GRADER BIAS IN CATTLE MARKETS?
EVIDENCE FROM IOWA

BRENT HUETH, JOHN LAWRENCE, AND PHILIPPE MARCOUL

ABSTRACT. Participants in U.S. markets for live cattle increasingly rely on federal grading standards to price slaughtered animals. This change is due to the growing prominence of "grid" pricing mechanisms that specify explicit premiums and discounts contingent on an animal's graded quality class. Although these changes alter the way cattle are priced, the technology for sorting animals into quality classes has changed very little: human graders visually inspect each slaughtered carcass and call a "quality" and "yield" grade in a matter of seconds as the carcass passes on a moving trolley. There is anecdotal evidence of systematic bias in these called grades across time and regions within U.S. markets. We examine whether such claims are supported in a sample of loads delivered to three different midwestern packing plants during the years 2000-2002. Overall, results indicate that indeed there is a bias, and that grading standards vary significantly across years and packing plants. Results also are consistent with a behavioral model where grader accuracy is inversely related to carcass quality.

GRADER BIAS IN CATTLE MARKETS? EVIDENCE FROM IOWA

Introduction

The pricing of slaughter cattle in U.S. markets increasingly depends on carcass quality measures used in so-called grid-pricing mechanisms (e.g., McDonald and Schroeder, 2003; Whitley, 2002). The carcass attributes most widely employed for this purpose include U.S. Department of Agriculture (USDA) "quality" and "yield" grades. These two attributes are intended to reflect meat palatability and the quantity of usable meat on an animal in relation to total carcass weight, respectively. Although the adoption of explicit grid-based pricing mechanisms is a relatively recent occurrence,[1] very little has changed in the way the USDA grades carcasses. In the majority of commercially graded cattle, grading is accomplished through visual inspection of each carcass, without the aid of physical measurement.

The rate of slaughter in a modern packing plant typically allows a grader less than 10 seconds to assess the quality and yield attributes of an individual carcass. Naturally, there is grading error in this process. However, so long as this grading error is in some sense unbiased, or at least consistently biased across time and location, it may be reasonable to expect little effect on efficiency or the distribution of returns in marketing cattle. One purpose of this paper is to examine whether bias exists in a sample of loads delivered to three different midwestern packing plants during the years 2000 to 2002, and, if so, to assess the direction and economic consequence of this bias. At a more abstract level, this work also proposes a framework to study decisions that involve an important element of human judgement.

The data (which we will describe in more detail below) that make this analysis possible include USDA *called* quality and yield grades, and additional carcass measures that allow

---

[1]In the absence of an explicit *ex ante* premium structure for yield and quality grade, there is nevertheless implicit pricing for these attributes embedded in the bids of packing plant field representatives who source live cattle. Schroeder et al. (2000) note how in the absence of an explicit grid, packing plants often provide order buyers with grid sheets indicating price premiums and discounts to award (or penalize) various expected (based on the buyer's *visual inspection*) quality outcomes.

computation of the *true* yield grade. In attempting to measure grading bias, we are interested in the difference between the distributions of called and true grades. Of course, it is possible to test whether the relevant aggregate empirical distributions are statistically different. However, this would not be quite what we want, because at present no grading technology can costlessly measure true grade without error (at least in the time frame demanded by current processing standards).[2] In other words, such a comparison imposes too high a standard on how graders should perform. Moreover, from a positive perspective, looking only at aggregate differences between true and called grade distributions would ignore information contained in the distribution of grader error. Thus, it is necessary to develop some reasonable model of how we think graders should perform given the imperfect information they have at the time grades are called, and to compare the distribution of called grades with the distribution predicted by this model. If these two distributions differ systematically, we will then say that grading is "biased."

Assuming some form of bias exists, we can think of at least three possible sources. First, a bias might be the result of the way training and incentives are provided to graders. As we discuss in more detail below, graders are trained in a specific technique for calling grades, and development of this technique by the USDA presumably was based on an implicit loss function with respect to different forms of grading error. Bias as we measure it might be "optimal" with respect to this loss function. Relatedly, graders are human beings, and thus potentially exhibit some form of perceptual bias. Finally, it seems plausible that packing plant staff (or cattle producers) either directly or indirectly influence grader calls. For the measure we consider (carcass "yield grade"), financial interests between cattle producers and packers tend to run in opposite directions. Packers tend to want graders to underestimate true yield, while cattle producers want graders to do the opposite. Only the last of these potential sources of bias is *possibly* an example of "corrupt" bias. That is, bias as we

---

[2]A mechanized yield grade instrument is currently under testing in a live production setting, but has yet to receive approval for use in grading services provided by the USDA (Meadows, 2005).

measure it is not necessarily the result of a premeditated or intentional distortion in grade calls. Moreover, if any bias is known by all industry participants, it is not clear what, if any, economic consequence it should have. In any case, although it may be possible to distinguish empirically among these (and other) sources of bias, our more modest objective in this paper is to consider whether a bias exists according to some reasonable baseline model of grader behavior. We leave determination of the source of any bias for future research.

Briefly, we consider four, mostly nonnested, models of grader behavior. In the first model, graders receive a signal of true grade and report their signal to make a grade call. This model is the simplest of all the models we consider because it assumes graders do not use any prior information they may have about the underlying distribution of the yield index to make a grade call. The second model assumes graders are more sophisticated in that they implicitly compute an expected grade and use this number to make a grade call. We assume that graders know the underlying distribution of the yield index, so that in computing this expectation, they effectively compute the mean yield index conditional on the signal they receive. The third model builds on this idea by assuming that a grader chooses the grade that is mostly likely to be the true grade, given the signal he or she receives. The last of our models is a hybrid of the others that imposes less structure on the manner in which grades are called.

Agricultural markets are typically governed by third-party grading of product attributes (see Gardner (2003) for a brief survey and historical overview). As a result, grading has been the subject of a considerable quantity of research by agricultural economists. While by no means an exhaustive list, previous research has considered topics such as the informational content of grades (e.g., Espinosa and Goodwin, 1991; Freebairn, 1973), the welfare consequences of minimum quality standards (Bockstael, 1984), the equilibrium provision of grading in markets with adverse selection (e.g, Hollander et al., 1999; Hennessy, 1995) and moral hazard (e.g., Ligon, 2002; Bogetoft and Olesen, 2003), and the potential for strategic use of grades among firms in an oligopoly market (e.g., Marette and Crespi, 2003; Chalfant

and Sexton, 2003). Our work is more closely related to the recent articles by Chalfant et al. (1999) and Diaz et al. (2002). In the first of these, the authors observe that the (purely mechanical) instrumentation used to grade California prunes is subject to an asymmetric error. In the second, the authors show how rounding rules for grading in U.S. peanut markets lead to excessive regrading. They further evaluate the economic impact of rounding errors and show that the excessive regrading it induces could be easily reduced by rounding with greater precision. Similar to this last pair of articles, we measure the economic impact of grading error. However, our work differs in its focus on a decision making process that is based purely on human perception. Diaz et al. (2002) contains treatment of "infrequent human errors," but these errors are essentially mistakes in "recording, transcribing, and computing weights." The mistakes we study are the result of human judgement unaided by any form of instrumentation. This is especially relevant in agricultural markets, because USDA grades are typically defined over product attributes that cannot easily be measured mechanically (e.g., internal color, texture, extent of damage).[3] Moreover, such decision-making problems are arguably pervasive in agriculture even outside commodity grading and involve, for example, a grain producer making a "seat of the pants" assessment of the maturity of his or crop and its readiness for harvest, a cattle buyer assessing quality and yield attributes of a live animal before making a purchase bid, or a show judge evaluating the quality of a live animal. Our data, which include both called and true yield grades, provide an opportunity to shed on light on these sorts of "judgement" decisions.

