# ABSTRACT

Belasco, Eric Joseph. Three Essays on Modeling Risk in Fed Cattle Production. (Under the direction of Barry K. Goodwin).

This study examines issues that arise when modeling risk associated with fed cattle production. While research concerning crop yield and revenue risks are numerous, studies focusing on production risk in livestock are much less frequent. The first essay evaluates the relationship among four variables associated with the health and performance of feedlot cattle, and the resulting production and profit risk. The four variables of interest include feed conversion rates, average daily gain, veterinary costs, and mortality rates, which are conditional on characteristics that are known when the pen is placed into a commercial feedlot. Conditional variables include gender, average weight, feedlot location, and season of placement. A multivariate Tobit model is used to characterize the relationship among the four dependent variables, where each element in the covariance matrix is conditional on placement characteristics. The second essay focuses on modeling cattle mortality rates, alone and as part of a system. A zero-inflated log-normal distribution is developed and shown to have advantages in model fit and prediction tests with this data set, relative to classical methods. A simulated data set is also utilized to assess the potential bias from assuming a Tobit model when the data are more accurately characterized through a mixture model. The third essay quantifies the amount of risk inherent in cattle feedlot operations through the use of simulation techniques. More specifically, *ex-ante* profit risks are evaluated under scenarios that utilize varying levels of price protection through the use of forward contracts and the options market.

# Three Essays on Modeling Risk in Fed Cattle Production

by

**Eric J. Belasco**

A Dissertation submitted to the Graduate Faculty of
North Carolina State University
in Partial Fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2007

APPROVED BY:

| | |
|---|---|
| Sujit K. Ghosh | Walter N. Thurman |
| Nicholas E. Piggott | Barry K. Goodwin, Chair |

# Dedication

To Megan, Analise, and Emery

# Biography

Eric Joseph Belasco was born in San Diego, CA on February 21, 1979 to Hugh Belasco and Maureen Cochran. While attending West Hills High School, he played varsity soccer and golf for three years and enjoyed playing the drums. Eric attended Saint Mary's College of California, where he received a Bachelor's of Science degree in Economics in May of 2001. Eric married Megan Brockhagen on May 5, 2001 in Moraga, CA and they have been blessed with two wonderful children, Analise and Emery.

# Acknowledgments

There are many people I wish to show my sincerest gratitude to who have influenced my life and career in so many ways. First, my advisors have been top notch and a pleasure to work with. Barry Goodwin has played such a major role in developing the ideas for this work and also helped me to cultivate professionally. He has shown enthusiasm, support, and been a good friend throughout this process. A special thanks also to Sujit Ghosh, for contributing so much to my understanding of statistics and professionalism; Nicholas Piggot, for his career advice and expertise on this topic; and Wally Thurman, for his attention to detail with my work and for demonstrating excellent teaching. I would also like to thank Ted Schroeder for his expertise and friendship, which have contributed so much to this document.

I also feel fortunate to have spent my time at NC State with such a wonderful group of students. I would like to especially thank Allen Klaiber and Mykel Taylor for their constant academic curiosity and willingness to drink coffee. I learned so much from both of them.

My family is also owed a debt of gratitude for so many reasons. I would like to especially thank my parents, grandparents, and siblings. My parents raised me with a clear understanding of the importance of unconditional love and support, for myself, my family, as well as others. My grandparents inspired a strong work ethic in me by demonstrating throughout my life to expect great things through hard work. I'd like to also thank my siblings for keeping me grounded and always ready to share a laugh.

From the bottom of my heart, I would like to thank my wife, Megan, for

sharing her life with me. She has given me a beautiful family and a life of love, laughter, and affection. She is beautiful in so many ways and I feel fortunate that she shares all of her wonderful qualities with me. Our children, Analise and Emery, have injected new life into my veins and given to me so much to look forward to. Watching them grow has been truly remarkable.

Eric J. Belasco

*North Carolina State University*
*October 2007*

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Background on Cattle Feeding Risk

This section is intended to offer background on risks in the cattle feeding industry. The first section begins with a discussion on agricultural risk and available forms of risk management tools. The discussion then outlines the direction of past and present farm insurance programs, as well as current cattle insurance offerings. The final section focuses on describing characteristics unique to cattle feedlots in order to understand their position in U.S. agriculture and the cattle industry.

## 1.1 Overview of Agricultural Risk

While risk and risk management strategies are nothing new to agriculture, federal insurance policies are still attempting to find better ways to insulate farmers from extreme risks. The formation of the Federal Crop Insurance Corporation (FCIC) by Congress in 1938 was the first major step towards insuring farmers against risk using federal funds. The goal of this program was to protect farmer's income from crop failure or price collapse. The Great Depression and major droughts throughout the 1930s led to the formation of federal crop insurance offerings to help farmers manage risk. This program was originally plagued with low participation rates and large losses, leading to the reduction of such programs (Goodwin and Smith, 1995).

Major reforms came in 1980 and 1994, resulting in private/public partnerships and the attempted elimination of *ad hoc* disaster relief. Since then, crop insurance programs have benefitted from higher participation rates and more precise premium calculation. As with any insurance program, whether it be public or private, information regarding the riskiness of the insured agent is critical. Some of the first federal crop insurance products insured against weather-related loss of yields. Weather related events leading to lower yields might include drought, hail, or extreme temperatures. In order to insure against these events at an actuarially fair rate, the principle insurer must understand the probability of a loss event and the implications of such an event. An actuarially fair rate occurs when the price of the policy is equal to the expected loss (Mas-Colell et al., 1995), at which point the risk averse agent will insure and the program is actuarially sound.

In this way, crop insurance contracts offer farmers a form of risk management to allow them to manage operations in spite of bad weather. Additional subsidies further support farm operations to lower the prices of insurance policies. Typically, farmers have the option of guarding against a minimum level of expected yields, where expected yields are a function of past performance. In the case where the realized yields fall below a specified percentage of expected yields, an indemnity payment is collected by the farmer to make up the difference.

Over the last 10 years, insurance policies have evolved to include revenue insurance products and area wide plans. Revenue insurance guards against farm revenue, which is a function of both price and yields. The idea behind this plan is that price and yields are negatively related so that when yields are low, supply falls, and prices rise. This causes fewer indemnity payments, relative to yield insurance. Area-wide plans work to minimize moral hazard distortions by basing expected and realized yields on an entire area, rather than a single farm. Moral hazard is one of the largest problems in insurance and occurs when insured agents adjust their behavior, based on level of insurance. In farm operations this includes less chemicals or fertilizer as a result of the safety net offered by insurance. Fraud can also be minimized through the use of are-wide programs.

This rebirth of federal crop insurance programs over the past 20 years has inspired an extensive amount of literature aimed at characterizing agricultural yields. This motivation can be found in much of the crop insurance literature that attempts to characterize conditional mean yield densities to evaluate the risks involved with crop management and accurately price crop insurance premiums.

Some research has focused on the use of the Normal distribution to characterize crop yields (Just and Weninger, 1999). They argued that using a normal distribution is not unreasonable, given their inability to reject normality. Atwood, Shaik and Watts (2003) reiterate the importance of not overlooking the normal distribution and argue in favor of proceeding with caution when dealing with heteroscedastic errors. Ramirez et al. (2003) find that corn and soybean yields are non-normally distributed due to the negative skewness. Sherrick et al. (2004) use goodness-of-fit measures to test the economic differences between different distributional assumptions and find that the Beta and Weibull distributions best characterize corn and soybean yields, while normal and log-normal distributions fail to describe the sample data.

Other research has focused on the Beta distribution as an alternative parametric measure of crop yields due to its flexibility in allowing for skewness and kurtosis (Coble et al., 1996; Nelson, 1990; Nelson and Preckel, 1989). Skewness and kurtosis are commonly found in yield data. Ker and Coble (2003) use Illinois corn data to show that the Beta outperforms the normal in small samples, while the opposite holds in larger samples. However, they point out that most agricultural data samples tend to be insufficient in size to statistically validate almost any reasonable parametric model. Gallagher (1987) uses a Gamma distribution in his study and proposes a technique for dealing with heteroscedastic yield data.

A major hurdle in most parametric tests is that data is rarely sufficient enough to eliminate the use of some reasonable parametric forms, while accepting others. Additionally, a parametric assumption will cause biased estimates when the assumed distribution is not the true density. To mitigate this problem, both semi-parametric and Bayesian methods have been proposed by Ker and Coble (2003) and Ozaki et al. (2006), respectively. A survey of these issues is further examined by Goodwin and Ker (2002).

## 1.2 The Evolution of Federal Insurance From Crops to Livestock

The Agricultural Risk Protection Act of 2000 mandated the development of livestock insurance plans. Currently, producers of hogs, fed cattle, and feeder cattle can purchase an insurance policy to guard against unexpected declines in prices with Livestock Risk Protection (LRP) and Livestock Gross Margin (LGM) insurance. LRP protects fed cattle producers from unexpected declines in market value, while LGM offers protection against loss of gross margin (market value of livestock minus feeder cattle and feed costs). To be more specific, the LRP program protects against adverse declines in live cattle prices, while the LGM program jointly protects against adverse swings in gross margin, which is a function of live cattle prices and feed prices. According to these policies, beef producers may insure up to 4,000 head of heifers and steers per crop year, weighing between 1,000 and 1,400 pounds. The length of insurance contract can be from 13 to 52 weeks. Coverage can range from 70% to 95% of the expected ending value.

This program offers advantages over purchasing futures and options on the Chicago Mercantile Exchange (CME) since the federal insurance contracts can be specially tailored to the unique needs of each producer (Hart, Babcock and Hayes, 2001). Additionally, insurance participation rates may eventually be higher than the amount of producers who currently use futures and options market to manage risk. Specific knowledge needed to participate in the futures and options market is not necessary in order to participate in federal livestock insurance programs.

The offering of livestock insurance products involves a much different set of risk variables when compared to traditional crop insurance offerings. Most crop programs are based on insuring against adverse swings in yields, as yields are the largest risk factor. These programs are intended to offset the risks associated with extreme weather or crop disease. Livestock products are much less sensitive to extreme weather conditions or shocks that affect yields. While yields can certainly be variable with livestock, most risk comes in the form of price changes (Schroeder et al., 1993). This is most notably seen as demand can change dramatically due to changing consumer preferences, animal disease outbreaks[1], or trade restrictions with major trading partners. Other risk factors with fed cattle production include feeding performance, mortality, veterinary costs, and corn prices. One of the major questions facing the early pilot livestock insurance programs is which risk factors should be covered under these programs when risks can come from production and price risk.

A recent update of existing federal livestock insurance programs was offered by Babcock (2005). During this hearing it was pointed out that all states involved in LRP or LGM programs have less than 2% of the eligible feeder cattle insured. The low participation rates may be the result of a new insurance program that has yet to fully gain acceptance among cattle owners, coupled with the closure of both programs in 2003 due to the discovery of BSE in the United States. In regard to the

[1]The most notable outbreaks that have affected the demand for beef products include foot-and-mouth disease (FMD) and Bovine Spongiform Encephalopathy (commonly referred to as BSE or “mad cow” disease). The effects from food safety outbreaks on meat demand has been analyzed by Piggott and Marsh (2004).

development of such a program, crop insurance programs illustrated that acceptance of new insurance products takes time. The number of head insured by existing cattle insurance programs is shown in Table 1.1.[2] In 2006, 29,593 fed cattle were insured under existing federal insurance plans, which is tiny relative to the nearly 12 million cattle on feed in the U.S. as of December 1, 2006 on 1,000+ head capacity feedlots.[3]

Table 1.1: Head of cattle insured through existing federal livestock insurance programs, by State and Plan Type

| | LRP - Fed Cattle | | | LRP - Feeder Cattle | | | LGM |
|---|---|---|---|---|---|---|---|
| State | 2004 | 2005 | 2006 | 2004 | 2005 | 2006 | 2006 |
| Texas | 0 | 37 | 46 | 7,364 | 4,702 | 5,257 | 0 |
| Kansas | 0 | 3,023 | 4,612 | 27,754 | 22,565 | 32,563 | 3,300 |
| Nebraska | 67,884 | 7,267 | 9,505 | 4,396 | 13,295 | 22,340 | 2,613 |
| Colorado | 0 | 530 | 0 | 210 | 1,650 | 2,340 | 0 |
| South Dakota | 0 | 3,700 | 2,016 | 12,586 | 18,828 | 27,864 | 823 |
| Iowa | 28,123 | 4,406 | 4,827 | 252 | 2,105 | 3,793 | 15,273 |
| Oklahoma | 0 | 50 | 242 | 14,587 | 16,923 | 25,087 | 1,951 |
| Other | 944 | 6,606 | 8,345 | 1,174 | 22,232 | 27,566 | 1,695 |
| Total U.S. | 96,951 | 25,619 | 29,593 | 68,323 | 102,300 | 146,810 | 25,655 |

Source: FCIC Summary of Business (as of 01/15/2007)

While much of the success of this program has yet to surface and will be revealed over the next few years, it will admittedly avoid some of the moral hazard problems associated with some crop policies. This is due to the fact that prices are based on futures prices that cannot be influenced by the actions of individual policy holders. Specifically, the live cattle prices are from the Chicago Mercantile Exchange (CME), while feed costs are from the Chicago Board of Trade (CBOT). In order for livestock insurance policies to be sound and successful, it is essential

[2]Total head of cattle insured under Livestock Risk Protection plans include both feeder and fed cattle. Livestock Gross Margin plans became available in 2006.

[3]Estimated figure published in *Cattle on Feed* by the National Agricultural Statistics Service, which was released on December 22, 2006. Cattle on feed for the current month are computed by starting with the cattle on feed for day one of the prior month and adding the difference over the past month. This difference is the sum of new placements minus marketed and other disappearances, which includes mortalities and transfers to other feedlots.

(as with crop insurance) to accurately model risks. This research attempts to further studies into cattle feeder risk by accurately modeling risk components using a multivariate framework, then combining the risk components together to ultimately assess profitability risks associated with fed cattle production.

## 1.3 Overview of Cattle Feeding

The primary function of commercial cattle feedlots is to serve as an intermediate producer between the period of time when cattle graze in pastures and are send to packers for slaughter. During this span of time, pens of cattle are fed a high energy diet, intended to rapidly add weight. Typically beef cattle are transported to begin the feedlot phase of their life cycle at 11-14 months of age. Upon arriving at the feedlot, pens of cattle are weighed and given necessary vaccinations and other health checks. Most cattle weigh between 500 to 900 pounds and are placed on feed for 3 to 6 months where they will usually gain over 3 pounds per day. The amount of time cattle are on feed is usually set upon entrance to the feedlot, but can vary slightly due to extreme weather around the expected marketing dates. The number of days a pen of cattle is placed on feed is largely dependent on the average weight of the pen. For transportation and feeding reasons, it is desirable to have relatively homogenous cattle placed together in the pen.

Because feed is largely corn-based, there is natural synergy between corn growers and cattle feeders. Historically, the Central Corn Belt[4] served as the dominant cattle feeding region due to its close proximity to corn(Williams and Stout, 1964). This dominant position began giving way to the plain states over the 1960s and 1970s, who utilized irrigation rows to grow supplies of grain sorghum that can be used to feed cattle. Grain sorghum, also known as milo, is more adaptable to

[4]The Central Corn Belt is defined as Iowa, Illinios, Indiana, and Ohio by Williams and Stout (1964).

extreme weather and different types of soil, when compared to corn. The major growing region for grain sorghum consists of Kansas, Nebraska, Oklahoma, northern Texas, western Colorado, and western New Mexico (Thompson and O'mary, 1983). This region currently holds a large majority of the cattle on feed[5] shown in Table 1.2 because of its grain sorghum production and its relatively mild and dry temperatures.

Table 1.2: Average Monthly Number of Cattle on Feed in 2005, by State

| State | Cattle on Feed (1,000 Head) | Proportion of Total US (%) |
|---|---|---|
| Texas | 2,803 | 25.8 |
| Kansas | 2,318 | 21.4 |
| Nebraska | 2,118 | 19.5 |
| Colorado | 969 | 8.9 |
| California | 525 | 4.8 |
| Iowa | 443 | 4.1 |
| Oklahoma | 335 | 3.1 |
| Other States [a] | 1,344 | 12.4 |
| Total US | 10,854 | |

[a] Other States include Arizona, Idaho, South Dakota, Washington, and New Mexico

Source: *Cattle on Feed* released by the National Agricultural Statistics Service (NASS)

The strong spatial relationship between corn and cattle feedlots illustrates their interdependence. This relationship is further investigated by Anderson and Trapp (2000) where a corn price multiplier effect is developed. Feed costs typically make up more than half of total operating costs on commercial feedlots. Feed components are purchased throughout the feeding period on a cash or contract basis, then mixed and distributed to fence-line bunks for consumption. Most feedlots typically use a corn-based feed that is comprised of grains, roughage, and protein supplements. Major grains include corn and milo, where a mixture of the two is used

[5]The USDA defines *cattle on feed* as "animals being fed a ration of grain, silage, hay and/or protein supplement for slaughter market that are expected to produce a carcass that will grade select or better. It excludes cattle being 'backgrounded only' for later sale as feeders or later placement in another feedlot."

in the southern plains. Roughage components are the second largest ingredients in cattle feed, which can comprise of corn silage, sorghum silage, or alfalfa hay. The exact combination of feed used largely depends on location in an effort to minimize cost and utilize home markets. Once cattle weight over 700 pounds, their diet typically consists of protein that accounts for 11% of total feed matter (PSU, 2001). Protein supplements are comprised of many ingredients, including soybean meal, urea, ammonia, brewers grains, and cottonseed meal (Ritchie, 1994). Upon arrival to the feedlot, cattle are continued on a diet of grass and quickly transitioned to a high-energy diet over a 2-3 week period. This transition takes place with the gradual decrease of hay and replacement of feed diet. Feed used for finishing diets are roughly three-fourths corn.

The largest source of risk within fed cattle production is the vulnerability to changes in cattle prices (Schroeder et al., 1993). This risk is a major reason why cow/calf breeders sell ownership of their cattle before they are placed on feed. When cattle are ready to be placed on feed, the cattle owner faces a decision concerning whether to retain ownership of the pen or sell for the feeder cattle price. Changes in this price are highly correlated with the price of cattle at the end of feeding, which is the fed cattle price. While both cattle prices change significantly throughout the year, lighter-weight cattle of the same grade cost more per pound. Major tools for managing this risk comes from the purchasing of futures and options through the Chicago Mercantile Exchange or newly released insurance products from the federal government. Profits from cattle ownership can largely influence incomes when the market for fed cattle takes a dramatic turn.

There is a distinct difference between owning feedlot facilities and owning cattle on feed. Custom feeding of cattle is a popular offering at larger feedlots. Over the last 20 years, large-scale feedlots have played a growing role in the cattle industry as shown in Table 1.3.

Table 1.3: Proportion of U.S. fed cattle sales by feedlot size

| | Feedlot capacity (number of head) | | | | | | |
|---|---|---|---|---|---|---|---|
| Year | Less than 1,000 | 1,000-1,999 | 2,000-3,999 | 4,000-7,999 | 8,000-15,999 | 16,000-32,00 | More than 32,000 |
| 1985 | 19.0 | 4.0 | 6.1 | 7.3 | 15.0 | 19.7 | 29.0 |
| 1990 | 15.6 | 4.1 | 7.0 | 7.5 | 14.5 | 23.0 | 28.2 |
| 1995 | 9.7 | 4.1 | 5.3 | 8.1 | 14.2 | 21.1 | 37.6 |
| 2000 | 14.2 | 3.2 | 4.6 | 7.6 | 11.1 | 19.4 | 39.8 |

Note: Figures are reported as percentages

Source: United States Department of Agriculture (2001)

Custom feeding contracts split risk between the cattle owner and the feedlot owner to differing degrees, depending on the type of contract. By custom feeding ones herd of cattle, the cattle owner can retain ownership in order to reap the benefits from genetic advantages or market changes, while taking advantage of the feeding expertise of the commercial feedlot. There are three major types of contracts between the feedlot and cattle owners; a yardage fee plus feed costs (YF) contract; a yardage fee plus feed costs and a markup on feed costs (YFMU); and a guaranteed cost-of-gain (GCOG) contract (Weimar and Hallam, 1990). These contracts differ in the way price, production, and cost of feed risk is distributed between the cattle owner and feedlot operator. The cattle owner is sometimes the cow/calf breeder, but can also be an outside investor, the feedlot itself, or the packer.[6]

Historically, most fed and feeder cattle are sold on a cash basis through livestock auctions (RTI, 2007). Here the price is based on the average weight of the entire pen, which ignores any quality differentiation.[7] More recently, the beef industry has moved to offer additional options when it comes to marketing live cattle. These new offerings include grid pricing and dressed weight pricing. Dressed weight alternative options are further explained by Fausti and Feuz (1995) as well as

[6]This is an issue that was investigated by United States Department of Agriculture (2001).

[7]This paper focuses its attention to production risk and assumes that cattle are marketed on a cash basis in its profit simulations. However, an area of increasing interest is the modeling of quality risk to accommodate grid pricing formulas.

Marsh (1999). Alternatively, grid pricing has been discussed by Fausti and Qasmi (2002) among others.

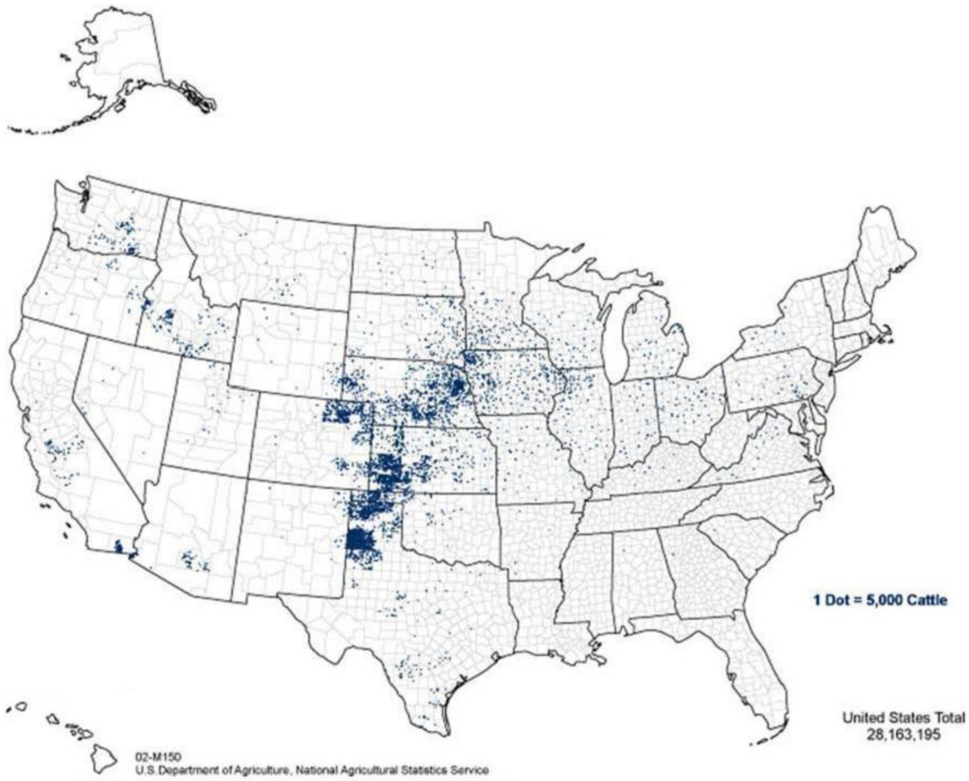



Figure 1.1: Location of Cattle Feedlots, 2002

# Chapter 2

# A Multivariate Evaluation of *Ex-ante* Cattle Feeding Risks

## 2.1 Introduction

Cattle feeding can be a risky venture. From the time of cattle placement to finishing, which usually lasts 3-5 months, the value and profitability of cattle can change immensely. Most of this risk comes in the form of fed and feeder cattle price risk, but can also come from large swings in feed prices. Both of these factors, which pose more than half of the variability in cattle feeder profits, are out of the cattle owners' control. In addition, the overall productivity of the pen can present risks that are akin to yield risks with crops.

Research from the crop insurance literature has indicated that agricultural yields can be modeled in a number of different ways. These differ in the restrictions that are imposed on the data. For example, parametric methods assume a particular distributional assumption, which is efficient when the form is correct, but biased when the assumption is incorrect. Different distributions have been argued to be the most accurate characterization of crop yields, which include the normal, log-normal, beta, gamma, and weibull distributions to name a few.

In modeling the *ex-ante* risks, variables known at the time of placement that may affect the expected mean or variance must be taken into account. Past cattle feeder research has shown that cattle feed conversion, average daily gains, mortality rates, and health are significantly affected by variables such as gender, location of the feedlot, average weight of the pen, and time of year the pen is placed. By conditioning on these variables, each pen of cattle can be modeled using a multivariate regression model.

Cattle yields present some additional complexities when compared to crop yields. The first difference is that production risk can be represented by four separate measures. These four yield components are usually highly correlated and have a dynamic relationship. A recent study by Belasco et al. (2006) modeled each of the four yield measures separately then using the parameter estimates computed the covariance matrix. Significant efficiency can potentially be gained in a multivariate framework, with the additional information that can be learned on the dynamic nature of the yields relationship.

The second complexity associated with cattle production yields is the introduction of a censored variable into the set of yield factors. Mortality rates in cattle feeding present a significant degree of censoring as nearly half of all pens in the data used in this research, had no mortalities. More specifically, mortality rates can be modeled as a latent variable where the variable is observable for positive values and unobservable for small positive values of the distributional realizations. To account for this relationship, a dynamic multivariate Tobit model will serve to model the latent mortality variable.

Once the dynamic multivariate relationship is characterized with the previously mentioned yield variables, profits will be simulated based on random draws from these yield variables. The profit function will consist of the six previously mentioned random variables. In order to model profits, the random variables will need to be jointly modeled by allowing for conditional covariance between the yield variables. Once the profit model is identified and characterized, profit simulations will provide insights into the effects from shocks on profits and revenue.

## 2.2 Literature Review

Studies investigating the risk factors associated with cattle feeding have keyed on the fact that risk comes from many different sources. To add to the complex nature of this risk, the variability changes as many of the key variables change.

One of the earlier studies focusing on cattle feeding profitability came from Swanson and West (1963). This study regressed returns to cattle feeding as a function of the gains from the price margin (difference between fed and feeder cattle prices times the purchase weight) and the value of feeding margin gain. To measure the influence of each independent variable to the overall variability, the coefficient of separate determination method was employed. This method was originally suggested by Wright (1921) and has the admitted problem of erratic estimates resulting from correlation between the independent variables. In spite of this limitation, this method allows for a direct measure of the variable's effect on overall variation. This study found that variation in returns are partially explained by price margin variation (38%) and feeding margin gain (44%), while 18% of the variation was unexplained.

This methodology provided a strategy to isolate the effects of cattle feeder risks. The next step was to estimate a profit function with more detail so that risks can be isolated to more specific components of cattle feeding. This step was taken in a study by Schroeder et al. (1993) where over 6,000 pens of steers from two major Kansas feedlots were evaluated. They concluded that 70 to 80 percent of the variation in cattle feeder profits came from variation in fed and feeder cattle prices, while the price of corn explained 6 to 16 percent of the variation, and cattle performance (which included average daily gain and feed efficiency) accounted for less than 10 percent.

Langemeier, Schroeder and Mintert (1992) included both steer and heifer pens in their sample and found that fed and feeder prices accounted for 50 and 25 percent of variability in cattle feeding profits, respectively. Meanwhile, corn price variability explained up to 22 percent of variability and animal performance explained less than 1 to 3.5 percent variability. In addition to echoing the results from Schroeder et al. (1993), this research also identified variables that affect the expected value and variability of profits. More specifically, 22 percent of the difference in steer and heifer profits are directly attributed to differences in feed conversion. Also, the impact from average daily gain on profits increases with placement weight. It is important to point out that not only does profit variability come from a few different sources, but these impacts change with different pen characteristics.

With the two previously mentioned studies of cattle feeder variability using data from large Kansas feedlots, a study by Lawrence, Wang and Loy (1999) utilized data from smaller Midwest feedlots. With their data set consisting of 223 different feedlots and over 1,600 pens, their hypothesis was that with feedlot locations spread over 5 states with wider climatic variations and smaller operations would result in a greater role from animal performance on profit risk. They essentially replicated earlier studies again using the coefficient of separate determination, but now to include a much different data set. Fed and feeder cattle prices together still explained around 70 percent of profit variability. As expected, animal performance (average daily gain) played a larger role in profit risk in explaining between 6 to 15 percent of the overall profit variation. The effects from corn variation fell below that of animal performance. Two possible explanations are given for the differences with previous research in animal performance's impact on profit variation; (1) climactic variation in Iowa and its surrounding states are higher than in Kansas; and (2) greater differences amongst feedlot operations.

Up to this point, research has focused on the amount of variability coming from prices and animal performance, while not explicitly evaluating the effect that pen characteristics might have on changes in variability and expected profits. [1] In an updated study by Mark, Schroeder and Jones (2000), an enhanced data set included over 14,000 pens from two major Kansas feedlots. The focus here is to evaluate the relationship between fed price, feeder price, corn price, interest rate, feed conversion, and average daily gain with expected profits and the variability of profits using standardized beta coefficients. The method of standardized beta coefficients can be summarized in the following manner. If we define the following function

$$Y = f(X_1, X_2, ..., X_k) \tag{2.1}$$

then by regressing the normalized independent variables on the normalized dependent variables, a unitless coefficient, $\beta_j^*$ is obtained from equation (2.2)

$$\frac{Y_i - \bar{Y}}{S_Y} = \sum_{j}^{k} \beta_j^* \frac{X_{ij} - \bar{X}_j}{S_{X_j}} + \epsilon \tag{2.2}$$

where S indicates the standard deviation. This method allows for the estimation of the relative impacts each independent variable contributes to the overall variability of the dependent variable. The previously mentioned research indicates that variability can change for different values of placement weight, season, and gender. Additionally, Mark and Schroeder (2002) explicitly show that the season of placement has significant effects on profitability and animal performance.

Using a different strategy, Buccola (1980) builds a model of break-even feeder cattle prices to investigate the influence of important supply and demand factors on feeder cattle price differentials. Emphasis is placed on the price differences from

[1]The exception to this statement is the discussion of the differences between steer and heifer pens (Langemeier, Schroeder and Mintert, 1992).

weight and sex categories. Twenty years of Virginia state-graded feeder cattle transaction data were used and separated by year, sex, and season. Sale prices were then regressed or conditioned on these variables. Results indicated that feed price changes impacted light-weight feeder cattle prices more than heavy-weight feeder cattle prices, suggesting corn prices have less influence on profit per head as placement weight increases. An important point to this paper is that a strictly linear relationship between ration cost, live cattle futures price, and feeder-cattle price is not the most appropriate representation of the feeder-cattle market.

To advance the point that profits cannot be linearly determined, Anderson and Trapp (2000) demonstrate that changes in the price of corn can lead to changes in placement weight, slaughter weight, and feed conversion rates. They also state that a break-even price model strategy allows for the multiplicative effects of the variables to be taken into account.

With the previously mentioned studies in mind, we can point to a few facts that run throughout the literature. First, most cattle feeder profit risk stems from swings in fed and feeder cattle prices. With this being said, any risk management strategy must begin with managing the possibility of cattle prices dramatically dropping during the feeding period. Second, animal performance factors significantly contribute to risk and are the only identified sources of risk that the feeder can affect through operational and placement decisions. Third, variables such as gender, placement weight, and time of placement can have significant effects on expected profits and variability.

## 2.3 Yield Modeling Framework

This section will lay out the framework used for modeling the four identified yield factors in fed cattle production. The first section will lay out the general strategy of modeling yields within an *ex-ante* framework, describing which variables are conditioned on at the beginning of the feeding cycle. The next section will make the case for a multivariate modeling strategy. Then a proposed method for estimation will be explained that allows for censored observations.

### 2.3.1 ***Ex-ante*** Conditional Modeling Strategy

Within the framework of insuring risks, current *ex-ante* information and past *ex-post* information is usually all that is available. In the case of insuring owners of cattle on feed, it is important to note certain variables that can have an effect on both the expected value and variance of profits. These characteristics include pen characteristics such as gender, weight upon entrance to the feedlot, location of the feedlot, and the season of placement. Conditioning on variables known to influence variable outcomes focus attention on risks associated with activity that occurs throughout the production process. Past studies have shown that these conditioning variables can have a significant effect on the first two moments of expected profits and each of the yield factors (Belasco et al., 2006). Each pen with its own unique characteristics should be expected to bring a different set of risks and expected values. The strategy of this research is derive a probabilistic measure for the four yield factors, in order to jointly model production aspects, profits, or revenue.

This framework utilizes the information known when a contract is made with the insured party. This strategy allows the model to control for information that is known at the time of placement, while characterizing the distributional characteristics of the error components that reflect the areas of risk during the feeding cycle. As with any type of insurance, characteristics of the insured will affect the premium rate for coverage. This occurs because certain characteristics make certain events more likely. For example, mortality rates should be lower for pens that enter with higher weights due to their superior maturity and shorter stay on the feedlot. The point here is that it is important to grasp onto these variables that affect the mean and covariance of performance.

### 2.3.2 The Case for Multivariate Modeling

One major difference between modeling crop yields and that of fed cattle production is that one can identify four separate correlated yield measures, of which each contribute to overall yields and profitability. In the case where each of the four yields are unrelated, it would not be problematic to model each yield factor separately to estimate the marginal effects on the mean and variance. However, if the yields are correlated with one another, then a framework that accounts for this dependency will potentially improve model efficiency.

To illustrate, consider the extreme weather case. Belasco et al. (2006) demonstrate that pens placed in the fall and fed as the temperatures drop into colder winter temperatures, typically have higher feed conversion rates and mortality rates. Additionally, veterinary costs are insignificant, however a higher mortality rate may allow for less of the pen to be cared for. Based on this, an extremely cold winter may exaggerate these relationships, leading to correlated error residuals among the yield factors. Efficiency gains can be made in this case by modeling the system of equations simultaneously.

Another example is the strong relationship between the feeding efficiency and daily gains for cattle. As cattle are able to convert pounds of feed into weight gain more efficiently, the speed of weight gain will increase. Both examples illustrates the interdependence of the four yield factors. Shocks that are not controlled for by the conditioning factors are likely to affect each one of the conditioning variables.

To take this a step further, it will also be important to control for any dynamic effects on the correlation structure. For example, some entry characteristics may lead to stronger or weaker relationships between the variables. If there is shown to be a relationship between the data and the correlation between dependent variables, then it is important that it is captured as part of this model.

### 2.3.3 Econometric Model

The variables that introduce production yield risk into the cattle feeder profit function are dry matter feed conversion (DMFC), average daily gain (ADG), the mortality rate (MORT), and veterinary cost per head (VCPH). As seen in past studies, these variables are influenced by pen characteristics such as gender, location, average in-weight, and season of placement. These factors affect both the expected value and variance for each yield measure. Additionally, there are significant correlations between each yield measure. The modeling strategy to follow will account for each of these complexities and characterize the probabilistic models of the cattle yield factors involved with fed cattle production. Additionally, the proposed model accounts for cross correlation between yields by using a multivariate normal distribution, rather than the uncorrelated univariate normals that was originally proposed by Belasco et al. (2006).

