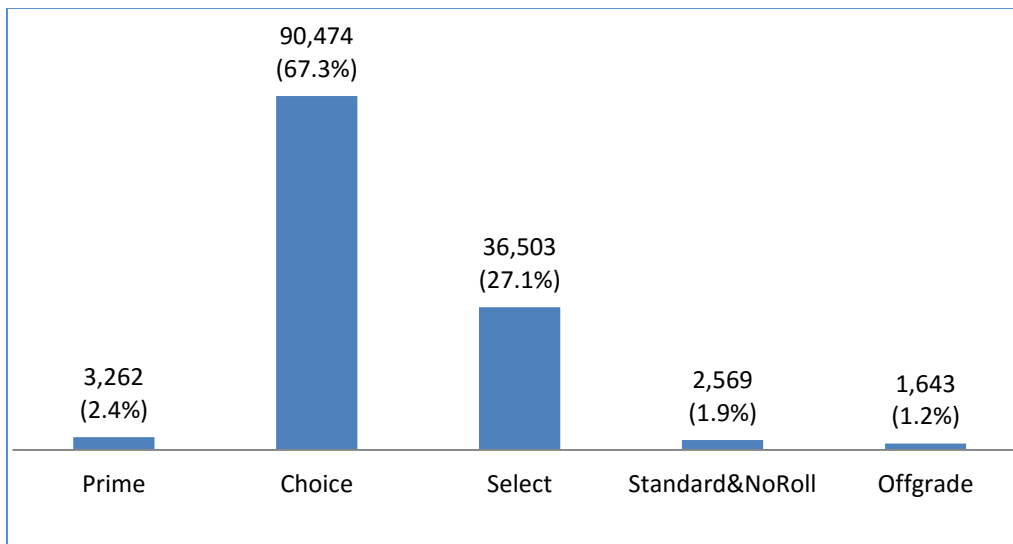
**Table 5.** Premiums and Discounts of ‘Camera-graded’ and ‘Called’ Quality Grade

Year	‘Called’ Quality Grade	Premiums and Discounts ‘Camera- graded’ Quality Grade (A)		Premiums and Discounts ‘Called’ Quality Grade (B)		Difference (A-B)	
		Sum	Ave.	Sum	Ave.	Sum	Ave.
2005	Prime	235	3.4	760	11.0	-525	-7.6
	Choice	-7,124	-4.3	0	0.0	-7,124	-4.3
	Select	-8,310	-12.9	-4,274	-6.7	-4,036	-6.3
	Standard	-327	-10.9	-339	-11.3	11	0.4
	Total	-15,527	-6.4	-3,853	-1.6	-11,674	-4.8
2006	Prime	435	2.4	2,080	11.6	-1,645	-9.1
	Choice	-32,408	-7.3	0	0.0	-32,408	-7.3
	Select	-25,490	-15.8	-21,376	-13.2	-4,114	-2.5
	Standard	-951	-18.3	-951	-18.3	0	0.0
	Total	-58,413	-9.3	-20,246	-3.2	-38,167	-6.1
2007	Prime	553	6.3	1,182	13.4	-629	-7.1
	Choice	-4,895	-1.2	0	0.0	-4,895	-1.2
	Select	-13,318	-6.9	-14,296	-7.4	977	0.5
	Standard	0	0.0	0	0.0	0	0.0
	Total	-17,660	-2.9	-13,114	-2.1	-4,546	-0.7
2008	Prime	454	7.8	601	10.4	-147	-2.5
	Choice	-98	0.0	0	0.0	-98	0.0
	Select	-1,838	-2.6	-2,579	-3.6	741	1.0
	Standard	0	0.0	0	0.0	0	0.0
	Total	-1,482	-0.5	-1,978	-0.6	496	0.2
Total	Prime	1,677	4.2	4,623	11.7	-2,946	-7.5
	Choice	-44,524	-3.5	0	0.0	-44,524	-3.5
	Select	-48,956	-10.0	-42,524	-8.7	-6,431	-1.3
	Standard	-1,278	-15.6	-1,290	-15.7	11	0.1
	Total	-93,082	-5.1	-39,191	-2.2	-53,891	-3.0

