EXPLORING THE EXISTENCE OF GRADER BIAS IN BEEF GRADING

Ju Won Jang
Graduate Research Assistant
Department of Agricultural Economics
Texas A&M University
junyoung73@tamu.edu

Ariun Ishdorj
Assistant professor
Department of Agricultural Economics
Texas A&M University

David P. Anderson
Professor and Extension Economist
Department of Agricultural Economics
Texas A&M University

Tsengeg Purevjav
Consulting Research Scientist, PhD
INTI Service Corp.
College Station, TX 77845

Garland Dahlke
Research Scientist
Department of Animal Science
Iowa State University

Selected Paper prepared for presentation at the Southern Agricultural Economics Association (SAEA) Annual Meeting, Atlanta, Georgia, January 31-February 3, 2015

ABSTRACT

The United States Department of Agriculture (USDA) beef grading system plays an important role in marketing and promoting of beef. USDA graders inspect beef carcasses and determine quality grade within a few seconds. Although the graders are well-trained, the nature of this grading process may lead to grading errors. In this paper we examine whether systematic grader bias exists in calling quality grade. Using data from a large-scale packing plant in Midwest we find that seasonality of beef demand and macroeconomic events influence the grading behavior of USDA graders. Producers gain financially from grade called by USDA graders rather than measured by camera.

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

JEL Classifications: Q13, Q18

I. Introduction

The United States Department of Agriculture (USDA) updated the *Official United States Standards for Grades of Carcass Beef* in 1997 (USDA, 1997). The USDA beef grading standards are comprised of USDA quality and yield grade designed to assess eating quality and amount of lean edible meat from a carcass, respectively. Livestock producers can predict the market value of their products by using these standardized grades. They also have an incentive to produce high-quality beef cattle since the system is designed to guarantee financial reward for higher beef quality. Consumers are able to make their informed purchasing decisions using beef grades and labels. In short, the system makes marketing process simpler and communication between producers and consumers easier (Field, 2007).

The effectiveness of the beef grading system is guaranteed by an accurate and precise grading. In reality, however, graders employed by USDA determine grades through a visual inspection within a few seconds. Although USDA graders are well-trained and independent of both producers and packers, the nature of the grading process might lead to grading errors. These errors could diminish the farmer's incentive to produce a high-quality product (Chalfant *et al*., 2002) and reduce the efficiency of the marketing process.

In 2006, two camera-based grading systems were approved by the USDA in order to improve the beef carcass grading accuracy and uniformity within the industry[1]. In August 2014, the USDA Agricultural Marketing Service (AMS) sought for public input on possible revisions to the *U.S. Standards for Grades of Carcass Beef* to help adjust for recent improvements and trends in animal raising and feeding. Although AMS has been working on improving the accuracy of beef grading, there are relatively few studies that looked at the presence and sources

[1] 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).

of grader errors. Hueth *et al.* (2007) used a behavior model and found the existence of grader bias in assigning yield grade. Mafi *et al*. (2013) found cameras/instruments are more accurate and consistent than USDA graders in assessing marbling score to determine quality grade. They also found that camera grading can reduce grader-to-grader and plant-to-plant variations.

The accuracy of the grading is crucial for the efficiency of the beef marketing system and the promotion of beef quality. The impact of grading errors on the efficiency and promotion can be minimized, if the errors are not systematically biased across time and locations (Hueth e*t al.*, 2007). When the errors are inconsistently biased across time or locations, users of the system cannot trust the system. The unreliable system cannot play its role effectively. We, thus, focus on investigating if grading errors are systematically biased across time.

Our study builds on previous literature by suggesting the evidence of the existence of grader bias and possible sources of grader bias using data from a large-scale Midwest packing plant for the 2005 through 2008 period. The data on quality grade called by USDA graders ("called" quality grade) and measured by cameras ("measured" quality grade) of each carcass are provided along with year, month, and day of the week when cattle was processed.

The specific objectives of this study are threefold. First, we analyze the difference between "called" and "measured" quality grade and using these given quality grades we estimate the cutoff points for each quality grade and then compare these estimated cutoff points to the USDA Standard cutoff points. We expect that the propensity of grading behavior itself can be another source of grader bias. Educationalists have studied rating errors in evaluating student's performance. 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 examine the grader bias patterns across time to account for seasonality in beef demand and macroeconomic events. In this paper, we extend the existing literature by analyzing seasonal bias patterns. The analysis includes comparing the patterns of grading errors across seasons and years, and examining the difference in grading patterns before/after the 2008 financial crisis. Graders as human beings are possibly influenced by some form of cognitive bias. If they have prior belief about beef demand or economic situation, their grading behavior can be biased due to the prior belief. Because USDA graders' grade influences the beef price, we expect, a priori, some effect on bias during the recession. If different grading behaviors are found during or after the financial crisis, we can conclude that macroeconomic events such as financial crises may also be another possible source of grader bias.

Finally, since USDA intends to better utilize its camera grading system in the future, it is worth the discussion and analysis of the impact of upcoming changes on producers and packers. Another addition to the literature is our prediction of possible results of the policy change. We measure the change in financial gains/losses of livestock producers and packers from reduced use of USDA graders. For this analysis, weekly weighted averages of premiums and discounts for each quality grade are collected from USDA AMS 5-Area Weekly Direct Slaughter Reports. These premiums and discounts data along with "measured" and "called" quality grade allow us to measure the financial impact of utilizing camera grading system instead of USDA graders. If grading errors are systematically biased, one of two groups has a financial advantage from the policy change. For instance, if USDA graders tend to call higher yield grade than the camera grading system, packers would have a financial benefit when USDA graders are replaced by a camera grading system. If, thus, we find the change of financial rewards between two groups, we can conclude that the policy influences their earnings.