The model is specified as follows:

$$Y_i = X_i B + \varepsilon_i \tag{2.3}$$

$$E[\varepsilon_i | X_i] = 0 \tag{2.4}$$

$$Var[\varepsilon_i | X_i] = \Sigma_i = \Sigma(X_i) \tag{2.5}$$

where $\varepsilon_i$ are independent, $Y_i = [DMFC_i, ADG_i, MORT_i, VCPH_i]$ and B is a px4 matrix containing the marginal effects of each conditioning variable on all four yield factors. $X_i$ is a 1xp matrix containing the following seven conditioning variables for each observation that include a constant term, the log of average in weight, and binary variables indicating gender, location, and season of placement.

$$X_i' = \begin{bmatrix} 1 \\ Steers \\ Mixed \\ KS \\ Log(Inwt) \\ Winter \\ Fall \\ Spring \end{bmatrix} \tag{2.6}$$

The 1x4 vector of errors, $\epsilon_i$ is assumed to have mean zero and covariance matrix, $\Sigma_i$. $\Sigma_i$ contains the covariance elements and is a 4x4 unknown positive definite (p.d) matrix. Notice that the covariance matrix is allowed to vary by observation. We propose to model $\Sigma_i = \Sigma(X_i)$ using the following unique decomposition of a p.d. matrix (Lau, 1978):

$$\Sigma_i = T_i' D_i T_i \tag{2.7}$$

where $T_i$ is upper triangular with ones along the main diagonal and $D_i$ is diagonal matrix with positive diagonal entries. More specifically,

$$T_i = \begin{pmatrix} 1 & t_{12i} & t_{13i} & t_{14i} \\ 0 & 1 & t_{23i} & t_{24i} \\ 0 & 0 & 1 & t_{34i} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.8}$$

and

$$D_i = \begin{pmatrix} d_{1i} & 0 & 0 & 0 \\ 0 & d_{2i} & 0 & 0 \\ 0 & 0 & d_{3i} & 0 \\ 0 & 0 & 0 & d_{4i} \end{pmatrix} \tag{2.9}$$

Upper off-diagonal elements of $T_i$ are unrestricted while the diagonal elements of $D_i$ are restricted to positive values. We use a linear regression to model $T_i$ and $D_i$ as follows:

$$log \begin{pmatrix} d_{1i} \\ d_{2i} \\ d_{3i} \\ d_{4i} \end{pmatrix} = GX_i' \tag{2.10}$$

where G is a 4xp matrix of regression coefficients. The off-diagonal terms within $T_i$ are also a linear function of the conditioning variables and have the following relationship

$$\begin{pmatrix} t_{12i} \\ t_{13i} \\ t_{14i} \\ t_{23i} \\ t_{24i} \\ t_{34i} \end{pmatrix} = AX_i' \tag{2.11}$$

where A is a 6xp matrix of regression coefficients. Covariance terms are first-order fully flexible within the regression framework. This model is an improvement from the two-step method used in Belasco et al. (2006).

The maximum likelihood method to obtain parameter estimates uses the following likelihood function:

$$L(B, A, G|Y, X) = \prod_{i=1}^{n} |\Sigma_i|^{-\frac{1}{2}} exp\left(-\frac{1}{2}\sum_{i=1}^{n}(Y_i - X_iB)\Sigma_i^{-1}(Y_i - X_iB)'\right). \quad (2.12)$$

Given n observations, this leads to the following negative log likelihood function (up to an additive constant) that is minimized with respect to elements within B, G, and A.

$$LL = \frac{1}{2}\sum_{i=1}^{n} \ln|\Sigma_i| + \frac{1}{2}\sum_{i=1}^{n}(Y_i - X_iB)\Sigma_i^{-1}(Y_i - X_iB)' \quad (2.13)$$

By modeling yields in the preceding manner, the model is flexible enough to allow for expected values and covariances between the yields to vary with the conditioning variables. These are key components to modeling the nature of risk in fed cattle production where the expected distributional properties can change, given the characteristics of the pen.

Thus far, the model assumes that all variables are completely observable across observations and free from any censoring or truncation bias. However, in the case of mortality rates, these values are censored at zero. This problem may cause us to underestimate mortality rates due to biased parameter estimates caused by the censoring mechanism (Greene, 2003). The most widely accepted solution to regressing censored dependent variables in the univariate case was first proposed by Tobin (1958) and is known as the Tobit Model. This method essentially assumes there is a latent variable, $y_i^*$, which linearly depends on the associated independent variables, $X_i$, where we observe $y_i$ in the following way:

$$y_i = \begin{cases} 0 \texttt{ if } y_i^* \leq c_i \\ y_i^* \texttt{ if } y_i^* > c_i \end{cases} \quad (2.14)$$

where $c_i$ is the censored value. Notice that when $y_i$ is censored at zero, the only information we have is that $y_i^* \leq 0$. So the likelihood function becomes

$$L(\beta, \sigma^2|x) = \prod_{i:y_i>0} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right)\frac{1}{\sigma}\prod_{i:y_i=0} \Phi\left(-\frac{x_i\beta}{\sigma}\right) \tag{2.15}$$

where $\Phi$ is the CDF and $\phi$ is the PDF of a standard normal distribution, respectively. In the case of censored dependent variables, the use of maximum likelihood estimation has been shown to result in estimators that are consistent and asymptotically normal (Amemiya, 1973), provided the assumed parametric model is correct. This method has been useful in applications spanning consumption, production, and income.

Censored univariate regressions have been extended to multivariate modelds and shown to posses the same attractive asymptotic properties as in the univariate case (Amemiya, 1974; Lee, 1993). The multivariate Tobit model has been considered in a number of recent studies. Cornick, Cox and Gould (1994) formulate a multivariate Tobit model in order to analyze fluid milk consumption expenditures and account for the correlations across milk types. Eiswerth and Shonkwiler (2006) investigate the success of plant seeding that follows wildfire on arid rangeland, where all types of grass do not typically grow together simultaneously due to geographical differences. Also, Chavas and Kim (2004) use a dynamic multivariate Tobit model to evaluate price dynamics when price floors exist in a given market. The dynamic component plays an important role in this analysis as the data is evaluated over time, where the correlations between prices adjust over different time periods. While the covariance matrix changes over time, it is held constant for differing values of the other conditioning variables. Here, we expand on these studies by allowing for the interdependence between the dependent variables to be a function of the data. The idea is to model the latent variables through the use of a multivariate Tobit model, using a dynamic multivariate sampling distribution under conditional heteroskedasticity while allowing for interdependence between the residuals.

In the univariate case, each observation can fall into one of two possible regimes where the dependant variable is either censored or not. However, within the framework of a generalized multivariate Tobit model, the possible censoring regimes increase to $2^m$ where m is the number of censored dependent variables. For our purposes, four dependent variables lead to 16 possible regimes. Due to the fact that only one variable is censored, only two regimes are possible. For observations with multiple censored dependent variables, integration becomes more complex by adding a dimension for each censored variable. As long as this dimension is not greater than three, standard maximum likelihood methods can be used (Chavas and Kim, 2004).

To obtain the likelihood function, each observation must be ordered as censored or noncensored variables for each regime. To this end, $Y_i$ will be partitioned into its censored variables, $y_i^{(1)}$, and uncensored variables, $y_i^{(2)}$, under each regime:

$$Y_i' = \begin{pmatrix} y_i^{(1)} \\ y_i^{(2)} \end{pmatrix} \tag{2.16}$$

Further, the sample log-likelihood function corresponding to each individual, can be expressed as follows:

$$LL = \sum_{i:Mort>0} \{ln\,[\phi(Y_i; \mu_i, \Sigma_i)]\} + \sum_{i:Mort=0} \left\{ln \left[\phi(y_i^{(2)}; \mu_i^{(2)}, \Sigma_{22i})\right] + ln\,[\Phi(0; \gamma_i, \lambda_i)]\right\} \tag{2.17}$$

where $\phi(y; \mu, \Sigma)$ refers to the multivariate normal probability density function with mean vector, $\mu$, and variance covariance matrix, $\Sigma$, while $\Phi(0; \gamma, \lambda)$ denotes the univariate cumulative distribution function evaluated at 0 with mean, $\gamma$, and variance, $\lambda$. The censored variable is modeled based on a multivariate normal density and is a function of the observable variables within the same observation. For this reason, the conditional mean and variance for $y_i^{(1)}$ given $y_i^{(2)}$ are respectively:

$$\begin{aligned} \gamma_i &= \mu_i^{(1)} + \Sigma_{12i}\Sigma_{22i}^{-1}\left(y_i^{(2)} - \mu_i^{(2)}\right) && (2.18) \\ \lambda_i &= \Sigma_{11i} - \Sigma_{12i}\Sigma_{22i}^{-1}\Sigma_{21i} && (2.19) \end{aligned}$$

where $\Sigma_i$ is decomposed into the following components:

$$\Sigma_i = \left(\begin{array}{c|c} \Sigma_{11i} & \Sigma_{12i} \\ \hline \Sigma_{21i} & \Sigma_{22i} \end{array}\right) \tag{2.20}$$

where $\Sigma_{11}$ corresponds to the censored variables and $\Sigma_{22}$ corresponds to the non-censored variables. This illustrates the major difference between the univariate and multivariate Tobit models, in that the expected mean and variance are a function of the other observed dependent variables. The negative log likelihood function is now given by

$$\begin{aligned} LL &= \frac{1}{2}\sum_{i:Mort_i>0}\left\{ln\left(|\Sigma_i^{-1}|\right) + (Y_i - X_iB)\,\Sigma_i^{-1}\,(Y_i - X_iB)'\right\} + \\ &\quad \frac{1}{2}\sum_{i:Mort_i=0} ln\left(|\Sigma_{22i}^{-1}|\right) + \left(y_i^{(2)} - X_i^{(2)}B^{(2)}\right)\Sigma_{22i}^{-1}\left(y_i^{(2)} - X_i^{(2)}B^{(2)}\right)' + \\ &\quad ln\left(\Phi\left[-\frac{\gamma_i}{\sqrt{\lambda_i}}\right]\right) && (2.21) \end{aligned}$$

where $B$ can be broken into two components containing the parameter estimates for the censored variable (e.g., MORT), $B^{(1)}$, and the parameter estimates for the uncensored variables (e.g., DMFC, ADG, and VCPH), $B^{(2)}$.[2]

[2]One drawback from using the model described above is that the optimization routine necessary to estimate all parameters may take quite long to converge. The next essay will focus on a Zero-inflated regression model that may work to shrink the computational burden from estimating so many parameters, as well as improve estimation efficiency.

## 2.4 Data

The proposed model is fitted to a comprehensive set of data collected from five commercial cattle feedlots located in Kansas and Nebraska. Proprietary production and cost data were obtained for 11,397 pens of cattle from 1995 to 2004. Table 2.1 contains summary statistics from the data sample.

Table 2.1: Variable Descriptions and Summary Statistics

| Variable | Description | Mean | Std Dev | Min Value | Max Value |
|---|---|---|---|---|---|
| DMFC | Dry matter feed conversion (lbs feed / lbs gain) | 6.19 | 0.72 | 4.00 | 24.00 |
| ADG | Average Daily Gain (lbs gain / day) | 3.36 | 0.48 | 0.74 | 5.78 |
| VCPH | Veterinary cost per head ($) | 11.83 | 6.25 | 0.00 | 60.00 |
| MORT | Percentage of pen that die | 0.93 | 1.53 | 0.00 | 25.83 |
| | Uncensored observations only | 1.71 | 1.73 | 0.16 | 25.83 |
| | Percentage of censoring (%) | 46.60 | | | |
| InWeight | Average weight per head of cattle upon entrance (lbs) | 737.50 | 87.22 | 500.00 | 900.00 |
| Winter | Binary variable equal to 1 if entry between Dec-Feb | 0.25 | | | |
| Spring | Binary variable equal to 1 if entry between Mar-May | 0.23 | | | |
| Fall | Binary variable equal to 1 if entry between Sep-Nov | 0.25 | | | |
| Steers | Binary variable equal to 1 if entire pen were Steers | 0.51 | | | |
| Mixed | Binary variable equal to 1 if pen was mixed gender | 0.12 | | | |
| KS | Binary variable equal to 1 if Kansas feedlot location | 0.80 | | | |

Total sample size n=11,397 pens of cattle

Dry Matter Feed Conversion (DMFC) measures the pounds of dry feed required per pound of live weight gain. To compute the average DMFC for a given pen, total dry feed consumed is divided by the total weight gained during the feeding cycle. Average daily gain (ADG) captures the average pounds gained throughout the feeding period. Veterinary costs per head (VCPH) are calculated by dividing the total dollar amount spent on veterinary services by the pen size upon entry. The mortality rate (MORT) is a percentage calculated as the number of death losses during the feeding period divided by the number of head initially placed on feed. Figure 2.1 shows the extent of correlation among the dependent variables.

The size of a pen of cattle averaged 134 head with an average placement weight of 738 pounds and an average finished weight of 1,178 pounds. *InWeight* is measured as the average weight per head in each pen upon placement on feed.[3] The log of InWeight is used as a conditioning variable. To capture seasonal effects, binary variables are constructed to denote Winter, Spring, Summer, and Fall seasons. In the U.S. there are distinct seasonal trends to placing feeder cattle on feed. Feeder cattle marketings peak in the fall when the calf crop is weaned and transported, while young cattle go on grass in the spring, which coincides with a slow-down in feedlot placements. While there is evidence in favor of placement time of year having an effect on production yield factors (Mark, Jones and Mintert, 2002), cattle need to be placed at all times of year so that feedlots remain near full capacity. Additionally, to keep feedlots near full capacity, pens of cattle must move along their feeding cycles keeping with a relatively tight schedule. Table 2.2 illustrates the unique pen characteristics that are associated with pen placements at different times of year.

Typically, pens of cattle finish near their scheduled dates for two major reasons; in order to keep cattle moving through the feedlots so that new pens can enter and contracts with packing plants restrict the flexibility feedlots have to release pens of cattle whenever optimal. Finishing days can adjust slightly and are mainly due to extreme weather. In this data, fall placements tend to be lighter than any other season, while spring placements tend to bring heavier weights (see Table 2.3).

[3]Pens with average placement weights below 500 pounds and above 900 pounds were excluded from our sample.

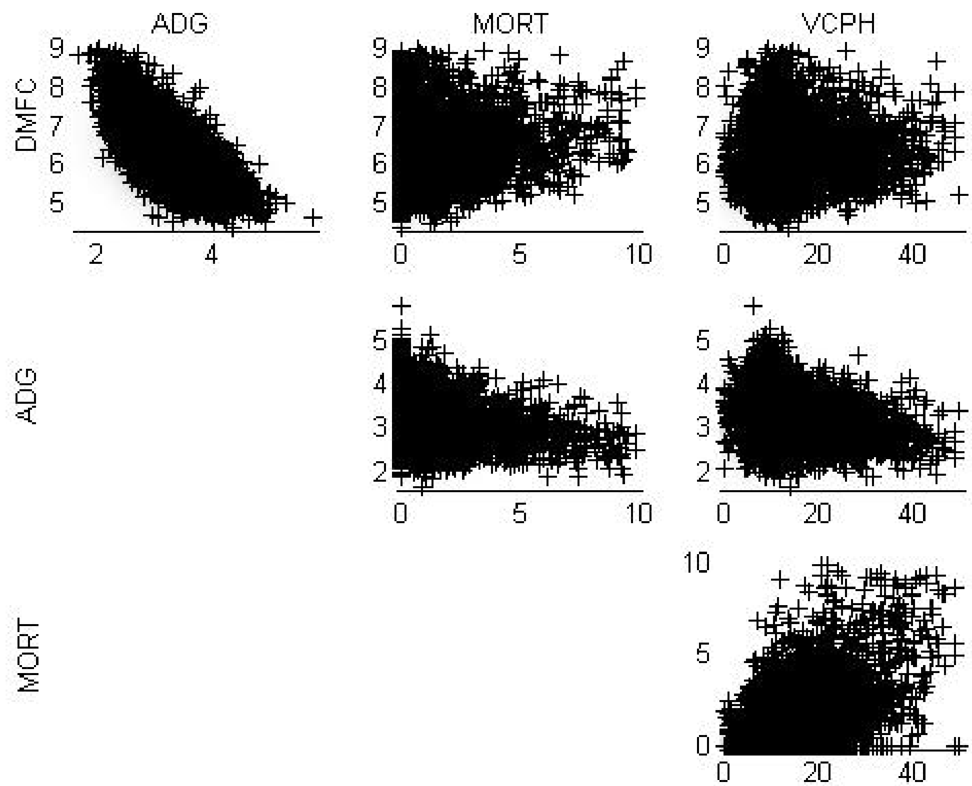



Figure 2.1: Scatter plots of dependent variables

Table 2.2: Pen characteristics of different placement seasons

| Variable | fall | winter | spring | summer |
|---|---|---|---|---|
| Observations | 2,884 | 2,898 | 2,615 | 3,000 |
| DMFC | 6.44 | 6.10 | 6.07 | 6.15 |
| ADG | 3.22 | 3.29 | 3.45 | 3.48 |
| Intake | 20.48 | 19.89 | 20.73 | 21.11 |
| VCPH | 13.11 | 11.50 | 10.64 | 11.96 |
| MORT | 1.14 | 0.92 | 0.70 | 0.95 |
| InWt | 719.69 | 730.77 | 760.65 | 740.93 |
| OutWt | 1,160.91 | 1,170.64 | 1,187.43 | 1,192.96 |
| Days on Feed | 134.66 | 131.52 | 120.90 | 127.55 |
| **Proportion of sample:** | | | | |
| Steers | 0.51 | 0.56 | 0.47 | 0.50 |
| Heifers | 0.36 | 0.32 | 0.41 | 0.39 |
| Mixed | 0.13 | 0.12 | 0.12 | 0.11 |

Table 2.3: Comparison of Different Weight Classes

| Variable | 500-600 | 600-700 | 700-800 | 800-900 |
|---|---|---|---|---|
| Observations | 815 | 2,876 | 4,545 | 3,061 |
| DMFC | 6.22 | 6.13 | 6.19 | 6.25 |
| ADG | 2.93 | 3.19 | 3.40 | 3.56 |
| Intake | 18.03 | 19.30 | 20.81 | 22.00 |
| VCPH | 17.79 | 13.86 | 10.91 | 9.73 |
| MORT | 1.97 | 1.22 | 0.79 | 0.61 |
| InWt | 561.09 | 656.93 | 749.90 | 841.34 |
| OutWt | 1,091.14 | 1,124.00 | 1,181.86 | 1,245.66 |
| Days on Feed | 176.23 | 143.01 | 123.78 | 110.55 |
| **Proportion of sample:** | | | | |
| Winter | 0.26 | 0.28 | 0.26 | 0.22 |
| Spring | 0.10 | 0.17 | 0.25 | 0.29 |
| Summer | 0.25 | 0.25 | 0.27 | 0.28 |
| Fall | 0.39 | 0.31 | 0.22 | 0.21 |
| Steers | 0.28 | 0.35 | 0.50 | 0.74 |
| Heifers | 0.52 | 0.50 | 0.38 | 0.19 |
| Mixed | 0.21 | 0.15 | 0.12 | 0.07 |
| KS | 0.62 | 0.80 | 0.84 | 0.81 |

On average, pens spent 129 days on feed, meaning most pens are fed throughout more than one season. For example, a pen placed in the pleasant fall months will likely be on feed while temperatures drop towards colder winter averages. Binary variables are also used to differentiate pens by gender (Steers, Heifers, and Mixed) where pens comprised of all steers make up more than half of the data sample. Steers are more often used for feeding due to the faster pace with which they put on weight, whereas heifers put on weight slower, max out at a lower weight, and are used for reproduction.

Binary variables are also used to differentiate feedlot location by state, which include Kansas and Nebraska. Feedlot locations could additionally be split by feedlot, however within Nebraska and Kansas each feedlot does not appear to be significantly different. The two Kansas feedlots are relatively larger than the Nebraska feedlots. Additionally, the Nebraska feedlots keep their pens for more days on feed, resulting in lower DMFC and wider weight swings. Table 2.4 illustrates the major differences among feedlots located within Kansas and Nebraska in our data set.

Table 2.4: Comparison of Kansas and Nebraska Feedlots

| Variable | Kansas | Nebraska |
|---|---|---|
| Observations | 9,157 | 2,240 |
| DMFC | 6.04 | 6.79 |
| ADG | 3.40 | 3.20 |
| VCPH | 11.34 | 13.85 |
| MORT | 0.929 | 0.952 |
| InWt | 741.6 | 720.8 |
| OutWt | 1,171.9 | 1,202.6 |
| Days on Feed | 124.0 | 148.7 |

Total sample size n=11,397 pens of cattle

The Kansas feedlots also have set up contracts with backgrounding operations (also known as stocker operations), where pens of cattle are more slowly transitioned from feeding on grass to the protein-rich diets they receive once at the feedlot. Background operations are advantageous to larger feedlots since lighter weight cattle are riskier in production. Lighter weighted cattle are shown to have significantly higher mortality rates and lower rates of average daily gain, as shown in Table 2.3. To minimize this risk, background operations prepare lighter weight cattle to go on feed by grazing for a period of time before they are placed on feed.

Histograms of the dependent variables and entry weight are shown in figure 2.2. Here the positively skewed nature of DMFC, VCPH, and MORT are quite apparent. For this reason the log of DMFC and VCPH is taken in order to symmetrize the variables. Unfortunately, there is no mechanism to take the log of mortality rates since so many observations are zero. Alternatively, ADG is already distributed similar to a normal distribution centered at 3.4. These histograms also illustrate the importance in recognizing the yield factors are not fixed and should be thought of as components of risk in order to more accurately describe fed cattle production.

Additional data from the CBOT and CME databases may be used in order to utilize expected fed cattle and corn prices. More specifically, based on the placement date, a 5 month live cattle futures price reflects the expected selling price, given the cattle are feeding for 5 months. Additionally, estimated feed costs can be estimated with a 2 month corn futures price to average the expected price of corn over the 5 month feeding cycle.

An additional complexity is the censored nature of mortality rates within the data. Nearly 40% of the observations contained no death losses, while some observations had mortality rates above 20%. With a large portion of the observations containing zero values, the parameter estimates will be biased. Belasco et al. (2006) used a Tobit model to correct for censoring, however this was in a univariate setting. The Tobit model was expanded by Chavas and Kim (2004) to account for a dynamic multivariate Tobit model under conditional heteroscedasticity. This research employs a similar strategy, with the additional flexibility that the correlations

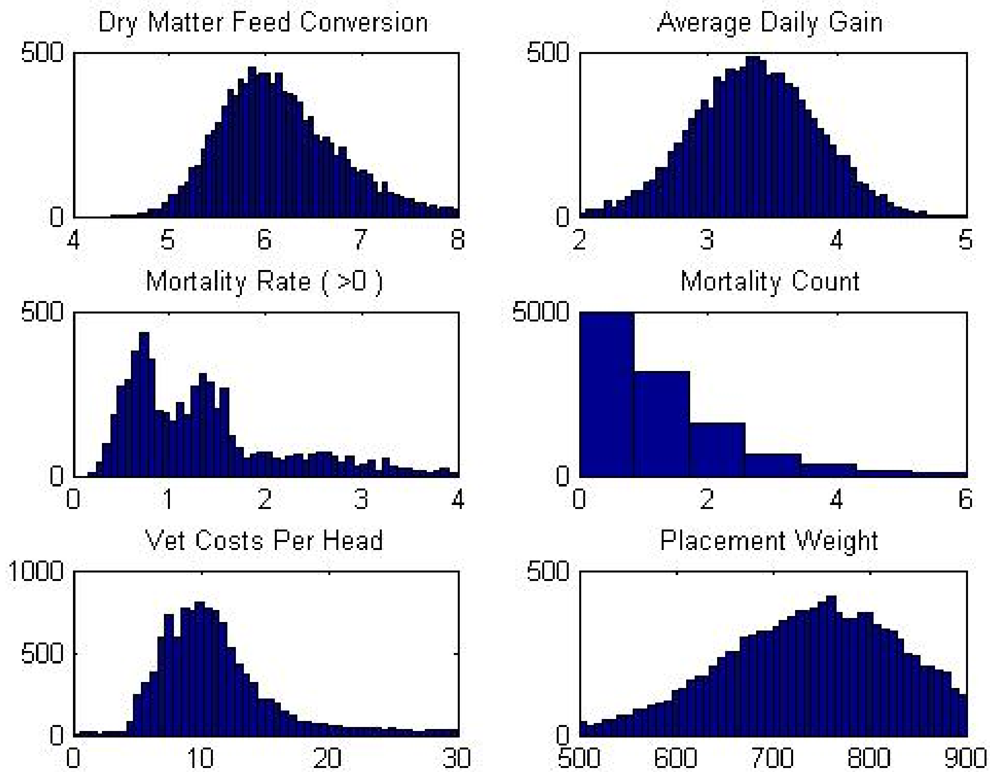



Figure 2.2: Histograms of quantitative variables

between variables are allowed to vary, given the unique conditioning variables.

## 2.5 Estimation Results

The four cattle production yield variables must be estimated by taking into account the finding that each conditioning variable has an effect on the mean and covariance terms. Due to this, it is necessary to discuss the results from each conditioning variable in the context of all yield measures. These parameter estimates contained in matrices B and G are displayed in Table 2.5. This section begins with an interpretation of the parameter estimates on the mean components of the system. The next section deals primarily with the covariance parameter estimates. The estimation results accompanying this discussion can be found in Table 2.6.

Table 2.5: Maximum likelihood parameter estimates

| | DMFC | | ADG | | MORT | | VCPH | |
|---|---|---|---|---|---|---|---|---|
| Variables | coeff. | se. | coeff. | se | coeff. | se | coeff. | se |
| Intercept: | 0.675* | 0.046 | -3.445* | 0.213 | 24.049* | 1.225 | 10.708* | 0.224 |
| Steers: | -0.069* | 0.002 | 0.300* | 0.008 | 0.191* | 0.045 | 0.065* | 0.009 |
| Mixed: | -0.027* | 0.003 | 0.128* | 0.013 | 0.597* | 0.084 | 0.214* | 0.015 |
| Kansas: | -0.123* | 0.002 | 0.169* | 0.010 | -0.102 | 0.048 | -0.221* | 0.008 |
| Log(inwt): | 0.193* | 0.007 | 1.002* | 0.033 | -3.610* | 0.187 | -1.241* | 0.034 |
| Winter: | -0.003 | 0.002 | -0.171* | 0.010 | -0.088 | 0.054 | -0.081* | 0.010 |
| Fall: | 0.050* | 0.002 | -0.221* | 0.011 | -0.032 | 0.060 | 0.003 | 0.010 |
| Spring: | -0.018* | 0.002 | -0.041* | 0.010 | -0.228* | 0.055 | -0.080* | 0.011 |
| **Heteroskedasticity:** | | | | | | | | |
| Intercept: | -9.067* | 0.739 | -8.804* | 0.723 | 12.716* | 0.886 | 7.168* | 0.788 |
| Steers: | -0.060 | 0.030 | 0.058 | 0.030 | -0.042 | 0.038 | -0.527* | 0.031 |
| Mixed: | 0.481* | 0.044 | 0.143* | 0.044 | 0.583* | 0.055 | -0.265* | 0.046 |
| Kansas: | -0.127* | 0.034 | -0.038 | 0.034 | 0.133* | 0.044 | 0.413* | 0.035 |
| Log(inwt): | 0.646* | 0.113 | 0.890* | 0.110 | -1.760* | 0.136 | -1.438* | 0.120 |
| Winter: | 0.013 | 0.037 | 0.085 | 0.037 | -0.109 | 0.048 | 0.466* | 0.038 |
| Fall: | 0.356* | 0.037 | 0.186* | 0.037 | 0.312* | 0.047 | 0.340* | 0.038 |
| Spring: | -0.351* | 0.038 | 0.127* | 0.038 | -0.117 | 0.052 | 0.769* | 0.041 |

*Denotes the estimate is statistically significant at the 0.05 level

### 2.5.1 Performance Effects From Gender

Gender differences are known to play a large role in cattle feedlot performance. Steer cattle are known to gain weight at a much faster rate than heifer cattle and are commonly marketed at a higher weight. Additionally, some heifers need to be used to stock new generations of cattle. For these reasons, steer cattle are more prevalent in feedlots than their heifer counterparts. This is also true within the given data set where steer pens compose 51% of the pens placed on feed.

To assist in capturing the effect that gender has on production, pens were identified as entirely steer, entirely heifer, or some mixture of the two. For estimation purposes, binary variables were developed for each type of pen. Results shown in Table 2.5 are relative to heifer pens. Not surprisingly, both steer and mixed pens have lower feed conversion rates and higher rates of average daily gain. More specifically, pounds of feed are converted into pounds of weight gain more efficiently by 6.9% and 2.7% for steer and mixed pens, respectively. This superior ability to convert feed into weight gain directly assists in the higher rates of ADG for steer and mixed pens, relative to heifer pens. The data suggests that steer pens gain weight faster than heifer pens by 0.30 pounds per day.[4] Results from Mark, Schroeder and Jones (2000) indicate that steer pens had similar performance advantages with a feed conversion that was 4% lower than heifer pens, while gaining an average of 0.34 pounds more per day.

While ADG and DMFC results make steer pens more desirable than heifer pens, the results from the regression equations for MORT and VCPH indicate that steer pens are inferior to heifer pens in general health measures. The percentage of mortality losses while on feed are higher for steer and mixed pens by 0.10 and 0.32, respectively.[5] Given the higher mortality rates for steer and mixed pens, it is not

[4] While 0.30 pounds per day may appear to be a small gain, it amounts to 39 extra pounds over 130 days on feed. This amounts to a 3% gain in out weight over the average heifer pen

[5] To interpret the marginal effects within the Tobit model, MLEs must be multiplied by the

surprising that veterinary costs are higher for these types of pens.[6] Steer pens incur 6.5% higher veterinary costs than heifer pens, while mixed pens are quite expensive with 21% higher veterinary costs.

The heteroskedasticity parameter estimates offer insight into the influence conditioning variables have on the variance. A parameter estimate that is positive indicates an increasing effect on the variance. Variance increases when $\gamma_k > 0$ and decreases when $\gamma_k < 0$. While steer pens do not differ significantly with variance, mixed pens bring on higher variance parameters for most variables, with the exception of veterinary costs.

### 2.5.2 Performance Effects From Location

As previously mentioned, the data contain results from five major cattle feedlots where two reside in Kansas and three in Nebraska. Differences in location are identified with binary variables indicating the state of residences. The main reason for this distinction is due to the geographic closeness of the feedlots within the same state and the similar management practices discussed earlier.[7] The binary location variable is then intended to control for any differences due to different weather systems as well as different management practices.[8]

One of the most distinguishing characteristics of the data is the higher entry

proportion of non-censored observations in the sample (Greene, 2003, pg. 766), which is 53.404% within the data.

[6]Veterinary costs at cattle feedlots can be incurred due to precautionary checks, which are typically performed at the beginning of the feeding cycle and consist of vaccinations and health checks, and visits due to deteriorating health. While this data does not distinguish between the two, it is assumed that all feedlots incur similar expenses for precautionary visits, so that any variation in veterinary costs can be linked to the health of the pen.

[7]Initially binary variables were used to distinguish between each feedlot until it was found that significant differences between the feedlots can be found by differentiating by state, since feedlots were typically not significantly different relative to feedlots within similar states.

[8]Management practices within this data do not represent state-wide practices. In 1996, there was an estimated 670 feedlots with a 1,000+ capacity within the state of Nebraska (NASS, 1997, pg. 109). While fewer feedlots reside within Kansas, a higher proportion of the feedlots are operations with capacities over 32,000 head. (NASS, n.d.).

weights and lower days on feed associated with feedlots residing in Kansas. This may be due to the practice of backgrounding on Kansas feedlots. According to Neville and McCormick (1981), calves that are weaned at an early age and well-fed need less time at the feedlot. The other more obvious reason is that cattle with higher placement weights need less time to reach the desire marketing weight. This finding appears consistent within our data where the Kansas lots have their cattle backgrounded to prepare them for the diet at feedlots. The results in Table 2.5 indicate that DMFC is 12% lower and ADG is higher by 0.17 in Kansas feedlots. Cattle feeding in Kansas feedlots do not have a significantly different rate of mortality, however veterinary costs are lower due to the less days on feed. Vet costs per head per day, which can be computed by dividing VCPH by days on feed, are roughly similar for each state at $0.09. Kansas feedlots within this data sample have mixed influences on variance for each dependent variable.

### 2.5.3 Performance Effects From Entry Weight

Entry weight is the only quantitative conditioning variable. This allows parameter estimates to be interpreted as elasticities for logged dependent variable regressions. The coefficient from the regression on DMFC implies that a 10% increase in entry weight corresponds to an increase in feed conversion by 1.9%. Similar results have been concluded by past studies (Mark et al., 2000; Schroeder et al., 1993). Increases in entry weight by 10% lead to an increase in ADG by 0.10. The increase in daily gain for heavier pens of cattle is partly due to the less time these types of pens spend on the feedlot. Higher feed conversion and daily gains imply an increase in intake for heavier placed cattle.[9]

Pens of cattle that are more mature in age and weight tend to have less health problems. According to Smith (1998), cattle mortalities in feedlot settings

[9] Intake (lbs of feed / day) = DMFC (lbs feed / lbs gain) $\times$ ADG (lbs gain / day).

come mostly from respiratory diseases and digestive disorders. A great majority of health problems occur in the first 90 days cattle are on feed. This is also shown by the results contained within regression results for MORT and VCPH. Mortality rates fall significantly for heavier weights as a one percentage point increase in placement weight is associated with a decrease in the rate of mortality by 0.04 percentage points. Intake per head per day increases steadily with entry weight from 18.03 pounds for the smallest weight class (500 - 600 lbs) to 22.00 for the highest weight class (800 - 900 pounds). These results are shown in Table 2.3. Additionally, feed conversion does not appear to increase linearly as it is maximized for extremely low and high entry weight classes.

Different weight classes also appear to strongly affect other characteristics such as time of placement and gender. Heavier placements appear to be dominated by steer pens, while the lighter placements comprise mostly of heifer pens. More than 50% of the pen placements with average weight below 700 pounds are heifers, while 74% of the heaviest placements (800-900 pounds) consists of steer pens. Also, lighter pens are introduced more typically in the fall months and rarely in spring months. Pens on the heavier side (>700 pounds) are mostly placed in spring or summer months. Entry weight has an increasing effect on the variability of DMFC and ADG, while the variability of mortality and veterinary costs diminishes for higher placement weights.

### 2.5.4 Performance Effects From Placement Season

Changes in temperature can have dramatic changes in cattle feedlot performance. Mark and Schroeder (2002) point out that optimal cattle performance typically occurs between 40 to 60 degrees. Deviations from this range, as well as variability in weather or precipitation, can lead to lower performance. Higher temperatures often result in less weight gain due to lower rates of consumption, while colder temperatures can lead to less efficient feeding as energy is used to maintain body heat. Feeding usually lasts anywhere from 3 to 5 months, meaning most pens will enter in one season and leave in the next.

Parameter estimates for the DMFC regression imply that fall placements have the highest feed conversion rate and are significantly different from summer. This is not surprising given the fact that pens placed in the fall months are fed as the temperatures drop, so that the coldest months are likely near the end of their feeding cycle. A binary variable indicating summer placements is left out of the regression so that parameter estimates are relative to that season. The feed conversion rate for spring placements are significantly lower than summer placements, while winter placements are not significantly different. All seasons experienced significantly lower gains on a daily basis, relative to summer.

Pens placed in the spring months appear to have the fewest health problems as indicated by the significant negative parameter estimates in both MORT and VCPH. Fall placements are not statistically different from summer concerning the mortality rate, while winter placements incur fewer veterinary costs.

### 2.5.5 Conditioning Variable Effects on Covariance Terms

One major benefit of the large set of data available for this research is the chance to allow covariance terms to be a function of the data. In a recent study by Belasco et al. (2006) covariance terms in this system of equations were assumed to be constant for all observations. However, OLS regressions indicated that the cross product residuals were correlated with the conditioning variables.[10] The individual-specific covariance matrix as defined in equation (2.7) allows for the added flexibility in the

[10] One way to test for heteroskedasticity is to regress the squared residual on the variables as defined by the White test (Greene, 2003). If variables are found to significantly effect an error term that is assumed to be independent, then heteroskedasticity must be controlled for. Alternatively, one may also take the cross product of residuals from a system of equations to determine if the covariance terms are in fact independent. This was the preliminary strategy which led to the finding that covariance terms were significantly effected by the conditioning variables.

off-diagonal elements in equation (2.8) to be a function of the data.

Not all variables are expected to be highly correlated with one another, however one can make a strong case for a few relationships to be strongly correlated. For example, feed conversion rates and rates of average daily gain certainly complement one another, while veterinary costs and mortality rates can both arise with unhealthy or sick pens. Each of these examples are shown to have almost all conditioning variables significantly effecting the level of covariance as seen in Table 2.6.

Covariance elements for an individual observation are contained in the matrix $\Sigma_i$, which is a product of two separate matrices, $T_i$ and $D_i$, and are linearly determined by the conditioning variables and parameter estimates from matrices A and G. To illustrate, the covariance between DMFC and ADG for a given observation, i, is estimated through the following equation:

$$\sigma_{12i} = d_{1i}t_{12i} = \left(a_1 x_i'\right) e^{\left(g_1 x_i'\right)} = \left(a_{11}x_{i1} + \ldots + a_{1p}x_{ip}\right) e^{\left(g_{11}x_{i1} + \ldots + g_{1p}x_{ip}\right)} \qquad (2.22)$$

where $\sigma_{12i}$ indicates the element within $\Sigma_i$ and $a_1$ is the first row of $A$ and is a $1 \times 8$ matrix containing the parameter estimates for the covariance level, based on the full sample. This form allows the unique characteristics of each pen to imply a different set of covariance parameters. The covariance terms account for effects that concurrently effect the cattle production yields. The high frequency of significant variables indicate the importance of including this flexibility.