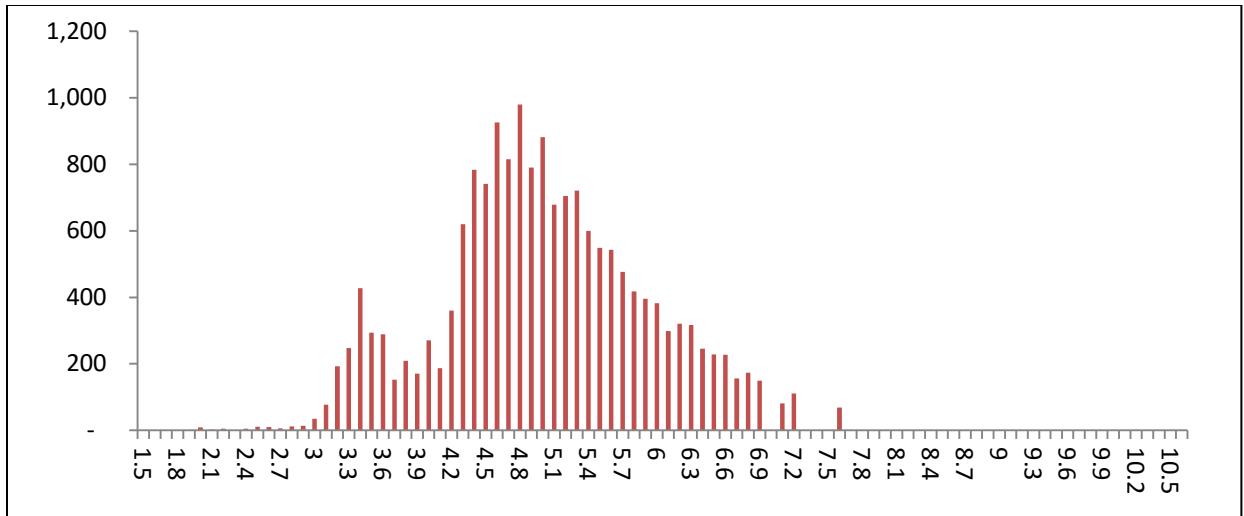
Note: All the values are reported in dollars per hundredweight (\$/cwt).