To our knowledge this is the first study investigating seasonality as a possible source of grader bias. This work will contribute some of the first research measuring the impact of increased utilization of the camera grading system on livestock producers and packers. Those analyses are possible, because the data sample contains a much larger number of observations than earlier studies. This advantage enables the analysis of grader biases in a new ways and thus allows the extension to general implications of the study.

II. Model

There are eight USDA quality grades for beef carcasses: USDA Prime, Choice, Select, Standard, Commercial, Utility, Cutter, and Canner. The factors used to determine quality grade are the degree of marbling and the maturity class which are classified into 9 and 5 different levels[2], respectively. USDA graders subjectively determine both maturity and marbling based on descriptions and illustrations provided in official USDA beef grading standards and practical work experiences.

The degree of marbling and the maturity class are combined to determine the final quality grade (Hale *et al.*, 2015). USDA Prime, Choice, Select, and Standard grades most commonly represent younger cattle. 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 (Garcia et al., 2008). Hence, in this study we assume that maturity class was A (9-30 month) or B (30-42 month). Given the maturity class, the primary determinant of quality grade will be marbling score as

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

reported in Table 1. In our analysis, thus, we include the beef carcasses which are graded as Prime, Choice, Select, and Standard. Given this exclusion, the model that used marbling score as the determinant of quality grade can be built.

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

(1) Quality Grade =

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

Let c_i be the USDA grader "called" quality grade, m_i be a "measured" quality grade, and t_i be a "true" quality grade for carcass i. "True" quality grade is unobserved. We do not assume that "measured" quality grade is equal to "true" quality grade because camera grading system also can make grading errors like human grading (Mafi *et al*, 2013). Using these definitions, the "called" and "true" quality grade can be expressed as follows:

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

where $u_i,\ v_i$ are error terms for "called" and "true" quality grade, respectively. We assume that error terms follow normal distribution with zero mean and standard deviation (σ_u, σ_v). This assumption allows us to build the likelihood function to estimate cutoff points and standard errors.

The USDA Standard Marbling Score Intervals ($\widehat{MSI}_k$) for each quality grade, as shown in Table 2, 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 MSI for Prime means that USDA graders should call Prime when marbling score is larger than 8.0. Other quality grades were called in a similar way, in that a grade is "called" when marbling score falls within the indicated range. Our data indicate that "called" quality grade is not the same as "measured" quality grade. Hence, it is possible to presume that USDA graders have different marbling score intervals from the USDA standards. Using this premise, we can define USDA grader's Marbling Score Intervals ($\widetilde{MSI}_k$) with implicit cutoff points ($\mathrm{C_k}$, k = Prime, Choice, Select, and Standard) such as

$$\widetilde{\mathrm{MSI}}_{\text{prime}} = [\mathrm{C}_{\text{prime}},\ +\infty),\quad \widetilde{\mathrm{MSI}}_{\text{choice}} = [\mathrm{C}_{\text{choice}},\ \mathrm{C}_{\text{prime}}),$$

$$\widetilde{\mathrm{MSI}}_{\text{select}} = [\mathrm{C}_{\text{select}},\ \mathrm{C}_{\text{choice}}),\ \text{ and }\ \widetilde{\mathrm{MSI}}_{\text{standard}} = (-\infty, \mathrm{C}_{\text{select}})$$

If these implicit cutoff points are different from the USDA standards across time then we can conclude there is grader bias and it is systemically biased across time.

If we assume that the joint probability of two events, graders call quality grade (k) and "true" quality grade is same with the "called" quality grade (k), then the likelihood function that graders want to maximize to call "true" quality grade can be defined as follows:

$$\text{(3)}\qquad \mathrm{L^i}(\mathrm{c_i}, \mathrm{m_i} \mid \sigma_\mathrm{u}, \sigma_\mathrm{v}, \mathrm{C}_{\text{prime}}, \mathrm{C}_{\text{choice}}, \mathrm{C}_{\text{select}})$$

$$= \mathbf{I}(\mathrm{c_i} = \text{standard})\left\{\Phi\left(\frac{\mathrm{C}_{\text{select}} - \mathrm{m_i}}{\sigma_\mathrm{u}}\right) \times \Phi\left(\frac{4 - \mathrm{m_i}}{\sigma_\mathrm{v}}\right)\right\}$$

$$\times\, \mathbf{I}(\mathrm{c_i} = \text{select})\left\{\left[\Phi\left(\frac{\mathrm{C}_{\text{choice}} - \mathrm{m_i}}{\sigma_\mathrm{u}}\right) - \Phi\left(\frac{\mathrm{C}_{\text{select}} - \mathrm{m_i}}{\sigma_\mathrm{u}}\right)\right] \times \left[\Phi\left(\frac{5 - \mathrm{m_i}}{\sigma_\mathrm{v}}\right) - \Phi\left(\frac{4 - \mathrm{m_i}}{\sigma_\mathrm{v}}\right)\right]\right\}$$

$$\times\, \mathbf{I}(\mathrm{c_i} = \text{choice})\left\{\left[\Phi\left(\frac{\mathrm{C}_{\text{prime}} - \mathrm{m_i}}{\sigma_\mathrm{u}}\right) - \Phi\left(\frac{\mathrm{C}_{\text{choice}} - \mathrm{m_i}}{\sigma_\mathrm{u}}\right)\right] \times \left[\Phi\left(\frac{8 - \mathrm{m_i}}{\sigma_\mathrm{v}}\right) - \Phi\left(\frac{5 - \mathrm{m_i}}{\sigma_\mathrm{v}}\right)\right]\right\}$$

$$\times\, \mathbf{I}(\mathrm{c_i} = \text{prime})\left\{\left[1 - \Phi\left(\frac{\mathrm{C}_{\text{prime}} - \mathrm{m_i}}{\sigma_\mathrm{u}}\right)\right] \times \left[1 - \Phi\left(\frac{8 - \mathrm{m_i}}{\sigma_\mathrm{v}}\right)\right]\right\}$$

where, **I**( ) is an indicator function, and Φ( ) is the cumulative density function of the standard normal distribution. The likelihood function is derived from the assumptions that USDA graders call quality grade to maximize the probability of calling "true" quality grade by using their own implicit intervals. Since "true" quality grade is unobserved, USDA graders call quality grade based on the result of their own visual inspection. The USDA graders use their own implicit cutoff points to determine quality grade when "true" quality grade is unobservable. The estimated cutoff points can provide us with information on how USDA graders decide quality grade.