The covariance matrix can be converted into an empirical correlation matrix in order to understand the correlation structure between two variables for a set of conditioning variables. To illustrate, correlation between DMFC and ADG can be estimated as follows:

$$\rho_{12i} = \frac{\sigma_{12i}}{\sqrt{\sigma_{11i}\sigma_{22i}}} \qquad (2.23)$$

Table 2.6: Maximum likelihood covariance parameter estimates

| | ADG | | MORT | | VCPH | |
|---|---|---|---|---|---|---|
| Variables | coeff. | se | coeff. | se | coeff. | se |
| **with DMFC:** | | | | | | |
| Intercept: | 0.119 | 1.391 | 0.367 | 10.142 | -0.123 | 2.466 |
| Steers: | -0.477* | 0.058 | -0.099 | 0.482 | -0.287* | 0.101 |
| Mixed: | 0.166* | 0.076 | 0.170 | 0.768 | 0.117 | 0.144 |
| Kansas: | 0.166* | 0.065 | 1.605* | 0.521 | 0.159 | 0.093 |
| Log(inwt): | -0.603* | 0.213 | 0.909 | 1.545 | 0.055 | 0.375 |
| Winter: | 0.490* | 0.070 | 0.413 | 0.586 | -0.102 | 0.111 |
| Fall: | 0.224* | 0.067 | -0.380 | 0.603 | -0.010 | 0.104 |
| Spring: | 0.136 | 0.083 | 1.281 | 0.668 | -0.352* | 0.143 |
| **with ADG:** | | | | | | |
| Intercept: | | | 3.273 | 4.712 | -0.930 | 0.907 |
| Steers: | | | -0.452* | 0.174 | 0.013 | 0.037 |
| Mixed: | | | -0.509 | 0.320 | -0.039 | 0.058 |
| Kansas: | | | -0.020 | 0.187 | -0.132* | 0.033 |
| Log(inwt): | | | -0.543 | 0.713 | 0.112 | 0.137 |
| Winter: | | | -0.088 | 0.215 | 0.404* | 0.040 |
| Fall: | | | 0.213 | 0.236 | 0.094* | 0.040 |
| Spring: | | | -0.156 | 0.216 | 0.184* | 0.045 |
| **with MORT:** | | | | | | |
| Intercept: | | | | | 0.701* | 0.103 |
| Steers: | | | | | -0.028* | 0.005 |
| Mixed: | | | | | -0.010 | 0.006 |
| Kansas: | | | | | 0.023* | 0.005 |
| Inwtlog: | | | | | -0.095* | 0.016 |
| Winter: | | | | | -0.004 | 0.006 |
| Fall: | | | | | 0.005 | 0.005 |
| Spring: | | | | | 0.015* | 0.007 |

*Denotes the estimate is statistically significant at the 0.05 level

Next, the empirical correlation matrix is computed for each observation in order to illustrate the distributional characteristics associated with the correlation between dependent variables, which is shown in Figure 2.3.

The empirical correlation matrix evaluated at the mean levels of X is also shown below in Table 2.7.

Table 2.7: Correlation matrix relationship evaluated at the means

| Variable | DMFC | ADG | MORT | VCPH |
|---|---|---|---|---|
| DMFC | 1.000 | -0.801 | 0.341 | 0.026 |
| ADG | | 1.000 | -0.319 | -0.064 |
| MORT | | | 1.000 | 0.363 |
| VCPH | | | | 1.000 |

It is no surprise to see high levels of correlation between vet costs / mortality rates and feed conversion / average daily gain for the reasons stated earlier. There also exists a high degree of positive correlation between feed conversion and mortality rates. This can be explained by the higher feed conversion rates that come from unhealthy cattle, while healthy cattle are more efficient at gaining weight. Almost all correlation terms are above 20%, with the exception of VCPH with DMFC and ADG.

Due to the heterogeneous nature of the data, a unique covariance matrix corresponds to every unique set of variables, implying a unique correlation matrix. To illustrate this fact, two hypothetical pens are chosen at approximately one standard deviation from the mean entry weight. The results below in Table 2.8 demonstrate that these two observations have a distinct correlation structure.

Pen A corresponds to a pen that is placed into a Kansas feedlot in the fall, comprised fully of steers. Conversely, Pen B corresponds to a pen that is comprised of heifers and was placed into a Kansas feedlot during the summer months. Three major differences between these observations include the dramatic difference

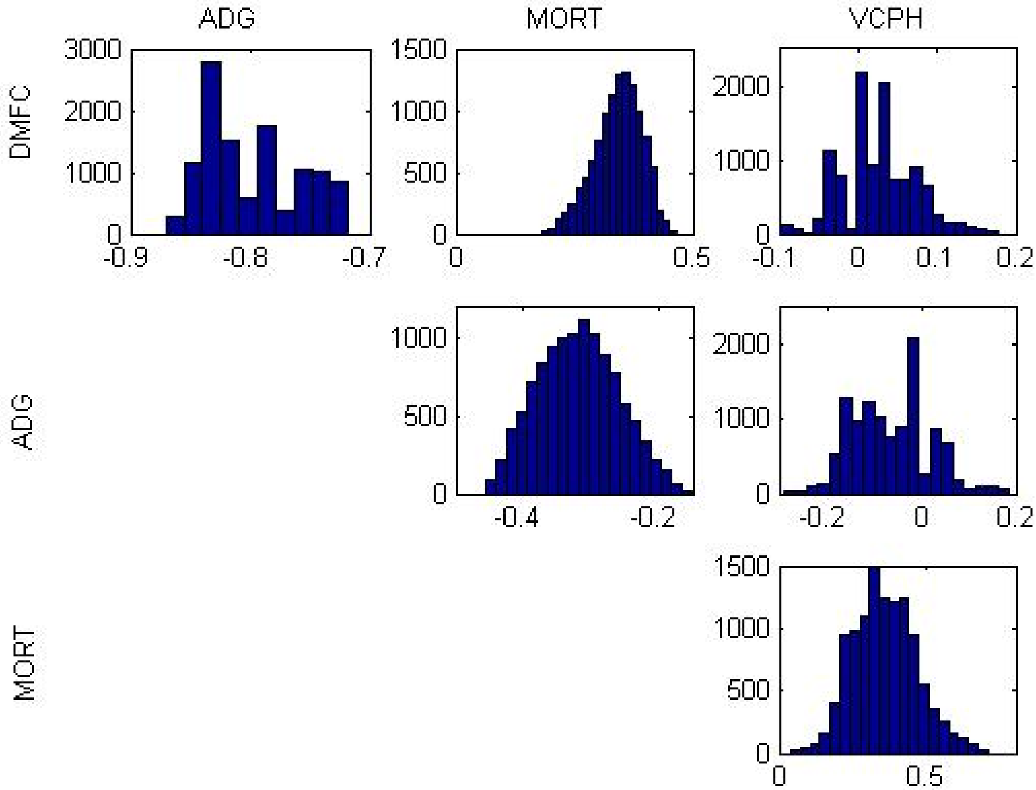



Figure 2.3: Histogram of Predicted Empirical Correlations Between Dependent Variables

Table 2.8: Comparison of correlation matrices for two separate pens

| | Pen A[a] | | | | Pen B[b] | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | DMFC | ADG | MORT | VCPH | DMFC | ADG | MORT | VCPH |
| DMFC | 1.000 | -0.834 | 0.283 | 0.020 | 1.000 | -0.811 | 0.385 | 0.090 |
| ADG | | 1.000 | -0.263 | -0.086 | | 1.000 | -0.342 | -0.181 |
| MORT | | | 1.000 | 0.459 | | | 1.000 | 0.390 |
| VCPH | | | | 1.000 | | | | 1.000 |

[a]Pen A represents a pen entering with a low average weight of 650 pounds

[b]Pen B corresponds to a heavier pen with an average entry weight of 815 pounds

in weight, as well as different placement months and gender.[11] The result is two correlation matrices with very different off-diagonal elements.

To further illustrate the point that changes in the conditioning variables influence elements within the covariance matrix and the implied correlation between two variables, average entry weight can be changed from the mean entry weight (737.5 lbs.) and increased by 10 percent (811.3 lbs.) while maintaining mean values for all other conditioning variables, resulting in Table 2.9.

Table 2.9: Percentage change in correlation matrix elements from increasing entry weight by 10% from mean

| Variable | DMFC | ADG | MORT | VCPH |
|---|---|---|---|---|
| DMFC | – | 0.13% | 11.34% | 16.95% |
| ADG | | – | 12.74% | 7.91% |
| MORT | | | – | -11.04% |
| VCPH | | | | – |

Table 2.9 provides interesting insights into how correlations between the defined dependent variables change for different levels of placement weights. Most notably, a 10 percent increase in weight from the mean results in increased correlations between MORT with DMFC and ADG by 11.34% and 12.74%, respectively. This relationship demonstrates the changing relationship between performance and health factors. In fact, we can broadly say that health factors and performance factors become more correlated as entry weight increases. The reason for this movement has much to do with the point that pens placed at heavier weights typically spend less time on feed. Because of this limited time on feed, sick animals have less time to recover from a performance stand point. Alternatively, a pen that is placed on feed at a light entry weight and gets sick upon arrival still has time to attain higher performance levels upon recovery.

[11] Pen A contains an entry of 650 pounds, which is lower than 79% of the placement weights in the data sample. Also, Pen B weighs in at 815 pounds, which is lower than 18% of the observations.

Additionally, with an increase in placement weight by 10 percent, there is a weakened relationship between MORT and VCPH. An explanation for this movement includes the point that lighter cattle are more prone to high mortality rates, in which situations veterinary treatments may not be as effective. Therefore, at lighter weights, feedlots are likely to experience high VCPH and MORT with a pen that arrives sick. Alternatively, as pens mature the expected mortality rate decreases and they are more adaptable to different occurrences, such as adverse weather. Because of this, veterinary car is likely to be effective and not result in mortality, which points to a weaker relationship between the variables.

The same figures can be computed to understand the relationship between variables for different pen genders. In Table 2.10 a steer pen is compared to a heifer pen, holding all other variables at their means. Biological differences between steer and heifer beef cows imply different relationships between dependent variables. It is interesting to note that the sign of correlation between DMFC and VCPH switches depending on the gender of the pen. Evaluating a heifer pen at mean values, this correlation is 0.046, which changes by -115.07% to -0.007 for a steer pen. Accounting for the differences that arise from different pen characteristics is an important addition from this model.

Table 2.10: Percentage change in correlation matrix elements in a steer pen, relative to heifer pen

| Variable | DMFC | ADG | MORT | VCPH |
|---|---|---|---|---|
| DMFC | – | 2.52% | -2.28% | 115.06% |
| ADG | | – | 11.58% | -47.36% |
| MORT | | | – | -11.97% |
| VCPH | | | | – |

The proportion of statistically significant covariance parameter estimates provide evidence in favor of including these variables. A restricted case of this model is where covariance parameters are constant across individuals. This restriction is typical in studies of this nature due to the elimination of many parameter estimates (Belasco et al., 2006; Chavas and Kim, 2004). For our purposes, a restriction that assumes a constant covariance structure across observations increases degrees of freedom by 42, as the parameters estimates drops from 112 to 70. To test the effectiveness of this restriction on the given data set, a likelihood ratio test can be applied.

Within the likelihood ratio test framework, the flexible model described above will be the unrestricted case, while the restricted model will assume constant covariance terms. The restricted model reduces equation (2.11) to include only the covariance terms, which now are not a function of the data. Formally, the restriction can be stated as follows:

$$H_0 \quad : \quad a_{j,k} = 0 \tag{2.24}$$

$$H_A \quad : \quad a_{j,k} \neq 0 \tag{2.25}$$

for all $j = \{1, 2, ..., 6\}$ and $k = \{2, 3, ..., 8\}$, where $a_{j,k}$ elements are contained within the matrix A. Two components are important when we compare the restricted and unrestricted model performance, which include model fit and predictive power. In order to test both components we split the data into thirds, using a randomly selected two-thirds of the data for estimation and the final third to examine out of sample prediction. In order to test model fit, a likelihood ratio test to examine the loss in explanatory power and the gain in degrees of freedom when moving from the unrestricted to the restricted model. The null hypothesis stated above, tests whether the conditioning variables have a significant impact on the covariance terms. The results from the implied likelihood ratio test are shown in Table 2.11.

These results strongly reject the notion of constant covariances within this data set. However, this restriction may be helpful when evaluating more homogenous production. The heterogeneity of cattle herds is an important aspect to this research and

Table 2.11: Likelihood Ratio Test Results based on estimation portion of data

| Model | Log Likelihood Values | P | Statistic | Critical Value ($\alpha = 0.05$) | p-value |
|---|---|---|---|---|---|
| Unrestricted | -15,683.8 | 112 | | | |
| Constant Covariance | -15,748.2 | 70 | 128.8 | 55.8 | $< .001$ |

led to the usage of covariance measures that were not constant across observations.

To examine predictive power based on out of sample observations, we use Mean Squared Prediction Error (MSPE) which is computed based on the following relationship:

$$MSPE = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2. \tag{2.26}$$

Based on prediction results, based on the two tests it is not clear which model possess superior predictive ability. The results based on the out of sample observations can be found below in Table 2.12

Table 2.12: Mean Squared Prediction Error Results based on out of sample prediction

| Model | DMFC | ADG | MORT | VCPH |
|---|---|---|---|---|
| Unrestricted | 0.008 | 0.182 | 2.691 | 0.174 |
| Restricted | 0.009 | 0.232 | 2.617 | 0.216 |

Superior predictive performance on the unrestricted model is found based on smaller MSPE. An interesting result is the fact that the only variable without a superior fit is that of mortality rates, which happen to be censored. However, with the exception of this variable, the unrestricted model is found to better characterize within and out of sample observations. This finding offers further support for the dynamic multivariate Tobit model proposed in this chapter. The next section will focus on using the estimation results from this model to simulate *ex-ante* cattle feeding profits.

## 2.6 Modeling Profits

With accurate distributional modeling of cattle production yield characteristics, coupled with assumed distributional characteristics of price variation, and taking into account the joint correlation between these variables, conditional *ex-ante* measures of profits can be computed. These profits will be a function of the unique characteristics of each pen, so that each set of characteristics lead to a unique profit distribution. This will be important in analyzing the extent of risk involved in overall profits. Simulation methods are used to incorporate the estimated distributional characteristics of yields, the assumed distributional characteristics of prices, and the marginal and joint effects from the conditioning variables. Here different shocks can occur that may affect the expected profits, variability, and covariance. An example of such a shock would be to the variability of fed cattle prices, or corn prices. It is also worth mentioning that shocks to fed cattle and corn prices will be independent of yield shocks.[12] Based on daily cash prices from 1980 - 2005, a correlation of -0.14 was used to characterize the relationship between fed cattle and corn prices.[13]

### 2.6.1 The Profit Function

In order to model profitability risk, the following *ex-ante* profit function takes into account both the revenue and costs specific to cattle feeding. A similar profit function was used by Belasco et al. (2006). Based on the fact that it is *ex-ante* conditional profits that we are interested in, profits are a function of characteristics that are known at the time a pen is placed. Based on this information, inferences can be

[12]This is mostly a simplifying assumption and may need further analysis. An argument has been formulated by Anderson and Trapp (2000) that changes in the price of corn can cause feedlots to substitute away from corn and towards other grains, like wheat. While this substitution will help the feedlots to keep their costs down, it may have an effect on feed conversion. Additionally, changes in corn prices may also effect the characteristics of cattle placed on feed. For example, placement weight may increase as a way to minimize days on feed.

[13]To compute the correlation coefficient, daily cash prices were obtained for #2 yellow corn at Central Illinois and TX/OK live cattle spot prices.

made concerning different areas of production risk that have previously been mentioned, in order to make inferences about expected profit distributions. Following is the set of equations that explain the fed cattle production profit function. Per head cattle feeding profits are the net difference between revenue and costs accrued during the cattle feeding period.

$$P = TR - FDRC - YC - FC - IC - VCPH \tag{2.27}$$

where $P$ are per head profits, $TR$ is the total revenue per head from cattle feeding, $FDRC$ is the per head cost of purchasing feeder cattle, $YC$ is the per head fixed cost (yardage cost) of feeding cattle, $FC$ is the per head feed cost, $IC$ is an interest cost, and $VCPH$ are the per head costs associated with veterinary care. $TR$ is defined as

$$TR = FP \times (0.96) \times CSW \times (1 - MORT) \tag{2.28}$$

where $FP$ is the price per hundred weight (\$/cwt) of fed cattle and $CSW$ is the average sell weight of the finished cattle, which is estimated based on the following equation

$$CSW = CPW + ADG \times DOF. \tag{2.29}$$

$CPW$ is the average weight of the feeder cattle at placement and $DOF$ is the number of days the pen of cattle is in the feedlot.

$TR$ is adjusted for death loss using the MORT variable and a standard 4% live-weight shrinkage factor is applied to reflect the expected loss in weight during transport from the feedlot to the packing plant. Sell weight is a function of a random performance variable ($ADG$) and therefore is not fixed. This profit function allows days on feed to be specified, while allowing sell weight to be determined by the average weight upon entry, $ADG$, and the length of time on feed. Cattle are assumed to be marketed on a cash basis as opposed to a price based on dressed weight or grid pricing.[14] To capture the expected *FP* at the time of placement, the futures price from the CME can be used to proxy the price for the expected end date. *FDRC* is defined as

$$FDRC = FRP \times CPW \tag{2.30}$$

where *FRP* is the price per hundred weight of feeder cattle. This cost is a large portion of total costs and reflects the value of the cattle upon entering the feedlot. On a per pound basis, *FRP* is greater than *FP*. *YC* is defined as

$$YC = (0.40) \times DOF \tag{2.31}$$

which assumes that \$0.40 is a typical per head day cost for feedlots in Kansas and Nebraska. *FC* is defined as

$$FC = CP \times \left\{ \frac{DMFC}{0.88} \Big[ CSW \times (1 - MORT) - CPW \Big] \right\} \tag{2.32}$$

where *FC* is the price per bushel of corn and is divided by 56 to convert this price into a per pound measure. The expected price of corn is based on the futures price for corn from the CBOT halfway through the feeding period. The reason for this timing is to capture an average price of corn over the entire feeding period. Further, dry matter is multiplied by the corn-based feed ration, which is assumed to 12% moisture. *DMFC* is adjusted to reflect the "as fed" feed conversion. *IC* is defined as

$$IC = \left\{ \frac{1}{2} \Big[ YC + FC + VCPH \Big] + FRC \right\} \times IntRate \times \frac{DOF}{365} \tag{2.33}$$

where *IR* is the interest rate. This expression assumes that an interest charge is

[14] For cattle sold on a grid, quality risk must enter the profit function. For the purposes of this research, quality risk is not taken into account. Cash prices are based on the average weight of the pen, without regard for the quality of the carcass. Evaluating quality risk remains an area of future research.

applied to the full amount of the feeder cattle cost, FRC, and half the total cost of yardage, feed, and veterinary fees. This assumption is based on the need to purchase feed throughout the feeding period, while the feeder cattle must be entirely purchased at the beginning of the feeding period.

It will be important to compare this function with *ex-post* profits, which is contained within the data set. To this end, the data set contains the actual feed charges and vet costs. Additionally, through the use of local Dodge City weekly prices for fed and feeder cattle, *ex-post* profits can be obtained.

However, these two profit measures will clearly differ due to variation that occurs throughout the feeding period. This variation will come through the random variables in this model. Analyzing the difference in *ex-post* and *ex-ante* will be informative in identifying the reasons for departures from expected values.[15] Additionally, once we model profits, one might realize that revenue and net margins are also contained within the profit function, which are measures used for livestock insurance purposes. Modeling these aspects will be important for the next part of this research.

### 2.6.2 Simulation of Profits

The formulated model in this study offers a flexible multivariate regression that characterizes production risk in fed cattle production by accounting for characteristics that impact the mean and covariance of the four defined cattle production yield factors. The implied estimation results assist understanding placement characteristics of pens that affect different areas of production. These areas of production range from the health of the pen to performance measures. This estimation allows for better prediction of yield characteristics. With knowledge of the mean and co-

[15] This research does not attempt to sort out the origins of unexplained variation, however an area of future research might include this endeavor. Major sources of variation are likely to come from unexpected or extreme weather or unobservable genetic traits.

variance of each particular yield factor, predictions for the system of equations can be estimated with relative accuracy. This accuracy is largely due to the flexibility offered by the large data set. Random draws from a multivariate normal distribution simulate a collection of predictions for each of the yield factors.

Random draws from a multivariate normal distribution simulate a collection of predictions for each of the yield factors. Given this information, the profit model described in equations (2.27)-(2.33) can be simulated, conditional on entry pen characteristics. In practice, this profit function can serve as a means for cattle owners or those in the cattle industry to understand expected profits that are a function of the unique characteristics of a pen of cattle placed on feed. To illustrate, a sample pen consists of its own unique characteristics, such as location, gender, entry weight, and placement season. This information influences the inferences made on production yield factors that were modeled using a multivariate normal dynamic regression model that describes production risk. The four dimensional multivariate dynamic regression model allows the mean values to change as a function of the parameter estimates and also the variance to change as a function of covariance parameter estimates. Simulated realizations of the cattle production yield variables and prices are plugged into the profit function (P) to obtain a distribution of conditional *ex-ante* profits.

In addition to production risk, the model must also account for price risk. The expected prices and variance for fed cattle and corn can be obtained by using futures and options measures from the CME and CBOT. This is all the information necessary to characterize the profit function for simulation. Repeated random draws are taken in order to illustrate profits as a distribution.[16] To further illustrate, the characteristics denoted in Table 2.13 are used to emulate a pen of cattle entering a Kansas feedlot on February 14th of 2007 (02/14/2007).

[16] For purposes of this study, 100,000 random draws were found to be enough to obtain a suitable profit distribution.

Table 2.13: Characteristics of Simulated Pen of Cattle

| Placement Characteristics | | Price Characteristics[a] | |
|---|---|---|---|
| Date | 2/14/2007 | Corn Futures Price | 4.25 |
| Weight(lbs) | 750 | Fed Cattle Futures Prices | 93.60 |
| Gender | Steer | Feeder Cattle Cash Price (DC) | 97.98 |
| Location | Kansas | Corn Volatility | .30 |
| Season | Winter | Fed Cattle Volatility | .20 |

[a]Expected prices based on 3 and 4 month futures price for corn and fed cattle, respectively.

Futures prices are used to approximate price expectations for corn and fed cattle. For corn, a 3-month futures price is used to approximate the average cost of corn over the entire feeding period. Fed cattle price expectations are estimated using a 4-month futures price, to denote the expected value of a fed steer when it is ready to be marketed. The Fed cattle volatility measure was assumed to be 20%, while the rise in corn volatility over the past year has led to the higher rate of 30%.[17]

While this simulation has aspects unique to the current production situation, inputs can be changed for other purposes. Figure 2.4 shows the expected distribution of profits, given the previously mentioned inputs. It is not surprising that profits are near zero, with wide tails. Expected profits are centered at $17.14 per head, with a standard deviation of $251.71. Recent increases in corn prices add to rising feeding costs, which negatively influence profits. One thing to note is that higher corn prices typically cause the demand for heavier feeder cattle to increase as it is seen as desirable to send pens to the feedlot for fewer days.

The distribution of *ex-ante* profits exhibits two distinct characteristics. First, the mean of profits are centered close to zero. The second characteristic is the large tails that indicate a small probability of heavy gains or significant losses. Some of this variation may be unobservable in the data, but observable by the producer. For example, a particular breeder may have a superior genetics program that is known

[17]The rise in corn price and volatility over the last year is largely the result of a strong summer drought in the midwest during 2006, coupled with the increased demand for ethanol production.

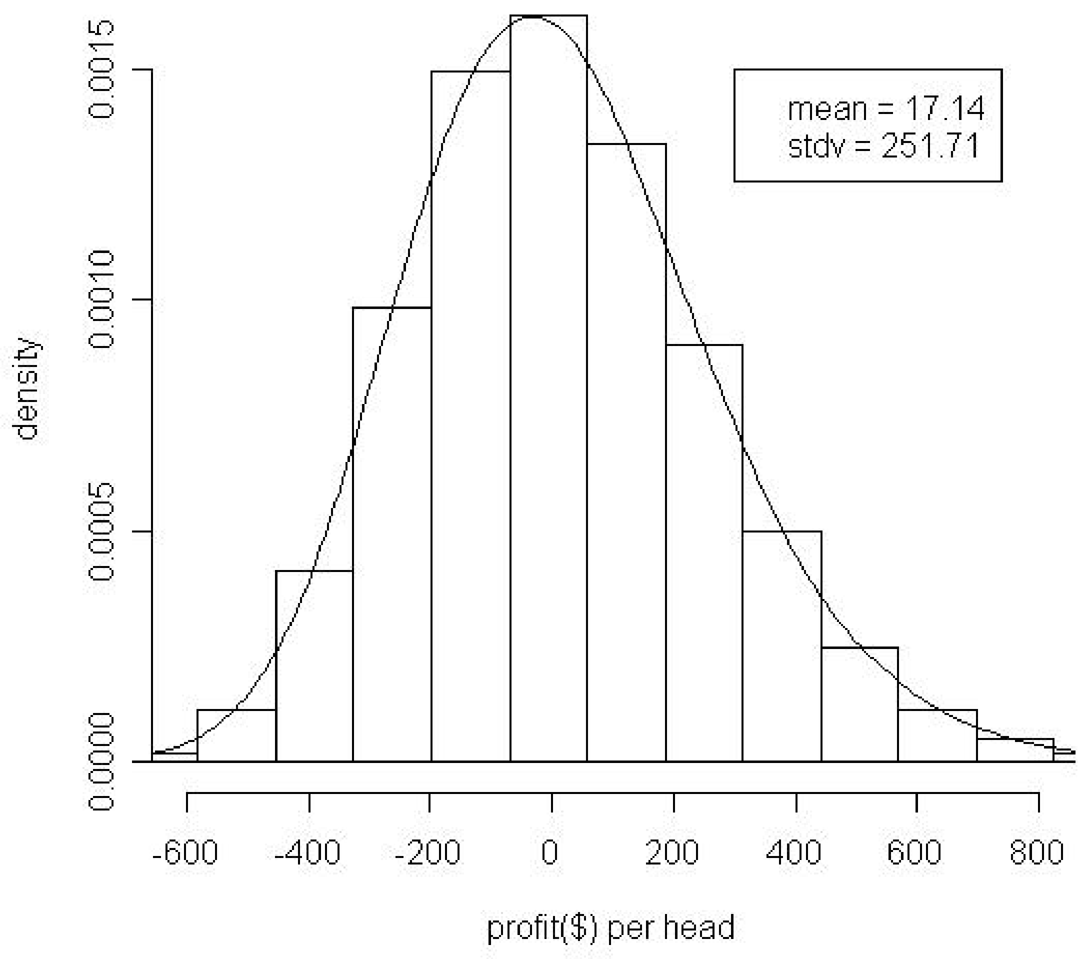



Figure 2.4: Distribution of *ex-ante* conditional profits

in the industry. This producer could demand a premium the pen for sale, which would be expected to outperform average pens. Alternatively, some owners may pay less for a pen with low expectations in hopes of making a profit with higher than expected performance. So, even though this pen performs poorly, relative to the average, it still may be profitable. The point here is that the cattle owner may have more information about a particular pen than is in this data set. This asymmetric information between the econometrician and the producer might lead to wider variation for the econometrician.

## 2.7 Conclusions

Risks involved with agricultural production distinguish these enterprises from most other ventures. Particular emphasis has been placed on characterizing production risks in crop production through the use of a conditional probability density function. The same techniques do not transfer directly when evaluating live animals, such as cattle. One complication is that productive efficiency can be characterized by a set of dependent variables, which are usually highly correlated. The second complication is the censoring mechanism found in one of these variables (MORT). To efficiently characterize cattle production yield risk, both complexities must be accounted for in a modeling strategy.

To this end, the cattle production yield risks are modeled through the use of a dynamic, multivariate Tobit model that consistently estimates the latent variable and captures the correlated nature of the dependent variable. The construction of this model leads to a positive definite hessian matrix, which guarantees that a maximum value will be uniquely defined.

A comprehensive data set, including 11,397 pens of cattle place on feed in Kansas and Nebraska feedlots from 1995 - 2004, were used to examine the risks associated with feeding cattle. The results indicate that conditioning variables, such as average placement weight, gender, location, and season of placement, had a significant effect on both the mean and covariance estimates for the four cattle production yield variables. Additionally, an empirical correlation matrix was computed from sample covariance matrices to illustrate the finding that correlations change significantly across observations.

Using a conditional multivariate normal distribution to characterize yield risks, along with implied price risks from the futures and options market, simulations were conducted to compute *ex-ante* conditional profit distributions. These profit distributions aid in our understanding of cattle feeding risks that can change over time (price risks) or for varying entry characteristics (yield risks).

# Chapter 3

# Modeling Censored Data Using Zero-Inflated Regressions with an Application to Cattle Production Yields

## 3.1 Introduction

The expanded use and availability of micro-level data sets has led to an increased demand for methods that model limited dependent variables efficiently. The traditional and most popular method of estimating censored data is to use the Tobit model. This method assumes there is a latent variable, $y_i^*$, which linearly depends on the associated independent variables, $x_i$, and is typically modeled by using maximum likelihood estimation. Recent studies have focused on extending the Tobit model into multivariate cases. However, a major limitation of the Tobit model is its assumption of normality. Inconsistent estimation results arise when residuals are positively skewed. Additionally, maximum likelihood estimation becomes complicated with a system of equations when censoring occurs in multiple equations because of the problem of integrating more than three integrals in the likelihood function. Furthermore, the most restrictive assumption in the Tobit model is that censored and non-censored observations come from the same set of functional relationships.

In this study we consider the use of a mixture model to characterize censored dependent variables as an alternative to the Tobit model. This particular mixture model arises from a generalized zero-inflated model. This model will be shown to nest the Tobit model, while major advantages include the flexibility in distributional assumptions and an increased efficiency in situations involving a high degree of censoring. Mixture models characterize the censored dependent variable as a function of two distributions in the following manner:

$$Y = VB. \tag{3.1}$$

First, $B$ measures the likelihood of zero or positive outcomes, which have been characterized in the literature using Bernoulli and Probit model specifications. Then, the positive outcomes are independently modeled as $V$, which have traditionally consisted of count models, such as the Poisson distribution. For our purposes, we examine the use of a zero-inflated log-normal and a zero-inflated Gamma distribution to evaluate the observed positive continuous dependent variables. This model is then extended to accommodate multivariate situations where one variable is censored and the others are not, which naturally extends to allow for situations where multiple variables are censored.

Data will be simulated to test the ability of each model to fit data and predict out of sample observations. Results from the zero-inflated mixture model will be compared to Tobit results through the use of goodness-of-fit and predictive power measures. By simulating data, the two models can be compared in situations where the data generating process is known, which is rarely the case when dealing with real world data.

In addition, a comprehensive data set will be used that includes proprietary cost and production data from 5 feedlots in Kansas and Nebraska, amounting to over 11,000 pens of cattle during a 10 year period. Cattle mortality rates on a feedlot, which are provided in this data set, provide valuable insights into the profitability and performance of cattle on feed. Additionally, cattle mortality rates may be more accurately characterized by a mixture model that takes into account the positive skewness of mortality rates, as well as allowing censored and non-censored observations to be modeled independently. Moreover, a zero-inflated specification is used rather than other mixture specifications, such as the Hurdle model, to more accurately capture measures of cattle production yields. In both univariate and multivariate situations, the proposed mixture model more efficiently characterizes the data. Placement characteristics, such as gender, average entry weight, date, and location, which significantly impact the health and performance of pens of cattle are also included in this data.

Additionally, a multivariate setting can be applied to these regression models by taking into account other variables that describe the health and performance of feedlot cattle. These variables include dry matter feed conversion (DMFC), average daily gain (ADG), and veterinary costs (VCPH). DMFC is measured as the average pounds of feed a pen of cattle require to add a pound of weight gain, while ADG is the average daily weight gain per head of cattle. VCPH is the amount of veterinary costs that are incurred over the feedlot stay. Three unique complexities arise when modeling these four correlated yield measures. First, the conditioning variables potentially influence the mean and variance of the yield distributions. Since variance may not be constant across observations, we assume multiplicative heteroskedasticity within our model and model conditional variance as a function of the conditioning variables. Second, the four yield variables are usually highly correlated, which is accommodated through the use of multivariate modeling. Third, mortality rates present a censoring mechanism where almost half of the fed pens contain no death losses prior to slaughter. This clustering of mass at zero presents biases when traditional least squares methods are used.

This research focuses on the third complexity, while accommodating the first two. While the focus of this particular research deals with cattle production yield modeling, a large proportion of censored observations are also commonly found in other production and consumption data sets. Modeling consumption, the spread of animal disease, or production processes that contain multivariate relationships, can also be the beneficiary of such research.

This paper provides two distinct contributions to existing research. The first is to develop a continuous zero-inflated model as an alternative modeling strategy to the Tobit model and more traditional mixture models. This model will originate in a univariate case, then be extended to allow for multivariate settings. The second contribution is to more accurately describe production risk for cattle feeders by examining model performance of different regression techniques. Mortality rates play a vital role in cattle feeding profits, particulary due to the skewed nature of this variable. A clearer understanding of mortality occurrences will assist producers as well as private insurance companies, who offer mortality insurance, in managing risk in cattle operations. Additionally, production risk in cattle feeding enterprises play a significant role in profit variability, but is currently uninsured by current federal livestock insurance programs. An accurate characterization of production risk plays an important role in addressing risk for producers or insurers.

A review of research examining different ways to model censored data are described in section 3.2. Section 3.3 describes the Tobit model and develops the zero-inflated mixture model to be used for estimation in this research. In both cases, the univariate model will precede the development of a multivariate model. The next section (3.4) will simulate data based on the given Tobit and zero-inflated models to evaluate the difference in assuming one model when the data comes from a different process. This evaluation will consist of both how well the model fits the data and the predictive accuracy. The advantage to this section will be to evaluate the performance of the two postulated models in situations where the data generation process is known. This will lead into an application where we evaluate data from commercial cattle feedlots in Kansas and Nebraska, which is discussed in section 3.5. Results from estimating the given data set using a Tobit and zero-inflated model will be assessed using both univariate and multivariate models in section 3.6. Finally, section 3.7 will provide the implications of this study and provide avenues of future research.

## 3.2 Literature Review

Censored dependent variables have long been a complexity associated with micro data sets. The most common occurrences are found in consumption and production data. In studying consumption data sets, households typically do not purchase all of the goods being evaluated in every time period. Similarly, a study evaluating the number of defects in a given production process will likely have outcomes with no defects. In both cases, ordinary least squares parameter estimates will be biased when applied to these types of regressions (Amemiya, 1984).

The seminal work by Tobin (1958) was the first to recognize this bias and offer a solution that is still quite popular today. In his study, household expenditures on durable goods were evaluated by focusing on the fact that observed expenditures cannot be negative. The resulting model later became known as the Tobit model, due to its similarities with the Probit model. This method essentially assumes there is a latent variable, $y_i^*$, which linearly depends on the associated independent variables, $x_i$. In the case of left censoring, the observable dependent variable, $y_i$, can be linked to the latent variable through the following relationship

$$y_i = max(y_i^*, s_i) \tag{3.2}$$

where $s_i$ indicates the location of censoring, which in many applications is 0. However, variables need not be restricted to censoring at 0. For example, prices may be censored at market equilibrium or at some federally defined price floor. Alternatively, it is also common to find right censoring, where the latent variable falls at a point above $s_i$. Combining these two concepts, interval censoring can also be found where observable data is bounded with both upper and lower limits.

Amemiya (1973) demonstrates that the parameters of the Tobit model can be estimated through the use of maximum likelihood, while restoring consistency and asymptotic normality. Univariate models have been extended, under a mild set of assumptions, to include multivariate settings (Amemiya, 1974; Lee, 1993). While empirical applications in univariate settings are discussed by Amemiya (1984), multivariate applications are becoming more frequent (Chavas and Kim, 2004; Cornick et al., 1994; Eiswerth and Shonkwiler, 2006).

While the Tobit model has had a large impact on modeling censored dependent variables, it is not without limitations. The two major assumptions made by the Tobit model in its original derivations included the assumption of normality and the point that both the observable and unobservable variable levels come from the same distribution. The assumption of normality has made the Tobit model inflexible to data generating processes outside of that major distribution (Bera et al., 1984). Additionally, Arabmazar and Schmidt (1982) demonstrate that random variables modeled by the Tobit model contain substantial bias when the true distribution is non-normal and has a high degree of censoring.

Cragg (1971) generalized this model to allow for the probability of a limit observation and the regression for the observed data to be independent processes.