In what follows, we briefly describe the grading procedure that is carried out in U.S. cattle markets and develop a formal model of bias definition and estimation. We then describe our data in greater detail and present results. The final section concludes and discusses directions for future research.

---

[3]Casual perusal of USDA grading standards (see `http://www.ams.usda.gov/standards`, accessed May 10, 2005) suggests that purely mechanical grading is in fact exceptional.

Grading Procedures in U.S. Cattle Markets

Grading services offered by the USDA are voluntary and provided "at cost" to meat packing firms upon request. Meat grading was originally intended to be used solely for meat-market reporting but it subsequently developed into a trading standard (U.S. Department of Agriculture, 1997). At present, most large-scale meat packing firms employ USDA graders in their facilities and rely on federal grade standards for the majority of their trading.

The *quality grade* of a beef carcass can be one of (in decreasing order of meat palatability) the following: Prime, Choice, Select, Standard, Commercial, Utility, Cutter, and Canner. Technically, the quality grade of an animal is determined as a function of the degree of marbling (intramuscular fat) observed on the carcass, ranging from "practically devoid" to "slightly abundant," and the maturity (or age) class of the animal, the definition of which varies across animal type (e.g., steer versus cow). In practice, however, maturity is not known by graders, and marbling is not physically measured. Both traits are assessed by focusing on the visual state of certain key parts of the carcass. In the definition of the Prime quality grade, for example, an animal's maturity class is determined in part by observing "...slightly red and slightly soft chine bones and cartilages on the ends of the thoracic vertebrae that have some evidence of ossification." Similarly, for this particular maturity class, Prime requires "a minimum slightly abundant amount of marbling, and moderately firm ribeye muscle" (U.S. Department of Agriculture, 1997, p. 13). Thus, in addition to the subjectivity inherent in the definition of quality grade, the grading process itself is subject to additional subjectivity in the form of graders' sensory assessment of the relevant carcass attributes.

Yield grade is in principle defined more precisely than quality grade. The USDA defines yield grades "1" (best) through "5" (worst) using a well-established prediction equation relating various carcass characteristics to a yield index (where "yield" refers to usable meat in relation to total carcass weight), together with a grading rule that specifies the intervals of the yield index that represent a distinct yield grade. The equation used by USDA to define

its yield-grade standard is

(1) yield index $= 2.50 + 2.5 \times$ fat thickness $+ 0.20 \times$ kph

$$+ 0.0038 \times \text{weight} - 0.32 \times \text{ribeye area},$$

where "kph" refers to kidney, pelvic, and heart fat. Yield grade is then 1 for predicted yield strictly less than 2.0; 2 for predicted yield between 2.0 and 2.99; 3 for predicted yield between 3.0 and 3.99; 4 for predicted yield between 4.0 and 4.99; and 5 otherwise.[4] In practice, however, this equation is never used during commercial grading operations. Instead, graders are trained to assess visually key points of the carcass and to call a grade, 1 to 5, in a matter of seconds. In particular, graders are trained to implement these standards according to a specific technique that involves calling a "preliminary yield grade" according to the amount of external fat at the 12th rib, and then by making various specific adjustments to this preliminary call (Hale et al., 2006). Although this technique is well documented, conversations with industry participants suggest the process is more art than science, and that it is learned over time (Meadows, 2005). Graders must participate in a two-year training period before making unsupervised calls in a commercial packing plant, and are regularly monitored for consistency and accuracy.

Thus, as with quality grade, there is significant subjectivity in the call of an animal's yield grade. Nevertheless, an interesting difference between quality and yield grade is the ability, at least in principle, to precisely measure yield-grade given adequate time and resources, and hence to compare graders' called yield grades with actual yield grades.[5] Such a comparison is provided in Figure 1, where, in the upper leftmost subplot, we plot histograms of called and

---

[4] Abraham et al. (1980) study the relation between equation (1) and actual "cutability," or trimmed retail meat as a fraction of total carcass weight. This equation was developed in the late 1950's by Murphy et al. (1960) and has been the basis for the USDA yield grade since its inception in 1965.

[5] Throughout this paper we refer to the yield-grade prediction generated from equation (1) as the "actual yield grade." This is a bit misleading in that equation (1) is itself a prediction equation for measuring cutability (see footnote 4), or true yield (i.e., not yield *grade*). However, the USDA yield-grade standard is defined in reference to equation (1) so that any measure of "bias" needs to be defined in terms of this measure, rather than true yield.

actual grades for the data in our sample. We describe the source and generation process for these data in a subsequent section, but for now it is sufficient to note the apparent difference between called and actual yield-grade distributions. USDA graders tend to call slightly fewer 1's, 2's, and 4's than actually occur, but also tend to call more 3's. Yield-grade 5's rarely occur, nor are they called by graders often.

Figure 1: Unconditional called versus actual yield-grade distribution, and conditional error distribution.

Of course, the histograms in the first subplot of Figure 1 mask important information regarding the nature of grading error. For example, it is impossible to tell from this plot if graders call more 3's than actually occur because they mistakenly call 3 when yield grade is actually 2, or if instead they mistakenly call 3 when yield grade is actually 1, 4, or 5. Moreover, we cannot conclude anything about grader "bias" from looking at this first subplot. Graders cannot be expected to call grades without error, and the fact that more

3's are called than actually occur may be the result of a rational (and unbiased) decision process.

To better understand the nature of grading error in our sample, we also plot the distribution of called grades conditional on each actual grade. Throughout the paper, we will refer to this set of conditional distributions as the "conditional error distribution" of called grades. Figure 1 clearly demonstrates that graders error away from the endpoints of the grade distribution support. When yield grade is actually 1 or 5, graders tend more often to call 2 and 4, respectively. Similarly, graders are far more likely to call 3 rather than 1 when yield grade is actually 2, and to call 3 rather than 5 when yield grade is actually 4. The unconditional mean yield index of animals in our sample is 2.88, so another way to interpret the observed behavior of graders is to say that they error toward the mean.

This is perhaps not surprising given the task that graders perform. With limited time to call yield grade, and with no supporting physical measures of the relevant carcass attributes, graders must *estimate* the yield index, and then mentally perform the relevant calculation converting the yield index into yield grade (or possibly use some other "rule of thumb"). In carrying out such an estimation, and given graders' prior knowledge of the yield-grade distribution, it may be "efficient" in some sense to err in the manner observed. Thus, without specifying a model of how we think graders *ought* to behave, it is impossible to conclude that graders are "biased," or to evaluate graders' efficacy in any other fashion. Moreover, thus far we have documented apparent differences in the distributions of called and actual yield grades without saying anything about the economic and statistical significance of these differences. In the next section, we develop a behavioral model of grading that will allow us to address these questions.