III. Data

The data used in the analysis provide information on "called" and "measured" quality grade of the beef carcasses from May 2005 to October 2008. As contained in Table 3 and Figure 1, 70.2% and 27.2% of carcasses were graded as Choice and Select, respectively by USDA graders. These numbers indicate that most of beef carcasses are graded as Choice or Select. However, more beef carcasses were graded as Select or Standard when camera grading system is used to grade. While USDA graders grade 27.2% (0.5%) of beef carcasses as Select (Standard), cameras grade 35.8% (12.0%) of them as Select (Standard). This shows the evidence for the existence of grader bias.

AMS announces the national summary of meat graded at the beginning of each year. From 2005 to 2008, 2.9%, 58.1%, 38.8%, and 0.2% of beef are, as Table 3 contains, graded as Prime, Choice, Select, and Standard, respectively. Both data from USDA and Midwest packing plants show that most of the carcasses were graded as either Choice or Select.

Figure 1 illustrates the distributions of “called” and “measured” quality grades. The distribution of “measured” quality grade has fatter right hand side. This difference indicates that USDA graders tend to call more Choice than camera grading system. It is not possible to know whether USDA graders are more generous than camera from the difference. But the difference shows that one of them could be biased.

The distribution of “measured” quality grade given “called” quality grade in Figure 2 illustrates the difference in USDA graders and camera grading system. Figure 2 indicates that, when 2.2% of beef carcasses were graded as Prime by USDA graders, only 36.5% of them were graded as Prime and the rest, 63.5%, were graded as Choice by cameras. If we assume that “measured” quality grade is close to “true” quality grade[3], then the distributions will reflect the grading error of USDA graders. In case of Prime grade, only 75.0% of beef carcasses have the identical “called” and “measured” quality grade. The 24.5% and 0.5% of them are incorrectly graded as Choice and Select, respectively. Other grades are similar with Prime. It implies the existence of grading errors on quality grades.

Figure 2 shows that 63.5% (33.4%) of the beef carcasses which graded as Prime (Choice) by USDA graders were graded as Choice (Select) by camera. This implies that USDA graders were more generous on grading the high-quality beef than camera. In case of Select, 11.9% (42.6%) of the beef carcasses graded as Select by USDA graders were graded as Choice (Standard) by camera. This distribution represents USDA graders more severe than camera when they grade low quality beef. The difference of “called” and “measured” quality grade for the same beef carcass might imply the existence of grader bias. However, this distribution analysis is not sufficient to confirm the existence of grader bias. Thus, we will estimate the implicit cutoff

[3] Since Mafi *et al.* (2013) suggest that camera grading system is more consistent than human grading, it is more reasonable to assume “measured” quality grade is close to true quality grade.

points for each quality grade to show the existence of grader bias and reveal possible sources of the bias.

IV. Results

1. Whole Sample Period Analysis

We estimated cutoff points for each quality grade using equation (3) to identify the implicit USDA graders' interval. The existence of grader bias can be checked by comparing the estimated and USDA Standard cutoff points. As shown in Table 4, estimated cutoff point for Prime, 8.904, is larger than the USDA Standard cutoff point, 8. This indicates that the estimated interval for Prime, $[8.904, +\infty)$, is significantly different from the USDA Standard, $[8.000, +\infty)$. This estimated interval also suggests that USDA graders call Choice although marbling score is larger than 8.000. According to the USDA Standard, when marbling score is located between 8.000 and 8.904, USDA graders should call Prime, but they actually call Choice. This result means that USDA graders have a higher standard for Prime grade.

Table 4 also shows the estimated cutoff point for Choice, 4.501, is smaller than the USDA Standards, 5.000. The difference between two cutoff points represents that USDA graders call Choice instead of Select when marbling score is less than 5.000[4]. The estimated cutoff point also means that the estimated implicit interval for Choice is [4.501, 8.904). The interval is much wider than the USDA Standard, [5.000, 8.000). The variation between two intervals suggests that USDA graders have a tendency to call Choice even though beef carcasses have more or less marbling than the USDA standards.

[4] If USDA graders follow the USDA Standard, they should call Select when marbling score is less than 5 and greater than or equal to 4.

The estimated cutoff point for Select is 3.177. The estimated value is smaller than the USDA standard, 4. Using the cutoff point, we can find out that the estimated intervals for Select and Standard: [3.177, 4.501) and (-∞, 3.177). These intervals indicate that USDA graders call Select although marbling score is less than the USDA Standard cutoff point, 4.000. This behavior means that USDA graders are generous on grading beef carcasses with less marbling.

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

The narrow estimated intervals for Prime and Standard means USDA graders tend to avoid calling extreme grades. And the wide intervals for Choice and Select show that graders like to call quality grade around average marbling score, 5.104 (Table 5). These results indicate that USDA graders tend to call central grades and avoid calling extreme grades, Prime and Standard. These grading patterns can be evidences of the central tendency bias in beef grading.

A reason of the central tendency bias in beef grading can be found from the financial impact of quality grade to producers and packers. Under the grid price mechanism, producers get a premium or discount based on the quality grade of a beef carcass. As shown in Table 6, only Choice grade carcasses do not receive any premium or discount. Under this mechanism, calling Choice is the way to make less impact upon the financial reward/losses of producers and packers. Moreover, calling central grades (Choice and Select) can 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 request could be higher. Since USDA graders are independent from producers and packers, they have no intention of affecting profits of producers or packers through their grading. Thus, unless marbling score is extremely high or low, USDA graders could have a tendency to call central grades. Based on these findings, thus, we could conclude that the nature of grading behavior itself can be a possible source of grader bias.