This model is commonly referred to as the hurdle model, which nests the standard Tobit formulation. The hurdle model is formulated so that once a hurdle has been crossed, an outcome can be characterized by a truncated-at-zero density function. This model allows for zero observations to come from one data-generating process and positive realizations to come from another. An extension of the hurdle model is the double-hurdle model, which adds a second hurdle to the process. For example, Jones (1989) uses a double-hurdle model to characterize cigarette consumption by arguing that participation and consumption are two separate processes. First, the consumer chooses whether to participate in the activity of smoking or not. Then, a regression follows to model the amount of cigarette consumption. This model allows for a participant defined as a "smoker" to consume a non-negative amount of cigarettes in any given period. The second hurdle in this case in the consumption of a positive amount of cigarettes. The important point here is that variables influencing the probability of a non-zero value need not also increase the conditional mean of the positive values in the same way (Yen, 1993).

In most applications of the hurdle model, a Probit regression model is used to model the probability of a non-zero value. The Probit regression model assumes that the probability of a non-zero outcome can be modeled as

$$Pr(d = 1|x) = \Phi(x'\beta) \tag{3.3}$$

where $d$ is a binary outcome variable, $x$ is a set of regressors, and $\Phi$ is the CDF of a standard normal distribution. Given a simple model with a latent variable, $y^*$, where $y^* = x'\beta + \epsilon$ and $\epsilon$ is normally distributed with constant variance, $d$ is equal to one when $y^*$ is positive and zero otherwise, which can be written as follows:

$$\begin{aligned} d &= 1 \texttt{ if } y^* > 0 \\ &= 0 \texttt{ otherwise} \end{aligned} \tag{3.4}$$

Another important model that is also used to characterize observations containing many zeros is the zero-inflated models. These models have been rarely used in econometrics. The Poisson distribution is commonly placed into a zero-inflated framework and is appropriately called the zero-inflated Poisson (ZIP) model. The advantage to using this type of model is again that it recognizes that decisions or production output processes are part of a two step process. However, this model is slightly different than the hurdle model in practice. Here the only possible observation is 0 with probability $\rho$, while there is a $1 - \rho$ probability that a random variable is observed from a Poisson distribution, which can still take on the value of zero (Cameron and Trivedi, 2005). Zero-inflated models use a binary process to model the likelihood of an outcome, leading to unbiased estimates of the conditional mean, which is typically not the case with hurdle models. When applied to continuous data, the zero-inflated and hurdle models can be generalized to be similar. As pointed out by Gurmu and Trivedi (1996), hurdle and zero-inflated models can be thought of as refinements to models of truncation and censoring. Hurdle models typically use truncated-at-zero distributions, but are not restricted to truncated distributions. For example, Cragg (1971) recommends the use of a log-normal distribution to characterize positive values. However, most applications of the hurdle model assume a truncated density function.

The probability of a non-censored outcome in zero-inflated models can be computed using a Bernoulli distribution. The Bernoulli distribution is a binary probability density function which takes the value of 1 with probability $\rho$ and the value 0 with probability $1 - \rho$. A Bernoulli density function is nested within a Binomial distribution with 1 trial ($n = 1$).

A common application has been to model the number of defects in a production process. In this case, a zero defect count is highly probable when the production state is nearly perfect. Conversely, if the production process is highly imperfect, then defective outcomes become more likely. The interesting point here is that a predicted outcome may still be zero, even though it is drawn randomly from a Poisson distribution. The binary component models the likelihood that the dependent variable can be described by the assumed distribution, which may still take on zero values.

Lambert (1992) extended the classical ZIP model to include regression covariates, where covariate effects influence both $\rho$ as well as the nonnegative outcome. This study modeled defect counts on printed writing boards as a function of variables that characterize the manufacturing process. Maximum likelihood estimation was used to regress the covariate effects for the ZIP model, which are further derived in Hall (2000). Li et al. (1999) developed a multivariate zero-inflated Poisson model that is motivated to evaluate production processes involving multiple defect types.

A further extension to zero-inflated models is introduced by Ghosh et al. (2006), where a flexible class of zero-inflated models can be applied to discrete distributions within the class of power series distributions. The study also finds that Bayesian methods have more desirable finite sample properties than maximum likelihood estimation, with their particular model. Computationally, a Bayesian framework may have significant advantages over classical methods. In classical methods, such as maximum likelihood, parameter estimates are found through numerical optimization, which can be computationally intensive in the presence of many unknown parameter values. Alternatively, Bayesian parameter estimates are found by drawing realizations from the posterior distribution. Within large data sets the two methods are shown to be equivalent through the Bernstein-von Mises Theorem (Train, 2003), which was originally derived by contributions from Bernstein (1917) and von Mises (1931). This property allows Bayesian methods to be used in place of classical methods, which are asymptotically similar and may have significant computational advantages. In addition to asymptotic equivalence, Bayes estimators, in a Tobit framework, have been shown to converge faster than maximum likelihood methods (Chib, 1992).

Bayesian methods also possess important philosophical advantages to classical methods. First, in spite of all statistical models having subjective information within the experiment, Bayesian techniques implement these components systematically (Robert, 2001). Within a Bayesian framework, subjective information is summarized within the prior distribution $\pi(\theta)$. In this way, the posterior distribution ($\pi(\theta \mid x)$) is a mixture of the prior distribution and a sampling distribution ($f(x \mid \theta)$).

$$\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta) \tag{3.5}$$

This illustrates the way in which Bayesian estimates are a mixture of prior information and updated information from the sample data. This is a notable point in agricultural applications given the fact that arguments for many different distributions have been made to characterize agricultural yields.

The second point involves the way in which subjective information is introduced into the model. In maximum likelihood methods, assumptions are made concerning the parametric characterization of the function. For example, crop yields have been argued to be distributed as Normal, Beta, Gamma, and Weibull. Since yield data are usually insufficient to reject any reasonable distributional assumption (Ker and Coble, 2003), the argument can be made for a number of alternative measures. In Bayesian analysis, prior information is less intrusive in that it can be introduced as a tool by providing a summary of available prior knowledge. With little or no prior knowledge, this element can be noninformative or vague. Even with little prior information available, Bayesian estimators have been shown to outperform classical estimators in many areas, as discussed by Zellner (1985).

A major argument against Bayesian techniques is that prior distributions are subjective and their form is debatable. While this subjectivity cannot be avoided using classical parametric methods, the prior can be noninformative so as to minimize the subjectivity within a model. It is worth noting that as data samples grow, the sampling distribution overpowers the prior distribution, so that prior distribution assumptions likely have little effect on parameter estimates. In this particular application the data consists of a sufficient number of observations so that the priors will have a minimal impact on the posterior inference.[1]

## 3.3 Modeling Censored Data

This section develops two strategies for modeling censored data sets. The Tobit model is first described in both univariate and multivariate cases. The second method is a zero-inflated mixture model which may possess characteristics that make it more appropriate when censored variables are non-normal or have a two-step process. The mixture model will be shown to nest the Tobit model and extend to allow additional flexibility in modeling the probability of a non-zero outcome.

### 3.3.1 The Tobit Model

When faced with dependent variables that are censored, there are many models a researcher can use. No method is more widely used in economics than the Tobit model. In this section, a univariate Tobit model is developed that allows for heteroskedastic errors, which will be generalized to accommodate a system of equations.

The standard likelihood function for a given random dependent variable $y_i$ with $n_0$ censored observations and $n_1$ uncensored observations is written as follows:

$$L(\beta, \sigma_i | y_i) = \prod_{i=1}^{n_0} \left[ 1 - \Phi\left(\frac{x_i'\beta}{\sigma_i}\right)\right] \prod_{i=1}^{n_1} \sigma_i^{-1} \phi \left[\left(\frac{y_i - x_i'\beta}{\sigma_i}\right)\right] \tag{3.6}$$

[1]The complete data set consists of 11,397 observations, which will be split into thirds so that two-thirds of the data (n = 7,598) is used to characterize the model, while the final third (n = 3,799) is used to test predictive accuracy. A random selection process was used to place each observation into one of these two categories.

where $\Phi$ is the cumulative density function (CDF) and $\phi$ is the probability density function (PDF) of a standard normal. Notice here that errors can also be assumed to be homoskedastic by assuming that $\sigma_i = \sigma$ for $\forall i$. Statistical inference essentially involves an inversion process where parameter estimates are obtained from the data. For maximum likelihood, the above likelihood function is used to maximize the log of the likelihood function, with respect to the parameters.[2] Notice here that the Tobit model is flexible to allow for heteroskedastic errors (Greene, 2003).

From the likelihood function described in equation (3.6), we obtain the logarithmic likelihood function

$$LL = \sum_{i=1}^{n_0} log \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma_i}\right)\right] - \frac{1}{2}\sum_{i=1}^{n_1}\left[log\left(\sigma_i^2\right) + \frac{(y_i - x_i'\beta)^2}{\sigma_i^2}\right] \tag{3.7}$$

where $\sigma_i^2$ can be written as a linear function of the data, expressed as

$$\sigma_i^2 = exp(x_i'\alpha) \tag{3.8}$$

where the first element of $\alpha$ contains a constant variance term ($\log(\sigma^2)$). Equation (3.7) can be rewritten to include (3.8) with the following equation

$$LL = \sum_{i=1}^{n_0} log \left[\Phi\left(\frac{-x_i'\beta}{exp(x_i'\alpha)^{-.5}}\right)\right] - \frac{1}{2}\sum_{i=1}^{n_1}\left[x_i'\alpha + \frac{(y_i - x_i'\beta)^2}{exp(x_i'\alpha)}\right] \tag{3.9}$$

This equation is maximized with respect to values of $\beta$ and $\alpha$ to obtain the MLEs and can be used in a variety of settings. These parameter estimates (MLEs) are then used in prediction for the latent variable in the following way:

$$\hat{y}_i^* = x_i'\hat{\beta} \tag{3.10}$$

[2]The maximization of this function can be quite cumbersome and time-consuming given the use of an iterative process such as the Newton-Raphson procedure. This complexity is magnified in a multivariate framework, as computation time increases quickly when additional parameters are added to the function.

such that $\hat{y}_i = max(\hat{y}_i^*, 0)$.

Next, we consider the case of multiple related equations, where one or more of the variables consists of censored variables, while the remaining variables are not censored. Within this system of equations, an array of censoring regimes may occur. For example, in a 4 equation system, there are 16 possible combinations of censoring types. Additionally, when one variable is censored within a vector of dependent variables, for a single observation, the uncensored variable levels offer information about the censored observation. Therefore, when a single variable is censored, the uncensored variable levels can enhance efficiency when they are accounted for while computing the expected mean and variance for a given censored observation.

Recall that the univariate Tobit model above includes heteroskedastic errors, which will be extended to include an appropriate covariance matrix within the multivariate setting. This covariance matrix must account for both variance and correlation effects between the dependent variables.

Equation (3.9) can be extended into the following multivariate form

$$\begin{aligned} LL \quad = \quad & -\frac{1}{2}\sum_{i=1}^{n_0}\left\{ln\left(\left|\Sigma_i^{-1}\right|\right) + \left(Y_i - X_iB\right)\Sigma_i^{-1}\left(Y_i - X_iB\right)'\right\} + \\ & -\frac{1}{2}\sum_{i=1}^{n_1} ln\left(\left|\Sigma_{22i}^{-1}\right|\right) + \left(y_i^{(2)} - X_i^{(2)}B^{(2)}\right)\Sigma_{22i}^{-1}\left(y_i^{(2)} - X_i^{(2)}B^{(2)}\right)' + \\ & \qquad ln\left(\Phi\left[-\frac{\gamma_i}{\sqrt{\lambda_i}}\right]\right) \end{aligned} \tag{3.11}$$

where $Y_i$ is a $1 \times j$ matrix and B is a $k \times j$ matrix containing marginal effects. In this specification, j corresponds to the number of dependent variables and k is the number of parameters. $B$ can be broken into two components containing the parameter estimates for the censored variable, $B^{(1)}$, and the parameter estimates for the uncensored variables, $B^{(2)}$. When censoring occurs at an observation for one or more dependent variables, it will also contain observable non-censored variable levels. Components from $Y_i$, within an observation, will be segregated into two vectors that differentiate the variables between censored, $y_i^{(1)}$, and uncensored, $y_i^{(2)}$. For variables with censoring, the censored variables will have a standard normal CDF applied, denoted as $\Phi$, with a mean and variance respectively shown below

$$\gamma_i = \mu_i^{(1)} + \Sigma_{12i}\Sigma_{22i}^{-1}\left(y_i^{(2)} - \mu_i^{(2)}\right) \tag{3.12}$$

$$\lambda_i = \Sigma_{11i} - \Sigma_{12i}\Sigma_{22i}^{-1}\Sigma_{21i} \tag{3.13}$$

Notice here that the expected mean and variance of the censored variables account for the variable levels of non-censored variables. The covariance matrix $\Sigma_i$ can be broken into the following components where $\Sigma_{11i}$ refers to the covariance matrix of censored variables, $\Sigma_{22i}$ refers to the covariance matrix of uncensored variables, while $\Sigma_{12i}$ and $\Sigma_{21i}$ measure the cross-covariances between the two components, at observation i.

$$\Sigma_i = \left(\begin{array}{c|c} \Sigma_{11i} & \Sigma_{12i} \\ \hline \Sigma_{21i} & \Sigma_{22i} \end{array}\right) \tag{3.14}$$

Additionally, $\Sigma_i$ is a $j \times j$ covariance matrix that is characterized as $\Sigma_i = T_i' D_i T_i$ to ensure global concavity when the following specification is used.

$$T_i = \begin{pmatrix} 1 & t_{12i} & t_{13i} \\ 0 & 1 & t_{23i} \\ 0 & 0 & 1 \end{pmatrix} \tag{3.15}$$

and

$$D_i = \begin{pmatrix} d_{1i} & 0 & 0 \\ 0 & d_{2i} & 0 \\ 0 & 0 & d_{3i} \end{pmatrix} \tag{3.16}$$

where upper off-diagonal elements of $T_i$ are unrestricted while the diagonal elements of $D_i$ are restricted to non-negative values. Here we illustrate with a 3 parameter system that can easily be expanded to include more dependent variables. A linear regression is used to model $D_i$ as follows:

$$log \begin{pmatrix} d_{1i} \\ d_{2i} \\ d_{3i} \end{pmatrix} = GX_i' \tag{3.17}$$

where the 3xk matrix, G, is used to calculate the variance for each observation, conditional on the unique variable levels. Elements of $T_i$ are assumed to be constant across all observations in this study. Elements of $T_i$ can also be conditional on the conditioning variables, which has been derived in the previous essay.

As mentioned previously, there exists two limitations with the Tobit model in both multivariate and univariate cases. First, the assumption of normality, for the latent variable, severely limits distributional assumptions that can be utilized. The second limitation is that variables and parameters that influence the conditional mean of positive values must also influence the probability of a non-zero observation. This is also hypothesized to limit the model efficiency in cases where a high degree of censoring occurs. In the next subsection, a zero-inflated mixture regression model is developed to overcome the two major limitations of the Tobit model.

### 3.3.2 Zero-Inflated Mixture Regression Model

Here we focus on the two limitations of the Tobit model and derive a zero-inflated mixture regression model that can be generalized to use any continuous distribution to model non-censored observations. The univariate mixture model will be shown to nest the Tobit model, while naturally extending to include the log-normal and Gamma specifications and allowing for more flexible conditions. The next step will be to extend the univariate log-normal model into a multivariate model. This particular mixture model will be derived from a generalized zero-inflated model, which will also have many of the same characteristics as a generalized hurdle model. Typically, zero-inflated regression models have been confined to count data. However, there may be significant advantages in modeling censored data as part of this particular two-part process. This is particularly true in cases where participation in the positive outcome group can be partially determined by the observed conditioning variables.

A major difference between this model and the Tobit model is that unobservable, censored observations are not directly estimated. Instead, the probability that an observation takes on a positive value is regressed using a Bernoulli distribution, conditional on the conditioning variables. In the case where the variable takes on a positive value, the distribution $V$ is used.[3] A generalized zero-inflated model can be characterized as follows:

$$\begin{aligned} f(y|\theta) &= 1-\rho(\theta) \qquad y=0 \\ &= \rho(\theta)g(y|\theta) \quad y>0 \end{aligned} \tag{3.18}$$

where $\int_0^\infty g(y|\theta)dy = 1 \quad \forall\theta$. This formation includes the standard univariate Tobit model when $\theta = (\mu, \sigma)$, $\rho(\theta) = \Phi\left(\frac{\mu}{\sigma}\right)$, and $g(y|\theta) = \frac{\phi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} I(y>0)$. Notice that in the log-normal and Gamma zero-inflated specifications to follow, $\rho$ is modeled independently of mean and variance parameter estimates, making them more flexible than the Tobit model.

The above formulation may also be compared to a typical hurdle specification when $g(y|\theta)$ is assumed to be a zero-truncated distribution and $\rho_i(\theta)$ is represented by a Probit model. The hurdle model as specified by Cragg (1971) is not limited to the above specification. In fact, the hurdle model can be generalized to include

[3]One potential drawback to this model is the loss in degrees of freedom due to the additional parameters that estimate the likelihood of a censored outcome. However the additional parameters do appear to more efficiently characterize the model given the high degree of censoring. This impact is thought to decrease as the degree of censoring falls.

any standard regression model that takes on positive values for $g(y|\theta)$ and any decision model for $\rho_i(\theta)$ that takes on a value between 0 and 1. In their generalized forms, both the hurdle and zero-inflated models appear to be similar, even though applications for each have differed.

Next, we develop two univariate zero-inflated models that include covariate variables, which then can be extended to allow for multivariate cases. Since only the positive outcomes are modeled through the second component, the log of the dependent variable can be taken. Taking the log of this variable works to symmetrize the dependent variable that was originally positively skewed. Using a log-normal distribution for the $V$ random variable and allowing $\rho$ to vary based on the conditioning variables, we can transform the basic zero-inflated model into the following form that can be generalized to include continuous distributions. We start by deriving the normal distribution to model the logarithm of the dependent variable outcomes, also known as the log-normal distribution, of the following form

$$\begin{aligned} f(y_i|\beta, \alpha, \delta) &= 1 - \rho_i(\delta) && for \;\; y_i = 0 \\ &= \rho_i(\delta)\frac{1}{y_i}\phi\left(\frac{log(y_i) - x_i'\beta}{\sigma_i}\right) && otherwise \end{aligned} \tag{3.19}$$

where

$$\rho_i = \frac{1}{1 + exp(x_i'\delta)} \tag{3.20}$$

$$\sigma_i^2 = exp(x_i'\alpha) \tag{3.21}$$

which guarantees $\sigma_i^2$ to be positive and $\rho_i$ to lie between 0 and 1 for all observations and all parameter values. Notice that this specification is nested within the generalized version in equation (3.18) where $g(y|\theta)$ is a log-normal distribution and $\theta = (\delta, \beta, \alpha)$.

In addition to deriving a zero-inflated log-normal distribution, we will also derive a zero-inflated Gamma distribution to demonstrate the flexibility of the zero-inflated regression models and perhaps improve upon modeling a variable that possesses positive skewness. Within a univariate framework, the sampling distribution can be easily changed by deriving $V$ as an alternative distribution in much the same way as equation (3.19). Following is the specification for the zero-inflated Gamma distribution, where V is distributed as a Gamma distribution where $\lambda_i$ is the shape parameter, and $\eta_i$ is the rate parameter.

$$\begin{aligned} f(y_i|\lambda_i, \eta_i, \delta) &= 1 - \rho_i(\delta) \qquad\qquad for \;\; y_i = 0 \\ &= \rho_i(\delta) \; \frac{y_i^{\lambda_i - 1} e^{-\eta_i y_i}}{\Gamma(\lambda_i)} \eta_i^{\lambda_i} \quad otherwise \end{aligned} \tag{3.22}$$

This function can be reparameterized to include the mean of Gamma, $\mu$, by substituting $\lambda_i = \mu_i \eta_i$, where $\eta_i = e^{(x_i' \kappa)}$ and $\mu_i = e^{(x_i' \gamma)}$. Within the Gamma distribution specification, the expected value and corresponding variance can be found through the following equations:

$$E(y_i) = \rho_i \mu_i \tag{3.23}$$

$$Var(y_i) = \rho_i (1 - \rho_i) \mu_i^2 + \rho_i \frac{\mu_i}{\eta_i} \tag{3.24}$$

Both the Gamma and log-normal univariate specifications allow for a unique set of mean and variance estimates to result from each distinct set of conditioning variables.

To model multiple dependent variables in a way that captures the covariance structure, we make a slight departure from the univariate version by utilizing the relationship between joint density functions and conditional marginal functions. More specifically, we use the following relationship when evaluating $y_1$ and $y_2$, where $y_1$

has a positive probability of taking on the value of 0 and $y_2$ is a continuous variable.

$$f(y_1, y_2) = f(y_1|y_2)f(y_2) \tag{3.25}$$

where $f(y_1, y_2)$ is the joint density function of $y_1$ and $y_2$, $f(y_1|y_2)$ is the conditional probability of $y_1$, given $y_2$, and $f(y_2)$ is the unconditional probability of $y_2$. In order to compare this model to that of the multivariate Tobit formulation, we derive a two-dimensional version of $y_2$, which can easily be generalized to fit any size. However, this model restricts $y_1$ to be one-dimensional under its current formulation.[4]

We begin by parameterizing, $Z_{2i} = log(y_2)$, which will be distributed as a multivariate normal, with mean, $X_iB^{(2)}$, and variance, $\Sigma_{22i}$, as defined in equation (3.11). The assumption of log-normality is often made due to the ease in which a multivariate log-normal can be computed and its ability to account for skewness. Using the previously defined notation, the function can be expressed as follows

$$Z_{2i} \sim N(X_iB^{(2)}, \Sigma_{22i}) \tag{3.26}$$

where $Z_{2i}$ is an n x j dimensional matrix of positive outcomes. This formulation allows each observation to run through this mechanism, whereas the Tobit model runs only censored observations through this mechanism.

The conditional probability of $y_1$ given $y_2$ is modeled through a zero-inflated modeling mechanism that takes into account the realizations from $y_2$ in the following way

$$Y_1|Y_2 = y_2 \sim ZILN(\rho_i, \mu_i(y_2), \sigma_i^2(y_2)) \tag{3.27}$$

where ZILN is a zero-inflated log-normal distribution, $\mu_i(y_2)$ is the conditional mean

[4]This will remain an area of future research. Deriving a model that allows for multiple types of censoring may be very useful, particularly when dealing with the consumption of multiple goods. Using unconditional and conditional probabilities to characterize a more complex joint density function with multiple censored nodes would naturally extend from this modeling strategy.

of $Z_{1i}$, which is defined as $Z_{1i} = log(y_1)$, given $Z_{2i}$, and $\sigma_i^2(y_2)$ is the corresponding conditional variance. These terms are similar to equations (3.12) and (3.13) and are shown below

$$\begin{aligned} \mu_i(y_{2i}) &= X_i B^{(1)} + \Sigma_{12i}\Sigma_{22i}^{-1}\left(y_{2i} - X_i B^{(2)}\right) && (3.28) \\ \sigma_i^2(y_{2i}) &= \Sigma_{11i} - \Sigma_{12i}\Sigma_{22i}^{-1}\Sigma_{21i} && (3.29) \end{aligned}$$

which leads to the following probability density function

$$\begin{aligned} f(Z_{1i}|Z_{2i}) &= 1 - \rho_i(\delta) && \texttt{for } y_{1i} = 0 \\ &= \rho_i(\delta)\frac{1}{y_{1i}}\phi\left(\frac{Z_{1i} - \mu_i(y_{2i})}{\sqrt{\sigma_i^2(y_{2i})}}\right) && \texttt{for } y_{1i} > 0 \end{aligned} \tag{3.30}$$

Ghosh et al. (2006) demonstrate through simulation studies that similar zero-inflated models have better finite sample performance with tighter interval estimates when using Bayesian procedures instead of classical maximum likelihood methods. Due to these advantages, the previously developed models will utilize recently developed Bayesian techniques. In order to develop a Bayesian model, the sampling distribution is weighted by prior distributions as shown in equation (3.5). The sampling distribution, $f$, is fundamentally equivalent to the likelihood function, $L$, where

$$L(\theta|y_i) \propto f(y_i|\theta) \tag{3.31}$$

and $\theta$ represents the estimated parameters, which for our purposes will include $\theta = (\beta, \alpha, \delta)$. While prior assumptions can have some effects in small samples, this influence diminishes with larger sample sizes. Additionally, prior assumptions can be uninformative in order to minimize any effects in small samples. For each parameter in the model, a non-informative normal prior will be assumed.

$$\pi(\beta_{kj}) \sim N(0, \Lambda_1) \tag{3.32}$$

$$\pi(\alpha_{kj}) \sim N(0, \Lambda_2) \tag{3.33}$$

$$\pi(\delta_{kc}) \sim N(0, \Lambda_3) \tag{3.34}$$

for $k = 1, ..., K$, $j = 1, ..., J$, and $c = 1, ..., C$, where K is the number of conditioning variables or covariates, J is the number of dependent variables in the multivariate model, and C is the number of censored dependent variables. Note here that the above formulation applies to univariate versions when $J = 1$ and $C = 1$. Additionally, $\Lambda$ must be large enough to make the prior relatively uninformative.[5]

Given the preceding specifications of a sampling density and prior assumptions, a full Bayesian model can be developed. Due to the difficulty in integrating a posterior distribution that contains many dimensions, Markov Chain Monte Carlo (MCMC) methods can be utilized to obtain samples of the posterior distribution using WinBUGS programming software. Chib and Greenberg (1996) provide a survey of MCMC theory as well as examples of its use in econometrics and statistics. MCMC methods allow for the computation of posterior estimates of parameters through the use of Monte Carlo simulation based on Markov chains that are generated a large number of times. The draws arise from a Markov chain since each draw depends only on the last draw, which satisfies the Markov property. As the posterior density is the stationary distribution of such a chain, the samples obtained from the chain are approximately generated from the posterior distribution following a burn-in of initial draws.

Predictive values within a Bayesian framework come from the predictive distributions, which is a departure from classical theory. In the zero-inflated mixture

[5] $\Lambda$ is assumed to be 1,000 in this study, so that a normal distribution with mean 0 and variance 1,000 will be relatively flat.

model, predicted values will be the product of two posterior mean estimates. Posterior densities for each parameter are computed from Markov Chain Monte Carlo (MCMC) sampling procedures using WinBUGS software. MCMC methods allow us to compute posterior density functions by sampling from the joint density function that combines both the prior distributional information and the sampling distribution (likelihood function).[6] Formally, prediction in the zero-inflated log-normal model is characterized as follows:

$$\hat{y_i} = v_i b_i \tag{3.35}$$

where $v_i$ and $b_i$ are generated from their predictive distributions. $log(v_i)$ is from a normal distribution with mean ($\mu_i = x_i'\hat{\beta}$) and variance ($\sigma_i^2 = exp(x_i'\hat{\alpha})$), while $b_i$ is from a Bernoulli distribution with parameter $\rho(\hat{\delta})$. Since many draws from a Bernoulli will result in 0 and 1 outcomes, the mean will produce an estimate that lies between the two values. To allow for prediction of both zero and positive values, the median of the Bernoulli draws was used for prediction, so that observations that contained more than 50% of 1 outcomes were given a 1 value and the rest were given 0. This allows for observations to fully take on the continuous random variable if more than half of the time it was modeled to do so, while those that are more likely to take on zero values, as indicated by the Bernoulli outcomes, take on a zero value.

## 3.4 Comparison Using Simulated Data

This section will focus on simulating data in order to evaluate model efficiency for the two previously specified models. The major advantage to evaluating a simulated set of data, is that the true form of the data generation process is known

[6] WinBUGS will fit an appropriate sampling method to the specified model to obtain samples from the posterior distribution. Typically this implies Gibbs sampling with Metropolis-Hastings steps.

prior to evaluation. This will offer key insights into what to expect when evaluating an application involving the cattle production data set to follow. Additionally, we will evaluate data that come from Tobit and mixture processes, which will help to assess the degree of losses when the wrong type of model is assumed. This will assist in identifying the type of data that the cattle production data set most closely represents. In simulating data, there will be two key characteristics that will align the simulated data set with the cattle production data set to be used in the next section. First, cattle production yield variables have been shown to possess heteroskedastic errors. To accommodate this component error terms will be simulated based on a linear relationship with the conditioning variables. While the first set of simulations will consist of homoskedastic errors, the remaining simulations will use heteroskedastic errors. Second, we are concerned with simulated data that exhibit nearly 50% censoring to emulate the cattle production data set to be used as an application in the next section. While there are many different ways to simulate data, the proposed method is aimed to offer guidance when evaluating cattle production yields.

Past research has focused on modeling agricultural yields, while research dealing specifically with censored yields is limited. The main reason for the lack of research into censored yields is because crop yields are not typically censored at upper or lower bounds. However, with the emergence of new livestock insurance products, new yield measures must be quantified in order for risks to be properly identified. In contrast to crop yield densities, yield measures for cattle health possess positive skewness, such as the mortality rate and veterinary costs. Crop yield densities typically possess a degree of negative skewness as plants are biologically limited upward by a maximum yield, but can be negatively impacted by adverse weather, such as drought. Variables such as mortality have a lower limit of zero, but can rise quickly in times of adverse weather, such as prolonged winter storms or disease. The simulated data set used for these purposes will possess positive skewness as well as a relatively high degree of censoring, in order to align with characteristics found in cattle mortality data for cattle on feed.

First, a simplified simulated data set will be examined with a varying number of observations. We assume in this set of simulations that that errors are homoskedastic. The simulated model will be as follows

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \tag{3.36}$$

$$\varepsilon_i \sim N(0, \sigma^2) \tag{3.37}$$

$$y_i = max(y_i^*, 0) \tag{3.38}$$

Data are simulated using *MATLAB* software. It is also important to emphasize the point that, in this scenario, the censored and uncensored variables come from the same data generation process. For each sample size, starting seeds were set in order to replicate results. Then, values for $x_i$ ranged from 1 to 10, based on a uniform random distribution.[7]

Error terms are distributed as a normal, centered at zero with a constant variance set to $\sigma$. $y_i^*$ is then computed from equation (3.36) and all negative values are replaced with zeros, in order to simulate a censored data set. The degree of censoring in these simulated data sets ranged from 49% to 62%.

Two thirds of this simulated data set is used for estimation, while the final third is used for prediction. This allows us to test both model fit measures as well as predictive power. In this study, Tobit regressions use classical maximum likelihood estimation techniques, while zero-inflated models use Bayesian estimation

[7]Values for $x_i$ might also be simulated using a normal distribution. A uniform distribution will more evenly spread values of $x_i$ from the endpoints, while a normal distribution would cluster the values near a mean, without endpoints (unless specified). Additionally, a uniform distribution will tend to result in fatter tails in the dependent variable due to the relatively high proportion of extreme values for $x_i$.

techniques. To derive measures of model fit we use the classical computation of Akaike's Information Criteria (AIC) (Akaike, 1974) and derive a similar measure for Bayesian analysis, the Deviance Information criteria (DIC)(Spiegelhalter et al., 2002). DIC results are interpreted similar to AIC in that smaller values of the statistics reflect a better fit. A major difference is that AIC is computed based on the optimized value of the likelihood function in the following manner:

$$AIC = -2logL(\hat{\beta}, \hat{\alpha}) + 2P \tag{3.39}$$

where P is the dimension of $\theta$, which is 3 in the case of homoskedastic errors and 6 with heteroskedastic errors. Alternatively, DIC is constructed by including prior information and is based on the deviance at the posterior means of the estimated parameters. A penalization factor for the number of parameters estimated is also incorporated into this measure. The formulation for DIC is as follows:

$$DIC = \bar{D} + p_D \tag{3.40}$$

where $p_D$ is the effective number of parameters and $\bar{D}$ is a measure of fit that is based on the posterior expectation of deviance, which is specified for our purposes as follows

$$\begin{aligned} \bar{D} &= E[-2logL(\delta, \beta, \alpha)|y] && (3.41) \\ p_D &= \bar{D} + 2logL(\tilde{\delta}, \tilde{\beta}, \tilde{\alpha}) && (3.42) \\ & && (3.43) \end{aligned}$$

taking into account the posterior means, $\tilde{\delta}$, $\tilde{\beta}$, and $\tilde{\alpha}$.

Robert (2001) reports that DIC and AIC are equivalent when the posterior distributions are approximately normal. The normality of all parameter estimates is supported by the posterior plots supplied by WinBUGS. Spiegelhalter et al. (2003) warns that DIC may not be a viable option for model fit tests when posterior distributions possess extreme skewness or bimodality. These do not appear to be problematic in this study.

To measure the predictive power within a modeling strategy, we compute the Mean Squared Prediction Error (MSPE) associated with the final third of each simulated data set. MSPE allows us to test out of sample observations to assess how well the model predicts dependent variable values. MSPE is formulated as follows:

$$MSPE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2 \tag{3.44}$$

where m is some proportion of the full data sample, such that $m = \frac{n}{b}$. For our purposes, $b = 3$, which allows for prediction on the final third, based on estimates from the first two thirds. This allows for a sufficient amount of observations available for estimation and prediction.

We estimate the simulated data set, given the above specifications, with the three models that have previously been formulated. MCMC sampling is used for Bayesian estimation with a burn-in of 1,000 observations and three Markov chains. WinBUGS uses different sampling methods, based on the form of the target distribution. For example, the zero-inflated Gamma distribution uses Metropolis sampling that fine tunes some optimization parameters for the first 4,000 iterations, which are not counted in summary statistics. MSPE calculations were conducted outside of WinBUGS, while predicted values were computed during WinBUGS regressions and compared with simulated data to align with the method used for predictions from the Tobit model.

Results from regressions on the simulated data set with homoskedastic errors can be found in Table 3.1. Based on AIC/DIC criteria, the zero-inflated log-normal regression model outperforms the Tobit model at all data sample levels. This is particularly interesting given the fact that simulation was based on a Tobit model. In cases where the degree of censoring is high, the parameter estimates that estimate the likelihood of censoring add more precision to the model. This impact is likely to diminish as the degree of censoring decreases and increases for higher degrees of censoring.

Table 3.1: Simulation results based on Tobit model with homoskedastic errors

| n | Model | MSPE | LL | AIC/DIC |
|---|---|---|---|---|
| 200 | Tobit | 0.773 | -86.044 | 184.088 |
| | ZILN | 23.982 | -58.162 | 122.934 |
| | ZIG | 4.453 | -165.280 | 337.138 |
| 500 | Tobit | 0.460 | -258.590 | 529.179 |
| | ZILN | 96.064 | -170.905 | 350.238 |
| | ZIG | 52.720 | -497.998 | 1,002.490 |
| 1,000 | Tobit | 0.466 | -508.902 | 1,029.805 |
| | ZILN | 94.809 | -400.526 | 809.331 |
| | ZIG | 47.098 | -985.960 | 1,977.590 |

The poor performance of the Gamma distribution highlights the problem associated with assuming an incorrect distribution. The Gamma distribution does particularly well with positive skewness, however, the degree of skewness in this model is not sufficient to overcome the incorrect distributional assumption.

Lower MSPE indicates that the prediction of the out of sample portion of the data set favors that of the Tobit model at all sample levels. MSPE penalizes observations with large residuals that tend to be more prevalent as the dependent variable value increases. Gurmu and Trivedi (1996) point out that mixture models tend to overfit data. By overfitting the data, model fit tests might improve, while prediction remains less accurate. This might explain part of the reason that mixture models appear to fit this particular set of simulated data better, while lacking prediction precision. Both zero-inflated models had particular trouble when predicting higher values of $y$, resulting in high MSPE values. Recall, MSPE refers to the average of the squared distance between predicted and actual values. It is also worth mentioning that the wide spread of MSPE values is largely a result of simulating data that contains a high level of variability and a relatively small number of observations.

It is also important to point out that the Tobit model assumes positive observations are distributed by a truncated normal distribution, while log-normal and Gamma distributions also take on only positive values but look very different than the truncated normal distribution. With roughly half of the sample censored, the density function from a truncated normal density would likely predict a larger mass near the origin, while the log-normal and Gamma distributions carry fatter tails.

As previously mentioned, another major difference between the hurdle and zero-inflated models in practice is the difference in modeling the binary decision variable using a Probit and Bernoulli distributions, respectively. Based on the simulated data, the binary variables were computed using both methods and shown below in Table 3.2. The Bernoulli method predicts just as well as the Probit model, despite the fact that the data are generated based on a Tobit model, which assumes that the binary process is based on a Normal distribution. For example, in the $n = 500$ simulation, 167 observations were used to compare predictive accuracy and the Bernoulli equation predicted 84 non-censored outcomes, while the Probit overestimated 89 non-censored outcomes.