Measuring Grader "Bias"

Suppose that graders, upon observing a carcass, receive an unbiased signal s of the true yield index y. Then we can write $s = y + \varepsilon$ with $E(\varepsilon) = 0$ and $E(\varepsilon^2) = \sigma_\varepsilon^2$. To make estimation tractable, we will assume that y is distributed independent of ε. In the empirical portion of our paper we will argue that it may be reasonable to think of the distribution of ε depending on y, and will test explicitly for one form of conditional dependence. We will further assume that y and ε are each normally distributed with the mean and variance of y given by β and σ_y^2, respectively. Then, s and y are jointly normally distributed with correlation coefficient $\rho = \sigma_y/(\sigma_\varepsilon^2 + \sigma_y^2)^{1/2}$.

Treating this signal generation technology as a maintained hypothesis, there are a variety of ways in which we might expect graders to behave. At a minimum, they ought to make use of their signals in forming an estimate of y. We consider four possibilities. To define these formally, we first let I_k represent the kth grade interval, where $I_1 = (-\infty, 2)$, $I_2 = [2, 3)$, $I_3 = [3, 4)$, $I_4 = [4, 5)$, and $I_5 = [5, \infty)$. The USDA yield-grade standard is then given by the function $g(y) = \{k | y \in I_k, k = 1, \ldots, 5\}$.

Of course, graders do not observe y and must instead make a "called" grade based on the signal, s, they receive. One strategy a grader can pursue is simply to call grade directly on the signal s. We let the function $c_1(s) = \{k | s \in I_k, k = 1, \ldots, 5\}$ represent this rule. Note that graders do not need any prior information about the distribution of y to implement this rule.

If, as seems reasonable, graders *do* have prior beliefs about the distribution of y, we might instead imagine that graders use this information together with the signal they receive to form an estimate of the true yield index y. If graders know the true distribution of the yield index, they can compute the expected value of y conditional on s, as $\hat{\beta}(s) \equiv \mathrm{E}(y|s) = \rho^2 s + (1 - \rho^2)\beta$.[6] Having performed this computation, we can let the function

---

[6] As we describe below, our data do not contain information that identifies graders, nor do we have any information about the timing of grading other than the year in which it was carried out. Thus, our data do

$c_2(s) = \left\{ k | \hat{\beta}(s) \in I_k, k = 1, \ldots, 5 \right\}$ represent an alternative grading rule. Letting $s_1 = 2$, $s_2 = 3$, $s_3 = 4$, and $s_4 = 5$ represent USDA grade cutoffs, and substituting the definition of $\hat{\beta}(s)$ into $c_2(s)$, we see that using graders using this rule make calls as under the previous rule (i.e., by reporting their signal), but subject to an adjusted set of grade standards $s_k^2 = \beta + (k + 1 - \beta)/\rho^2$. Thus the yield grade standard implicit in $c_2(s)$ converges to the USDA standard as ρ converges to 1, and we have the relationship $s_k^2 \gtrless s_k$ as $(k + 1) \gtrless \beta$.

If $c_2(s)$ is a reasonable description of how graders behave, then it should result in a called grade distribution similar to that in Figure 1. As noted in the previous section, we observe both the aggregate distribution of called-grade outcomes and the error distribution of called grades. To illustrate what the behavioral rule $c_2(s)$ implies for the called-grade distribution, suppose β is between 2 and 3. Then $s_k^2 < s_k$ for $k = 1$, while $s_k^2 > s_k$ otherwise. Thus, the intervals that define grades 1 and 5 get unambiguously smaller, while the interval for grade 2 gets bigger. In using the decision rule $c_2(s)$, graders will call more 2's, and fewer 1's and 5's, than actually occur. Without knowing more about specific parameter values for β and ρ, we cannot say what happens to the expected empirical frequency of grades 3 and 4.

While useful as starting points, neither of the behavorial rules so far involve much in the way of economic decision making. For example, it seems reasonable that grader attention is costly. In this context, graders potentially face a tradeoff between exerting costly effort (i.e., paying attention) and gaining some form of unspecified benefit from accurate grading. Similarly, given that grading error is inevitable, how would the USDA like graders to behave? One sensible and intuitive possibility is that graders behave as though they choose the grade that is most likely conditional on the information they receive from observing a carcass. In

---

not allow us to examine whether graders update their beliefs in some fashion. For example, one might imagine daily or weekly updates depending on the characteristics of carcasses that arrive in a given period. Also, there may be structural differences in the beliefs of relatively inexperienced and experienced graders. Both possibilities represent interesting potential future directions for further empirical work. In any case, graders typically observe hundreds of carcasses per day and presumably have a good sense of the distribution of the yield index. Assuming a degenerate prior on the distributions of β and σ_ε therefore seems like a reasonable starting point.

terms of our model, this rule can be formulated as

$$c_3(s) = \arg\max_k \left\{ \text{Prob}\left(g(y) = k|s\right), \ k = 1, \ldots, 5 \right\}.$$

To see what this rule implies regarding grader behavior relative to this previous two, first note that

$$\text{Prob}\left(g(y) = k|s\right) = \text{Prob}\left(y \in I_k|s\right) = \int_{I_k} f(y|s)dy,$$

where $f(y|s) = \text{N}\left(\hat{\beta}(s), \sigma_y^2(1 - \rho^2)\right)$. Thus, we can write

$$\text{Prob}\left(g(y) = k|s\right) = \int_{I_k} \phi\left(\frac{y - \hat{\beta}(s)}{\sigma_y\sqrt{1 - \rho^2}}\right) dy.$$

We are now in a position to determine cutoff values s_k^3, $k = 1, \ldots, 5$, analogous to those above for behavioral rules 1 and 2.

When $s = -\infty$ clearly grade 1 is the most likely true grade. As s increases, the conditional distribution of y shifts to the right and the probability that true grade is 1 falls. We define the point s_1^3 as the signal value such that grades 1 and 2 are equally likely,

(2) $$2\Phi\left(\frac{2 - \hat{\beta}(s_1^3)}{\sigma_y\sqrt{1 - \rho^2}}\right) = \Phi\left(\frac{3 - \hat{\beta}(s_1^3)}{\sigma_y\sqrt{1 - \rho^2}}\right).$$

As s increases beyond this point, grade 2 becomes the most likely true grade. Analogously, as s increases still further we can define the point s_2^3 such that the probability of grade 2 just equals the probability of grade 3. The grader will be indifferent between grade 2 and grade 3 when the conditional distribution of y has equal mass below 3 and above 3. Given that the normal distribution is unimodal and symmetric, this will happen $\hat{\beta}(s_2^3) = \rho^2 s_2^3 + (1 - \rho^2)\beta = 3$. Rearranging this expression, it is easy to see that $s_2^3 = s_2^2$. Analogously, $s_3^3 = s_3^2$, while s_4^3 is determined implicitly by the equation,