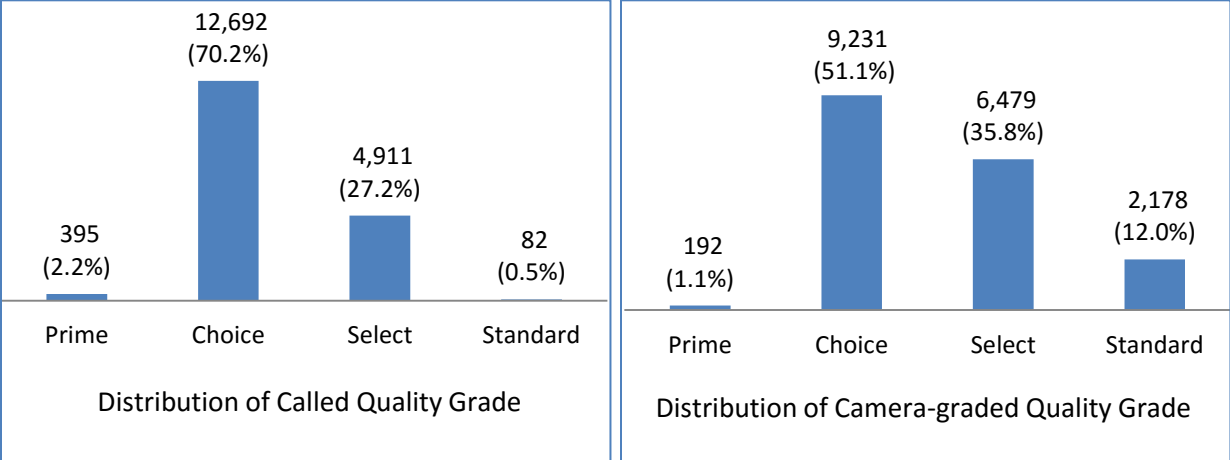
## Figures



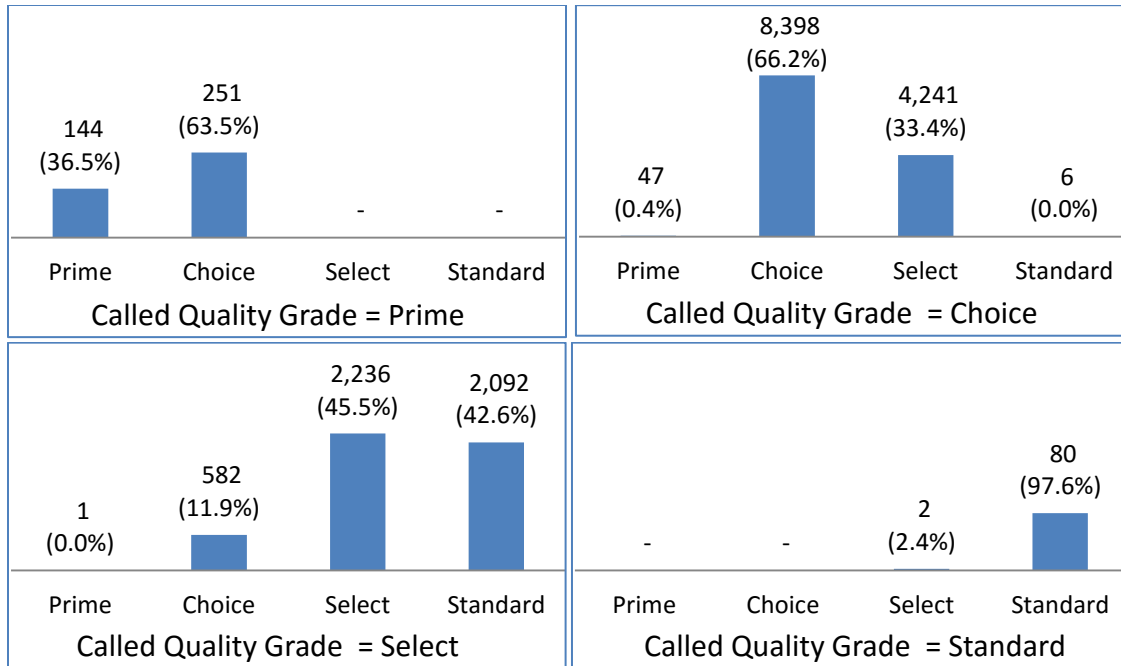
**Figure 1.** The Distribution of Quality Grade (n=134,451, the number of head, percent of total graded in parentheses)