As an alternative to the preceding simulation process, data can also be simulated using two separate data processes that emulate that of a mixture model more consistent with zero-inflated models. The distinction between this simulation and the previously developed Tobit-based data set, is that the probability of a censored outcome is modeled according to equation (3.20). Additionally, outcomes that are

Table 3.2: Comparison of binary methods of predicting positive outcomes from simulation based on Tobit model with homoskedastic errors

| n | m | Type | Actual | Probit | Bernoulli |
|---|---|---|---|---|---|
| 200 | 67 | Count | 41 | 43 | 42 |
| | | Percentage | 0.61 | 0.64 | 0.63 |
| 500 | 167 | Count | 86 | 89 | 84 |
| | | Percentage | 0.52 | 0.53 | 0.51 |
| 1,000 | 333 | Count | 160 | 165 | 166 |
| | | Percentage | 0.48 | 0.50 | 0.50 |

Note: m is the number of observations used solely for prediction

described by a probability density function must be positive, which is achieved by taking the exponential of a normal distribution.[8]

Results from the second simulation can be found in Table 3.3. Once again, the results indicate a superior model fit with the zero-inflated model, relative to the Tobit model. Additionally, the zero-inflated models possess a substantially lower MSPE, indicating better out of sample prediction performance. Both zero-inflated formulations are capable of accounting for positive skewness. For the larger sample sizes, the Gamma formulation shows superior prediction ability, while the log-normal formulation fits the data best. These results may come from the data generating process where the positive observations are generated from a log-normal distribution, however, the Gamma formulation predicts the outcomes that may come furthest from zero most accurately.

Accounting for both types of data generating processes, zero-inflated models are better able to fit data that contain a high degree of censoring. Prediction appears to depend on the data generation process. If the data comes from a zero-

[8]This is the same as assuming $y_i$ is distributed as a log-normal distribution. Alternatively, data could be generated using a truncated normal distribution where simulations on a normal continues until all values are positive through iterations that spit out negative values and keep only positive values. The two methods would generate two very different data sets.

Table 3.3: Simulation results based on mixture model with homoskedastic errors

| n | Model | MSPE | LL | AIC/DIC |
|---|---|---|---|---|
| 200 | Tobit | 1.953 | -95.535 | 203.069 |
| | ZILN | 0.280 | -2.583 | 11.510 |
| | ZIG | 0.303 | 6.905 | -7.865 |
| 500 | Tobit | 2.153 | -276.410 | 564.821 |
| | ZILN | 0.079 | 29.148 | -50.412 |
| | ZIG | 0.083 | 13.379 | -20.674 |
| 1,000 | Tobit | 1.189 | -499.676 | 1,011.351 |
| | ZILN | 0.044 | 21.211 | -33.565 |
| | ZIG | 0.046 | 82.563 | -158.735 |

inflated model, then prediction is more efficient when it is from a zero-inflated model. Alternatively, if all data comes from the same data generating process, then the Tobit may predict better than the proposed alternatives. One notable feature of the results generated from this simulation is that values for DIC appear to take on both positive and negative values. As mentioned previously, lower DIC/AIC indicate better fit measures. Therefore, a negative DIC is favorable to a positive AIC, which is the case under this scenario.

Tests were once again performed to compare the results based on the Probit and Bernoulli methods of predicting the binary choice component and are shown in Table 3.4. As shown below, there does not appear to be any significant differences between the two methods under this scenario.

The previous set of simulations assumed homoskedastic errors, which is a simplyfying assumption which has not been shown to hold in the application of cattle production yields. In order to more closely align with the given application, we now move to simulate a data set containing heteroskedastic errors. Heteroskedasticity is introduced into this data by constructing $\varepsilon_i$ by substituting equation (3.45) for equation (3.37) and accounting for the relationship between the error terms and the

Table 3.4: Comparison of binary methods of predicting positive outcomes from simulation based on a mixture model with homoskedastic errors

| n | m | Type | Actual | Probit | Bernoulli |
|---|---|---|---|---|---|
| 200 | 67 | Count | 45 | 45 | 45 |
| | | Percentage | 0.67 | 0.67 | 0.67 |
| 500 | 167 | Count | 95 | 96 | 96 |
| | | Percentage | 0.57 | 0.57 | 0.57 |
| 1,000 | 333 | Count | 193 | 194 | 193 |
| | | Percentage | 0.58 | 0.58 | 0.58 |

Note: m is the number of observations used solely for prediction

conditioning variables, as shown in equation (3.46).

$$\varepsilon_i \sim N(0, \sigma_i^2) \tag{3.45}$$

$$\sigma_i^2 = exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i}) \tag{3.46}$$

These equations impose a dependence structure on the error term, where the variance is a function of the conditioning variables. This specification is thought to better characterize cattle production yield measures, such as mortality rates, based on results from Essay 1.

Simulations were conducted in much the same manner as the previous set of simulations, with the addition of heteroskedastic errors. The results from generating data from a Tobit model with heteroskedastic errors are shown in Table 3.5.

The same general results apply with lower MSPE resulting from the Tobit model, with slightly higher MSPE from the zero-inflated models. Also, results from the simulation based on a mixture model are shown in Table 3.6.
Results from this simulation also align with the homoskedastic version.

Most current research concerning censored data is focused on multivariate systems of equations. This is because of the many applications that make use of

Table 3.5: Simulation results based on Tobit model with heteroskedastic errors

| n | Model | MSPE | LL | AIC/DIC |
|---|---|---|---|---|
| 200 | Tobit | 16.624 | -87.057 | 186.114 |
| | ZILN | 20.416 | -57.162 | 123.416 |
| | ZIG | 17.083 | -187.151 | 381.949 |
| 500 | Tobit | 6.205 | -248.387 | 508.774 |
| | ZILN | 10.156 | -168.842 | 348.661 |
| | ZIG | 12.018 | -546.32 | 1,100.390 |
| 1,000 | Tobit | 10.012 | -453.739 | 919.478 |
| | ZILN | 14.528 | -358.822 | 728.513 |
| | ZIG | 16.745 | -1,024.810 | 2,058.140 |

Table 3.6: Simulation results based on mixture model with heteroskedastic errors

| n | Model | MSPE | LL | AIC/DIC |
|---|---|---|---|---|
| 200 | Tobit | 77.813 | -40.024 | 92.049 |
| | ZILN | 63.386 | -26.204 | 60.732 |
| | ZIG | 64.052 | -76.506 | 160.651 |
| 500 | Tobit | 12.941 | -130.961 | 273.923 |
| | ZILN | 4.303 | -102.802 | 214.994 |
| | ZIG | 3.971 | -235.094 | 478.824 |
| 1,000 | Tobit | 4.610 | -272.204 | 556.407 |
| | ZILN | 2.948 | -183.034 | 376.549 |
| | ZIG | 2.468 | -430.032 | 867.893 |

multivariate relationships in current studies. For this reason, it will be important to simulate a multivariate data set that comes from both a Tobit and a two-step process. Since multivariate data comes in both forms, it will be important to evaluate each to see the potential bias from assuming, for example, that data are generated from a multivariate Tobit model when a two-step process is more appropriate. The Tobit process is constructed from a multivariate normal distribution that assumes a covariance matrix constructed from equations equations (3.15) and (3.16). Censoring in this case occurs when the censored dependent variable falls below a specified level. Alternatively, the two-step process uses a Bernoulli distribution to estimate the likelihood of a censored outcome, which is also a function of the conditioning variables. To be consistent, the same variables that increase the likelihood of a censored outcome in the two-step case, also decrease the mean of the variable so that they increase the likelihood of a variable being censored in the Tobit process. $Y_1$ and $Y_2$ are variables without censoring, while $Y_3$ contains censoring in nearly half of its observations.

The results from a simulated data set based on a multivariate Tobit model are shown in Table 3.8. Overall, the fit of both models appear to be more closely aligned with the Tobit model in the first two simulations, while the zero-inflated model more accurately fits the model with the largest sample size. Additionally, the Tobit model predicts more efficiently, as shown by the lower MSPE in most cases. This is consistent with the univariate results and again is not surprising, given the data were generated from a Tobit model.

It is surprising the closeness of model fit, when we compare the results from data simulated from a mixture model, as shown in Table 3.8. Here, the zero-inflated model strongly improves the model fit, relative to the Tobit formulation. It is surprising that while most MSPE measures are close, they tend to favor the Tobit model formulation. For an explanation of estimation and prediction of the zero-

Table 3.7: Multivariate simulation results based on Tobit model

| n | Model | $Y_1$ MSPE | $Y_2$ MSPE | $Y_3$ MSPE | LL | AIC/DIC |
|---|---|---|---|---|---|---|
| 200 | Tobit | 0.481 | 1.281 | 0.290 | -168.227 | 378.454 |
| | ZILN | 0.482 | 3.236 | 3.165 | -156.399 | 354.223 |
| 500 | Tobit | 0.328 | 0.789 | 0.312 | -477.028 | 996.056 |
| | ZILN | 0.327 | 2.182 | 5.057 | -499.801 | 1,043.380 |
| 1,000 | Tobit | 0.261 | 0.639 | 0.205 | -912.345 | 1,866.691 |
| | ZILN | 0.261 | 1.211 | 3.167 | -907.265 | 1,859.090 |

inflated multivariate model, see Appendix 1.

Table 3.8: Multivariate simulation results based on mixture model

| n | Model | $Y_1$ MSPE | $Y_2$ MSPE | $Y_3$ MSPE | LL | AIC/DIC |
|---|---|---|---|---|---|---|
| 200 | Tobit | 0.343 | 0.995 | 1.032 | -197.166 | 436.332 |
| | ZILN | 0.355 | 1.817 | 1.950 | -107.884 | 243.534 |
| 500 | Tobit | 0.376 | 0.858 | 2.395 | -2,636.920 | 5,315.839 |
| | ZILN | 0.373 | 3.713 | 3.205 | -347.296 | 739.812 |
| 1,000 | Tobit | 0.220 | 1.067 | 2.760 | -4,599.517 | 9,241.034 |
| | ZILN | 0.226 | 1.714 | 5.169 | -585.280 | 1,216.920 |

These simulations were conducted to compare the efficiency of the two given models in cases where the data are generated from a single data generating process, and that of a two-step process. Simulation results indicate that both models do relatively well in fitting the data, when the data come from a Tobit model. Alternatively, the model fit tests quickly move in the direction of the zero-inflated model in cases where the data comes from a mixture model, where non-zero observations are modeled using a multivariate log-normal distribution. Prediction of out of sample observations appear to be more efficiently characterized through the Tobit model. This is interesting given the fact that classical Tobit prediction uses only the optimized parameter values, while the zero-inflated model employs a Bayesian method that employs the entire posterior distribution of the estimated parameter values. Overall, the ZILN model tends to fit the data particularly well whether the data are generated from a Tobit or two-step process. This may result from the additional parameters that characterize the probability of a non-zero outcome in mixture models. However, this over-fitting does not assist in improving prediction, as prediction tends be more precise when the appropriate model is specified.

This section has offered some initial guidance into evaluating real data through the use of simulated data sets. The simulated data sets offer the opportunity to evaluate the performance of the Tobit and zero-inflated mixture models in situations where the true data generation process is known. The next section will look to evaluate the same postulated models in an application where the true data generating process is unknown.

## 3.5 Data

This section applies the preceding models to cattle production risk variables. Past simulation results offer initial guidance as we proceed to this particular application. The data set that will be used for this section possesses many of the same properties from the last section, such as a relatively high degree of censoring and positive skewness. The proposed zero-inflated mixture model is hypothesized to characterize censored cattle mortality rates better than the Tobit model because of the two part process that mortality observations are hypothesized to follow, as well as based on a visual inspection of positive mortality observations being more closely characterized by a log-normal distribution. Cattle mortality rates are thought to follow a two step process because pens tend to come from the same, or nearby, producers and are relatively homogenous. Therefore, a single mortality can be seen as a sign of a pen that is prone to sickness or disease. Additionally, airborne illnesses are contagious and can be spread rather quickly throughout the pen. Additional variables that describe cattle production performance are introduced and evaluated using the previously developed multivariate framework. In this setting, it is proposed that the mixture model improves upon the multivariate Tobit model used in the first essay of this research.

This research focuses on the estimation and prediction of cattle production yield measures. Cattle mortality rates from commercial feedlots are of particular interest due to their censored nature and importance in cattle feeding profits. Typically, mortality rates are zero or small, but can rise significantly during adverse weather, illness, or disease. The data used in this study consists of 5 commercial feedlots residing in Kansas and Nebraska, and includes entry and exit characteristics of 11,397 pens of cattle at these feedlots. Table 3.9 presents a summary of characteristics for different levels of mortality rates, including no mortalities

Particular attention will be placed on whether zero or positive mortality rates can be strongly determined based on the data at hand. The degree of censoring in this sample is 54%, implying that 46% of the observations contain no mortality losses. There is strong evidence that mortality rates are related to the previously mentioned conditioning variables, but we will need to determine whether censored mortality observations are from a different empirical process than observed positive values. Positive mortality rates may be a sign of poor genetics coming from a particular breeder or sickness picked up within the herd. The idea here is that the cattle within the pen are quite homogeneous. Homogeneity within the herd is desirable as it allows for easier transport, uniform feeding rations, medical attention, and the amount of time on feed. If homogeneity within the herd holds, then pens that have mortalities can be put into a class that is separate from those with no mortalities.

However, mortalities also may be caused suddenly and without warning for unknown reasons. Glock and DeGroot (1998) report that 40% of all cattle mortali-

Table 3.9: Comparison of pens with differing mortality losses

| | Mortality Rate (%)[a] | | | | | |
|---|---|---|---|---|---|---|
| Variable | 0 | 0.01 - 1 | 1 - 2 | 2 - 3 | 3 - 4 | >4 |
| Observations | 5,161 | 2,415 | 2,327 | 744 | 305 | 445 |
| DMFC | 6.05 | 6.27 | 6.21 | 6.34 | 6.42 | 6.85 |
| ADG | 3.49 | 3.35 | 3.28 | 3.11 | 3.06 | 2.82 |
| Intake | 20.96 | 20.82 | 20.15 | 19.52 | 19.49 | 18.88 |
| VCPH | 10.18 | 10.08 | 12.46 | 15.67 | 17.89 | 26.57 |
| InWt | 754.27 | 751.53 | 719.72 | 686.35 | 699.00 | 671.63 |
| OutWt | 1,188.72 | 1,179.90 | 1,168.75 | 1,152.95 | 1,158.40 | 1,144.66 |
| HeadIn | 120.24 | 182.04 | 126.19 | 123.14 | 114.72 | 110.57 |
| Days on Feed | 123.45 | 125.66 | 133.44 | 143.83 | 141.57 | 150.57 |
| **Proportion of sample:** | | | | | | |
| Winter | 0.24 | 0.27 | 0.26 | 0.29 | 0.29 | 0.21 |
| Spring | 0.26 | 0.24 | 0.21 | 0.16 | 0.17 | 0.10 |
| Summer | 0.27 | 0.26 | 0.26 | 0.25 | 0.24 | 0.30 |
| Fall | 0.23 | 0.23 | 0.27 | 0.30 | 0.30 | 0.39 |
| Steers | 0.53 | 0.56 | 0.49 | 0.43 | 0.42 | 0.35 |
| Heifers | 0.36 | 0.37 | 0.37 | 0.36 | 0.38 | 0.33 |
| Mixed | 0.10 | 0.07 | 0.14 | 0.20 | 0.20 | 0.32 |
| KS | 0.82 | 0.76 | 0.82 | 0.77 | 0.79 | 0.82 |

[a]Note: A mortality rate that results in a whole number is placed into the higher bins (ie, 3.00% is placed in 3-4 bin)

ties in a Nebraska feedlot study were directly caused by Sudden Death Syndrome.[9] However, the authors also point out that these deaths were without warning, which could be due to a "sudden death" or lack of observation by the feedlot workers. Smith (1998) also reports that respiratory disease and digestive disorders are responsible for approximately 44.1% and 25.0% of all mortalities, respectively. The high degree of correlation between dependent variables certainly indicates that lower mortality rates can be associated with different performance in the pen. However, the question in this study will be whether positive mortality rates significantly alter the performance. For this reason, we estimate additional parameters to examine the likelihood of a positive mortality outcome in the zero-inflated regression model.

A recent study by Belasco et al. (2006) found that the mean and variance of mortality rates in cattle feedlots are influenced by entry-level characteristics such as location of the feedlot, placement weight, season of placement and gender. These variables will be used as conditioning variables. By taking these factors into account, variations will stem from events that occur during the feeding period as well as characteristics that are unobservable in the data. The influence of these parameters will be estimated using the previously formulated models, based on two-thirds of the randomly selected data set where $n = 7,598$. The remaining portion of the data set, $m = 3,799$, will be used to test out of sample prediction accuracy. Predictive accuracy is important in existing crop insurance programs where past performance is used to derive predictive density functions for current contracts.[10]

After estimating expected mortality rates, based on pen-level characteristics, we will focus our attention to estimating mortality rates as part of a system of equations that includes other performance and health measures for fed cattle, such as dry matter feed conversion (DMFC), average daily gain (ADG), and veterinary

[9]Glock and DeGroot (1998) loosely define Sudden Death as any case where feedlot cattle are found dead unexpectedly.

[10]The most direct example of this is the Average Production History (APH) crop insurance program that insures future crop yields that are based on a 16-year average of production history.

costs (VCPH), which are additional measures of cattle production yields.

## 3.6 Estimation Results

The purpose of this section is to compare the results from the preceding models, using an extensive data set. A favorable model will be one that fits the data in estimation and is able to predict dependent variable values with accuracy. For these reasons, these models will be compared in a way similar to the simulated data sets. First, we begin with univariate results. Results from using a classical Tobit model with heteroskedastic errors to model cattle mortality rates can be found in Table 3.10.

Table 3.10: Univariate Tobit estimates of fed cattle mortality parameters

| | $\beta^{\#}$ | | $\alpha^{@}$ | |
|---|---|---|---|---|
| Variables | coeff. | se. | coeff. | se |
| Intercept: | 25.782* | 1.515 | 12.068* | 1.114 |
| Steers: | 0.168* | 0.055 | -0.021 | 0.048 |
| Mixed: | 0.307* | 0.124 | 0.984* | 0.073 |
| Kansas: | -0.068 | 0.058 | 0.292* | 0.058 |
| log(inwt): | -3.893* | 0.231 | -1.654* | 0.172 |
| Winter: | 0.065 | 0.068 | -0.243* | 0.061 |
| Fall: | 0.043 | 0.079 | 0.315* | 0.061 |
| Spring: | -0.089 | 0.069 | -0.278* | 0.065 |
| LL: | -11,389.353 | | | |
| AIC: | 22,790.706 | | | |
| MSPE: | 2.661 | | | |

*Denotes estimate is significant at the 0.05 level

#$\beta$ measures the marginal change on the mean of the latent variable

@$\alpha$ measures the relative impact on the variance

Tobit estimates for $\beta$ measure the marginal impact of changes in the conditional variables on the latent mortality rate.[11] For example, the coefficient corre-

[11] The Tobit specification assumes that the latent variable is a continuous, normally distributed

sponding to in-weight, states that a 10% increase in entry weight lowers the latent variable by 3.9%. McDonald and Moffitt (1980) show how Tobit parameter estimates can be decomposed into two parts, where the first part contains the effect on the probability that the variable is above zero, while the second part contains the mean effect, conditional on being greater than zero. Also, the estimates for $\alpha$ measure the relative impact on the variance. For example, the estimation coefficient corresponding to fall implies that a pen placed in that period is associated with a variance that is 32% higher than the base months containing summer. MSPE is computed as the average squared difference between the predicted and actual mortality rates.

The Tobit model employs maximum likelihood methods, which maximizes the log-likelihood (LL) value by optimizing parameter estimates for $\beta$ and $\alpha$. The LL is this optimized value, which is used to compute AIC. The high AIC results mainly from the amount of unobservable variation that is not captured in mortality rates. These values hold meaning when compared to other model specifications.

Next, we move to estimate the same set of data using the previously developed zero-inflated models in order to test our hypothesis that they will have a better fit. Before proceeding to estimation, there are a few notable differences when using classical and Bayesian methods. First, Bayesian point estimates are typically computed as the mean from Monte Carlo simulations of the posterior density function. This estimation process is done in two parts; first the likelihood of a zero value is modeled, followed by simulating the positive predicted realizations, based on a log-normal distribution. In addition to the mean value, additional characteristics of the posterior distributions are supplied, such as the median, 2.5 and 97.5 percentile values, and the standard deviation, as well as the Monte Carlo standard error of the

variable that is observed for positive values and zero for negative values. Marginal changes in the latent variable must then be converted to the marginal changes in the observed variable, in order to offer inferences on the observable variable. The marginal impact on mortality rates can be approximated by multiplying the marginal impact on the latent variable by the degree of censoring (Greene, 1981) mean. Results from the zero-inflated log-normal model are shown below in Table 3.11.

Parameter estimates in the zero-inflated model refer to two distinct processes. The first process includes the likelihood of a zero outcome or one described by a log-normal distribution. This process is estimated through $\delta$ utilizing equation (3.20). Based on this formulation, the parameter estimates can be expressed as the negative of the marginal impact of the conditional variable on the probability of a positive outcome, relative to the variance of the Bernoulli component:

$$\delta_k = -\frac{\partial \rho_i}{\partial x_{ki}} \cdot \frac{1}{\rho_i} \cdot \left[\frac{1 + exp(x_i'\delta)}{exp(x_i'\delta)}\right] = -\frac{\partial \rho_i}{\partial x_{ki}} \cdot \frac{1}{\rho_i(1-\rho_i)} \tag{3.47}$$

where the variance is shown as $\rho_i(1-\rho_i)$. For example, entry weight largely and negatively influences the likelihood of positive mortality rates. This is not surprising given that more mature pens are better equipped to survive adverse conditions, whereas younger pens tend to be more likely to result in mortalities. Alternatively, mixed pens have a negative $\delta$ coefficient which implies that there is a positive relationship, relative to heifer pens. Therefore, if a pen is mixed, it has a higher probability of incurring positive mortality realizations that can be modeled with a log-normal distribution.

The Tobit model assumes that estimates for $\beta$ and $\delta$ will work in the same way. For most variables, $\delta$ coefficients are negatively related to $\beta$ coefficients, which points to directional consistency. For example, increases to entry weight shift the mean of mortality rates downward and also decrease the probability of a positive outcome. This does not necessarily mean that the two processes work identically, as is assumed with the Tobit model, but rather tend to generally work in the same direction.

Parameter estimates for $\beta$ refer to the marginal impact that the conditioning variables have on the positive realizations of mortality rates. Interpretations for

Table 3.11: Univariate ZILN estimates of fed cattle parameters

| node | mean | sd | MC error | 2.5% | median | 97.5% | Parameter |
|---|---|---|---|---|---|---|---|
| Intercept | 6.047 | 2.193 | 0.160 | 0.334 | 6.466 | 9.076 | |
| Steers | 0.011 | 0.034 | 0.002 | -0.066 | 0.014 | 0.070 | |
| Mixed | 0.394 | 0.037 | 0.001 | 0.321 | 0.394 | 0.467 | |
| KS | 0.118 | 0.028 | 0.001 | 0.062 | 0.119 | 0.172 | |
| log(inwt) | -0.907 | 0.336 | 0.025 | -1.371 | -0.970 | -0.033 | $\beta^{\#}$ |
| Winter | -0.020 | 0.030 | 0.001 | -0.078 | -0.021 | 0.038 | |
| Fall | 0.087 | 0.032 | 0.001 | 0.025 | 0.086 | 0.151 | |
| Spring | -0.085 | 0.031 | 0.001 | -0.148 | -0.085 | -0.023 | |
| Intercept | -3.379 | 1.240 | 0.090 | -5.008 | -3.695 | 0.324 | |
| Steers | 0.059 | 0.051 | 0.002 | -0.043 | 0.059 | 0.157 | |
| Mixed | -0.274 | 0.072 | 0.001 | -0.414 | -0.273 | -0.133 | |
| KS | 0.078 | 0.060 | 0.002 | -0.034 | 0.078 | 0.199 | |
| log(inwt) | 0.618 | 0.191 | 0.014 | 0.049 | 0.666 | 0.877 | $\alpha^{@}$ |
| Winter | 0.163 | 0.061 | 0.001 | 0.046 | 0.163 | 0.280 | |
| Fall | -0.069 | 0.06 | 0.001 | -0.190 | -0.069 | 0.048 | |
| Spring | 0.229 | 0.065 | 0.001 | 0.100 | 0.229 | 0.357 | |
| Intercept | -11.380 | 3.070 | 0.224 | -16.160 | -11.610 | -4.508 | |
| Steers | -0.007 | 0.063 | 0.003 | -0.127 | -0.007 | 0.118 | |
| Mixed | -0.177 | 0.079 | 0.001 | -0.331 | -0.177 | -0.023 | |
| KS | 0.137 | 0.062 | 0.002 | 0.016 | 0.138 | 0.257 | |
| log(inwt) | 1.686 | 0.469 | 0.034 | 0.639 | 1.720 | 2.419 | $\delta^{!}$ |
| Winter | -0.066 | 0.065 | 0.001 | -0.196 | -0.066 | 0.060 | |
| Fall | -0.104 | 0.067 | 0.002 | -0.237 | -0.103 | 0.027 | |
| Spring | 0.169 | 0.068 | 0.001 | 0.036 | 0.169 | 0.302 | |
| LL: | -9,341.500 | | | | | | |
| DIC: | 18,742.300 | | | | | | |
| MSPE: | 2.399 | | | | | | |

$^{\#}\beta$ measures the marginal change on the mean of the latent variable

$^{@}\alpha$ measures the relative impact on the variance

$^{!}\delta$ measures the negative relative impact in the probability of a non-zero entry these parameters refer to the marginal increase in the log of mortality rate. For example, an increase in entry weight by 1.0% is associated with a reduction in mortality rates by 0.9% for the observations that experience a positive mortality rate.

It is interesting to note the different implications from parameter estimates from the Tobit and ZILN models. For example, an insignificant mean parameter estimate for the variable KS in the Tobit model implies that mortality rates are not significantly impacted by feedlot location. However, parameter estimates from the ZILN model infer that pens placed into feedlots located in Kansas have a lower likelihood of a positive mortality realization by 13.7%, relative to Nebraska feedlots. At the same time, pens placed in Kansas that have a positive mortality rate, can be expected to realize a rate that is 11.8% higher than Nebraska feedlots. This might seem strange to have significant impacts in opposite directions that influence both the likelihood of a mortality and the positive mortality rate, but by distinguishing between these processes we can isolate their respective impacts. One possible explanation might be that Kansas lots spend more time to prevent mortalities from occurring through vaccinations or backgrounding, but are not able to prevent the spread of disease as quickly as the Nebraska feedlots. This is a notable departure from the Tobit model which saw no significant influence since these impacts essentially canceled each other out.

Another notable difference is in seasonal impacts on the mean of mortality rates. While the none of the seasonal variables are significantly different than summer under the Tobit model, both Fall and Spring are significantly different under the ZILN specification. The ZILN results are more inline with expectations as Fall placement are put under stress from extremely cold weather, which is different from summer placements. In fact, most of the pens with mortality losses above 10% in this data sample come from pens placed in the fall months.

The zero-inflated log-normal models also demonstrate a superior ability to characterize and predict cattle feedlot observations. A DIC measure of $18,742.3$ demonstrates a closer fit, relative to the Tobit model, which as an AIC value of $22,790.7$. Additionally, MSPE is minimized when using the ZILN model. The likely explanation for these findings is due to the data generating process. Cattle mortalities appear to be part of a two part process where once a pen experiences a mortality, the rate of mortalities can be modeled using a distinct distribution from those observations without mortalities. This method may prove to be fruitful in situations where the data have some similar characteristics. Examples may be modeling the prevalence of animal disease, where a Bernoulli distribution characterizes the likelihood of an outbreak. Once an outbreak has occurred, a model describing its biological spread is needed. This strategy may also extend into areas of bio-security and food safety issues where biological processes may be allowed to spread within a population once contamination has occurred. Additionally, data that are characterized with a high degree of censoring can be efficiently characterized through the use of a zero-inflated model, as shown in earlier simulations.

The given data set was also modeled using a zero-inflated Gamma (ZIG) distribution. While the Bernoulli component is similar to the ZILN model, this model characterizes the positive observations using a Gamma distribution, which also can take into account highly skewed data. The results from the ZIG model are shown below in Table 3.12.

In the process of running this model in WinBUGS, it became apparent that the data did not result in any statistically significant variables used to estimate the rate parameter, $\eta$. Additionally, model fit tests were conducted with and without $\kappa$, which showed that the model did not lose any efficiency when estimating a single intercept variable for $\kappa$. Because there were no apparent advantages to estimating this set of variables, regressions were run using only an intercept term for $\kappa$, keeping

Table 3.12: Univariate ZIG estimates of fed cattle parameters

| node | mean | sd | MC error | 2.5% | median | 97.5% | Parameter |
|---|---|---|---|---|---|---|---|
| Intercept | 2.569 | 0.424 | 0.031 | 1.904 | 2.510 | 3.350 | |
| Steers | -0.054 | 0.033 | 0.002 | -0.115 | -0.054 | 0.010 | |
| Mixed | 0.379 | 0.044 | 0.002 | 0.295 | 0.379 | 0.466 | |
| KS | 0.014 | 0.034 | 0.002 | -0.054 | 0.015 | 0.079 | |
| log(inwt) | -0.410 | 0.066 | 0.005 | -0.533 | -0.402 | -0.310 | $\gamma$ |
| Winter | 0.035 | 0.040 | 0.002 | -0.040 | 0.034 | 0.118 | |
| Fall | 0.134 | 0.040 | 0.002 | 0.057 | 0.133 | 0.217 | |
| Spring | -0.188 | 0.044 | 0.002 | -0.270 | -0.190 | -0.094 | |
| | 1.187 | 0.028 | 0.001 | 1.134 | 1.187 | 1.243 | $\kappa$ |
| Intercept | -0.465 | 0.427 | 0.031 | -1.426 | -0.429 | 0.318 | |
| Steers | 0.114 | 0.052 | 0.003 | 0.010 | 0.113 | 0.221 | |
| Mixed | -0.168 | 0.078 | 0.003 | -0.322 | -0.166 | -0.019 | |
| KS | 0.167 | 0.055 | 0.003 | 0.063 | 0.165 | 0.280 | |
| log(inwt) | 0.020 | 0.063 | 0.005 | -0.102 | 0.015 | 0.156 | $\delta$ |
| Winter | -0.108 | 0.064 | 0.003 | -0.240 | -0.106 | 0.013 | |
| Fall | -0.145 | 0.065 | 0.003 | -0.280 | -0.145 | -0.018 | |
| Spring | 0.212 | 0.063 | 0.003 | 0.074 | 0.215 | 0.330 | |
| LL: | -10,985.850 | | | | | | |
| DIC: | 21,987.000 | | | | | | |
| MSPE: | 2.767 | | | | | | |

$\eta$ constant across all observations.[12]

Regressions from univariate mortality models offer information concerning the relative impacts each conditioning variable has on mortality rates. However, this variable is likely better characterized in a multivariate setting with other variables that explain the health and performance of cattle on feedlots, ultimately describing production risk in cattle feeding enterprises. To this end, the multivariate Tobit model and multivariate zero-inflated models were used to characterize these four variables, described earlier. The results from the multivariate Tobit model are shown in table 3.13.

Results from this estimation mostly appear to be in line with the estimation from Belasco et al. (2006), as well as the first essay. While the same data set was used, this study employs two-thirds of the data for estimation and the final third for out of sample prediction. Mortality rates contain the most variability in prediction, mostly due to the relative lack of explanatory power from the conditioning variables. While these *ex-ante* variables offer information on expected mortality rates, there does still appear to be a bit more unexplained variation than with the other variables. Performance variables, such as DMFC and ADG, are largely determined by observable biological traits. While not all of these biological traits are captured in these data, there does not appear to be a large portion unexplained by these variables.

The elements contained within the covariance matrix, $\Sigma_i = T_i' D_i T_i$, are described by equations (3.15) and (3.16) where estimates labeled as 'Heteroskedasticity' are contained within $D_i$ and covariance elements are contained within $T_i$. Since all diagonal elements of matrix $D_i$ are guaranteed to be positive by construction the sign of the covariance element describes the relationship between two variables. For

[12] This finding was interesting, given the fact that variance is not dependent on the conditioning variables for this specification. While heteroskedasticity is quite apparent for the other specifications, it does not seem to add modeling efficiency for the Gamma specification. This finding may be the result of the Gamma distribution accounting for the structure of the error terms.

Table 3.13: Multivariate Tobit estimates of fed cattle parameters

| | DMFC | | ADG | | VCPH | | MORT | |
|---|---|---|---|---|---|---|---|---|
| Variables | coeff. | se. | coeff. | se | coeff. | se | coeff. | se |
| Intercept: | -1.208* | 0.035 | -2.684* | 0.288 | 8.490* | 0.178 | 24.001* | 1.447 |
| Steers: | -0.092* | 0.001 | -0.083* | 0.012 | 0.431* | 0.007 | 0.141* | 0.052 |
| Mixed: | -0.028* | 0.002 | 0.152* | 0.020 | 0.134 | 0.011 | 0.556* | 0.098 |
| Kansas: | -0.137* | 0.002 | -0.253* | 0.013 | 0.251* | 0.007 | -0.006 | 0.061 |
| log(inwt): | 0.481* | 0.005 | 0.806* | 0.044 | -0.822* | 0.027 | -3.605* | 0.221 |
| Winter: | 0.011* | 0.002 | 0.003 | 0.014 | -0.269* | 0.008 | -0.041 | 0.064 |
| Fall: | 0.069* | 0.002 | 0.074* | 0.014 | -0.313* | 0.008 | 0.126 | 0.070 |
| Spring: | -0.022* | 0.002 | -0.164* | 0.016 | -0.020 | 0.008 | -0.215* | 0.065 |
| **Heteroskedasticity:** | | | | | | | | |
| Intercept: | -9.032* | 0.787 | -8.387* | 0.790 | 6.534* | 0.929 | 12.713* | 1.116 |
| Steers: | -0.038 | 0.036 | -0.563* | 0.036 | 0.117* | 0.036 | 0.007 | 0.048 |
| Mixed: | 0.450* | 0.054 | -0.139* | 0.056 | 0.242* | 0.056 | 0.605* | 0.070 |
| Kansas: | -0.269* | 0.042 | 0.236* | 0.042 | 0.014 | 0.041 | -0.041 | 0.057 |
| log(inwt): | 0.683* | 0.122 | 1.048* | 0.122 | -1.433* | 0.141 | -1.742* | 0.170 |
| Winter: | -0.094* | 0.045 | 0.075 | 0.046 | 0.082 | 0.046 | -0.130* | 0.061 |
| Fall: | 0.326* | 0.045 | 0.170* | 0.046 | 0.196* | 0.046 | 0.209* | 0.060 |
| Spring: | -0.355* | 0.046 | 0.511* | 0.051 | 0.126* | 0.047 | -0.124 | 0.065 |
| **Covariance(t):** | | | | | | | | |
| Cov(DMFC VCPH:) | | | 1.254* | 0.054 | | | | |
| Cov(DMFC MORT:) | | | 7.503* | 0.254 | | | | |
| Cov(DMFC ADG:) | | | -4.090* | 0.031 | | | | |
| Cov(VCPH MORT:) | | | 1.039* | 0.053 | | | | |
| Cov(VCPH ADG:) | | | -0.094* | 0.006 | | | | |
| Cov(ADG MORT:) | | | 0.003 | 0.002 | | | | |
| LL: | | | -15,748.2 | | | | | |
| AIC: | | | 31,636.5 | | | | | |
| DMFC MSPE: | | | 0.009 | | | | | |
| VCPH MSPE: | | | 0.232 | | | | | |
| ADG MSPE: | | | 0.216 | | | | | |
| MORT MSPE: | | | 2.617 | | | | | |

*Denotes the estimate is statistically significant at the 0.05 level

example, DMFC and ADG are negatively related with a coefficient of -4.09, since a healthy pen of cattle will be expected to have a low feed conversion rate and a high rate of gain. Additionally, MSPE is broken out by dependent variable. MORT has the highest MSPE, which illustrates the lack of predictive power with that variable.

Next, the multivariate zero-inflated model is applied to the cattle feedlot data set and results are shown in Table 3.14. Estimates displayed here are consolidated, relative to the univariate table due to space constraints. In a Bayesian framework, confidence intervals are typically computed using the highest posterior density region, which will be different from a classical confidence interval when posterior distributions are bi-modal or asymmetric. Since the posterior estimates do not show bi-modal attributes, we proceed by taking the interval between the 2.5 and 97.5 percentiles to test whether the variable is significantly different from zero. While this is a departure from Bayesian theory, it nearly aligns with significance tests for the multivariate Tobit model. For example, in the zero-inflated model, if the posterior density function does not cross zero in the given interval, which includes 95 percent of the posterior density, then it is said to be significant at the 5% level.