(3) $$2\Phi\left(\frac{5 - \hat{\beta}(s_4^3)}{\sigma_y\sqrt{1 - \rho^2}}\right) = 1 + \Phi\left(\frac{4 - \hat{\beta}(s_4^3)}{\sigma_y\sqrt{1 - \rho^2}}\right).$$
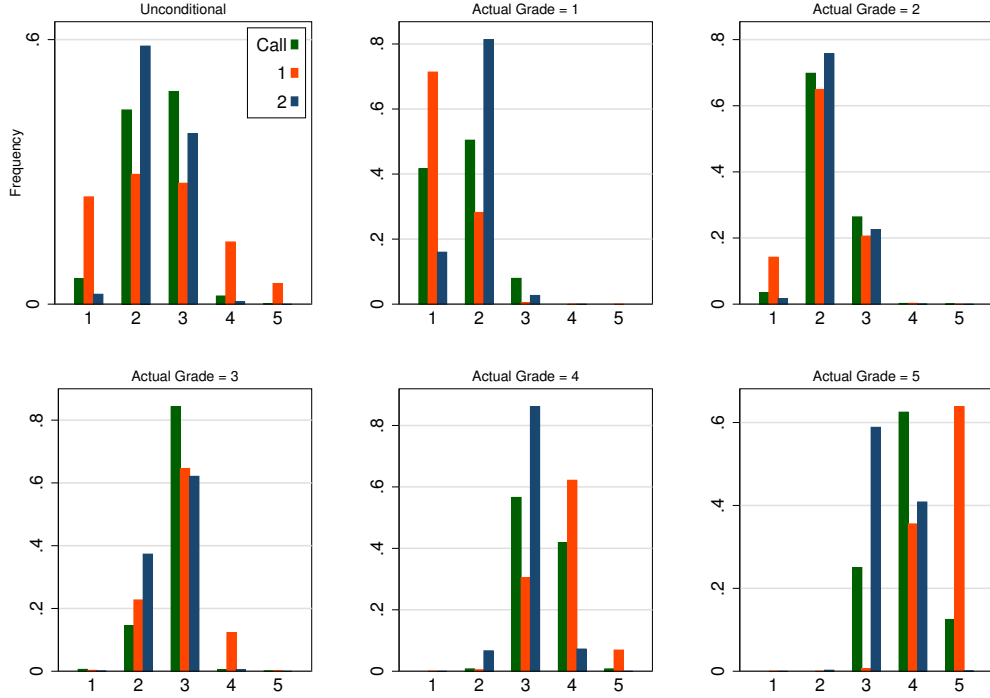
11

In summary, we see that assuming graders choose the "most likely" grade results in a similar, though not identical, behavioral rule relative to $c_2(s)$. Under $c_3(s)$ graders evaluate extreme grades differently than under $c_2(s)$.

Behavioral rules 2 and 3 can be interpreted as modified versions of $c_1(s)$, where graders effectively choose a different set of grade cutoffs for s. However, in both cases we impose significant structure on the way those cutoffs are determined. Possibly some other model that implies a different set of cutoff criteria can better account for grader behavior. Thus, one natural specification test for our models is to compare each with an alternative model where we do not impose *any* structure on how the grade cutoffs are determined. That is, suppose graders use the rule $c_4(s) = \left\{ k | s \in \hat{I}_k, k = 1, \ldots, 5 \right\}$, where $\hat{I}_k$ are considered unknown. In this last model, we can estimate s_k directly, letting the data determine what cutoffs graders most likely use.

It is informative to compare the conditional error distributions generated by $c_m(s)$ with that in Figure 1. The relevant conditional probabilities are given by $\text{Prob}\left(c_m(s) = k \mid g(y) = j\right)$, and can be computed given a particular parameterization for the joint distribution of y and s. Figure 2 graphically displays these probabilities for models 1 and 2, together with the called-grade conditional-error distribution from Figure 1. We let $\sigma_\varepsilon = 1$ and, for model 2, assume graders are unbiased with β and σ_y set equal to the sample mean and standard deviation of y.[7] Graders using $c_1(s)$ call to many extreme grades, while those using $c_2(s)$ do not call enough. Overall, model 2 performs well conditional on actual grades 2 and 3, relative to other actual grades. However, note how even here model 2 calls too many 2's and not enough 3's. Thus, increasing β in model 2—or in other words, introducing "grader bias"—represents one way to better match the distribution of graders' calls. Doing so will lead the model to predict more 3's and fewer 2's.

---

[7] Computing predictions for model 4 requires setting arbitrary values for s_k^4, and so is not informative with respect to the effect of grader bias. Model 3 results in nearly identical predictions as those in model 2.

12

Figure 2: Actual versus predicted called grade distributions. Predictions based on $\beta = 2.88$.

More generally, given specification of distributions for y and ε, each of the models we have specified generates different predictions regarding the joint sampling distribution for called grade c and true yield y. Thus, we can use observation on c and y to estimate the parameters in each model. There is only a single parameter in model 1, and no sense in which we can determine whether graders are biased from estimation of this single parameter. We present this model only as a benchmark where graders do not use prior information regarding the distribution of y. In models 2 and 3, bias can be assessed by comparing the sample mean and standard deviation of y with estimates of the same parameters (β and σ_y) from each respective model of grader behavior. As we explain below, it turns out that σ_y cannot be identified in our estimation, but we *can* obtain unbiased estimates of a representative grader's (degenerate) implied prior for β. Model 4 is used as a specification test against models 2 and 3. In effect, we examine to what extent these two models are missing something relative to

13

some other model of grader behavior. However, given that we do not take a stand on how grading standards are determined in model 4, there again is no sense in which we can make a claim about bias one way or another from this model alone.

Empirics

Data

The data for our analysis come from a sample of loads delivered to three different packing plants in the U.S. midwest during the years 2000 to 2002. During this period, a small group of southern Iowa cattle producers paid a third party to collect carcass attributes of their slaughtered animals beyond those reported in grid-pricing closeout sheets. In particular, for each animal delivered by participating producers, the third party measured and reported (among other things) the characteristics needed to compute the actual yield index according to equation (1).[8] Additionally, our data include the USDA quality and yield grades attributed to these same animals, and details of the relevant grid payment structure.

This sample is not randomly selected and thus potentially not representative of all loads delivered to the relevant plants. Results subsequently reported should be interpreted with this caveat in mind. Additionally, the presence of third parties (paid by cattle producers) could alter graders' behavior. However, discussions with plant and USDA representatives suggest this is not a concern. The total number of animals delivered by the producers in our sample represents a small fraction of the total animals graded on any given day, and it is not uncommon for producers to request measurement of carcass attributes beyond yield and quality grade. Additionally, if anything, one might expect that the presence of a third party should reduce bias, if any exists.

---

[8]For more detail on the rationale for producers to collect such data, and on other activities of the relevant group of growers, see Hueth and Lawrence (2006).

Table 1: Distribution of called grades across plants and years.

Percent with	Plant A			Plant B			Plant C		
Yield Grade	2000	2001	2002	2000	2001	2002	2000	2001	2002
1	0.0714	0.1538	0.0768	0.0084	0.0710	0.0500	0.0331	0.0860	0.0492
2	0.4670	0.4615	0.4852	0.4202	0.6557	0.3625	0.3995	0.3937	0.4007
3	0.4267	0.3776	0.4057	0.5462	0.2732	0.5750	0.5496	0.5204	0.5431
4	0.0330	0.0070	0.0323	0.0210	0.0000	0.0055	0.2250	0.0000	0.0036
5	0.0018	0.0000	0.0000	0.0042	0.0000	0.0000	0.0018	0.0000	0.0000
Total Animals	546	143	742	238	183	80	1119	221	569

Table 1 summarizes the data in our sample. Overall, we have observations on $n = 3,841$ graded carcasses, including 1,431 animals delivered to Plant A, 501 delivered to Plant B, and 1,909 delivered to Plant C. With the exception of deliveries to Plant B in 2001, yield-grade 2 accounts for roughly 40 percent of graded carcasses. There is slightly more variation across years and plants for yield-grade 3 calls, but in over half of the cases reported in Table 1, yield-grade 3 accounts for at least 50 percent of graded carcasses. Yield-grade 1 is called on between 1 and 15 percent of graded carcasses, and yield-grade 4 is called on 3 percent or less of graded carcasses, with a notable exception in Plant C during the year 2000. Yield-grade 5 is called on an almost negligible number of graded carcasses. In the following section, we evaluate whether the called grades summarized in Table 1 and Figure 1 can plausibly be generated by graders' who employ one of the behavioral rules $c_m(s)$, $m = 1, \ldots, 4$, discussed in the previous section, and explicitly test for the presence of grader bias.