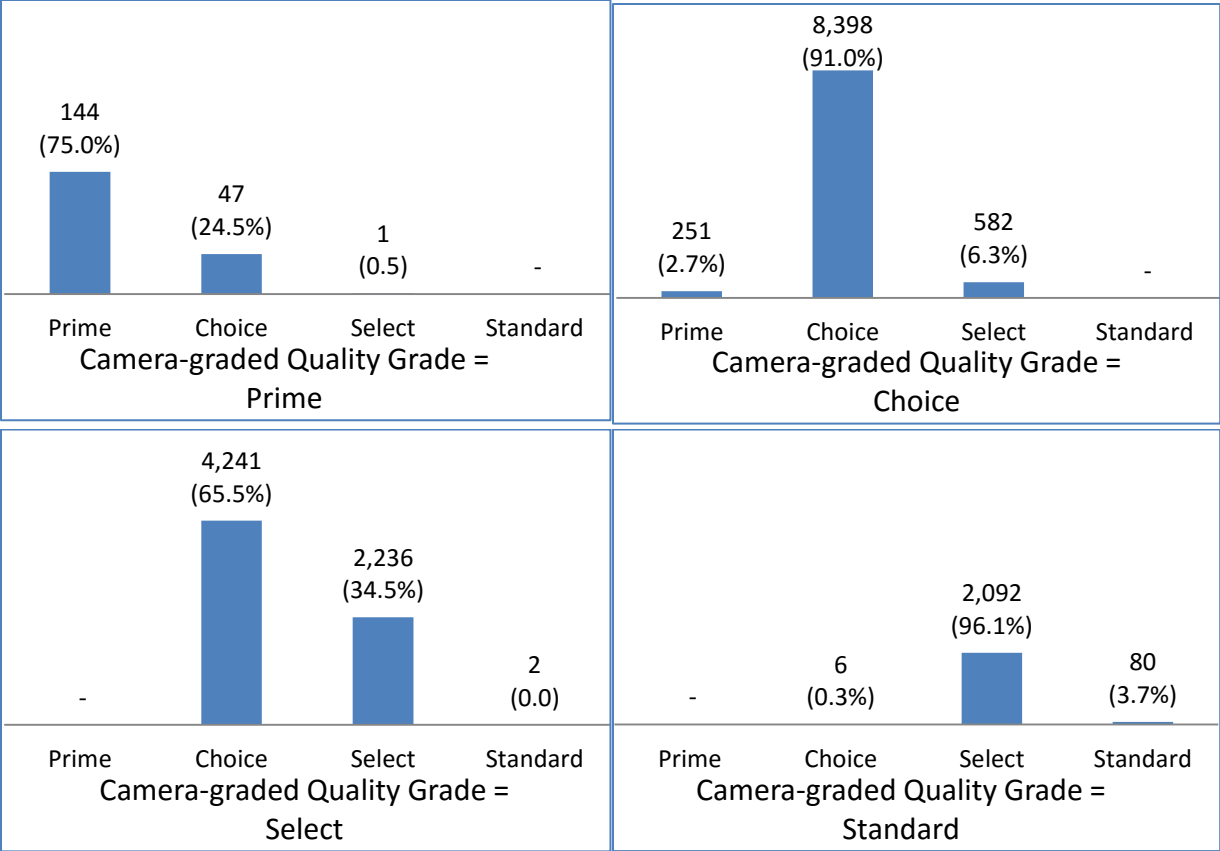
**Figure 2.** The Distribution of the Numeric ('Camera-graded') Marbling Score (n=18,080)