Many of the same estimates appear in the zero-inflated table, with the addition of 'Delta' terms, which describe the negative of the relative impact on the probability of a non-zero outcome. These estimates are computed in the same way as the univariate version of this model, leading to many of the same inferences. Additionally, parameter estimates corresponding to DMFC, ADG, and VCPH are mostly the same between the Tobit model and ZILN models. Parameter estimates corresponding to MORT are different between the two models, as discussed with the univariate model.

The zero-inflated model does a superior job of fitting the data and in terms of prediction accuracy, relative to the Tobit model. DIC for the ZILN model is

Table 3.14: Multivariate ZILN estimates of fed cattle parameters

| | DMFC | | ADG | | VCPH | | MORT | |
|---|---|---|---|---|---|---|---|---|
| Variables | coeff. | se. | coeff. | se | coeff. | se | coeff. | se |
| Intercept: | 0.767* | 0.048 | -2.091* | 0.160 | 11.000* | 0.215 | 11.220* | 0.144 |
| Steers: | -0.069* | 0.002 | 0.322* | 0.010 | 0.073* | 0.012 | 0.094* | 0.021 |
| Mixed: | -0.030* | 0.004 | 0.150* | 0.016 | 0.218* | 0.019 | 0.359* | 0.034 |
| Kansas: | -0.121* | 0.002 | 0.186* | 0.012 | -0.207* | 0.012 | 0.111* | 0.027 |
| log(inwt): | 0.178* | 0.007 | 0.794* | 0.025 | -1.289* | 0.033 | -1.707* | 0.022 |
| Winter: | -0.002 | 0.003 | -0.193* | 0.013 | -0.073* | 0.013 | -0.020 | 0.027 |
| Fall: | 0.052* | 0.003 | -0.242* | 0.014 | 0.016 | 0.013 | 0.047 | 0.027 |
| Spring: | -0.017* | 0.003 | -0.054* | 0.013 | -0.083* | 0.014 | -0.044 | 0.028 |
| **Heteroskedasticity:** | | | | | | | | |
| Intercept: | -1.606* | 0.251 | -7.182* | 0.196 | -2.983* | 0.250 | -0.357 | 9.925 |
| Steers: | -0.397* | 0.026 | 0.253* | 0.066 | -0.045 | 0.060 | -0.002 | 10.210 |
| Mixed: | 0.123* | 0.040 | 0.365* | 0.075 | 0.092 | 0.080 | -1.067 | 10.210 |
| Kansas: | 0.148* | 0.030 | -0.155* | 0.051 | -0.207* | 0.053 | -1.876 | 9.530 |
| log(inwt): | -0.504* | 0.038 | 0.790* | 0.028 | -0.271* | 0.032 | -9.971* | 5.932 |
| Winter: | 0.160* | 0.032 | -0.189* | 0.056 | -0.120 | 0.063 | 0.322 | 10.720 |
| Fall: | 0.313* | 0.032 | 0.194* | 0.056 | 0.047 | 0.063 | -0.018 | 9.713 |
| Spring: | 0.332* | 0.037 | -0.300* | 0.069 | -0.164* | 0.070 | -0.167 | 9.724 |
| **Delta:** | | | | | | | | |
| Intercept: | | | | | | | -16.060* | 0.360 |
| Steers: | | | | | | | -0.062 | 0.049 |
| Mixed: | | | | | | | -0.181* | 0.081 |
| Kansas: | | | | | | | 0.108 | 0.062 |
| log(inwt): | | | | | | | 2.402* | 0.055 |
| Winter: | | | | | | | -0.051 | 0.064 |
| Fall: | | | | | | | -0.082 | 0.067 |
| Spring: | | | | | | | 0.152* | 0.065 |
| **Covariance(t):** | | | | | | | | |
| Cov(DMFC VCPH:) | | | 4.696* | 0.061 | | | | |
| Cov(DMFC MORT:) | | | 1.841* | 0.113 | | | | |
| Cov(DMFC ADG:) | | | -1.742* | 0.185 | | | | |
| Cov(VCPH MORT:) | | | 6.720* | 0.584 | | | | |
| Cov(VCPH ADG:) | | | 0.483 | 0.401 | | | | |
| Cov(ADG MORT:) | | | 0.059 | 0.105 | | | | |
| LL: | | | -9,617.8 | | | | | |
| DIC: | | | 19,348.0 | | | | | |
| DMFC MSPE: | | | 0.008 | | | | | |
| VCPH MSPE: | | | 0.182 | | | | | |
| ADG MSPE: | | | 0.170 | | | | | |
| MORT MSPE: | | | 1.296 | | | | | |



*Denotes the estimate is statistically significant at the 0.05 level

substantially lower than AIC in the Tobit model, mainly due to the more accurate fit for Mortality rates, which contributes quite a lot of unexplained variability to the system of equations. The more efficient modeling of mortality rates stem from the ability of the zero-inflated model to more accurately represent MORT by taking into account the two part process inherent in mortality rates. MSPE measures are approximately the same for each of the non-censored variables, largely because they are modeled in a similar fashion. However, more information about mortality rates in the zero-inflated model add to more accurately model the other variables. In fact, the multivariate zero-inflated model predicts every dependent variable with more precision, leading to large gains in both prediction and model fitting.

We can decompose the total DIC and LL values from the multivariate zero-inflated model into dependent variable components, which is shown in Table 3.15.[13] This table is helpful in breaking down model fit measures to identify the performance of the model on each variable. MORT is more accurately characterized in a multivariate setting because of the effects from other non-censored variables. Recall, that the expected value and variance of MORT accounts for the uncensored variable levels in the multivariate setting. D represents the estimation on the parameters of 'Delta', which performs roughly similar in both multivariate and univariate situations, since it uses the same modeling mechanism in each case. DMFC is modeled very tightly, as shown by the negative DIC, while ADG and VCPH leave some variability unestimated. The results from the total line are reported with the full model results in Table 3.14.

Given the fact that the zero-inflated model outperforms the Tobit model for the given feedlot data set, we next replicate the profit simulation conducted in essay 2 to quantify any distinct difference in expected *ex-ante* profits resulting from the

[13]LL values are computed by dividing Dhat by -2, since $Dhat = -2 * LL(\hat{\beta}, \hat{\gamma}, \hat{\delta})$. This aligns with LL values in the Tobit model which are computed based on optimized values. Along the same lines, Dhat is computed by using the optimal posterior means.

Table 3.15: Multivariate ZILN model fit estimates, by element

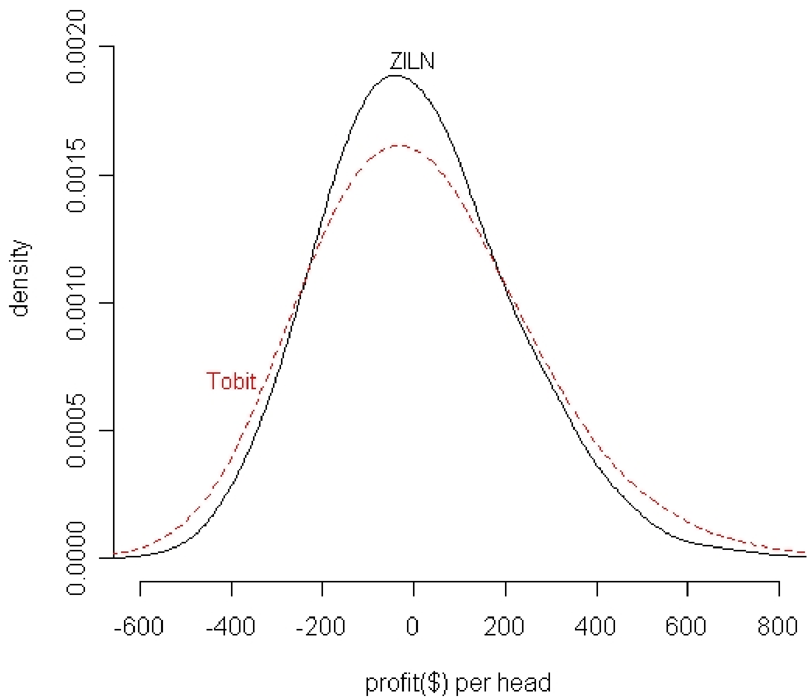
| Variable | Dbar | Dhat | pD | DIC | LL |
|---|---|---|---|---|---|
| ADG | 7,923.6 | 7,911.5 | 12.1 | 7,935.7 | -3,955.8 |
| DMFC | -15,002.6 | -15,013.3 | 10.7 | -14,991.8 | 7,506.7 |
| MORT | 7,195.6 | 7,179.1 | 16.5 | 7,212.0 | -3,589.6 |
| VCPH | 8,964.0 | 8,954.1 | 9.9 | 8,973.9 | -4,477.1 |
| D | 10,211.2 | 10,204.2 | 7.0 | 10,218.2 | -5,102.1 |
| total | 19,291.8 | 19,235.6 | 56.2 | 19,348.0 | -9,617.8 |

different functional forms. More specifically, we are interested in evaluating the differences between the model assumed in essay 2 and that of the zero-inflated log normal distribution. The two *ex-ante* profit density functions are plotted on the same axis below in Figure 3.1.

The most notable difference between the densities is that the zero-inflated model results in a reduction in the standard deviation by 16.4%. This is a meaningful difference shown by the change in the first quartile by $20 per head. The additional insight obtained by including this particular two-part model adds an important component to modeling cattle production yields and ultimately adding precision to simulated *ex-ante* cattle feeding profit densities.

## 3.7 Implications and Recommendations

Modeling censored data sets remains a large problem in economics. While use of the Tobit model may be well-justified in certain instances, the results from both simulated and actual cattle feeder data sets suggest the use of a zero-inflated modeling mechanism. This is particularly true in instances where data come from a mixture model. While two-step processes have been applied to hurdle models, zero-inflated models have largely been ignored in economic studies. This is mainly a result of the past limitation of zero-inflated models to count data. In this essay, a




| Scenario | Mean | Sd | 25% |
|---|---|---|---|
| Tobit | 17.14 | 251.71 | -158.18 |
| ZILN | 6.98 | 210.34 | -138.03 |

Figure 3.1: Distribution of *ex-ante* conditional profits per head based on multivariate Tobit and zero-inflated log normal density functions

zero-inflated model is developed that can handle both univariate and multivariate situations rather efficiently, in addition to nesting the standard Tobit model. Additionally, the inherent parametric flexibility allows for distributional assumptions to change based on the data on hand, rather than strictly using truncated or normally distributions. Here we use a log-normal distribution to capture the positively skewed nature of cattle feedlot mortality rates, which gives the zero-inflated model significant advantages over the Tobit model. Advantages in model fit for the ZILN model stem from the ability of the zero-inflated model to isolate the impacts from observing a positive mortality rate and the level of mortality rates.

Production risk in cattle feeding enterprises is inherently complex, given the many areas risk can originate. This study develops a zero-inflated model, in contrast to the more commonly used Tobit model. Results from this research demonstrate the potential gains from using this particular mixture model. Before applying this model to the data, simulations were conducted to test the model's ability to predict and fit data generated in different forms. These simulations provided results that concluded the advantages of the mixture model, in both prediction and model fit, when the data is from a two-step process. Additionally, the mixture model demonstrated a strong ability to fit the data, even when the data are generated based on a Tobit model. These results are in general agreement with the results obtained within our application of cattle feeding.

A solid understanding of cattle production risks is limited by our ability to characterize variability. The proposed model takes a step forward in developing a modeling strategy that can be used to measure other livestock or live animal productive measures. By more accurately characterizing these risks insurance companies, animal producers, and operators can better understand the risks involved with animal production. Additionally, the flexibility of this model allows for uses outside of live animal yields. The major flexibility in the proposed model lies in the ability to make different distributional assumptions. Distributional assumptions typically need to be made in cases when data cannot fully explain variability. However, non-parametric and semi-parametric methods may be of particular interest when large data sets are evaluated, since they allow empirical data to create a unique density. With more data available on live animal yields, augmenting this model to include these types of density functions may provide additional precision.

# Chapter 4

# The Impact of Price Risk Management on Overall Fed Cattle Profit Risk

## 4.1 Introduction

Agricultural production is distinct from most other enterprises due to its inherent risky nature. Risk in agriculture generally originates from both yield and price. Yield risk refers to the variation in productive efficiency, which is often expressed with a conditional probability density function. Productive efficiency varies mostly due to uncertainty in weather, disease, or pests. However, with live animals, production can also vary due to differences in placement characteristics such as weight, gender, and season, as well as unknown genetic differences. Additionally, fluctuations in price result from changes in world market demand or supply conditions.

Since 1980, the federal government has taken large strides to help farmers reduce risk vulnerability by offering subsidized insurance. Historically, these programs have been mostly focused on crops and based on reducing yield risk exposure. However, since 1997, revenue-based insurance programs have become the federal government's main tool in assisting farmers to manage risk. The major advantage to revenue-based insurance programs is that they account for risk originating from yield and price variability and ultimately guarding farmers' income more precisely. Additionally, prices and yields tend to be negatively correlated, which leads to lower indemnity payment levels and revenue variability relative to traditional yield-based insurance products.

The recent interest in revenue-based programs within crop production leads to the question of similar insurance offerings within cattle industries. Recently, the federal government introduced two insurance programs to insure fed and feeder cattle through Livestock Gross Margin (LGM) and Livestock Risk Protection (LRP) policies. Both of these products insure cattle owners against adverse swings in prices. These products are based on the finding that a large majority of risk in cattle industries originate in prices. Alternatively, yield risk has been excluded from existing federal cattle insurance policies which is likely because it is thought to contain a relatively small amount of risk to profits. This research will quantify the amount of risk arising from cattle production yield risks, once price risk is accounted for through different risk management techniques, such as forward-pricing and options contracts. Within this context, profit risk will be illustrated under different scenarios where elements of price risk are stripped from profit risks.

The characterization of risk arising from production yield measures in livestock will be the first step in assessing the significance of production risk in cattle feedlot operations. To comprehensively characterize profit risks, we must begin with a clear understanding of risks that can be accounted for in both yield and price areas. Using simulated and actual cattle feeding data, production risks will be quantified in the presence of forward and option contracts on both live cattle and corn prices. In using both risk management techniques, *ex-ante* conditional profit density functions can be evaluated when price risk is no longer a risk factor. This will isolate cattle production yield risks, which will allow us to quantify the amount of risk currently left uninsured.

There are three major contributions from this research. First, this study will evaluate the amount of risk that can be attributed to different risk components in cattle feeding production. Specifically, the amount of profit risk involved in cattle feeding when components of price risk are peeled away. Second, the amount of production risk that remains after price risk management efforts have been made will be quantified. Four sources of production risk are identified and through the use of simulation techniques describe profit risks. The amount of profit risk will be illustrated through the use of probability density functions and cumulative density functions. The major question this study asks is whether production risk is a significant amount of risk to cattle feeders. This question is relevant to current policy that only insures prices involved in livestock production. A previous study by Knoeber and Thurman (1995) evaluated the relative contributions of price and yield risk in the profitability in broiler production. The third contribution will be a sensitivity analysis that examines the impact different price and production risk factors have on expected profits.

The next section will move to discuss the current place of federal revenue insurance programs in risk management strategies within crop farming. Following, section 4.3 will discuss the current status of livestock insurance and its potential need for insurance that accounts for production yield risk. Section 4.4 quantifies the amount of risk currently uninsurable through federal insurance programs by using probabilistic density functions that characterize *ex-ante* profits in cattle feeding enterprises, as well as a sensitivity analysis in order to quantify the realistic bounds on profitability that result from extreme price and yield adjustments. The final section summarizes the findings from the previous sections and briefly describes potential frameworks for dealing with moral hazard and adverse selection problems when insuring cattle production risk and points to future directions of research in this area.

## 4.2 The Rise of Revenue Insurance Programs

This section is intended to offer some background information regarding the use of revenue-based insurance products in managing risk in crop farming. Particular emphasis will be placed on the relationship between yield and price risk in crop production.

As of 1996, crop insurance programs had been confined to insuring crop yield-risk, based on past performance. The most popular offering was the Actual Production History (APH) program, where indemnity payments are triggered when yields fall below a specified percentage of the average performance. Throughout the 80s and early 90s, the federal crop insurance programs were highly criticized for the low participation rates, inability to eliminate disaster aid relief, and serious adverse selection problems, as discussed by Goodwin and Smith (1995). Low participation rates during this time may have been related to the problem associated with adverse selection (Miranda, 1991).

The two major issues in most insurance products are moral hazard and adverse selection. Adverse selection is problematic since it can lead to a limited pool of insured agents who possess more information than the insuring body. This pool of agents is typically populated by producers who participate only because they expect to receive more indemnity payments than pay through premium rates. Glauber (2004) discusses the difficulty of overcoming adverse selection problems in the United States from the 1981 to 1993, where agents received twice as much in indemnity payments as their premium rates.[1]

Moral hazard occurs when an insured agent take actions once insured that

[1] Within this period of time the U.S. experienced a major drought in 1988 that covered approximately a third of the country at its peak, as well as a dramatic flood in 1993 that affected the Midwest region, which increased loss ratios in 1998 and 1993 to nearly 2.5 and 2.3, respectively.

affect the probability of an adverse event. For example, in crop production this might include the use of less chemicals (Goodwin et al., 2004). An increase in monitoring and surveillance through the RMA has been the major tool in minimizing moral hazard. Additionally, new insurance products that are based on county-level yields, such as the Group Risk Plan (GRP) and Group Risk Income Protection (GRIP) programs, minimize moral hazard.

In 1996, revenue-based insurance plans were first offered to more directly insulate farming incomes from risk that can originate from yield and price variability. Since adverse price and yield events tend to be negatively correlated, indemnity payments decrease with lower variability in revenue risk. The introduction of revenue-based products have started to replace traditional yield-based programs, such as APH, as the major tool the federal government uses to guard farming incomes. For example, 57% of the total amount of acres insured by federal crop insurance offerings in 2006 were revenue-based products[2], compared to 28% insured through the APH product (FCIC, 2007). The total share of liability from revenue-based and APH products was 59% and 26% in 2006, respectively. Hennessy et al. (1997) found that revenue-insurance offerings provide additional benefits at a cost lower than existing crop insurance programs. Pricing revenue-based products is complicated due to the bivariate nature of prices and yields. Because of this relationship, prices and yields must be quantified as well as their covariate relationship, which is discussed further by Goodwin and Ker (2002) as well as Mahul and Wright (2003).

## 4.3 Current State of Cattle Insurance

The Agricultural Risk Protection Act (ARPA) of 2000 declared a fundamental shift in federal insurance involvement. One major change was the expansion of insurance

[2]Revenue-based products offered by the Federal Crop Insurance Company (FCIC) currently include Crop Revenue Coverage (CRC), Group Risk Income Protection (GRIP), Income Protection (IP), and Revenue Assurance (RA).

offerings to include new livestock products. Livestock offerings are expected to double the size of existing crop insurance programs (Glauber, 2004). In addition to the previously mentioned LGM and LRP productions, a new Group Risk Plan (GRP) Pasture, Rangeland, Forage (PRF) Insurance product is currently offered and provides risk protection to farmers and ranchers who use pasture, rangeland, or forage for haying or grazing. Rangeland and pastures amount to almost 600 million acres in the United States. Rangeland production is tied to two separate weather indices based on rainfall and a vegetation index[3].

Additionally, livestock insurance products include the insulation from price risk in the form of LRP and LGM products. While the LRP product protects against adverse swings in fed cattle prices, it is not much different from a put option on the Chicago Mercantile Exchange (CME). Fackler (1989) discusses the relation between futures, options, and government subsidies. The LGM product insures against both output and input prices, and is formulated in Hart et al. (2001). The authors point out that production risk is excluded from insurance due to its small magnitude, relative to price risk. Past research has generally agreed that most cattle revenue risk comes from price variability (Belasco et al., 2006; Lawrence et al., 1999; Mark et al., 2000). The LGM product allows for cattle pens to be insured throughout the year, while indemnity payouts are based on the total gross margin throughout the year. The major advantage of both livestock insurance products is that they do not allow producers to influence the probability of an indemnity payout by tying prices to the futures market. This offers a convenient way for this particular type of insurance to minimize problems associated with moral hazard.

However, Hall et al. (2003) report that drought and cattle price variability were the greatest two concerns to ranch income by cattle ranchers in Texas and Ne-

[3]The rainfall index is maintained by NOAA's Climate Precipitation Center, while the vegetation index is the U.S. Geological Survey's Earth Resources Observation and Science (EROS) normalized difference vegetation index (NDVI)

braska, followed by extremely cold weather and disease. Of the four largest concerns, only price variability is covered by existing federal insurance programs. Drought has an indirect effect on the feed supplies and prices, since lower corn yields generally decrease supplies and increase the price of corn used to feed cattle. Extreme weather, in the form of drought or extreme cold, can negatively impact the performance and health of cattle in all stages of life. Disease can have different effects depending on its severity. Disease impacting the food supply can have extremely large impacts on the entire livestock industry as past outbreaks have led to temporary stoppages with important trading partners, as well as a hesitation to consume products that are found to have disease outbreaks. The less severe scenario includes a spread of disease throughout the herd, causing diminished performance and/or an increase in mortality rates.

Low participation rates in existing insurance offerings and reliance on other risk management methods, such as understocking pasture and storing hay, may indicate inadequate insurance offerings or a lack of educational outreach with existing programs. The exclusion of production risk from existing federal livestock insurance plans may also be due to the relative lack of data evaluating production risks, when compared to the wealth of available information on livestock prices (Hart, 2006). In response, this study utilizes a comprehensive data set to evaluate cattle production risks and their impact on profit risks.

Establishing that production risk is significant in overall profit risk does not necessarily make it an insurable component. Issues of moral hazard and adverse selection are the biggest obstacles in protecting against production risk. With many different breeds and an array of genetic potential within herds of cattle, insuring performance and health becomes a difficult task. A successful insurance product will prevent the insurance pool from consisting strictly of the lower performing cattle that have characteristics that are unobservable to the insurer, but observable to the rancher. As a first step, this research moves forward with the goal of quantifying the relative amount of production risk in cattle feedlot operations.

## 4.4 Production Risk in Cattle Feeding

The purpose of this section is to quantify the relative amount of production risk in cattle feeding enterprises. This element of risk is uninsured under current federal livestock insurance policy. Here, we look to isolate the amount of risk that arises strictly from production variability that can originate in performance or health factors. Hart et al. (2001) point out that most of the production risk in livestock production originates from disease, mechanical failure, or variability in weight gain. After identifying four sources of production risk, we move to quantify this risk through the use of probability density functions that isolate production risk from price risk. Four different scenarios of coverage levels are used and include fed cattle price protection only, corn price protection only, both fed cattle and corn price protection, and no price protection. Using these four scenarios, we first define price protection through the use of forwarding contracts so that prices are fixed and no volatility is associated with the expected price. Next, the use of options contracts are utilized to manage adverse price risks.

Before quantifying the amount of production risk in cattle feeding, we must first understand the origins of this risk. We identify two main areas of production risk: performance and health. It is true that these two areas are highly correlated, but measure separate outcomes. To evaluate production risk, we once again utilize the feedlot data set from the previous essays. Additionally, data from the Chicago Mercantile Exchange (CME) and Chicago Board of Trade (CBOT) are used for price information.

First, cattle performance measures the ability of feedlot cattle to gain weight and can be thought of as similar to crop yield measure. Two measures are used to gauge performance which include Average Daily Gain (ADG) and Dry Matter Feed Conversion (DMFC). ADG measures the average pounds of weight gain the pen gains on a daily basis. Feeding rations can be adjusted to keep this measure near a normal level, which is usually above 3 pounds per head, per day. DMFC is measured as the amount of dry feed needed for a head of cattle to gain one pound. These two measures are negatively correlated and together describe the performance of a pen of cattle. This performance can vary due to extreme weather, unobservable genetics, or management. Variation in ADG directly impacts ending weight, which is a direct input into revenues. DMFC has a large impact on feed costs, since a high feed conversion rate will force the feedlot operator to use more feed in order to attain the desired weight results.

To characterize the health of a pen of cattle, veterinary costs per head (VCPH) and mortality rate (MORT) are used. These variables vary for many of the same reasons that performance varies and can be highly correlated. Disease can cause these variables to jump quickly as is demonstrated by the positively-skewed nature of these variables. Both variables show a significant amount of skewness as they can rise quickly when disease spreads or the pen has poor health. While VCPH is an independent cost that influences profits, MORT directly impacts the bottom line since cattle that perish during the feeding period are not sold, but have already been purchased as feeder cattle.

The next few sections will focus on different scenarios that can impact cattle feeder profits. First, a forward-pricing scenario allows a cattle owner to hedge by locking in on the expected prices of fed cattle and corn early in the feeding process. This section will be followed by a sensitivity analysis of these results when allowing expected values of the mean and volatility of prices to vary in high and low scenarios. Additionally, production risk will be quantified where all price risk is managed, yet production risk variables are adjusted in order to evaluate their relative impact on expected profits. The final section will simulate expected profits when the options market is used to hedge downward risk.

### 4.4.1 Forward-pricing contract

In order to quantify the amount of risk a cattle owner confronts when placing their pen on feed, we use previously developed simulation methods from chapter 2 to characterize *ex-ante* profit risks, conditional on entry level pen characteristics. These density functions characterize variability in profits under different scenarios. We have four scenarios of interest in this section, which are shown in Table 4.1. Each scenario is analyzed through the use of a forward-pricing contract as well as a contract in the options market. Here we illustrate the use of each scenario in the case of forward-pricing. The first scenario assumes a forwarding contract made on the price of fed cattle to be made at the end of the feeding period. According to RTI (2007), an average of 17.3% of cattle sold by "large" producers used forward contracts, while "small" producers used forward contracts only 3.0%. By locking in this price, the cattle owner essentially eliminates any price risk associated with fed cattle prices. The owner is no longer subject to downward price risk, however, also surrenders the likelihood of upward price risk. It is hypothesized that this forward contract will eliminate most profit risk. The second scenario assumes a forward-pricing contract made on the price of corn. This is of particular interest at this time, when the volatility for corn prices have increased with the additional demand for corn from federal mandates of increased ethanol production. The third scenario assumes forward-pricing contracts on both corn and fed cattle prices. This type of contract would likely include two separate contracts with two sellers. More importantly, for our purposes, it characterizes a situation where a cattle owner has eliminated all price risks and is left with only production risks arising from health and performance measures. The fourth scenario is a control scenario where the cattle owner assumes all risks arising from both prices and production.

Table 4.1: Risk management scenarios in this study

| Scenario | Description | Forward-pricing | Options market |
|---|---|---|---|
| 1 | Cattle Protection | Live Cattle | Put Option on Live Cattle |
| 2 | Corn Protection | Corn | Call Option on Corn |
| 3 | Full Price Protection | Both | Both |
| 4 | No Price Protection | None | None |

The performance and health measures for each of these scenarios will be based on a hypothetical pen, placed on July 16, 2007. The steer pen is placed into a Kansas feedlot with 150 head, with an average entry weight of 750 pounds. The anticipated marketing date is in 130 days (November 23, 2007). The price of the corn forwarding price is assumed to be equal to the September 2007 futures settlement of 3.346 per bushel, plus a basis adjustment. This price is intentionally placed at a midway point of the feeding period to measure an average price.

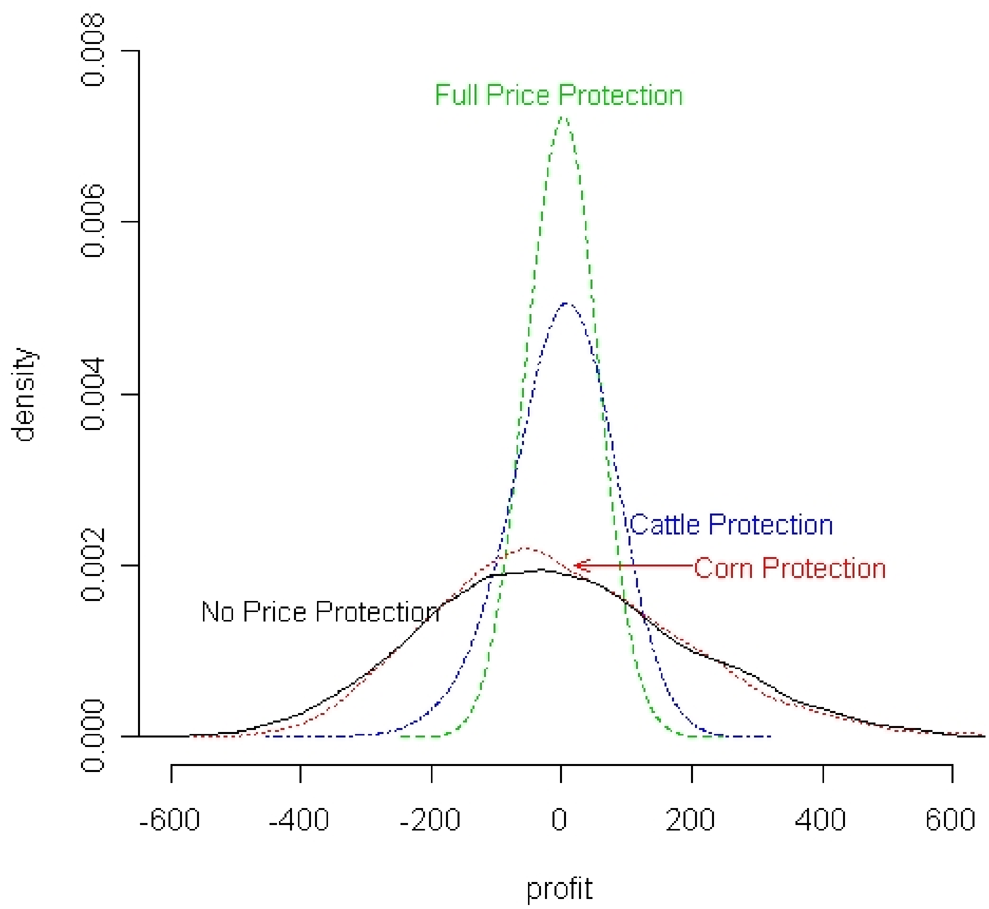
In order to simulate scenarios where forward-pricing contracts are not made, an implied volatility associated with each futures price needs to be computed. The implied volatility associated with a strike price of \$3.50 for a corn call option expiring in September was 30.59% and is computed based on the generalized Black-Scholes formulation. While this analysis isn't explicitly focused on price variability, a sensitivity analysis might be worthwhile to evaluate the impacts from different types of swings in prices. There is particular interest concerning corn prices over the past couple of years, which have shown an incredible amount of volatility.

The price of the fed cattle forward contract price is assumed to be equal to the December 2007 futures settlement of 97.675 per 100 pounds. The fed cattle price is meant to emulate the expected price near the marketing date. Once again, an implied volatility measure was calculated for a live cattle put option with a target price of \$92 that expires in December. The implied volatility resulting from this price was 16.42%. The information obtained from the options market will be used in the next set of exercises that employ the options market to help manage risk.

Simulations were conducted based on the previously mentioned information to emulate the four different scenarios discussed previously. To illustrate the results, a density function is used to characterize the *ex-ante* profit risk in each case, and can be found in Figure 4.1.

Probability density functions offer a flexible way to illustrate expected profits for cattle feeding enterprises, where we can visually inspect differences in the mean, variability, and skewness under the four highlighted scenarios. It is no surprise that when we eliminate price risks that the variability in expected profits decrease. We can see this as the tails thin out and the cluster of mass towards the mean increases. Alternatively, for the scenario where no price risk measures are taken, we can see that cattle feeding becomes a very risky business. In this scenario, there is a mean of -$7.36 in profits per head and a standard deviation of 202.34. The tails of this density expand with a smaller mass near the mean, implying a higher probability of large losses or gains. Ninety percent of this density lies between the interval between -$324.61 and $345.33 as measured in profits per head. The wide spread in this density is largely driven by large implied volatilities in both fed cattle and corn prices and demonstrates the importance of managing risk in cattle feeding. To visually inspect the profit levels associated with different percentile levels, cumulative density function plots are shown in Figure 4.2.

The recent federal mandates and consumer demand for ethanol production have caused feedlot producers to pay a higher price for corn used to feed livestock. The expected prices are also associated with very high volatilities over the last year and a half, that moved upward of 40%. In this scenario, the implied volatility for corn was 31%, which is a quarter of the amount associated with this options contract from one month previous. One way of managing the risks involved with




| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---|---|---|---|---|---|---|
| No Price Risk | 0.19 | 49.30 | -81.60 | -33.18 | 33.47 | 81.65 |
| Cattle Protection | -0.74 | 77.34 | -133.94 | -49.15 | 52.80 | 118.36 |
| Corn Protection | -2.03 | 193.53 | -291.19 | -137.58 | 118.60 | 340.45 |
| No Protection | -7.36 | 202.34 | -324.61 | -151.62 | 124.61 | 345.33 |

Figure 4.1: Distribution of *ex-ante* conditional profits under four types of risk coverage

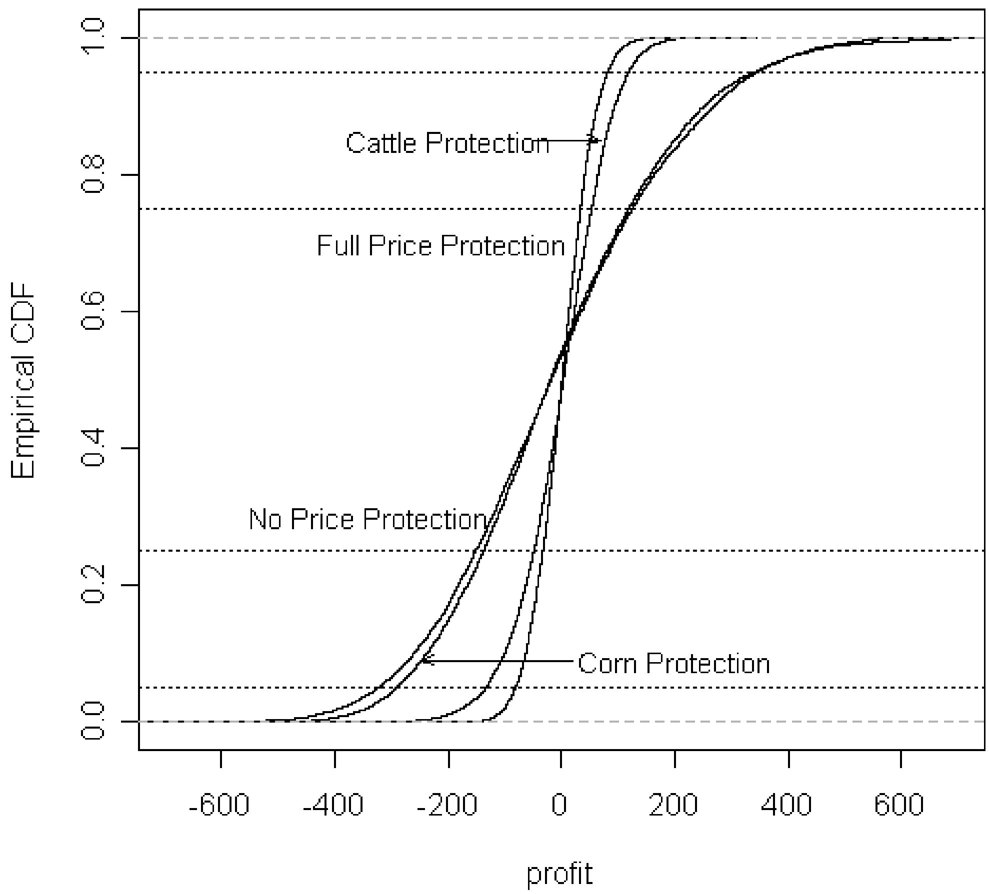



Figure 4.2: Cumulative Density of *ex-ante* conditional profits under four types of risk coverage

high corn volatility is to enter into a forward-pricing contract that assumes the futures settlement price of corn. This is scenario two, and is shown in the previously mentioned plot of densities. In spite of the high volatility for corn prices, fed cattle prices still drive much of the volatility, however, we do see a reduction in risk when this strategy is used. For example, the lower 5% of the density increased by $33, while the upper tail 95% is reduced by $15. However, in cases where the corn futures price and implied volatility measures are high, eliminating corn risk can be seen as a way to reduce more downward profit variability than upward profits. The reason for this is the positively skewed nature of prices, which are assumed to be distributed as a log-normal. Because of this positive skewness, there is a small probability that corn prices will go very high, which negatively impacts profits. While the mean is mostly unchanged and the the variation is slightly less, the lower end of the tail has been reduced.