Estimation

Our basic unit of observation is a graded carcass, and for each of these we observe the carcass the yield index y,[9] and a USDA called grade c. Of course, the signal s observed

---

[9]Actually we observe measures of each carcass attribute (fat thickness, kph, carcass weight, and ribeye area) that is used to define the USDA yield-grade standard, and then compute y from equation 1. Each of these attributes are presumably measured with some degree of error. Given the additive nature of y in our likelihood function below, we can accommodate this error by interpreting ε as the sum of errors due to

by the grader is an unobserved latent variable from the perspective of an econometrician. Each of our behavioral models can be formulated to have the same basic structure, with graders calling grade k when s falls in the kth interval, I_k^m, the definition of which varies across models. The parameters to estimate also vary across models. We let θ^m represent the relevant set of parameters for the mth model. Thus, we have $\theta^1 = \sigma_\varepsilon$, $\theta^m = (\sigma_\varepsilon, \sigma_y, \beta)$ for models 2 and 3, and $\theta^4 = (\sigma_\varepsilon, s_1^4, s_2^4, s_3^4, s_4^4)$.

For the mth model, we can compute the probability of observing a particular grade and yield index pair (c, y) conditional on given parameter vector θ^m as

$$
\begin{aligned}
\ell^m(c, y | \theta^m) &= \prod_{k=1}^{5} \mathbb{1}\{c = k\} \text{Prob}\left[s \in I_k^m | \theta^m\right] \\
&= \mathbb{1}\{c = 1\} \text{Prob}\left[\varepsilon < s_1^m - y | \theta^m\right] \times \mathbb{1}\{c = 5\} \text{Prob}\left[\varepsilon > s_4^m - y | \theta^m\right] \times \\
&\quad \prod_{k=2}^{4} \mathbb{1}\{c = k\} \text{Prob}\left[s_k^m - y < \varepsilon < s_{k+1}^m - y | \theta^m\right] \\
&= \mathbb{1}\{c = 1\} \Phi\left(\frac{s_1^m - y}{\sigma_\varepsilon}\right) \times \mathbb{1}\{c = 5\}\left[1 - \Phi\left(\frac{s_4^m - y}{\sigma_\varepsilon}\right)\right] \times \\
&\quad \prod_{k=2}^{4} \mathbb{1}\{c = k\}\left[\Phi\left(\frac{s_{k+1}^m - y}{\sigma_\varepsilon}\right) - \Phi\left(\frac{s_k^m - y}{\sigma_\varepsilon}\right)\right],
\end{aligned}
$$

(4)

where with a slight abuse of notation we have suppressed the dependence of s_k^m on the parameter vector θ^m. We estimate the unknown parameters θ^m by maximizing the log of the product of these likelihoods across all loads i delivered to each plant k during year t. We first estimate a base model where parameters are fixed across all plants and years, and then a model that allows for some degree of heterogeneity.

For model 1, s_k^1, $k = 1, \ldots, 5$, are fixed numbers and our only task is estimation of σ_ε. Model 4, which nests model 1, is more difficult numerically, but still relatively straightforward. From an estimation perspective, this model is very much like an ordered probit. For model 2 recall that $s_k^m = \beta + (k + 1 - \beta)/\rho^2$. The parameter σ_ε is difficult to estimate in

---

measurement of the relevant carcass attributes plus error due to graders' limited capacity for perceiving the true y.

this model because, from (4), it mostly just scales the other parameters. To see this more clearly, consider the normalization $\sigma_\varepsilon = 1$. Then we can rewrite $(s_k^2 - y)/\sigma_\varepsilon$ as $\beta_0 + \beta_1 k - y$, where $\beta_1 = 1/\rho^2$ and $\beta_0 = \beta_1 - \beta(1 - \rho^2)/\rho^2$. In this case, we can identify β_0 and β_1, and recover ρ and β. However, from our definition for ρ, we have $\sigma_\varepsilon = \sigma_y(1/\rho^2 - 1)^{1/2}$. Thus, choosing the normalization $\sigma_\varepsilon = 1$ generates a biased estimate for ρ that is below (above) its true value when the true value for σ_ε is greater (less) than 1. In principle, it may be possible to identify σ_ε and σ_y in model 2, though we have so far been unsuccessful in attempts to do so. Thus, in estimating models 2 and 3 we set $\sigma_\varepsilon = 1$ and treat ρ as a nuisance parameter. We focus on the estimated value of β in reporting our results.

Model 3 is further complicated by the fact that there are no closed form solutions for the cutoffs s_1^3 and s_4^3 implicitly defined in equations (2) and (3). In estimating the parameters in θ^3, it is necessary to find these cutoff values at each iteration on the likelihood function in (4). To accomplish this we construct a pair of approximations to the implicit functions $s_k^3(\rho, \beta)$ defined in (2) and (3) using the bivariate Chebyshev function approximation algorithm in Judd (1999, pg. 238). Using estimation results from model 2 as a guide, we compute each approximation (with accuracy of order 1e-9) on an interval for ρ between 0.3 and 0.8, and for β between 2 and 4.

Results

Results for our base model where grader behavior across all plants and years is assumed constant are reported in Table 2. The first row contains estimates of σ_ε based on the behavioral model $c_1(s)$. Over the entire sample, our estimate for this value is .449, and this compares with a sample standard deviation of y equal to .654. Thus, called grades are consistent with a signal generation technology where variation in y accounts for about 82 percent of the total variation in s. The last row reports results from model 4 which relaxes the restriction in model 1 that graders use USDA grade standards for y to make calls with

Table 2: Estimation results.

α	β	σ_ε	$\mathbf{s_1^*}$	$\mathbf{s_2^*}$	$\mathbf{s_3^*}$	$\mathbf{s_4^*}$	$\mathbf{\ln L}$
			Model 1				
na	na	.449	2	3	4	5	-2935.89
		(.007)					
			Model 2				
.456	3.034	na	.764	2.959	5.154	7.349	-2712.91
(.006)	(.016)						
			Model 3				
.451	3.038	na	.769	2.954	5.170	7.353	-2709.19
(.006)	(.016)						
			Model 4				
na	na	.552	1.543	2.864	4.680	5.549	-2462.89
		(.013)	(.029)	(.013)	.(042)	(.110)	

respect to realized values for s. In this model, our estimated value for σ_ε increases slightly. Estimated values for s_1 and s_2 are both below their respective USDA counterparts for y, while the values for s_3 and s_4 are both above their respective counterparts. As expected, graders effectively use relatively large intervals for grades 2 ($s_2 - s_1 = 1.321$) and 3 ($s_3 - s_2 = 1.816$) and small intervals for grades 1, 4 ($s_4 - s_3 = 0.869$), and 5.