2. Seasonal analysis

A large number of observations in our data set allow us to estimate the cutoff points across seasons. The estimated cutoff points for each season can help identify the effect of seasonality in beef demand on grading process. The implicit intervals of USDA graders affect the supply of each quality grade beef. For instance, when an implicit interval for Prime is narrow, it is highly possible that a supply of Prime beef is low. Graders are possibly conscious of their influence on the supply.

USDA graders may also have a priori belief about patterns of specific quality grade beef consumption over the seasons from their experience. Given the priori belief on specific quality grade beef demand, USDA graders could choose to match the supply and demand. When the demand for a specific quality grade beef is expected to be high, USDA graders could call that quality grade more.

Lusk *et al.* (2001) suggest that demand for both Choice and Select beef in spring and summer (April through September) is more price inelastic than one in fall and winter. In addition, as shown in Figure 3, beef per capita consumption increases in spring and summer, which are typically cookout seasons. In our sample, most of beef carcasses, as shown in Table 7, are also processed in spring and summer (March through August). This concentration can be related with high beef demand for these seasons and also the seasonality of production. Based on these findings, we can expect that the demand for high-quality beef increases in spring and summer. Given these expectations, we can analyze seasonal patterns in the bias by comparing the estimated intervals across seasons. If significant differences are found from the analysis, we can conclude that the seasonality in beef demand is attributable to the systematic grader bias across seasons.

Table 8 contains the estimated intervals for Prime, [7.967, +∞), and Choice, [4.321, 7.967), in summer. The data indicates that the summer interval is wider than those in other seasons. This estimation results indicate that USDA graders call more Prime and Choice in summer than other seasons. The results also mean that, in other seasons, USDA graders could call Select for carcasses which might be graded as Choice in summer.

The estimated intervals for Prime and Choice in spring, [8.585, +∞), [4.550, 8.585), are similar with those for fall, [8.536, +∞), [4.560, 8.536) while these two seasons' intervals for Prime and Choice are wider than those in winter, [8.794, +∞), [4.814, 8.794). The intervals for Prime and Choice in winter are narrowest among the seasons. These results mean that USDA graders are severe on grading high-quality beef carcasses in winter.

Our results could support Lusk et al. (2001)'s finding. They explained the price inelasticity for Choice and Select in spring and summer using the consumer preference change. In grilling seasons, consumers have a willingness to pay more for high quality beef (Lusk *et al.*, 2001). As shown in the estimation results, USDA graders call more Prime or Choice grades in summer and less Prime and Choice in winter. It is possible that high quality beef demand during the grilling seasons is contributing to systematical grader bias across seasons.

USDA Graders could be well aware of consumers' preference change across seasons. It is, however, difficult to say that USDA graders intentionally change their grading criteria to meet the market demand. It might be more reasonable that graders may be influenced by the seasonality of beef demand because graders could not entirely get rid of influences of surrounding environments and the judgments of human beings are naturally influenced by their environments.

3. Annual analysis

Using the estimation results of the equation (3), we can examine the existence of grader bias across years and before/after economic events. If major economic events which contribute to beef supply and demand occur during our sample period, we can measure the effect of these events on grading behaviors through comparing the estimated intervals before/after the events.

One of the major economic events during the sample period is the global financial crisis of 2007 and 2008. The crisis led to decline in wealth and consumption of goods. Beef consumption, as shown in Figure 3, also declined after the crisis. If this decline in beef consumption influences grading behavior, implicit intervals for each quality grade would be altered after the crisis.

Table 8 shows that estimated intervals for each quality grade are different from USDA standard across years. The estimated cutoff points for each quality grade in 2005 and 2006 are different from those in 2007 and 2008. The estimated intervals for Choice in 2007 and 2008: [4.882, 8.539) and [4.967, 8.527), are significantly narrower than those in 2005 and 2006: [3.873, 9.020) and [3.994, 9.212), respectively. The estimated interval for Select also has a similar pattern with Choice across years while one for Standard have an opposite pattern. The intervals for Standard in 2007 and 2008: (-∞, 3.535) and (-∞, 3.798), are wider than those in 2005 and 2006: (-∞, 1.942) and (-∞, 2.281). The wider implicit interval for Standard in 2007 and 2008 means that USDA graders could call Standard for the same carcass which could be graded as Select in 2005 and 2006. It is possible that these changes across years were induced by the financial crisis.

In order to examine the impact of the global financial crisis on grading behavior clearly, we estimated and compared implicit cutoff points before and after the crisis. As shown in Table 10, the intervals for Choice and Select after the crisis: [5.040, 8.469) and [3.927, 5.040), are significantly narrower than those before the crisis: [4.201, 8.279), [1.873, 4.201). While the interval for Prime after the crisis is almost identical with the one before the crisis, the estimated interval for Standard after the crisis, (-∞, 1.873), is significantly wider than before, (-∞, 3.927). Now it is clearer that USDA graders undervalued beef carcasses after the crisis. This change substantially results in more supply of low quality beef instead of high quality beef in a retail beef market.

The analysis clearly shows that macroeconomic event could affect beef grading behavior. In an economic recession, USDA graders expected that consumption of high-quality beef would decline and vice versa. USDA graders also know that increase in beef price during a recession is not desirable to stabilize a retail beef market because consumers do not want to spend more money on purchasing beef when their income decreases. If graders were aware of this, they could have an intention to call less expensive quality grade.

It is also possible that USDA graders do not consider supply and demand in a retail beef market. They, however, can acknowledge that people want to purchase cheaper beef cuts when economy is bad. This simple thought can influence their grading behavior without complicated economic logic. This reason might be more plausible to account for the undervaluation of beef carcasses after the crisis. Thus, without the strong intention to control a supply of high-quality beef, grading process can be influenced by macroeconomic events.