While managing corn price risk may be useful, it still leaves quite a bit of variability in profits uninsured. Because cattle prices are thought to be the main contributor to profit risk, the next scenario includes a forward-pricing contract on fed cattle prices. By essentially guaranteeing a marketing price per weight, the cattle owner is able to eliminate fed cattle price risk. By doing this, both upward and downward risks associated with the price of fed cattle are eliminated, leading to an implied volatility in this scenario of zero. The owner is still left with uncertainty from corn prices and production activities. This type of contract reduces the standard deviation by 62% when compared to the case where all risks are included. Here, 90% of the density lies within the interval of -$133.94 to $118.36. Here the position of the .05 and .95 percentiles shrink by 59% and 66%, respectively. The average of this density is mostly unchanged at -$0.74.

While this scenario has shrunk the profit density tails substantially, there still remains a lot of variability. For example, one might interpret the above results to mean there is a 5% chance that one might expect to lose at least \$133.94 per head. Given our hypothetical pen of 150 head, this would amount to a loss of more than \$20,091.00 in profits. The remaining variability is likely a major driver behind the LGM product, as an alternative to the LRP product.

By protecting cattle owners against price risks originating from corn and fed cattle prices, the LGM product is the most comprehensive cattle insurance product currently available. The final scenario focuses on the risk that is left uninsured under this type of plan. Here, both prices are fixed through the use of forward contracts to eliminate any price volatility. The remaining variability occurs due to variability within the production process, conditional on the entry level characteristics of the cattle pen.

The lower 5th percentile is located at -\$81.60, while the upper 95th percentile lies at \$81.65. As expected, this is a substantial difference from the prior scenarios that left some element of price risk uninsured. Even though this scenario presents substantially less risk, it is important to understand the amount of risk that is still present. The 5th percentile implies that there is a 5% probability of a pen of 150 head to lose at least \$12,240.

The large variability resulting from productive variability comes the previously identified sources. Health measures, such as veterinary costs and mortality rates, are each positively skewed and can take large values in pens that are sick or experience adverse weather conditions. Mortality rates that rise above 5% can very quickly make positive profits hard to realize. Performance measures, such as feed conversion rates and daily gain rates, can vary based on unobservable genetic potential or weather conditions. While daily gain rates can often be managed by feeding rations, feed conversion rates move upward quickly when pens are not in optimal form. Feed conversion rates are also positively skewed. While much of the variability is reduced through forward pricing strategies that eliminate price risk, a significant amount of variability is left over.

In the next section, a sensitivity analysis will be performed in order to identify the sensitivity of our results to variability in price and production risk factors. This allow us to make more general comments about the risks that face cattle feeding profits as well as their origins.

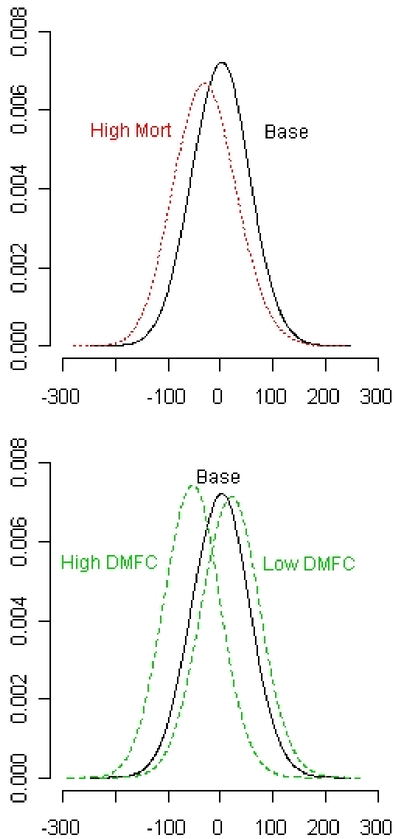
### 4.4.2 Sensitivity analysis

This section will focus on evaluating the sensitivity of our results by allowing for a low and high range of values pertaining to both price and production risk factors. In changing the mean values that characterize risk in our simulated results, we can evaluate the impact on overall profits per head, while holding all other factors constant. The mean of the distribution is shifted to the 5th and 95th percentile levels in order to evaluate upper and lower shocks to the distribution.

First, we will focus our attention on shifts in production risk factors under full price coverage, meaning a forward contract on both corn and fed cattle prices. With price risk set to zero in this case, we are able to evaluate the range in profits that result from low and high shocks in production risk factors. Under these scenarios, expected values for production risk factors were adjusted upward and downward to examine the impact on profits per head when these risk factors take on extreme values. The results can be found in Figure 4.3.

In the data used in this study, 5 percent of the observations contained mortality rates of at least 3.41%. To evaluate the impact a head loss of this magnitude would have on overall profits, simulations were conducted with a mean of 3.41% death loss. The result is a shift in the per head profit distribution, where the mean is reduced by $29 over the baseline case. It is no surprise that such a large mortality loss in the pen would result in a significant decrease in profits.

Next average daily gain is adjusted in both directions to develop a range of




| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---:|---:|---:|---:|---:|---:|---:|
| Baseline - No Price Risk | 0.19 | 49.30 | -81.60 | -33.18 | 33.47 | 81.65 |
| High MORT (3.41%) | -28.99 | 53.88 | -115.29 | -66.07 | 6.49 | 63.02 |
| Low ADG (2.56 lbs/day) | -73.78 | 44.15 | -146.15 | -103.79 | -44.27 | -0.86 |
| High ADG (4.14 lbs/day) | 32.95 | 51.61 | -53.03 | -1.93 | 67.88 | 117.90 |
| Low FC (5.26 lbs feed/lbs gain) | 18.73 | 49.96 | -64.43 | -14.97 | 52.40 | 101.13 |
| High FC (7.43 lbs feed/lbs gain) | -54.49 | 47.51 | -132.93 | -86.45 | -22.76 | 24.30 |

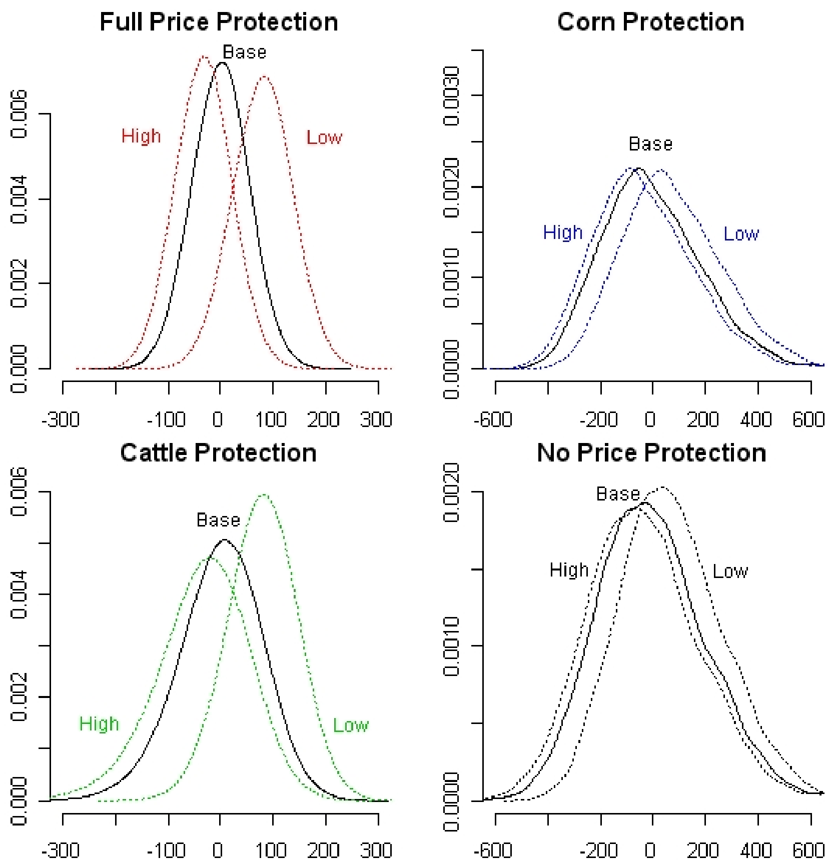
Figure 4.3: Distribution of *ex-ante* conditional profits from shocks to production risk factors, under full price coverage

profits that is based solely on shifting this variable. The impacts are dramatic as the mean shifts downward to -\$73.78 in the low scenario and \$32.95 in the high scenario. Average daily gain directly impacts the amount of weight that is available for sale at the end of the feed period. ADG rates of 2.56 and 4.14 demonstrate the lower and upper bounds of feeding efficiency, respectively. Lastly, feeding efficiency is also adjusted in both directions to contrast an efficient feeder with a conversion rate of 5.26, with that of an incredibly inefficient feeder at 7.43. While 90 percent of the data fall within this range, it once again offers a range in which profits will vary based solely on feed conversion rates.

Feed conversion rates directly impact the amount of corn that is used to add weight, meaning an inefficient feeder will cause feed costs to increase quickly. Once again there are dramatic effects as a pen that is expected to have a high feed conversion rate will have a profit density function that crosses the zero profit threshold just above the 5% level. Alternatively, the highly efficient pen has more than a 70% chance of positive profits. It is also worth mentioning that these results might be even more exaggerated with higher corn prices.

The next set of scenarios evaluate the sensitivity of our results to changes in corn prices and volatility in corn prices. Corn prices and volatilities have changed dramatically over the last few years. The main reasons include mandates for alternative fuel which have increased the demand for the commodity, as well as increases in transportation costs. The last few years have been characterized by extreme highs in cash prices as well as increased volatility. First, we focus on the changes in cash prices by assuming the high and low cash settlement prices, based on the 5th and 95th percentile price levels since 2005, for corn at \$3.93 and \$1.94 per bushel, respectively. The results can be found in Figure 4.4

In each scenario, one detects a strong shift at all levels with changes in the corn price. With corn still a main component of feed in livestock, this is not




| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---|---|---|---|---|---|---|
| Low Scenario (1.935 per bu.) | | | | | | |
| No Price Risk | 80.01 | 52.27 | -6.83 | 44.72 | 115.46 | 165.49 |
| Cattle Protection | 79.47 | 62.54 | -27.64 | 38.20 | 122.27 | 179.52 |
| Corn Protection | 77.79 | 194.34 | -212.55 | -58.02 | 198.62 | 421.78 |
| No Protection | 77.26 | 202.05 | -229.46 | -64.27 | 203.83 | 432.99 |
| High Scenario (3.9325 per bu.) | | | | | | |
| No Price Risk | -32.98 | 48.19 | -112.64 | -65.59 | -0.66 | 47.01 |
| Cattle Protection | -34.08 | 85.06 | -182.17 | -85.44 | 24.93 | 95.75 |
| Corn Protection | -35.20 | 193.24 | -324.75 | -170.24 | 85.27 | 307.60 |
| No Protection | -36.30 | 214.63 | -364.23 | -184.62 | 97.87 | 330.49 |

Figure 4.4: Distribution of *ex-ante* conditional profits from shocks to expected corn prices

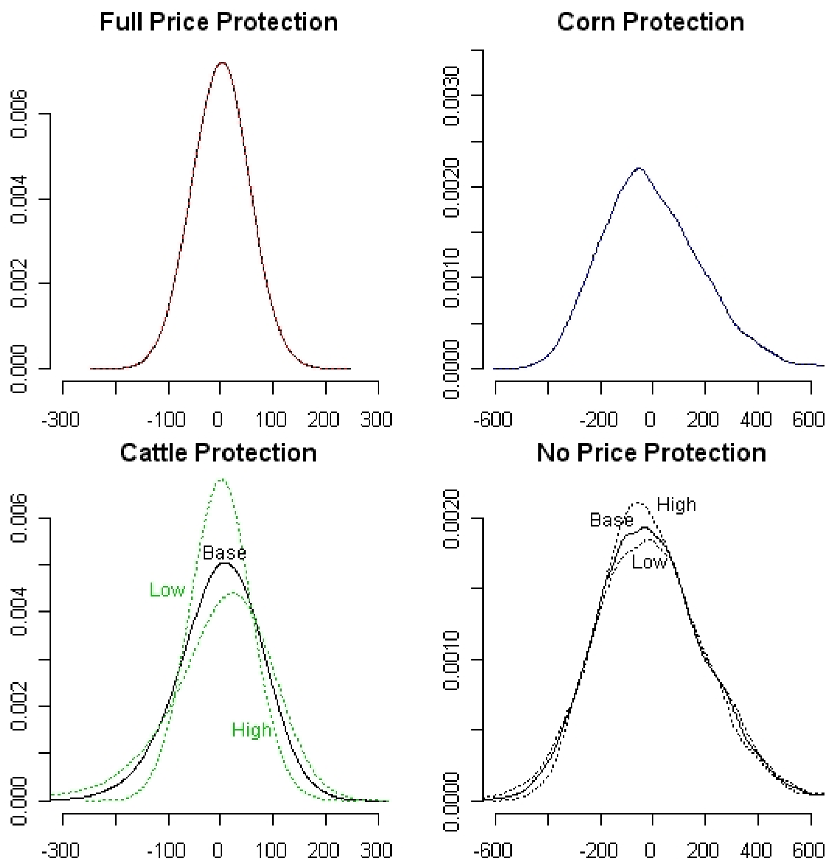
surprising. With mean profits jumping to nearly $90 in all scenarios concerning the low corn price of $1.94 per bushel, profits are strongly impacted by dramatic shifts in the expected price of corn. Alternatively, a high price of $3.93 per bushel strongly decreases profits to about-$35 per head, which is a dramatic change from scenario with the low corn price.[4]

Another reality of cattle feeders today, beyond the high corn prices, are the higher than normal implied volatility associated with options contracts for corn. To illustrate the impact wide swings in corn volatility has on cattle feeding profits, we use the same simulations as before with high and low volatility scenarios of 42.2% and 10.45%, respectively. In the previous cases we focused our attention mostly on the mean, as the adjustments mostly caused shifts in the density functions without changing variability too dramatically. However, as shown in Figure 4.5, the tails of the profit distributions fluctuate quite dramatically.

For example, when no forward pricing contracts are used, the lower bound of the 90% confidence interval changes from -$347.84 under the high scenario and -$298.47 in the low scenario. A difference of $50 that results merely from a change in the volatility. Most of the mean elements are mostly unchanged in these scenarios. Two scenarios that eliminate all corn price risk will be identical in these scenarios, which include corn protection and no price risk scenarios. It is interesting to note that even when fed cattle prices are hedged, the 90% interval still changes quite dramatically between the two scenarios in the tails.

In addition to corn price risk, fed cattle prices can also have very large impacts on profits from changes in level and volatility. In an *ex-ante* evaluation of profits, if we were to shock fed cattle prices upward and downward, we would need to similiarly shock feeder prices up and down as well. Because of this, results from shocking prices are less dramatic that previously seen, since really the margin

[4]High and low corn price scenarios were based on cash price information from 2000 - 2007.




| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---|---|---|---|---|---|---|
| Low Scenario (vol = 0.1045) | | | | | | |
| No Price Risk | 0.19 | 49.30 | -81.60 | -33.18 | 33.47 | 81.65 |
| Cattle Protection | -0.10 | 53.08 | -87.61 | -35.83 | 35.33 | 87.08 |
| Corn Protection | -2.03 | 193.53 | -291.19 | -137.58 | 118.60 | 340.45 |
| No Protection | -2.32 | 197.35 | -298.47 | -140.35 | 120.49 | 347.37 |
| High Scenario (vol = 0.4220) | | | | | | |
| No Price Risk | 0.19 | 49.30 | -81.60 | -33.18 | 33.47 | 81.65 |
| Cattle Protection | -1.15 | 97.60 | -177.22 | -54.68 | 65.98 | 138.74 |
| Corn Protection | -2.03 | 193.53 | -291.19 | -137.58 | 118.60 | 340.45 |
| No Protection | -3.37 | 221.42 | -347.84 | -155.33 | 135.63 | 370.76 |

Figure 4.5: Distribution of *ex-ante* conditional profits from shocks to expected corn price volatility

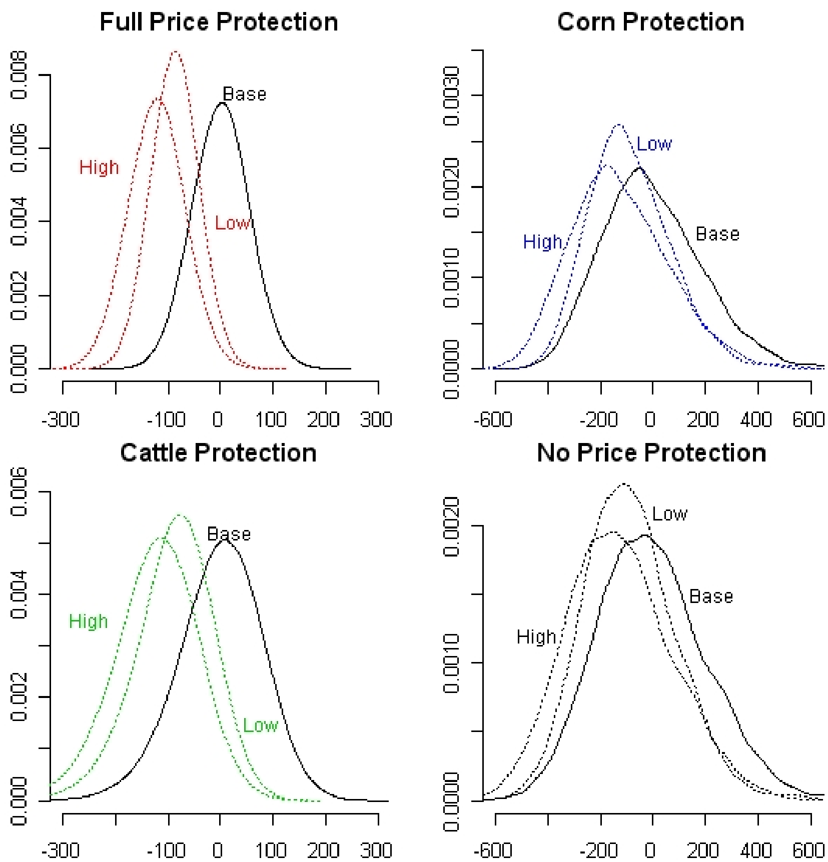
between feeder and fed prices is where profits are made and lost. Figure 4.6 shows the results from these scenarios where feeder and fed cattle prices were brought to \$98.32 and \$79.12 in the low scenario and \$128.38 and \$96.01 in the high scenario, respectively.

The large spread between fed and feeder prices in the above scenario keep the men of profits strongly negative. One notable feature of cattle prices is that volatility has dramatic impacts on overall profits per head (Belasco et al., 2006). To test this hypothesis, we conducted to same sensitivity test to cattle price volatility. Since feeder prices are known at the time of placement, this only impacts the distribution of expected fed cattle prices. These results can be found in Figure 4.7.

In spite of the range around cattle price volatility being somewhat less that corn price volatility, it still has a rather dramatic impact on the variance around profits. Again, we notice with volatility shocks that the mean values stay relatively unchanged, with notable differences in the tails. For example, with an implied volatility of 28% in fed cattle prices the lower 5% of the distribution decreases by \$250, while the upper 5% increases by almost \$400 when moving from an implied volatility of 12% to 28%. This dramatically shows why fed cattle prices have such an impact and are the focus of so many current risk management tools.

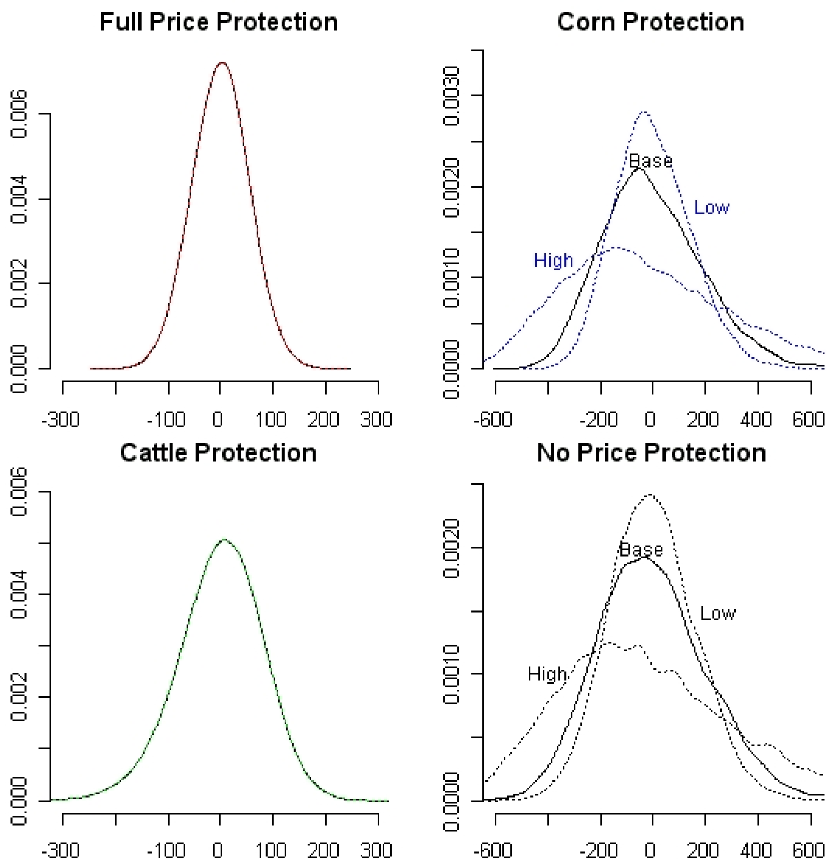
So far we have demonstrated how shocks to risk factors can impact the distribution of profits. For the final sensitivity analysis, weight is shocked so that we can see the impact on profits. Because placement weight is so closely tied to days on feed, days on feed is adjusted to reflected the average days on feed associated with the given weight. The high and low scenarios for weight will refer to average placement weights of 582 pounds and 872 pounds, which corresponds to 177 and 95 days on feed, respectively. The results for this sensitivity analysis can be found in Figure 4.8.

There appears to be a strong incentive to bring in lighter pens under this test.



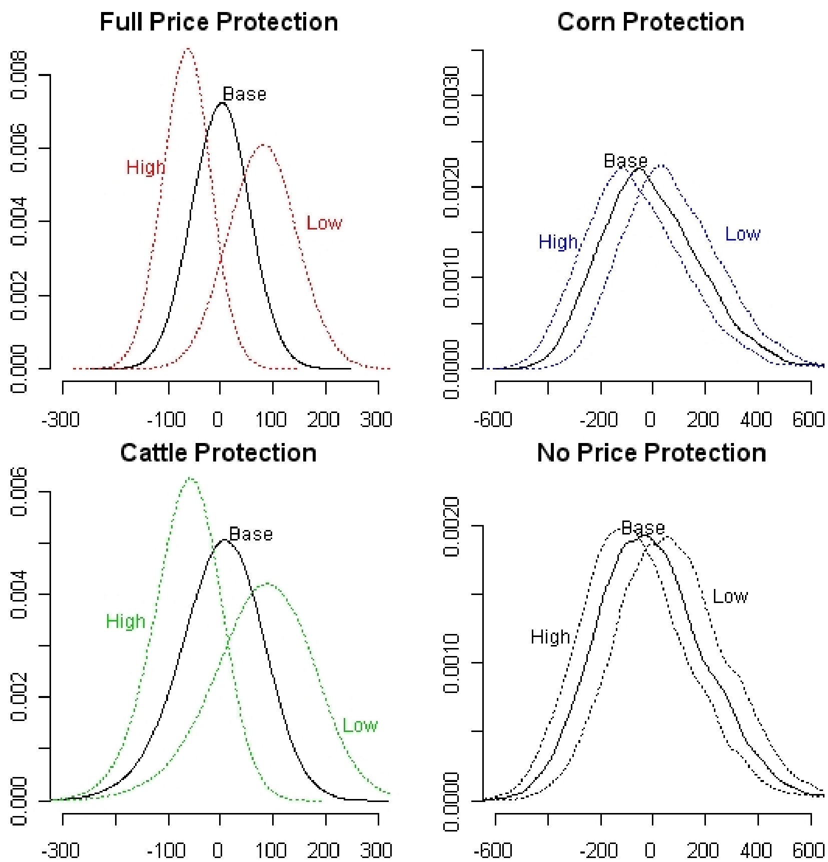

| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---|---|---|---|---|---|---|
| Low Scenario (98.32(fc) and 79.12(lc) per cwt) | | | | | | |
| No Price Risk | -87.48 | 38.96 | -152.04 | -113.72 | -61.38 | -22.86 |
| Cattle Protection | -88.41 | 71.24 | -213.57 | -131.26 | -39.21 | 19.87 |
| Corn Protection | -89.28 | 156.50 | -323.40 | -198.62 | 8.01 | 188.09 |
| No Protection | -90.21 | 175.15 | -359.50 | -211.57 | 20.07 | 208.99 |
| High Scenario (128.38(fc) and 96.01(lc) per cwt) | | | | | | |
| No Price Risk | -122.43 | 48.37 | -202.70 | -155.17 | -89.83 | -42.43 |
| Cattle Protection | -123.36 | 76.75 | -255.55 | -171.04 | -70.16 | -5.38 |
| Corn Protection | -124.61 | 190.21 | -409.00 | -257.76 | -6.02 | 212.15 |
| No Protection | -125.54 | 207.17 | -442.55 | -271.22 | 3.78 | 230.66 |

Figure 4.6: Distribution of *ex-ante* conditional profits from shocks to expected cattle prices




| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---|---|---|---|---|---|---|
| Low Scenario (vol = 0.12) | | | | | | |
| No Price Risk | 0.19 | 49.30 | -81.60 | -33.18 | 33.47 | 81.65 |
| Cattle Protection | -0.74 | 77.34 | -133.94 | -49.15 | 52.80 | 118.36 |
| Corn Protection | -1.40 | 144.99 | -223.39 | -102.68 | 91.28 | 248.41 |
| No Protection | -2.33 | 164.17 | -260.85 | -116.89 | 105.12 | 273.99 |
| High Scenario (vol = 0.28) | | | | | | |
| No Price Risk | 0.19 | 49.30 | -81.60 | -33.18 | 33.47 | 81.65 |
| Cattle Protection | -0.74 | 77.34 | -133.94 | -49.15 | 52.80 | 118.36 |
| Corn Protection | -4.07 | 350.58 | -479.09 | -251.12 | 189.07 | 634.89 |
| No Protection | -5.00 | 363.77 | -509.13 | -262.22 | 199.16 | 659.65 |

Figure 4.7: Distribution of *ex-ante* conditional profits from shocks to expected cattle price volatility




| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---|---|---|---|---|---|---|
| Low Scenario (weight=581, dof=177) | | | | | | |
| No Price Risk | 79.75 | 60.21 | -18.95 | 38.74 | 120.35 | 179.67 |
| Cattle Protection | 78.62 | 94.02 | -83.17 | 19.87 | 143.69 | 223.12 |
| Corn Protection | 77.62 | 189.14 | -204.22 | -55.86 | 195.76 | 410.24 |
| No Protection | 76.48 | 211.56 | -247.32 | -71.58 | 209.67 | 438.83 |
| High Scenario (weight=872, dof=95) | | | | | | |
| No Price Risk | -65.19 | 38.83 | -130.57 | -90.81 | -39.14 | -1.97 |
| Cattle Protection | -65.91 | 60.60 | -170.76 | -103.57 | -23.92 | 27.18 |
| Corn Protection | -67.42 | 193.01 | -356.38 | -202.51 | 51.99 | 274.96 |
| No Protection | -68.15 | 204.88 | -380.15 | -210.74 | 58.78 | 289.33 |

Figure 4.8: Distribution of *ex-ante* conditional profits from shocks to average placement weight

Even though feeder prices are lower for heaver weighted pens on a per pound basis, the value added doesn't get over this cost hurdle. However, it should be noted that this model assumes linear weight gain in pens of cattle, which is not necessarily true. Making inference on weight distinctions should be reserved until further analysis when a non-linear growth curve is assumed. However, by conditioning on weight, we do assume higher feed conversion rates which adds to per costs associated with feed. On average, we are able to capture growth linearly, but to be more specific, a non-linear growth path would be more accurate. For example, as beef cows get near their finishing weight, the do not put on weight as quickly as when they were younger. They begin to converge to a physical capacity before they are finished.

This section focused on conducting a sensitivity analysis in order to gauge the sensitivity of our results to changes in prices, which change over time, and production factors, which change for different characteristics. The implications of this research point to the use of flexible modeling mechanisms, like the one used here, in order to capture the dynamic effects that occur for changes in price and yield risks. Another implication is the fact that both should be taken into account when evaluating profits. As shown in the first analysis, even when all price risk is taken out, risk from production areas such as mortality, feed conversion, and average daily gain, can shift the distribution of profits dramatically.

The next section focuses on a risk management technique that is used almost as frequently as forward contracting and includes the use of options contract to help hedge risk. These contracts will add a new dimension to this research by including upside risk into a strategy that still eliminates downside risk. The advantage to this type of strategy is that if allow for upside profit gains, while eliminating large losses.

### 4.4.3 Options contract

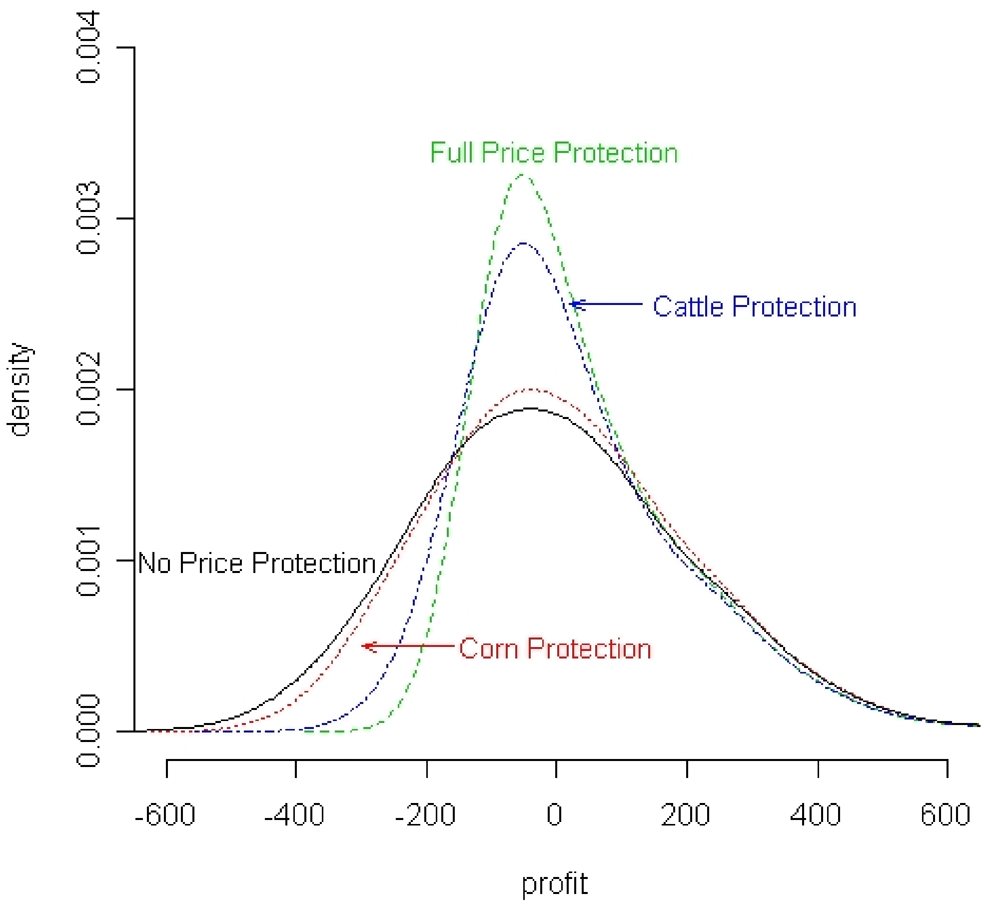
An alternative risk management strategy might include the use of the options market to manage price risk. As in the previous simulation, we will use the four scenarios of varying price coverage with a call option on corn and a put option on fed cattle prices. Simulations will be conducted in much the same way, with the exception of a bound on both prices that result from the options purchase. For corn price insurance, a call option offers the opportunity to purchase a futures contract at a specified price, protecting the buyer from corn prices above the strike price. Alternatively, for fed cattle price insurance, a put option offers the opportunity to sell at a specified price, protecting the buyer from falling fed cattle prices. While this strategy secures the contract holder from adverse price shocks, it leaves open the possibility of favorable shocks to prices.

The price paid for the options contract accounts for the premium associated with securing this risk as well as an assumed commission rate of \$.02 per bushel and \$.10 per 100 pounds for corn and fed cattle, respectively. Commission rates vary by commodity broker, but are relatively inexpensive compared to premium rates. The rates used in this study are found in Purcell and Koontz (1998). Expected profits from this set of simulations are shown in Figure 4.9. Since only downward variability is restricted, the left tails vary substantially based on the level of protection. The right tails converge, since an observation in that portion of the distribution experiences fortunate circumstances and likely do not utilize the protection established through options. For these observations, the only losses occur from the purchasing of options contracts though the premium rate. Premium rates are based on the end of day settlement prices from the CBOT and CME options contract information. For July 16, 2007, the settlement premiums for options on corn will be \$.091 per bushel with a strike price of \$3.50 and \$1.25 per hundred pounds of cut-weight with a strike price of \$92 for fed cattle. On average, the price of corn and live cattle coverage per head of cattle were \$6.19 and \$15.74, respectively, at the given prices. Given the cost of insurance, it was particularly surprising that the upper tails converged as quickly as shown in the given plots.

Because the lower tails diminish as the amount of protection increases, the mean shifts substantially. The lower tails diminish because of the insurance against adverse events, while the cost of insurance reduces the upper tail as protection increases. The mean shifts from -\$2.96 with no protection to \$39.33 with protection on both corn and fed cattle prices. The top two quantile levels illustrated in Figure 4.9 are approximately similar in all cases, which is quite different from the wide spread apparent in the lower two quantiles.

While the upper tails appear to converge rather quickly when using options as risk management tools, it should be mentioned that the majority of mass is stacked near the zero profits. For these components of the distributions, paying extra money for insurance may be the difference between turning a profit and being stuck with losses. In the upper part of the tail where large profits are realized, the relatively small amount of insurance appears to have an impact that is marginal at best. Under this scenario the lower tail under full price protection is much more negative than in the forward-pricing situation. More specifically, the 5% under full coverage lies at -\$138.42, which is almost twice as negative as under forward-pricing. The reason for this is that even though the contract holder is protected at a certain level, there still remains a small amount of downside risk in prices, whereas forward-pricing eliminates this risk. Under this scenario, one might still anticipate a 5% chance of losing over \$20 thousand for our hypothetical pen of 150 head, which is a loss of -10% of total selling value. Alternatively, one might expect to lose \$11,034.00 with 25% probability.