The next two rows report estimation results from models 2 and 3. The coefficient α refers to our biased estimate for ρ^2. The estimate of β from model 2 compares with a sample mean of y equal to 2.878. Standard errors on our estimate of β are extremely small so that this value lies well outside the 95% confidence interval for β given by [3.002,3.066]. This is not quite enough to conclude that there is a significant difference between the estimated values for β and $\bar{y}$. To construct a hypothesis test for this difference we need to take account of the fact that the sample mean of y is also an estimated value, and that it is correlated with our estimate of β. Thus, we also construct a 90 percent confidence interval for $\beta - \bar{y}$ based on a bootstrap with 1000 replications. This yields a 95% confidence interval of [.130,.180]. We therefore reject the hypothesis that graders are unbiased. The distribution of called grades,

conditional on realized values for the yield index y, suggests that graders use an estimate of the population value for β that is too large.[10]

In terms of point estimates for β and α, results for model 3 are virtually identical to those for model 2. This is perhaps to be expected given the similarity of the two behavioral rules. Although model 3 predicts different behavior at the extremes, most called grades are 2's and 3's so that much of the information used to estimate β comes from these observations. However, this similarity in point estimates does not preclude the possibility of distinguishing the two models statistically. We use the Vuong (1989) nonnested test for model specification to make a comparison. Letting $LR_{mp} = \sum_{i=1}^{n} [\ln \ell_i^m - \ln \ell_i^p]$, and $\hat{w}_{mp} = n^{-1} \sum_{i=1}^{n} [\ln (\ell_i^m / \ell_i^p)]^2 - [n^{-1} \sum_{i=1}^{n} \ln (\ell_i^m / \ell_i^p)]^2$, the relevant test statistic is computed as $V_{mp} = LR_{mp} / \hat{w}_{mp}$. A large value of V_{mp} supports model m, while a small value supports model p. Under the null that the two models are indistinguishable, V_{mp} is distributed standard normal.
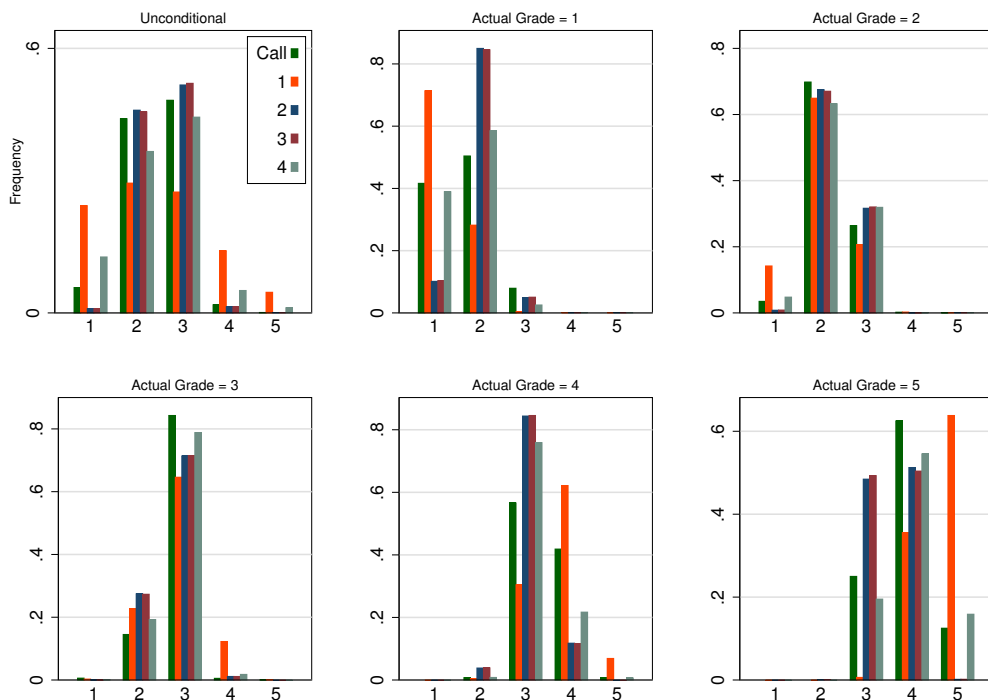
Comparing models 3 and 2, we get a value of $V_{32} = 10.45$ so that we reject the hypothesis that the models are the same in favor of model 3. Similarly, comparing models 3 and 1, we get a value of $V_{31} = 6.40$, so that model 1 is rejected in favor of model 3. Similarly, $V_{21} = 6.30$; model 2 also does better than model 1. However, model 4 does still better than models 2 and 3 with $V_{42} = 10.91$ and $V_{43} = 10.79$. Model 1 is nested in model 4 so we can use a simple unadjusted likelihood ratio test between these two models. In this case, we get test statistic of 946, so that model 4 clearly dominates model 1. In summary, models 2 and 3 do a much better job of explaining grader behavior than model 1. This suggests that graders do attempt a prediction that is based in some form on priors they hold regarding the underlying distribution of yield. Model 3 does better than model 2 so that graders seem

---

[10]A reviewer correctly pointed out that our results might also be intereperted as *sample* bias. That is, the data we use may not be representative of the cattle graders observe. We acknowledge this possibility, but also note that aggregate proportions of yield grades 1, 2, 3, 4, and 5 were 9.0, 41.9, 44.9, 3.7, and 0.5 percent for US beef markets in 2002 (U.S. Department of Agriculture, 2002). This compares with our data where the analogous percentages are 5.75, 44.05, 48.24, 1.85, and 0.10. Thus, our data contain slightly fewer yield grade 1's, 4's, and 5's, and slighly more yield grade 2's and 3's, but in general are quite similarly distributed across the various yield grades.

to make adjustments on extreme grades that reflect an attempt to choose the most likely grade. However, model 4 does significantly better than models 2 and 3 suggesting there are aspects of grader behavior that these models are not capturing.

Figure 3 presents results visually by plotting actual versus predicted called grade distributions for both aggregate calls, and conditional error distributions of calls across each of the behavioral models. First note that relative to Figure 2, model 1 still predicts far too many 1's and 5's relative to the calls graders actually make. Likewise, neither model 2 or 3 predict *enough* 1's and 5's. When actual grade is 1 or is 5, graders sometimes *do* call these grades. Only model 4 is able to account for this behavior. However, relative to Figure 2, models 2 and 3 do much better with respect to predictions for grades 2 and 3. Of course, this is the result of β being set to its estimated value from our grader model, 3.03, rather than the mean of y, 2.88.

Figure 3: Actual versus predicted called grade distributions. Predictions based on $\beta = 3.03$.

These observations suggest a model where behavior differs in some way at the extremes of the distribution for y. One simple way to accommodate such an effect is to allow σ_ε to vary with y. For example, conditional dependence of the distribution for ε on y might result from graders being "more careful" on carcasses that appear to be outliers.[11] Table 3 reports estimation results for models 2 and 4 where in model 2 we let α be an index of a constant plus a pair of indicator variables $\mathbb{1}\{y < 2\}$ and $\mathbb{1}\{y > 3.5\}$ that take on the value 1 for extreme values of y, and similarly for σ_ε in model 4. For model 2, it is not possible to interpret the sign of the coefficients for each indicator variable. The parameter α is a biased estimate for ρ^2 which itself is a nonlinear function of σ_ε. However, given the strong significance of the coefficient on the indicator for $y < 2$, there does seem to be some degree of conditional dependence. Moreover, allowing for this form of dependence does not alter estimation results with regard to grader bias. Our estimate of β in fact goes up slightly, relative to the base model results in Table 2.