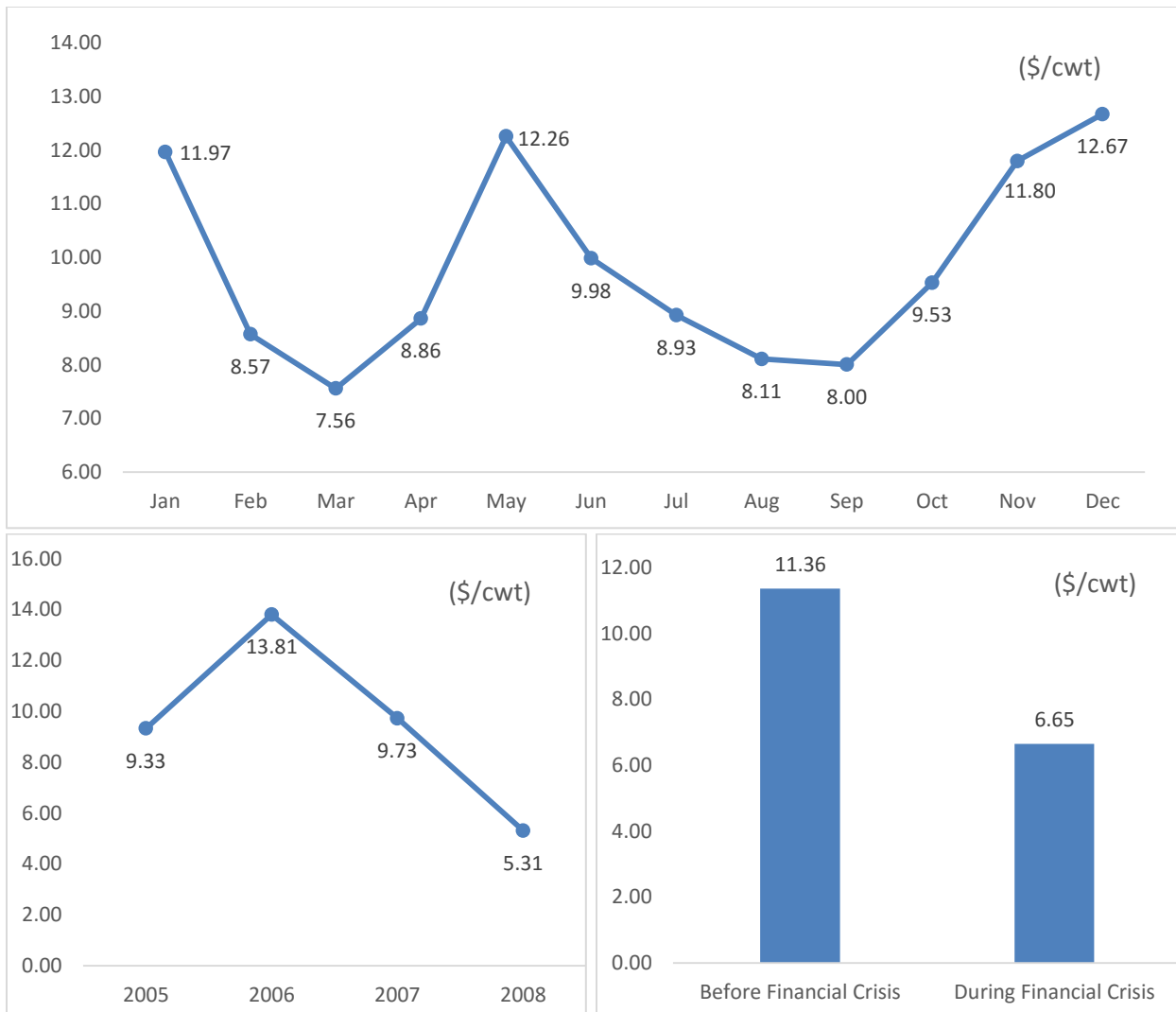
**Figure 3.** The Distribution of ‘Called’ and ‘Camera-graded’ Quality Grade (n=18,080, the number of head, percent of total graded in parentheses)



**Figure 4.** The Distribution of ‘Camera-graded’ Quality Grade given ‘Called’ Quality Grade (n=18,080, the number of head, percent of total graded in parentheses)

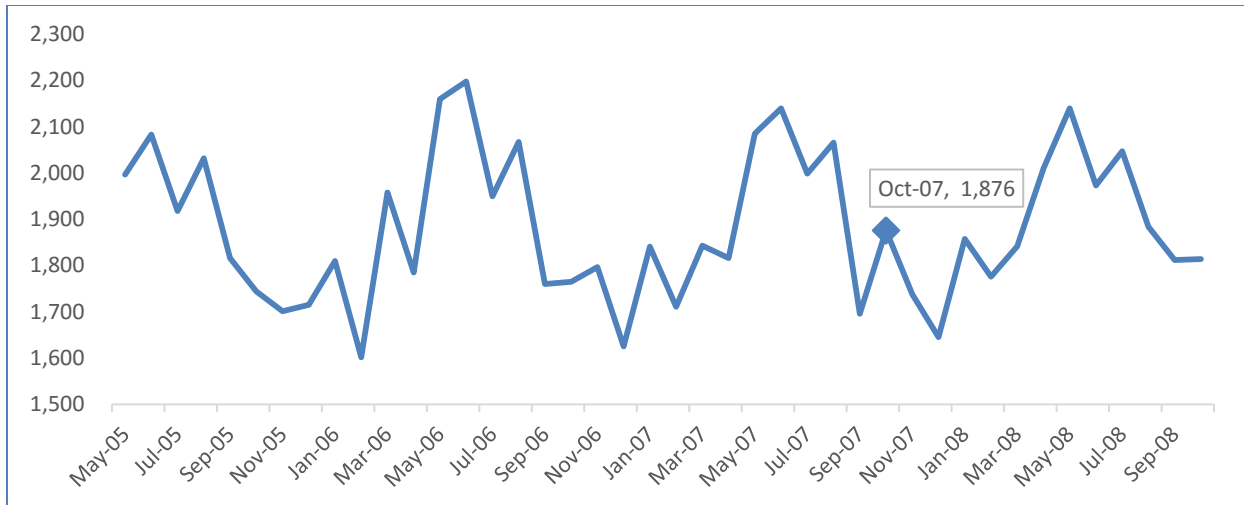


**Figure 5.** The Distribution of ‘Called’ Quality Grade given ‘Camera-graded’ Quality Grade (n=18,080, the number of head, percent of total graded in parentheses)