The estimation results also show that the estimated cutoff points for each quality grade, 3.927, 5.040, and 8.469, are similar with the USDA standard, 4.000, 5.000, and 8.000, after the crisis. This shows that, USDA graders try to avoid grading errors to prevent giving financial advantage to either producers or packers due to their misbehavior.

4. Premium/Discount Analysis

The trend of Premiums and Discounts for each quality grade during our sample period (from May 2005 to October 2008) is illustrated in Figure 4. The premium/discount for Choice is zero over the sample period, because it is a base quality grade for Prime, Select, and Standard. In 2008, as shown in Figure 4, premiums for Prime decreased, when discounts for Select and Standard also decreased. It means that the spread between Prime and Select (Standard) become narrow. In 2008, economy was down due to the financial crisis. Since the change of premiums and discounts relates with the consumer preference, the narrow spread in 2008 could imply that consumers chose to purchase low quality (or less expensive) beef instead of high quality beef as their income declines.

Our quality grade data includes a weight of each beef carcasses. It allows calculating premium/ discount of “measured” and “called” quality grade for each cattle. The difference of “measured” and “called” quality grade premium/discount indicates how much producers or packers gain/lose when USDA graders are replaced by camera grading system.

The difference of 53.9 million dollars as reported in Table 13 means that this is the amount of money which producers would lose if camera grading system were used instead of USDA graders. USDA graders give a favor to producers. This result is consistent with our analysis in the previous section. From the estimated interval for Choice, we can conclude that USDA graders were generous on grading. The generosity of USDA graders leads to producers’ financial benefit. We can, thus, expect that producers would lose financially when the number of USDA graders will reduced through using more camera grading system.

Since USDA is working on reducing human graders, it might be inevitable to use more camera grading system. It is foreseeable that producers would lose financially from the change.

V. Conclusion

The analyses in this paper used data from a large-scale Midwest packing plant. The data includes “called” and “measured” quality grade for each beef carcass from May 2005 to October 2008. We also use the weekly weighted averages of premiums and discounts for each quality grade along with two quality grades to measure the financial impact of the reduced use of USDA graders on livestock producers and packers.

The interval estimation analysis shows that USDA graders' called grades were significantly different from those measured by camera and identified possible source of grader bias using. The analysis also suggests that seasonality of beef demand and macroeconomic events such as financial crises are possible sources of grader bias. We also found that central tendency bias could be influencing the grading behavior of USDA graders from the analysis.

After we verified the existence of systematic grader bias across time, we investigated the possible impact of using camera grading methods and reduced use of USDA graders on financial gains/losses of producers and packers. When grading errors are systematically biased, the reduction of USDA graders' utilization can influence the financial reward of producers and packers. It is, thus, meaningful to analyze the gains/losses after confirming the existence of the bias. The results of the premium/discount analysis support the findings in the interval estimation analysis and show that USDA graders provide financial advantage to producers through their grading. The analysis also represents that the producers' earnings would decline when more camera grading is utilized in the beef grading system.

References

Chalfant, J.A., and R.I. Sexton. "Marketing orders, grading errors, and price discrimination." *American Journal of Agricultural Economics* 84(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(1994):93–112.

Field, T.G. *Beef Production Management Decisions,* Fifth ed. New Jersey: Prentice Hall, 2007.

Hale, D.S., K.J. Goodson, and J.W. Savell. "USDA Beef Quality and Yield Grades." 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(2007):890-903.

Leckie, G., and H. Goldstein. "Understanding uncertainty in school league tables." *Fiscal Studies* 32(2011):207–224.

Lusk, J., T.L. Marsh, T.C. Schroeder, and J.A. Fox. "Wholesale Demand for USDA Quality Graded Boxed Beef and Effects of Seasonality." *Journal of Agricultural and Resource Economics* 26(2001):91-106.

Mafi, G., B. Harsh, and J. Scanga. "Review of Instrument Augmented Assessment of USDA Beef Carcass Quality Grades." American Meat Science Association , 2013.

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(2009):371–389.

Saal, F.E., Downey, R.G., and Lahey, M.A. "Rating the ratings: Assessing the psychometric quality of rating data." *Psychological bulletin* 88(1980):413-28.

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.

1. TABLES:

Table 1. The Minimum Marbling Score Requirements for USDA Quality Grade within Each Final Maturity Group

		Maturity Score				
		A (0-30 mos.)	B (30-42 mos.)	C (42-72 mos.)	D (72-96 mos.)	E (>96 mos.)
USDA Quality Grade	$Prime^{+}$	Abundant	Abundant	-	-	-
	$Prime^{0}$	Moderate Abundant	Moderate Abundant	-	-	-
	$Prime^{-}$	Slightly Abundant	Slightly Abundant	-	-	-
	$Choice^{+}$	Moderate	Moderate	-	-	-
	$Choice^{0}$	Modest	Modest	-	-	-
	$Choice^{-}$	Small	-[a]	-	-	-
	$Select^{+}$	$Slight^{50}$	-[a]	-	-	-
	$Select^{-}$	$Slight^{00}$	-[a]	-	-	-
	$Standard^{+}$	Traces	Traces	-	-	-
	$Standard^{-}$	Practically Devoid	Practically Devoid	-	-	-
	$Commercial^{+}$	-	-	Moderate	Slightly Abundant	Abundant
	$Commercial^{0}$	-	-	Modest	Moderate	Slightly Abundant
	$Commercial^{-}$	-	-	Small	Modest	Moderate
	$Utility^{+}$	-	-	Slight	Small	Modest
	$Utility^{0}$	-	-	Trace	Slight	Small
	$Utility^{-}$	-	-	Practically Devoid	Traces	Slight

[a] Carcasses having B final maturity scores with Small and Slight marbling must grade U.S. standard. There is no U.S. Select grade for B maturity carcasses.