In model 4, coefficients on both indicator variables are highly significant and opposite in sign. Graders' signals are less noisy for low-yielding (high y) carcasses, and more noisy for high-yielding (low y) carcasses. This outcome is consistent with graders being exceptionally careful when making calls on carcasses that "look bad," but relatively careless when carcasses "look good."

Table 4 reports results from model 2 allowing for plant and year heterogeneity in β.[12] The point of results in this table is to examine whether "grading standards" vary across plants and years. There are three plants and three years for each plant. Overall, we find evidence of bias in all but one instance; for Plant B in 2001 we cannot reject unbiasedness. Interestingly,

---

[11] Although graders normally assess carcasses via visual inspection, one individual we spoke with noted how graders sometimes resort to using a ruler when they see an unusual carcass. Most modern packing plants employ a "tagger" whose job is to identify potential grader miscalls, and to request a regrade. Potentially, taggers demand relatively more regrades when grades 4 or 5 are called, relative to grade 1. If regrades are costly for graders, this sort of behavior by taggers may lead graders to pay more careful when grading low-yield (high yield-grade) carcasses. We thank an anonymous referee for calling our attention to the role of taggers.

[12] We attempted an analogous estimation using model 3, but encountered convergence difficulty in maximizing the relevant likelihood function.

Table 3: Estimation results: models 2 and 4 with signal generation technology conditional on y.

Parameter	Estimate	(Std. Err.)
Model 2, ln L = -2645.01		
β	3.076	(0.018)
α		
Coeff. on indicator, $y < 2$	0.237	(0.027)
Coeff. on indicator, $y > 3.5$	-0.002	(0.017)
Intercept	0.421	(0.008)
Model 4, ln L = -2422.79		
s_1	1.533	(0.034)
s_2	2.892	(0.013)
s_3	4.566	(0.038)
s_4	5.324	(0.097)
σ_ε		
Coeff. on indicator, $y < 2$	0.302	(0.052)
Coeff. on indicator, $y > 3.5$	-0.107	(0.024)
Intercept	0.536	(0.015)

the direction of bias reverses in Plant A moving from year 2000 to 2001, and remains reversed in 2002. For this plant, graders place too low a value on β. Results in each of the remaining plant-years are similar to those obtained from estimation without plant-year effects. There is evidence of bias across all years in Plant C, and during the years 2000 and 2002 for Plant B. The magnitude of bias is largest in these last two cases, however this is also where we have the fewest number of observations (though even here standard errors on our estimates are extremely low). On average across the years, graders in Plant 2 appear to be more strict than in the other plants; it is harder to achieve a "good" yield grade in this plant.

To this point we have said nothing about the potential economic importance of "grader bias." Toward this end, Table 5 reports average yield premiums under actual and called grade distributions for four alternative yield premium grids.[13] Although we find evidence of grader bias, expected yield premiums are remarkably close to actual yield premiums. With the exception of grid number 4, the empirical distribution for called grades results in an

---

[13]We do not associate the grids with plant-years to preserve plant anonymity.

Table 4: Estimation results: model 2 with plant and year heterogeneity.

Year	β	$\bar{y}$	$\beta - \bar{y}$ [95% C.I.]		s_1^*	s_2^*	s_3^*	s_4^*
Plant 1								
2000 (n=238)	2.961	2.881	0.019	0.144	-0.238	2.207	4.439	6.670
2001 (n=183)	2.675	2.801	-0.229	-0.017	0.294	2.525	4.756	6.987
2002 (n=80)	2.747	3.069	-0.369	-0.276	0.553	2.784	5.016	7.247
Plant 2								
2000 (n=546)	3.644	2.417	1.095	1.364	0.648	2.879	5.110	7.341
2001 (n=143)	2.713	2.696	-0.077	0.115	0.708	2.939	5.170	7.401
2002 (n=742)	3.386	2.631	0.584	0.928	0.816	3.048	5.279	7.510
Plant 3								
2000 (n=1119)	3.175	2.900	0.232	0.316	1.080	3.311	5.543	7.774
2001 (n=221)	3.098	2.717	0.287	0.477	1.123	3.354	5.585	7.816
2002 (n=569)	3.050	2.948	0.042	0.159	1.169	3.400	5.631	7.862
$\alpha = 0.448$, ln L=-2594.51								

expected yield premium that is within $.15/cwt of the expected premium computed with the true distribution of grades. For two grids, the distribution of called grades results in a lower average yield premium than under the actual distribution of grades, and for the other two grids, the reverse is true. During the last decade, net returns for cattle producers in Iowa has averaged about $25/head (Lawrence, 2006). Thus, for an animal weighing 1,250 lbs., an economic bias of $.10/cwt translates into about $1.25/head, or approximately 5% of net returns.

Note that the results in Table 5 suggest that graders may be unbiased based on some behavioral model other than the one we specify. For example, the relevant model might be minimization of a loss function of prediction error for yield premiums. Alternatively, it is conceivable that packers respond to grader bias by developing payment grids that minimize the effect of such bias. This latter case is important, because it suggest the possibility of important welfare consequences from grader error, despite the small difference between yield premiums expected under USDA grader calls versus an "unbiased" grader. Finally, the structure of yield premiums in a typical grid and the results reported in Table 3 suggest that graders may face "social pressure" (e.g., Garicano et al., 2005) in packing plants with

Table 5: Expected yield premiums under four alternative yield premium grids.

	Grade					Average	
Grid	1	2	3	4	5	Actual	Called
	——$/cwt——						
1	5.00	3.00	0.00	-15.00	-25.00	1.35	1.31
2	3.75	3.25	0.00	-15.00	-20.00	1.37	1.35
3	5.00	3.00	0.00	-20.00	-25.00	1.19	1.21
4	4.00	1.50	0.00	-20.00	-25.00	0.37	0.50

respect to yield grades 4 and 5. In particular, from Table 5, it seems clear that packers seek to discourage the realization of yield grades 4 and 5. Large penalties are imposed on growers who deliver animals that grade these values. Recognizing this, graders may feel less inclined to call these grades, knowing that doing so will call attention to their activities.

Conclusion

This paper uses data from a sample of loads delivered to three different packing plants in the U.S. midwest to examine the extent of USDA grader bias in cattle markets. Our data contain carcass characteristics sufficient to compute true yield grade, which we then compare with USDA called grades for each of 3,841 graded carcasses. In our analysis, we define "bias" in reference to four alternative behavioral grading rules that specify how graders call grades, given that they observe an imperfect measure of the true yield index.

Overall, we find that our behavioral rules are capable of explaining important qualitative features of the observed distribution of called grades. Most importantly, our models allow for grading error. When the actual grade is 1, for example, graders often call a grade of 2 or 3. Our behavioral models capture this feature. Our base model assumes graders receive an unbiased signal of the true yield index and report this value, with no attempt at "prediction" based on prior information about its distribution. Two additional models, increasing in degree of computation complexity, assume graders make a prediction using

knowing the true distribution of the yield index. For these two models, estimation results suggest that graders implicitly use an estimate for the mean of the distribution for the USDA yield index that is significantly above its sample mean. In this sense, we find evidence of "grader bias." The last of our models in some ways nests the others, and is more flexible in it's ability to capture grader behavior. However this flexibility comes at a cost, because there is no underlying behavioral model that can be used to make normative statements about graders' behavior. Although we are able to capture much of graders' behavior in models two and three, this fourth model statistically dominates all others. This is mostly the result of grader behavior at extreme values for the distribution of yield, or in other words at grades 1, 4, and 5.