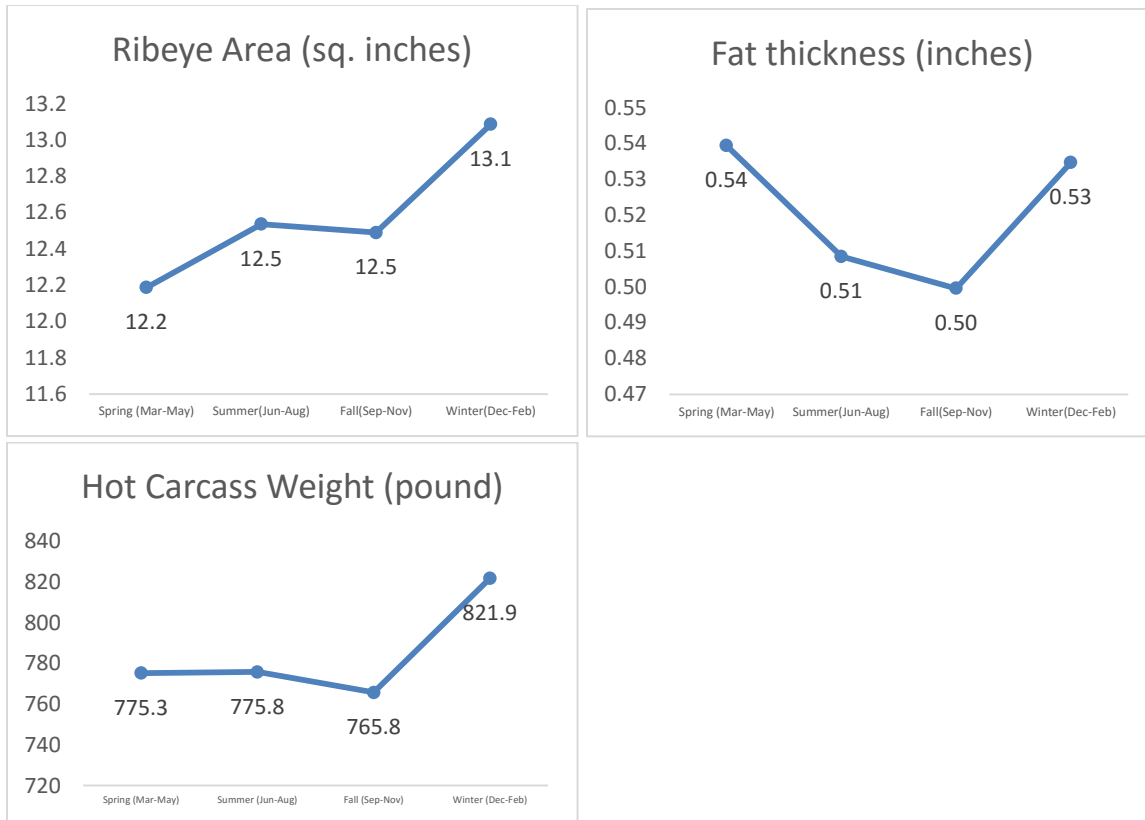
Source: USDA AMS 5-Area Weekly Direct Slaughter Reports