Source: Hale *et al.* (2015)

Table 2. Degree of Marbling, Marbling Score, and Quality Grade

Degree of Marbling	Marbling Score	Quality Grade
Abundant Moderately Abundant Slightly Abundant	10.0- 9.0-9.9 8.0-8.9	Prime
Moderate Modest Small	7.0-7.9 6.0-6.9 5.0-5.9	Choice
Slight	4.0-4.9	Select
Traces Practically Devoid	3.0-3.9 2.0-2.9	Standard

Table 3. National Summary of Meat Graded (Pounds in Thousands, 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 (100.0)	20,229 (100.0)	20,109 (100.0)	20,435 (100.0)	80,214 (100.0)

Table 4. Estimates of Standard Errors (σ_u, σ_v) and Cutoff Values (C_k) (Entire Sample Period)

Parameter	Estimate	Std Err	t Value	Pr > |t|	ln L
σ_u	0.893	0.035	25.19	0.000	
σ_v	0.996	0.008	125.22	0.000	
C_{Select}	3.177	0.013	242.43	0.000	-23,881.30
C_{Choice}	4.501	0.009	484.15	0.000	
C_{Prime}	8.904	0.148	60.13	0.000	

Table 5. Summary Statistics of Marbling Score (Whole Sample Period)

	Num. of obs.	Mean	Max	Min	Range	Std Dev
Marbling Score	18,080	5.104	10.6	1.5	9.1	1.037

Table 6. Average Premium and Discount (Entire Sample Period, dollars)

	Prime	Choice	Select	Standard
Premiums/Discounts	15.43	0.00	-9.78	-15.79

Table 7. Summary Statistics of Marbling Score (Seasonal)

Season	Num. of obs.	Mean	Max	Min	Range	Std Dev
Spring (Mar - May)	7,785	5.008	9.2	1.5	7.7	0.999
Summer (Jun - Aug)	8,160	5.107	9.4	2.0	7.4	1.050
Fall (Sep - Nov)	1,485	5.344	9.0	2.0	7.0	1.026
Winter (Dec - Feb)	650	5.674	10.6	3.3	7.3	1.080

Table 8. Estimates of Standard Errors (σ_u, σ_v) and Cutoff Values (C_k) (Seasonal)

Parameter	Spring	Summer	Fall	Winter
σ_u	0.887 (0.078)	0.720 (0.022)	1.115 (0.062)	0.662 (0.085)
σ_v	0.959 (0.011)	0.861 (0.009)	0.824 (0.022)	0.633 (0.056)
C_{Select}	3.130 (0.021)	2.243 (0.040)	1.345 (0.235)	3.638 (0.060)
C_{Choice}	4.550 (0.024)	4.312 (0.015)	4.560 (0.047)	4.814 (0.039)
C_{Prime}	8.585 (0.271)	7.967 (0.056)	8.536 (0.146)	8.794 (0.206)
ln L	-10,340.36	-9,669.40	-1,938.02	-536.32

Note: Standard errors in parentheses

Table 9. Estimates of Standard Errors (σ_u, σ_v) and Cutoff Values (C_k) (Annual)

Parameter	2005	2006	2007	2008
σ_u	0.847 (0.047)	0.930 (0.124)	0.645 (0.053)	0.669 (0.042)
σ_v	1.116 (0.010)	1.049 (0.008)	0.579 (0.020)	0.603 (0.028)
C_{Select}	1.942 (0.086)	2.281 (0.139)	3.535 (0.038)	3.798 (0.034)
C_{Choice}	3.873 (0.020)	3.994 (0.049)	4.882 (0.014)	4.967 (0.018)
c_{Prime}	9.020 (0.113)	9.212 (0.410)	8.539 (0.181)	8.527 (0.086)
ln L	-3,419.73	-8,911.72	-5,774.11	-2.656.56

Note: Standard errors in parentheses

Table 10. Estimates of Standard Errors (σ_u, σ_v) and Cutoff Values (C_k) (Before/After Financial Crisis)

	σ_u	σ_v	C_{Select}	C_{Choice}	C_{Prime}	ln L
Before	0.817 (0.042)	0.940 (0.007)	1.873 (0.085)	4.201 (0.019)	8.279 (0.122)	-18,002.3
After	0.660 (0.036)	0.611 (0.026)	3.927 (0.023)	5.040 (0.016)	8.469 (0.075)	-3,396.3

Note: Standard errors in parentheses

Table 11. Summary Statistics of Marbling Score (Producer)

Weekday	Num. of obs.	Mean	Max	Min	Range	Std Dev
Producer1	4,655	5.027	9.2	2.0	7.2	0.990
Producer2	2,816	4.667	8.9	2.0	6.9	0.917
Producer3	2,188	4.866	8.9	2.2	6.7	0.914
Producer4	1,105	5.282	10.6	2.6	8.0	0.946
Producer5	927	5.135	8.6	1.5	7.1	1.010
Producer6	912	5.586	9.3	2.9	6.4	0.998
Producer7	502	5.758	8.5	3.8	4.7	0.829

Table 12. Estimates of Standard Errors (σ_u, σ_v) and Cutoff Values (C_k) (Producer)

Parameter	P1	P2	P3	P4	P5	P6	P7
σ_u	0.717 (0.028)	0.851 (0.246)	0.903 (0.162)	0.770 (0.065)	0.886 (0.190)	0.727 (0.059)	0.565 (0.100)
σ_v	0.875 (0.011)	0.954 (0.015)	0.962 (0.017)	0.652 (0.028)	0.861 (0.047)	0.678 (0.033)	0.528 (0.054)
C_{Select}	1.703 (0.115)	1.805 (0.475)	2.059 (0.240)	1.657 (0.316)	3.299 (0.056)	1.993 (0.310)	3.985 (0.089)
C_{Choice}	4.307 (0.017)	4.306 (0.059)	4.145 (0.063)	4.591 (0.040)	4.922 (0.092)	4.378 (0.058)	4.902 (0.053)
C_{Prime}	7.851 (0.062)	9.060 (1.212)	9.686 (0.481)	8.269 (0.158)	8.848 (0.611)	7.883 (0.103)	8.486 (0.270)
ln L	-5,776.02	-4,100.72	-2,766.03	-1,207.40	-1,154.98	-837.67	-327.39