We investigate the possibility that some of this discrepancy is the result of graders changing their behavior conditional on carcass characteristics and find evidence to support this hypothesis. In particular, estimation results suggest that graders receive a more precise signal of true yield when yield is very bad (grades 4 or 5), and a less precise signal of yield when yield is very good (grade 1). Given that financial penalties are much higher for very bad yields grades than are corresponding premiums for very good grades, this outcome is consistent with the intuition that graders want to avoid drawing attention to their activities and exert extra effort to make sure their call is correct when calling grades 4 and 5. In any case, allowing for this effect does not alter the magnitude or statistical significance of our measure for "grader bias."

When we evaluate the economic importance of graders bias, it turns out that the errors graders make generally have a small impact on expected yield premiums and discounts. For example, comparing expected yield premiums under the actual distribution of yield grade and the distribution of yield grade induced by graders' calls, differences range between \$.13/cwt and \$.02/cwt, depending on the specific set of yield-grade premiums we consider. Though we do not evaluate the statistical significance of these differences, they are in relation to average net returns on the order of \$2/cwt, and so represents about 5% of average profit

per head. Nevertheless, there is still the possibility for market distortion if current grids are endogenous to the degree of grader bias and measurement error. We have no way to assess how grid structures (and hence, indirectly, production outcomes) might change in the presence of an unbiased measurement technology.

Finally, an obvious direction for future research is consideration of the source of grader bias. We note three potential directions for such research. First, graders are taught particular "rules of thumb" for calling grades, and are periodically evaluated for performance. Depending on the institutional objectives of the USDA in its grading program, this training *cum* oversight and incentives may induce some form of bias. Alternatively, it seems plausible that packing plant staff either directly or indirectly influence grader calls. In particular, most modern packing plants employ a "tagger" whose job is to identify potential grader miscalls, and to request a regrade. Thus, for example, if packers prefer that high-yield animals be classified as low yield, but not vis-versa, the tagger's presence may introduce bias. Finally, graders are human beings, and thus potentially exhibit some form of perceptual bias. For example, Tversky and Kahneman (1992) emphasize the importance of reference points in human evaluation of probabilities. In our framework, 90% of graded animals have an actual yield grade of 2 or 3 so that it is natural to interpret the mean yield index (2.88) as a reference point for graders. Thus, the evidence of bias we find may be partly explained by a prediction rule that puts "too much" weight on the central portion of the distribution for the yield index. Exploration of these and other hypotheses regarding the source of observed bias are the subject of ongoing research.

References

Abraham, H., C. Murphy, H. Cross, G. Smith, and W. Franks (1980). Factors Affecting Beef Carcass Cutability: An Evaluation of the USDA Yield Grade for Beef. *Journal of Animal Science 50*(5).

Bockstael, N. (1984). The Welfare Implications of Minimum Quality Standards. *American Journal of Agricultural Economics 66*, 231–250.

Bogetoft, P. and H. B. Olesen (2003). Incentives, Information Systems, and Competition. *American Journal of Agricultural Economics 85*(1), 234–247.

Chalfant, J., J. James, N. Lavoie, and R. Sexton (1999, July). Asymmetric Grading Error and Adverse Selection: Lemons in in the California Prune Industry. *Journal of Agricultural and Resource Economics 24*, 57–79.

Chalfant, J. and R. Sexton (2003, February). Marketing Orders, Grading Errors, and Price Discrimination. *American Journal of Agricultural Economics 84*(1), 53–66.

Diaz, E. F., B. Brorsen, K. Anderson, F. Richter, and P. Kenkel (2002). The Effect of Rounding on the Probability Distribution of Regrading in the U.S. Peanut Industry. *Journal of Agribusiness 20*(1), 103–116.

Espinosa, J. and B. Goodwin (1991). Hedonic Price Estimation for Kansas Wheat Characteristics. *Western Journal of Agricultural Economics* (1), 72–85.

Freebairn, J. (1973). The Value of Information Provided by a Uniform Grading System. *Australian Journal of Agricultural Economics 17*, 127–139.

Gardner, B. (2003, August). U.S. Food Quality Standards: Fix for Market Failure or Costly Anachronism? *American Journal of Agricultural Economics 85*(3), 725–730.

Garicano, L., I. Palacios-Huerta, and C. Prendergast (2005, May). Favortism Under Social Pressure. *Review of Economics and Statistics 87*(2), 208–216.

Hale, D. S., K. Goodson, and J. W. Savell (2006). Beef Quality and Yield Grades. `http://meat.tamu.edu/beefgrading.html#yield` (accessed April 19, 2006).

Hennessy, D. (1995). Microeconomics of Agricultural Commodity Grading: Impact on the Marketing Channel. *American Journal of Agricultural Economics 77*, 980–989.

Hollander, A., S. Monier, and H. Ossard (1999). Pleasures of Cockaigne: Quality Gaps, Market Structure, and the Amount of Grading. *American Journal of Agricultural Economics 81*(3), 501–511.

Hueth, B. and J. Lawrence (2006). Information Transmission in Cattle Markets: A Case Study of the Chariton Valley Beef Alliance. *Journal of Agribusiness 24*(1).

Judd, K. L. (1999). *Numerical Methods in Economics*. MIT Press.

Lawrence, J. D. (2006). Estimated Livestock Returns. `http://www.econ.iastate.edu/faculty/lawrence/EstRet/` (accessed April 24, 2006).

Ligon, E. (2002). *Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, Chapter Quality and Grading Risk. Kluwer Academic Publishers.

Marette, S. and J. M. Crespi (2003). Can Quality Certification Lead to Stable Cartels. *Review of Industrial Organization 23*, 43–64.

McDonald, R. A. and T. C. Schroeder (2003, April). Fed Cattle Profit Determinants Under Grid Pricing. *Journal of Agricultural and Applied Economics 35*(1), 97–106.

Meadows, L. (2005). Personal Communication. USDA, Market Grading and Certification Branch.

Murphy, C., D. Hallett, W. Tyler, and J. Pierce (1960). Estimating Yields of Retail Cuts from Beef Carcasses. Technical report. Presented at the 62nd Meet. of the Amer. Soc. of Anim. Prod., Chicago, November, 26.

Schroeder, T. C., T. L. Marsh, and J. Mintert (2000, March). Beef Demand Determinants: A Research Summary. Cooperative Extension Report MF-2457, Kansas State University.

Tversky, A. and D. Kahneman (1992). Avances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

U.S. Department of Agriculture (1997). *United States Standards for Grades of Carcass Beef*. Washington, D.C.: U.S. Department of Agriculture, Agricultural Marketing Service, Livestock and Seed Division.

U.S. Department of Agriculture (2002). National Summary of Meats Graded. http://www.ams.usda.gov/lsg/mgc/mgc-pubs.htm.

Vuong, Q. H. (1989, March). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica 57*(2), 307–333.

Whitley, J. E. (2002). The Political Economy of Quality Measurement: A Case Study of the USA Slaughter Cattle Market. *The Australian Journal of Agricultural and Resource Economics 46*(4), 515–538.