**Figure 6.** Average Choice-Select Spread during our sample period (May 2005 – Oct 2008)



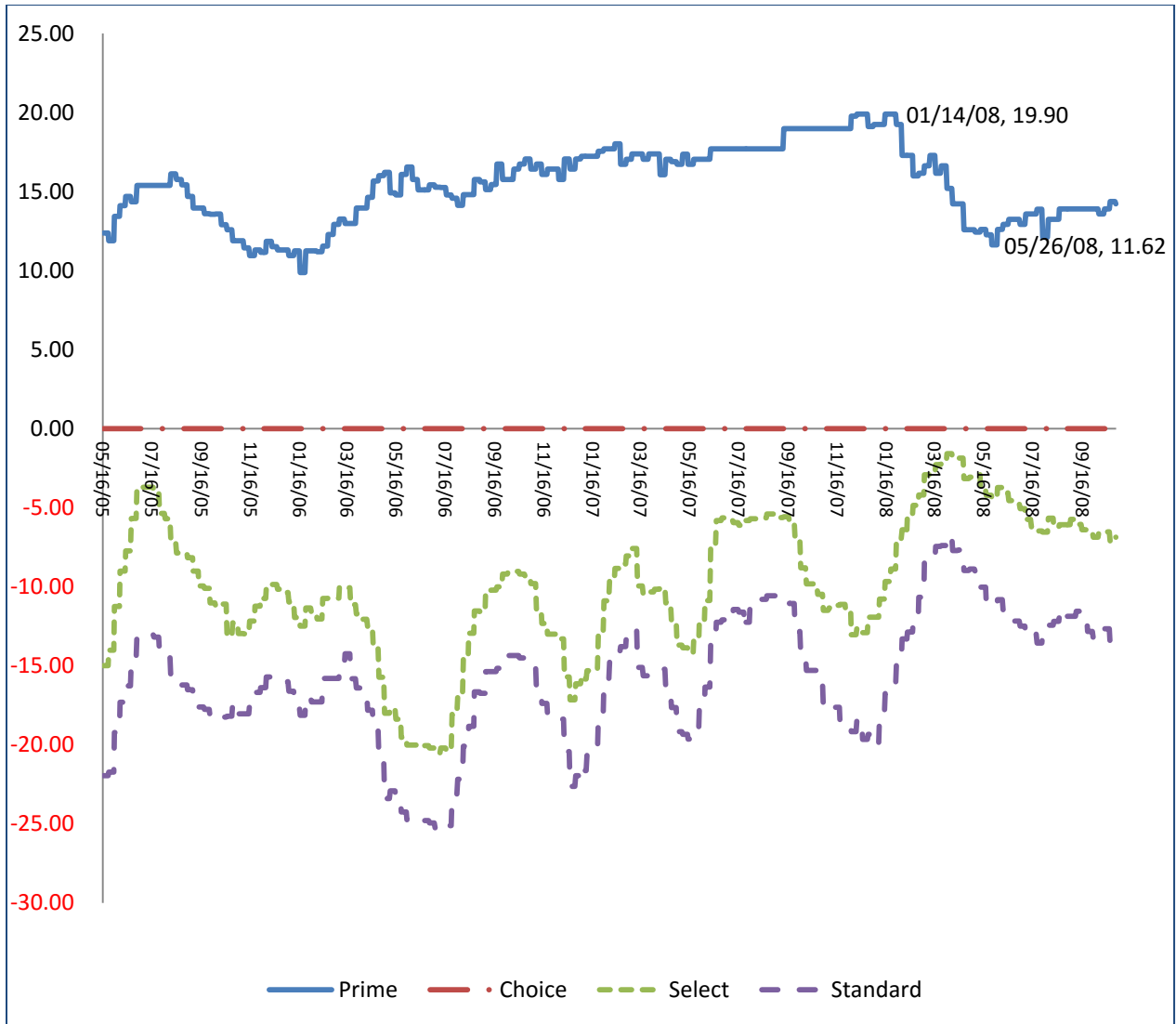
Source: USDA reports, Cattle on Feed  
 (<http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1020>)

**Figure 7.** Number of Fed Cattle Marketed on 1,000+ Capacity Feedlots – United States: May 2005 - October 2008 (Unit: 1,000 head)



**Figure 8.** Average Ribeye Area, Fat thickness, and Hot Carcass Weight from the Whole Sample (Seasonal, May 2005 – Oct 2008)





Source: USDA AMS 5-Area Weekly Direct Slaughter Reports

**Figure 9.** Premiums and Discounts – Weekly Average Direct Beef carcasses (\$ Per Cwt)