Note: Standard errors in parentheses

Table 13. Premiums/Discounts of "Measured" and "Called" Quality Grade (dollars)

Num. of obs.	Premiums/Discounts	Sum	Mean
18,080	Measured (A)	-93,081,696	-5,148.3
	Called (B)	-39,191,132	-2,167.7
	Difference(A-B)	-53,890,564	-2,980.7

2. FIGURES:

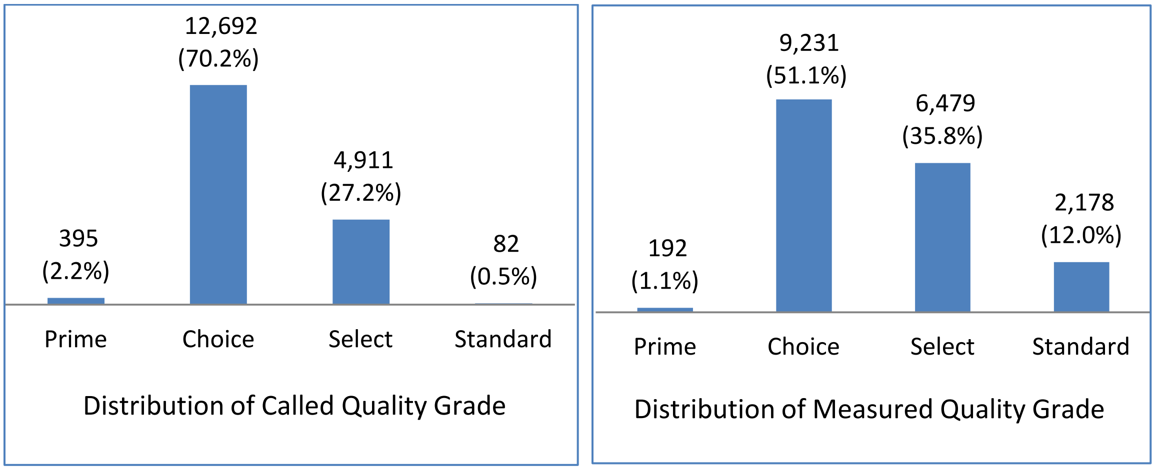

Figure 1. The Distribution of "Called" and "Measured" Quality Grade

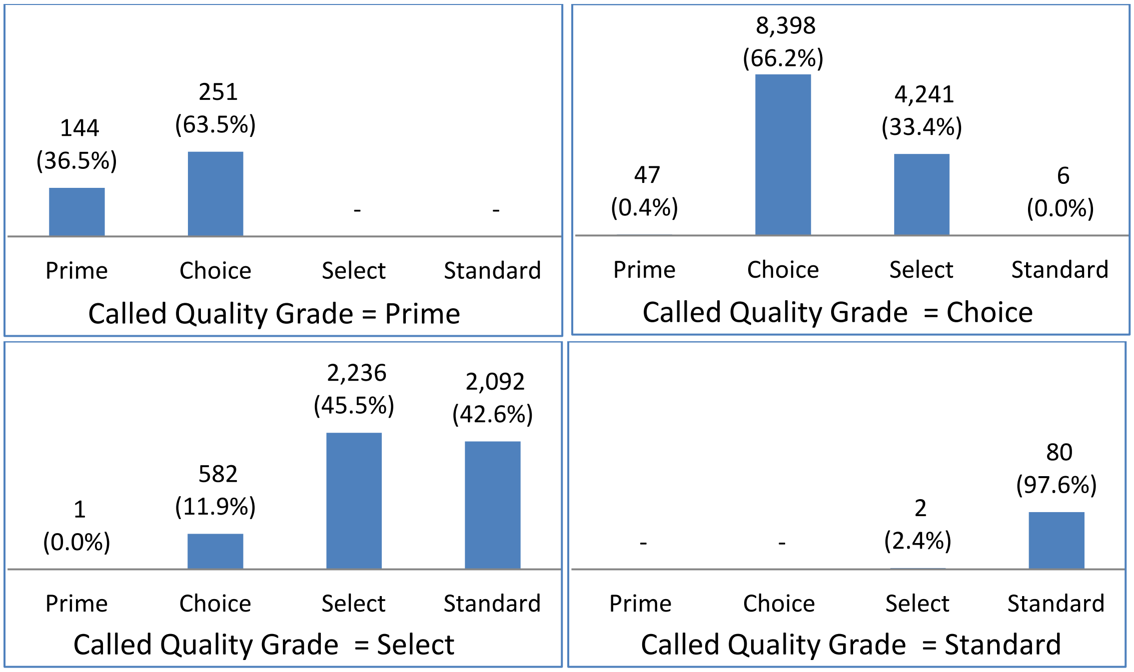

Figure 2. The Distribution of "Measured" Quality Grade given "Called" Quality Grade

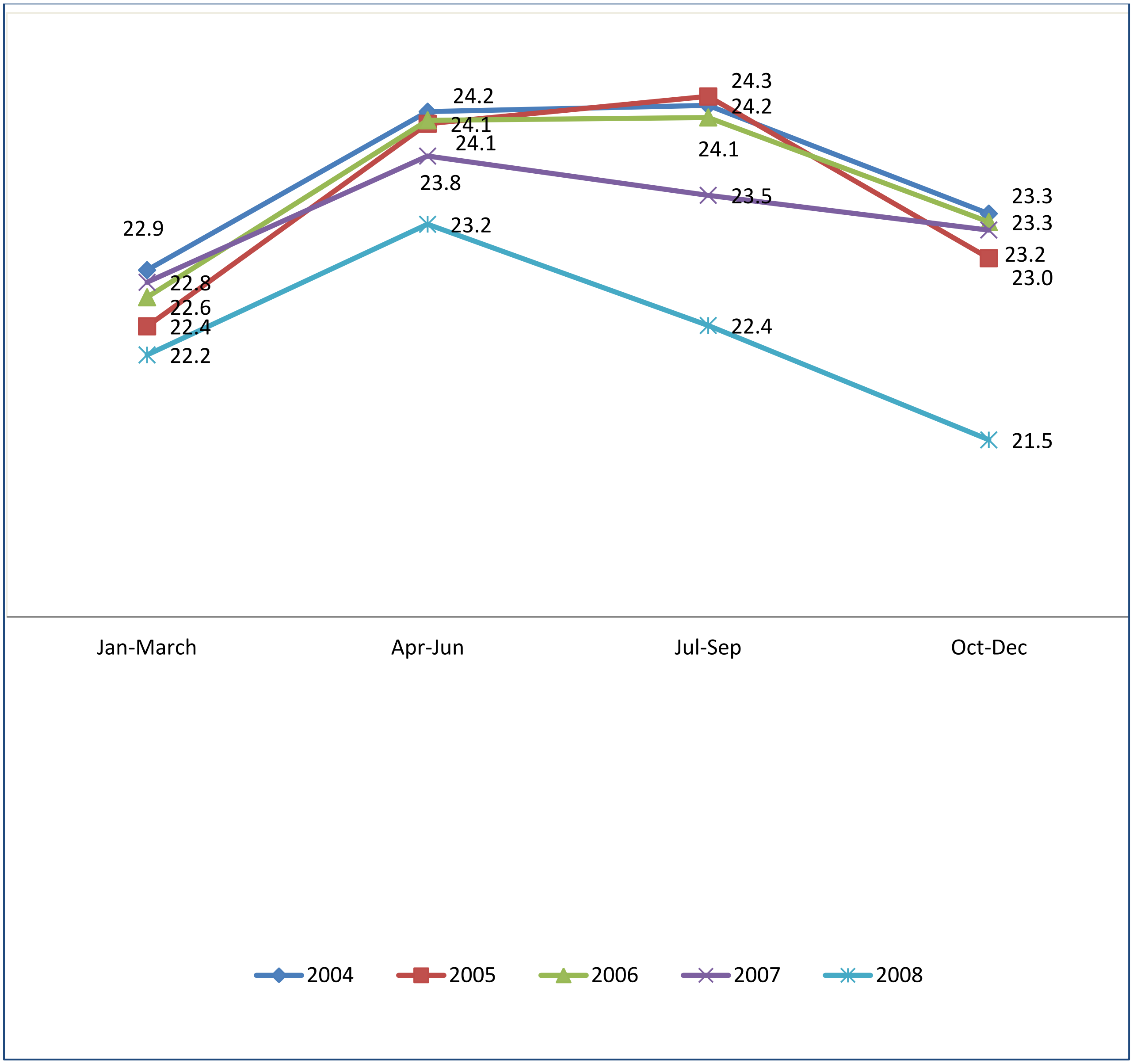

Figure 3. Per Capita Disappearance of Carcass Weight (Pound)

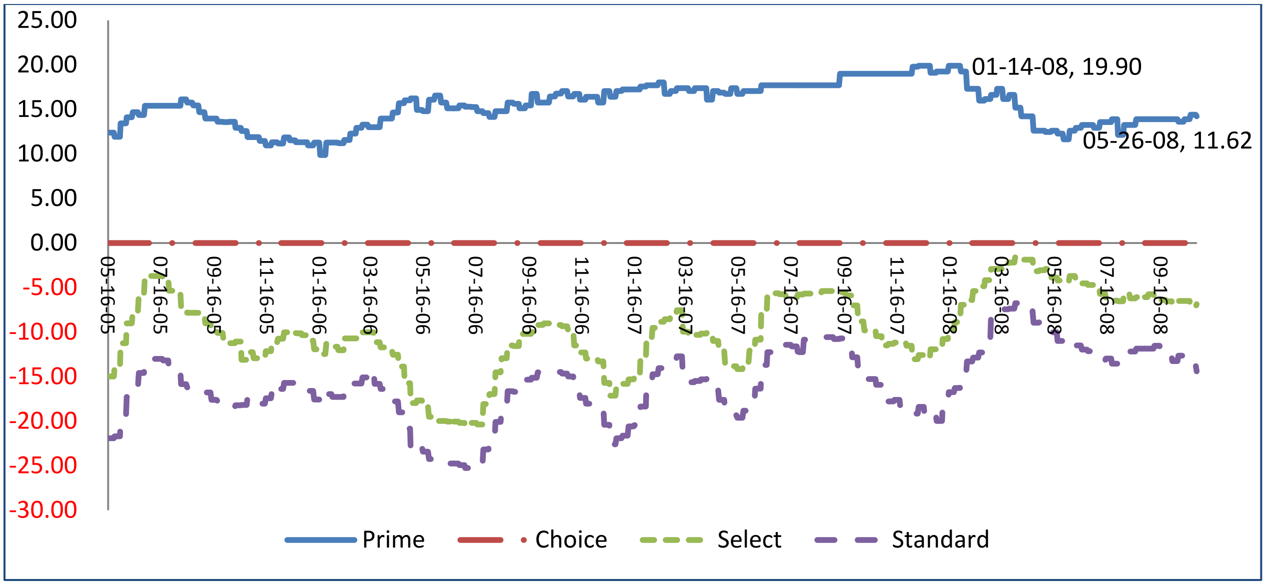

Figure 4. Premiums and Discounts – Weekly Average Direct Beef carcasses ($ Per Cwt)