This section has simulated profits based on a hypothetical pen and allowed for both forward-pricing contracts and options contracts to be used for risk manage-




| Scenario | Mean | Sd | 5% | 25% | 75% | 95% |
|---|---|---|---|---|---|---|
| No Price Risk | 39.33 | 155.98 | -138.42 | -73.56 | 118.21 | 347.42 |
| Cattle Protection | 25.47 | 167.12 | -188.50 | -89.32 | 112.41 | 343.22 |
| Corn Protection | 10.91 | 200.09 | -287.76 | -130.44 | 134.19 | 362.92 |
| No Protection | -2.96 | 210.36 | -324.73 | -150.93 | 128.26 | 359.55 |

Figure 4.9: Distribution of *ex-ante* conditional profits under four types of risk coverage using options

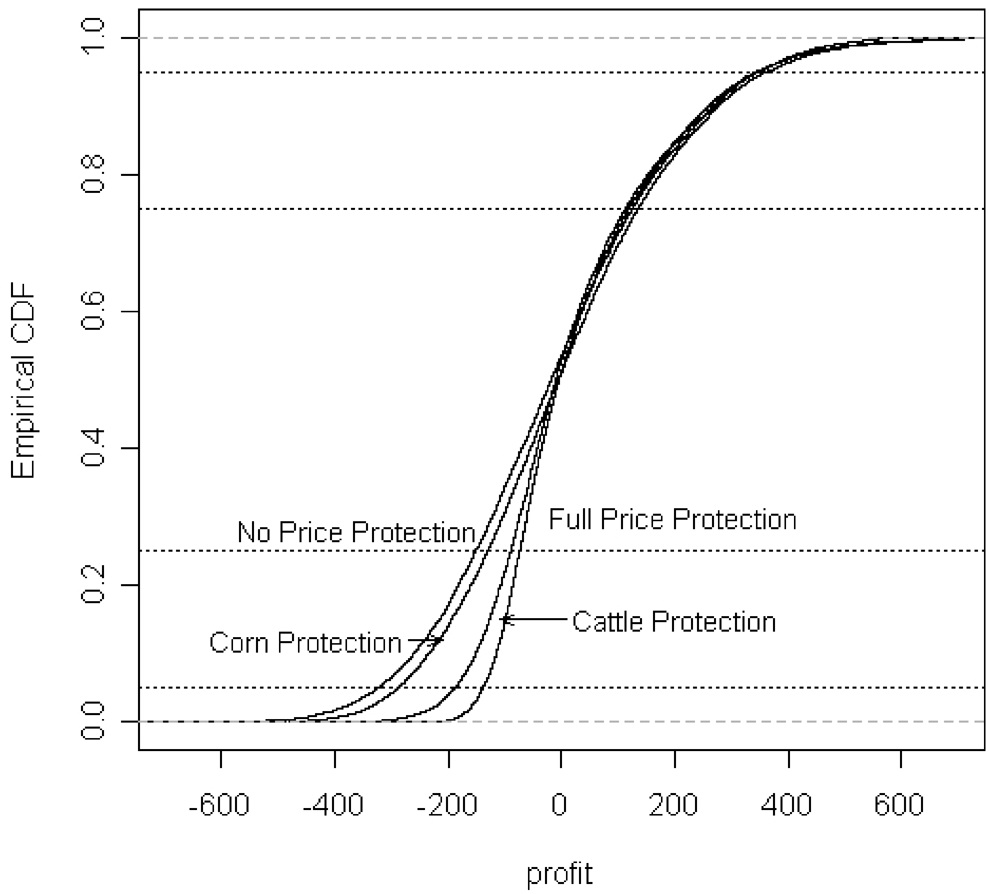



Figure 4.10: Cumulative Density of *ex-ante* conditional profits under four types of risk coverage using options

ment. Even under the scenario where both corn and live cattle prices are managed, there appears to be a substantial amount of downward risk to profits. The risk that is left after using risk management techniques can be mostly attributed to production risk. While insuring production risk is much more complicated than with prices, there does appear to be reason to consider such a product.

## 4.5 Concluding Comments

This research has focused on quantifying the amount of risk inherent in fed cattle production for different types of risk management strategies. Specific attention was focused on the degree of risk left uninsured as cattle owners purchase higher levels of insurance protection. Even though some current risk management strategies for managing price risk can eliminate a good portion of overall risk, a significant amount of risk still exists in production areas. In this study, we evaluate the use of forward-pricing contracts and options contracts as two risk management strategies. Currently, both allow cattle owners to eliminate a significant amount of risk. However, given the direction of crop insurance programs as well as the relatively new involvement in federal livestock insurance, the question remains as to whether a true livestock revenue insurance is appropriate. Major sources of uncertainty in this industry involve adverse weather, through drought and extremely cold weather. These are events not currently covered, outside of disaster aid relief, through federal insurance programs.

One limitation of this study is that it spans 5 distinct feedlot locations. Even though the feedlots are distinguished by binary variables that indicate the state of residency, some variability might be attributed to differences among the feedlots. Additionally, the five feedlots in this sample are not representative of the cattle feeding industry as a whole. For these reasons, a future study should include an evaluation of production risk by feedlot location. While controlling for placement characteristics should control for most of the covariate effects that control risk, a further study to evaluate the difference in variability amongst feedlots might produce interesting results. Currently, two methods are used in crop insurance settings, which include individual and county-level production history. Within each state, the feedlots are relatively close, making county-level aggregations a possibility. Future research should evaluate both alternatives and isolate any significant differences between the two methods.

The degree of comfort a cattle owner has, given this amount of risk, strongly depends on their aversion to risk. A future direction of work should include the evaluation of expected utility given differing levels of risk aversion. By evaluating a menu of different risk aversion coefficients may illustrate how different layers of protection may be more or less necessary to different types of participants in the cattle industry. For example, a speculator or investor might have a higher risk threshold and take on more risk, while a rancher would likely be more risk averse and prefer additional layers of coverage. An analysis of this type would also allow for simulations to approximate the likelihood of taking on production risk or demanding a product that protects against production risk.

Additionally, some future work is also needed to characterize production risk from an *ex-post* perspective. For example, an evaluation into the influence that weather has on performance and health factors may lead to a better understanding of the sources of risk throughout the production process. Additionally, weather outcomes are out of the producer's control and may be an appropriate variable to correlate with realized production outcomes. Miranda and Vedenov (2001) recommend the use of index-based derivatives in developing countries where production data are limited and administration costs must be kept low. In particular, weather-based indices can be designed to be highly correlated with productive outcomes, while not creating an incentive for the insured agents to change production efforts

(Turvey et al., 2002). Recently, a temperature-humidity index was proposed to protect against reduced milk from dairy production in times of extreme weather (Deng et al., 2007). This is the first attempt to utilize a weather-based index to characterize *ex-post* livestock productivity. The authors point out that since weather data is publicly available, the lack of information asymmetry work to minimize distortions from moral hazard or adverse selection. Future research should evaluate the impact of weather on beef cattle health and productivity.

# Chapter 5

# Conclusion

Agricultural production holds a unique position in the U.S. economy in the way that risk enters into production. Variability around prices and yields impose large uncertainties concerning profits. This relationship between risk and production is quite apparent in the area of cattle feeding. While large quantities of research have been devoted to understanding variability around crop yield risk, little research has focused on livestock. The 2000 Agricultural Risk Protection Act shifted the way risk is managed in livestock industries by encouraging federal insurance programs that cover livestock.

As a result of the mandate for new research evaluating livestock insurance, two major programs (LRP and LGM) have been developed and are currently expanding into new states. These programs, which are in line with most past research, have focused on guarding price risk in cattle feeding enterprises. In contrast to these studies, this research looks to evaluate cattle feeding profit variability by recognizing that risks originate in both price and yield areas.

When pens of cattle are transported to feedlots, where they will spend the next 3 - 5 months gaining incredible amounts of weight, financial plans must be made. In assessing the possibilities that may occur over the next few months, a cattle owner is left with an array of uncertainties. First, cattle prices can change immensely which can dramatically change the value of the pen and ultimately the realized profits. Because prices are determined based on global supply and demand, this is out of the control of the cattle owner. This is no more apparent than with the variability associated with corn prices, which is the second type of uncertainty. Corn prices largely impact the price of feed, since nearly 70% of feed used for beef cattle at the feedlot stage. Changes in the demand for corn in other areas, such as ethanol production, can dramatically impact the ability to make profits feeding cattle.

In addition to these two types of uncertainty, risk can also come from production yields, which in this industry include both performance and health measures. Concerning performance, the ability to gain weight and the efficiency with which weight is gained can also be thought of as random variables, particularly from an *ex-ante* perspective since unknown genetics or weather can play a strong role in determining performance outcomes. In addition, health and mortalities are major sources of mortality and can cause quick losses to profits in an industry where profits are already close to zero in most cases.

This research attempted to address some fundamental questions related assessing the risks related to fed cattle production. Chapter 2 specifically addressed the multivariate nature of the four elements that jointly affect production risk in cattle production. The first step in this process was to identify areas of *ex-ante* production risk, which includes average daily gain (ADG), dry matter feed conversion (DMFC), mortality rates (MORT), and veterinary costs (VCPH). Additionally, conditioning variable were identified that influence the mean and covariance between these four variables.

A multivariate Tobit model was used to capture the censored nature of mortality rates. Two complexities arose which demanded extending the traditional Tobit model; the individual variance associated with each performance risk variable changes for difference levels of the conditioning variables and the correlations between these variables change with the conditioning variables. For example, as placement weight increases the amount of variability around mortality rates decrease. This occurs as more mature beef cows are able to sustain through extreme winter storm or disease. Additionally, the correlation between mortality rates and veterinary costs weaken for higher placement weight levels. In order to take these effects into account, the Tobit model was extended to allow for all covariance elements to change with the conditioning variables. Likelihood ratio tests were used to examine the gains from adding the additional parameter estimates needed for our extended model and demonstrated the gains in efficiency. Further research in this area would be helpful in order to quantify the amount of bias to expected profits that result from imposing constant correlation between these variables.

In addition, conditioning variables significantly impacted the mean and covariance elements in most of the equations. The size of the data allows us to evaluate the impacts that variables, which are known when a pen is placed on feed, has on both the mean and covariance elements. These variables are useful in feedlot management, as well as an understanding of overall profit risk that accounts for all of the identified areas of risk. Additional impacts come from differences in gender, location, and season of placement.

Chapter 3 focused on modeling censored variables as an individual variable and as part of a system. This essay originated with the idea that mortalities may be generated as part of a two-step process, rather that a one-step as assumed by the Tobit model. Possible reasons for the two-step process fundamentally begin with the observation of homogeneity within a pen of cattle. Because a pen of cattle tend to come from the same producer and therefore have similar genetics programs, positive mortalities may be a sign of poor genetics inherent in the pen. In a more general way, variables that influence the level of positive mortality rates need not influence the probability that a non-zero amount of mortalities occur.

In order to test this hypothesis, a zero-inflated log-normal distribution is developed, using Bayesian theory, to evaluate mortality rates using an appropriate mixture model. A major advantage of this model is the flexibility in distributional assumptions, which is in contrast to the Tobit model. Additionally, past research has demonstrated the advantages of using Bayesian methods with censored data (Chib, 1992; Ghosh et al., 2006). First, data are simulated to examine the loss in efficiency from model fit and predictive power when assuming a Tobit model when the data is generated from a mixture model, and vice versa. The zero-inflated log-normal model performed extremely well when the data was generated from a mixture model and even had a better fit when the data was generated from a Tobit model with a high degree of censoring. The data demonstrated significant gains in model fit and predictive power in both the univariate and multivariate models that were developed and compared to the Tobit counterparts. In addition to added efficiency, variables are more accurately characterized. For example, location (KS) has no statistical impact on the mean of mortality rates when evaluated using the Tobit model. However, the zero-inflated log-normal distribution demonstrated a significantly negative relationship between KS and mean mortality rates, coupled with a positive relationship with KS and the positive amount of mortalities. The point here is that Kansas feedlots appear to be more successful in preventing any mortality rates, which may be a result of a strong backgrounding or vaccination program, while positive mortality rates result in higher rates. Higher positive mortality rates may be the result of more extreme weather or more contagious diseases in the Kansas feedlots. While the Tobit assumptions led to an insignificant variable estimate, this masked the true two-part relationship with that variable.

In addition to evaluating mortality rates, a multivariate model was developed and shown to perform more efficiently that the multivariate Tobit model. By recognizing the link between joint probabilities with conditional and marginal densities, a multivariate model is developed that is able to consistently regress more than three censored variable, which is a major limitation of frequentist methods.

Future research in this area is needed to better characterize mortality rates. Of the four variables evaluated in this study, mortality rates have the largest variance. One problem is the bi-modal and positively skewed empirical distribution, which might more closely be characterized with a log-normal distribution. One suggestion is to model this variable as a mixture of two log-normal distributions so that the bi-modal nature can be more accurately captured. Another suggestion is to utilize weather forecasts, that are known when the pen is placed, to evaluate the importance of weather forecasts in predicting mortality rates. If we are interested in an *ex-post* evaluation, there is an obvious link to using weather to evaluate the impact on all health and performance measures. For example, the development of an index that accounts for winter storms, heat, humidity, and other components that impact the production of cattle should be evaluated in order to understand this impact on production and ultimately on profits.

Profit risk is the focus of chapter 4, which evaluates this risk in the face of different shocks to production and prices. A major contribution of this chapter includes the quantification of profit risk under different price risk management strategies. Forward pricing and options contracts are used to manage profit risks under corn price protection, fed cattle price protection, both, and neither. Under full price protection, we are able to see that a significant portion of risk still exists simply from production risk components. An extension of this work might include a direct quantification of the proportion of variability that is eliminated with prices are controlled for. However, for the purposes of this research production risk does appear to be a significant amount. This research demonstrates that cattle feeding profits can be drastically impacted by shocks coming from many different areas.

A sensitivity analysis was conducted that allows for a visual representation of the impact that extreme shocks in prices and production have on overall expected profits. The purpose of this analysis was to provide a range of values associated with profit distributions that can occur given two opposing extremes.

This evaluation of profit risk under different scenarios is an area of research that needs more attention. For example, a quantification into the amount of *ex-post* profit risk that is accounted for when weather is accounted for would add a quite a bit to existing research. The major advantage to using weather to gauge production risk is that in an insurance perspective, it minimizes moral hazard and is widely available to the public. Future research evaluating different opportunities to offer insurance products against overall profit risk in cattle production appears to be an important area of research, given the current limitations in offerings.

# Bibliography


Akaike, H. (1974) 'A new look at the statistical model identification.' *IEEE Transactions on Automatic Control* 19, 716–723

Amemiya, T. (1973) 'Regression analysis when the dependent variable is truncated normal.' *Econometrica* 41(6), 997–1016

_ (1974) 'Multivariate regression and simultaneous equation models when the dependent variables are truncated normal.' *Econometrica* 42(6), 999–1012

_ (1984) 'Tobit models: A survey.' *Journal of Econometrics* 24, 3–61

Anderson, J. D., and J. N. Trapp (2000) 'The dynamics of feeder cattle market responses to corn price change.' *Journal of Agricultural and Applied Economics* 32(3), 493–505

Arabmazar, A., and P. Schmidt (1982) 'An investigation of the robustness of the tobit estimator to non-normality.' *Econometrica* 50(4), 1055–1064

Atwood, J., S. Shaik, and M. Watts (2003) 'Are crop yields normally distributed? a reexamination.' *American Journal of Agricultural Economics* 85(4), 888–901

Babcock, B. A. (2005). Statement Before the United States Senate Committee on Agriculture, Nutrition and Forestry. Washington, DC, June 28, 2005

Belasco, E. J., M. R. Taylor, B. K. Goodwin, and T. C. Schroeder (2006) 'Probabilistic models of yield, price, and revenue risk for fed cattle production.' Selected Paper. American Agricultural Economics Association Annual Meeting, Long Beach

Bera, A. K., C. M. Jarque, and L. F. Lee (1984) 'Testing the normality assumption in limited dependent variable models.' *International Economic Review* 25(3), 563–578

Bernstein, S. (1917) *Calcul des probabilities*

Buccola, S. T. (1980) 'An approach to the analysis of feeder cattle price differentials.' *American Journal of Agricultural Economics* 85(2), 291–304

Cameron, A. C., and P. K. Trivedi (2005) *Microeconometrics: Methods and Applications* (Cambridge University Press)

Chavas, J. P., and K. Kim (2004) 'A heteroskedastic multivariate tobit analysis of price dynamics in the presence of price floors.' *American Journal of Agricultural Economics* 86(3), 576–593

Chib, S. (1992) 'Bayes inference in the tobit censored regression model.' *Journal of Econometrics* 51, 79–99

Chib, S., and E. Greenberg (1996) 'Markov chain monte carlo simulation methods in econometric.' *Econometric Theory* 12(3), 409–431

Coble, K. H., T. O. Knight, R. D. Pope, and J. R. Williams (1996) 'Modeling farm level crop insurance demand with panel data.' *American Journal of Agricultural Economics* 78, 439–447

Cornick, J., T. L. Cox, and B. W. Gould (1994) 'Fluid milk purchase: A multivariate tobit analysis.' *American Journal of Agricultural Economics* 76, 74–82

Cragg, J. G. (1971) 'Some statistical models for limited dependent variables with application to the demand for durable goods.' *Econometrica* 39(5), 829–844

Deng, X., B. J. Barrett, D. V. Vedenov, and J. W. West (2007) 'Hedging dairy production losses using weather-based index insurance.' *Agricultural Economics* 36, 271–280

Eiswerth, M. E., and J. S. Shonkwiler (2006) 'Examining post-wildfire reseeding on arid rangeland: A multivariate tobit modeling approach.' *Econological Modeling* 192, 286–298

Fackler, P. L. (1989) 'A note on alternative market and governmental risk transference mechanisms.' *Southern Journal of Agricultural Economics* pp. 203–209

Fausti, S. W., and B. A. Qasmi (2002) 'Does the producer have an incentive to sell fed cattle on a grid.' *International Food and Agribusiness Management Review* 5, 23–39

Fausti, S. W., and D. M. Feuz (1995) 'Production uncertainty and factor price disparity in the slaughter cattle market: Theory and evidence.' *American Journal of Agricultural Economics* 77, 533–540

FCIC (2007) 'Crop statistics for 2006: Nationwide summary - by insurance plan'

Gallagher, P. (1987) 'U.s. soybean yields: Estimation and forecasting with nonsymmetric disturbance.' *American Journal of Agricultural Economics* 69, 798–803

Ghosh, S. K., P. Mukhopadhyay, and J. C. Lu (2006) 'Bayesian analysis of zero-inflated regression models.' *Journal of Statistical Planning and Inference* 136, 1360–1375

Glauber, J. W. (2004) 'Crop insurance reconsidered.' *American Journal of Agricultural Economics* 86, 1179–1195

Glock, R. D., and B. D. DeGroot (1998) 'Sudden death of feedlot cattle.' *Journal of Animal Science* 76, 315–319

Goodwin, B. K., and A. P. Ker (2002) 'Modeling price and yield risk.' In *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture,* ed. R. E. Just and R. D. Pope (Boston: Kluwer Academic Press) pp. 289–323

Goodwin, B. K., and V. H. Smith (1995) *The Economics of Crop Insurance and Disaster Aid* (The AEI Press)

Goodwin, B. K., M. L. Vandeveer, and J. L. Deal (2004) 'An empirical analysis of acreage effects of participation in the federal crop insurance program.' *American Journal of Agricultural Economics* 86, 1058–1077

Greene, W. (2003) *Econometric Analysis,* 5th ed. (Pearson Education)

Greene, W. H. (1981) 'On the asymptotic bias of the ordinary least squares estimator of the tobit model.' *Econometrica* 49(2), 505–513

Gurmu, S., and P. K. Trivedi (1996) 'Excess zeros in count models for recreation trips.' *Journal of Business and Economic Statistics* 14(4), 469–477

Hall, D. B. (2000) 'Zero-inflated poisson and binomial regression with random effects: A case study.' *Biometrics* 56(4), 1030–1039

Hall, D. C., T. O. Knight, K. H. Coble, A. E. Baquet, and G. F. Patrick (2003) 'Analysis of beef producers' risk management perceptions and desire for further risk management education.' *Review of Agricultural Economics* 25(2), 430–448

Hart, C. (2006) 'The current state of u.s. federally supported livestock insurance.' In *The Economics of Livestock Disease Insurance,* ed. S. R. Koontz, D. L. Hoag, D. D. Thilmany, J. W. Green, and J. L. Grannis (CABI Publishing, Massachusetts) pp. 164–180

Hart, C. E., B. A. Babcock, and D. J. Hayes (2001) 'Livestock revenue insurance.' *The Journal of Futures Markets* 21(6), 553–580

Hennessy, D. A., B. A. Babcock, and D. J. Hayes (1997) 'Budgetary and producer welfare effects of revenue insurance.' *American Journal of Agricultural Economics* 79, 1024–1034

Jones, A. M. (1989) 'A double-hurdle model of cigarette consumption.' *Journal of Applied Econometrics* 4(1), 23–39

Just, R. E., and Q. Weninger (1999) 'Are crop yields normally distributed?' *American Journal of Agricultural Economics* 81, 287–304

Ker, A. P., and K. Coble (2003) 'Modeling conditional yield densities.' *American Journal of Agricultural Economics* 85(2), 291–304

Knoeber, C. R., and W. N. Thurman (1995) 'Don't count your chickens: Risk and risk shifting in the broiler industry.' *American Journal of Agricultural Economics* 77, 486–496

Lambert, D. (1992) 'Zero-inflated poisson regression, with an application to defects in manufacturing.' *Technometrics* 34(1), 1–14

Langemeier, M. L., T. C. Schroeder, and J. Mintert (1992) 'Determinants of cattle finishing profitability.' *Southern Journal of Agricultural Economics* 24, 41–47

Lau, L. J. (1978) 'Testing and imposing monotonicity, convexity, and quasi-convexity constraints.' In *Production Economics: A Dual Approach to Theory and Applications,* ed. M. Fuss and D. McFadden (Amsterday: North-Holland) pp. 409–453

Lawrence, J. D., Z. Wang, and D. Loy (1999) 'Elements of cattle feeding profitability in midwest feedlots.' *Journal of Agricultural and Applied Economics* 31(2), 349–357

Lee, L. F. (1993) 'Multivariate tobit models in econometrics.' In *Handbook of Statis-*

*tics, Vol. 11,* ed. G. S. Maddala, C. R. Rao, and H. D. Vinod. chapter 6, pp. 145–173

Li, C. S., J. C. Lu, J. Park, K. Kim, P. A. Brinkley, and J. P. Peterson (1999) 'Multivariate zero-inflated poisson models and their applications.' *Technometrics* 41(1), 29–38

Mahul, O., and B. D. Wright (2003) 'Designing optimal crop revenue insurance.' *American Journal of Agricultural Economics* 85(3), 580–589

Mark, D. R., and T. C. Schroeder (2002) 'Effects of weather on average daily gain and profitability.' Technical Report, Catlemen's Day Report of Progress No. 890, Kansas Agricultural Experiment Station, March

Mark, D. R., R. Jones, and J. Mintert (2002) 'Seasonality in steer feeding profitability, prices, and performance.' Kansas State University Research and Extension Report MF-2547

Mark, D. R., T. C. Schroeder, and R. Jones (2000) 'Identifying economic risk in cattle feeding.' *Journal of Agribusiness* 18(3), 331–344

Marsh, J. M. (1999) 'Economic factors determining changes in dressed weights of live cattle and hogs.' *Journal of Agriculture and Resource Economics* 24, 313–326

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995) *Microeconomic Theory* (Oxford University Press)

McDonald, J. F., and R. A. Moffitt (1980) 'The uses of tobit analysis.' *The Review of Economics and Statistics* 62(2), 884–895

Miranda, M. J. (1991) 'Innovations in agricutural and natural disaster insurance.' *American Journal of Agricultural Economics* 73, 233–242

Miranda, M. J., and D. V. Vedenov (2001) 'Innovations in agricultural and natural disaster insurance.' *American Journal of Agricultural Economics* 83(3), 650–655

NASS, USDA (1997) 'Nebraska agricultural statistics 1996-1997.' available at http://www.nass.usda.gov/ne/nebrpubs.htm

_ 'Kansas farm facts.' available at http://www.nass.usda.gov/ks

Nelson, C. H. (1990) 'The influence of distribution assumptions on the calculation of crop insurance premia.' *North Central Journal of Agricultural Economics* 12, 71–78

Nelson, C. H., and P. V. Preckel (1989) 'The conditional beta distribution as a stochastic production function.' *American Journal of Agricultural Economics* 71, 370–378

Neville, W. E., and W. C. McCormick (1981) 'Performance of early- and normal-weaned beef calves and their calves.' *Journal of Animal Science* 52, 715–724

Ozaki, V. A., S. K. Ghosh, B. K. Goodwin, and R. Shirota (2006) 'Spatio-temporal modeling of agricultural yield data with an application to priceing crop insurance contracts.' unpublished

Piggott, N. E., and T. L. Marsh (2004) 'Does food safety information impact u.s. meat demand?' *American Journal of Agricultural Economics* 86(1), 157–174

PSU (2001) 'Feeding beef cattle.' Technical Report, Penn State College of Agricultural Services Research and Extension. available at http://agalternatives.aers.psu.edu

Purcell, W. D., and S. R. Koontz (1998) *Agricultural Futures and Options: Principles and Strategies,* 2nd ed. (Prentice Hall)

Ramirez, O. A., S. Misra, and J. Field (2003) 'Crop-yield distributions revisited.' *American Journal of Agricultural Economics* 85(1), 108–120

Ritchie, H. D. (1994) 'A review of applied beef cattle nutrition.' Technical Report, MSU Extension Beef Bulletins - 23310001, January

Robert, Christian P. (2001) *The Bayesian Choice: from decision-theoretic foundations to computational implementation,* 2nd ed. (Springer texts in statistics)

RTI, International (2007) 'Gipsa livestock and meat marketing study.' available at http://www.gipsa.usda.gov

Schroeder, T. C., M. L. Albright, M. R. Langemeier, and J. Mintert (1993) 'Factors affecting cattle feeding profitability.' *Journal of the American Society of Farm Managers and Rural Appraisers* 57, 48–54

Sherrick, B. J., F. C. Zanini, G. D. Schnitkey, and S. H. Irwin (2004) 'Crop insurance valuation under alternative yield distributions.' *American Journal of Agricultural Economics* 86(2), 406–419

Smith, R. A. (1998) 'Impact of disease on feedlot performance: A review.' *Journal of Animal Science* 76, 272–274

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn (2003) 'Winbugs user manual, version 1.4 mrc biostatistics unit, cambridge, uk.' http://www.mrc-bsu.cam.ac.uk/bugs

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002) 'Bayesian measures of model complexity and fit.' *Journal of the Royal Statistical Association, Series B* 64, 583–639

Swanson, E. R., and V. I. West (1963) 'Statistical analysis of feeder cattle returns.' *Journal of Farm Economics* 45(1), 146–150

Thompson, G. B., and C. C. O'mary (1983) *The Feedlot* (Lea and Febiger)

Tobin, J. (1958) 'Estimation of relationships for limited dependent variables.' *Econometrica* 26(1), 24–36

Train, Kenneth, E. (2003) *Discrete Choice Methods with Simulation* (Cambridge University Press)

Turvey, C. G., M. Hoy, and Z. Islam (2002) 'The role of ex ante regulations in addressing problems of moral hazard in agricultural insurance.' *American Journal of Agricultural Economics* 62(2), 103–116

United States Department of Agriculture, Grain Inspection Packers Stockyards Administration (2001) 'Assessment of the cattle and hog industries calendar year 2000.' available at http://archive.gipsa.usda.gov/pubs/packers/assessment2000.pdf

von Mises, R. (1931) *Wahrscheilichkeitsrechnung* (Springer-Verlag, Berlin)

Weimar, M. R., and A. Hallam (1990) 'Risk sharing in custom cattle feeding.' *North Central Journal of Agricultural Economics* 12(2), 279–291

Williams, W. F., and T. T. Stout (1964) *Economics of the Livestock-Meat Industry* (The Macmillan Company)

Wright, S. (1921) 'Correlation and causation.' *Journal Agricultural Research* 20, 557–585

Yen, S. T. (1993) 'Working wives and food away from home: The box-cox double hurdle model.' *American Journal of Agricultural Economics* 75, 884–895

Zellner, A. (1985) 'Bayesian econometrics.' *Econometrica*

# Appendix

# Appendix 1: Multivariate Zero-Inflation Regression WinBUGS code

### Overall Strategy

This section outlines the steps taken to characterize a multivariate Zero-inflated regression model which can be used in WinBUGS software. The code used for running this model will be supplied at the end of this section. We consider a setup where three variables are contained within the dependent variable $Y$, where

$$Y = \left[ \begin{array}{ccc} Y_1 & Y_{21} & Y_{22} \end{array} \right] = \left[ \begin{array}{cc} Y_1 & Y_2 \end{array} \right] \tag{1}$$

We assume that $Y_1$ has a positive probability of taking on the value of 0, while $Y_2$ contains two variables, $Y_{21}$ and $Y_{22}$, and is continuous. The first step is to formulate the joint density between $Y_1$ and $Y_2$, which can be written as

$$f(Y_1, Y_2) = f(Y_1|Y_2)f(Y_2) \tag{2}$$

$Y_2$ will be distributed as a multivariate normal, with the variance as described below and a standard mean. The conditional probability of $Y_1$ given $Y_2$ is modeled through a zero-inflated modeling mechanism that takes into account the realizations from $Y_2$ in the following way

$$Y_1|Y_2 \sim ZIN(\rho_i, \mu_i(Y_2), \sigma_i^2(Y_2)) \tag{3}$$

where ZIN refers to the zero-inflated normal density function, $\mu_i(Y_2)$ is the conditional mean of $Y_1$ given $Y_2$ and $\sigma_i^2(Y_2)$ is the corresponding conditional variance.

#### Modeling Non-censored values

This section will describe how the continuous variables are modeled within this framework. The first objective will be to model $Y_2$ as a multivariate normal, which consists of 2 continuous variables, $Y_{21}$ and $Y_{22}$. The specifications are laid out below for ease of use in WinBUGS software and is equivalent to a multivariate normal specification. These variables are utilized in the following manner,

$$\begin{aligned} Y_{21i} &\sim N(\mu_{1i}, \Sigma_{1i}) \\ Y_{22i} &\sim N(\mu_{2i}, \Sigma_{2i}) \end{aligned} \tag{4}$$

where the mean is specified as follows

$$\begin{aligned} \mu_{1i} &= X_i B_1 \\ \mu_{2i} &= X_i B_2 + \Sigma_{12i}(Y_{21i} - X_i B_1) \end{aligned} \tag{5}$$

and the covariance matrix is constructed from $\Sigma_i = T_i' D_i T_i$,
with

$$T_i = \begin{bmatrix} 1 & t_{12} & t_{13} \\ 0 & 1 & t_{23} \\ 0 & 0 & 1 \end{bmatrix} \text{ and } D_i = \begin{bmatrix} d_{1i} & 0 & 0 \\ 0 & d_{2i} & 0 \\ 0 & 0 & d_{3i} \end{bmatrix}$$

where $\Sigma_i = \begin{bmatrix} \Sigma_{1i} & \Sigma_{12i} & \Sigma_{13i} \\ & \Sigma_{2i} & \Sigma_{23i} \\ & & \Sigma_{3i} \end{bmatrix}$ is a symmetric and positive definite hessian matrix. Additionally, diagonal elements of $D$ are functions of $exp(X_i\gamma)$. Additionally, elements of $\Sigma_i$ are a function of elements from $T_i$ and $D_i$ and be written as follows:

$$\Sigma_i = \begin{bmatrix} d_{1i} & t_{12}d_{1i} & t_{13}d_{1i} \\ & t_{12}^2 d_{1i} + d_{2i} & t_{12}t_{13}d_{1i} + t_{23}d_{2i} \\ & & t_{13}^2 d_{1i} + t_{23}^2 d_{2i} + d_{3i} \end{bmatrix}$$

**Modeling Censored values**

This section will focus on modeling $Y_1$ as a zero-inflated normal distribution. The specification begins with the Bernoulli component that models the likelihood of a value that is modeled according to the specified distribution

$$\omega_i \sim Ber(\rho_i) \tag{6}$$

where $\rho_i$ is specified as follows:

$$\rho_i = \frac{1}{[1 + exp(X_i\delta)]} \tag{7}$$

For our purposes, the specified distribution will be a normal, specified as follows

$$Y_{1i} \sim N(\mu_{3i}, Cov_{3i}) \tag{8}$$

where the conditional mean is specified as follows

$$\mu_{3i} = X_i B_4 + \Sigma_{13i}\Sigma_{1i}^{-1}(Y_{21i} - X_i B_1) + \Sigma_{23i}\Sigma_{2i}^{-1}(Y_{22i} - X_i B_2) \tag{9}$$

and the element $\Sigma_3 i$ incorporates elements from noncensored variable outcomes, by construction above. Additionally, the accompanying conditional variance is shown below:

$$Cov_{3i} = \Sigma_{3i} - \Sigma_{13i}\Sigma_{1i}^{-1}\Sigma_{13i} - \Sigma_{23i}\Sigma_{2i}^{-1}\Sigma_{23i} \tag{10}$$

The code used for estimation in WinBUGS software follows:

```
model{
for(i in 1:n){
#Model the observed non-censored variables as desired:
ldmfc[i] ~ dnorm(mean1[i], cov1[i])
lvcph[i] ~ dnorm(mean2[i], cov2[i])
adg[i] ~ dnorm(mean3[i], cov3[i])
#Define the characteristics of the normal
mu1[i] <- beta1[1] + beta1[gen[i]] + beta1[loc[i]] + beta1[7]*inwtlog[i]
 + beta1[season[i]]
mu2[i] <- beta2[1] + beta2[gen[i]] + beta2[loc[i]] + beta2[7]*inwtlog[i]
 + beta2[season[i]]
mu3[i] <- beta3[1] + beta3[gen[i]] + beta3[loc[i]] + beta3[7]*inwtlog[i]
 + beta3[season[i]]

tau1[i] <- exp(gam1[1] + gam1[gen[i]] + gam1[loc[i]] + gam1[7]*inwtlog[i]
 + gam1[season[i]])
tau2[i] <- exp(gam2[1] + gam2[gen[i]] + gam2[loc[i]] + gam2[7]*inwtlog[i]
 + gam2[season[i]])
tau3[i] <- exp(gam3[1] + gam3[gen[i]] + gam3[loc[i]] + gam3[7]*inwtlog[i]
 + gam3[season[i]])

mean1[i] <- mu1[i]
mean2[i] <- mu2[i] + cov12[i]*(ldmfc[i] - mu1[i])
mean3[i] <- mu3[i] + cov13[i]*(ldmfc[i] - mu1[i]) + cov23[i]*(lvcph[i] - mu2[i])

cov1[i] <- 1/tau1[i]
cov2[i] <- 1/(t12*t12*tau1[i] + tau2[i])
cov3[i] <- 1/(t13*t13*tau1[i] + t23*t23*tau2[i] + tau3[i])
cov12[i] <- t12*tau1[i]
cov13[i] <- t13*tau1[i]
cov23[i] <- t12*t13*tau1[i] + t23*tau2[i]

#Model the observed zeros as bernoulli trials:
w[i] ~ dbern(p[i])
p[i] <- 1/(1 + exp(delta[1] + delta[gen[i]] + delta[loc[i]] + delta[7]*inwtlog[i] +
    delta[season[i]]))
}

#Model the observed non-zeros as desired:
for(i in 1:n1){
lmort[i] ~ dnorm(mean4[i], cov4[i])

mu4[i] <- beta4[1] + beta4[gen[i]] + beta4[loc[i]] + beta4[7]*inwtlog[i]
 + beta4[season[i]]
tau4[i] <- exp(gam4[1] + gam4[gen[i]] + gam4[loc[i]] + gam4[7]*inwtlog[i]
 + gam4[season[i]])

#Mean takes into account the conditional terms
mean4[i] <- mu4[i] + cov14[i]*cov1[i]*(ldmfc[i] - mu1[i]) +
```

```
    cov24[i]*cov2[i]*(lvcph[i] - mu2[i]) + cov34[i]*cov3[i]*(adg[i] - mu3[i])

cov14[i] <- t14*tau1[i]
cov24[i] <- t12*t14*tau1[i] + t24*tau2[i]
cov34[i] <- t13*t14*tau1[i] + t23*t24*tau2[i] + t34*tau3[i]
cov4[i] <- 1/(t14*t14*tau1[i] + t24*t24*tau2[i] + t34*t34*tau3[i] + tau4[i])
}

#PREDICTION
for(h in 1:m){

mu1.pred[h] <- beta1[1] + beta1[genp[h]] + beta1[locp[h]] + beta1[7]*inwtlogp[h]
 + beta1[seasonp[h]]
mu2.pred[h] <- beta2[1] + beta2[genp[h]] + beta2[locp[h]] + beta2[7]*inwtlogp[h]
 + beta2[seasonp[h]]
mu3.pred[h] <- beta3[1] + beta3[genp[h]] + beta3[locp[h]] + beta3[7]*inwtlogp[h]
 + beta3[seasonp[h]]

tau1.pred[h] <- exp(gam1[1] + gam1[genp[h]] + gam1[locp[h]] + gam1[7]*inwtlogp[h]
 + gam2[seasonp[h]])
tau2.pred[h] <- exp(gam2[1] + gam2[genp[h]] + gam2[locp[h]] + gam2[7]*inwtlogp[h]
 + gam2[seasonp[h]])
tau3.pred[h] <- exp(gam3[1] + gam3[genp[h]] + gam3[locp[h]] + gam3[7]*inwtlogp[h]
 + gam3[seasonp[h]])

mean1.pred[h] <- mu1.pred[h]
mean2.pred[h] <- mu2.pred[h] + cov12.pred[h]*(ldmfcp[h] - mu1.pred[h])
mean3.pred[h] <- mu3.pred[h] + cov13.pred[h]*(ldmfcp[h] - mu1.pred[h]) +
    cov23.pred[h]*(lvcphp[h] - mu2.pred[h])

cov1.pred[h] <- 1/(tau1.pred[h])
cov2.pred[h] <- 1/(t12*t12*tau1.pred[h] + tau2.pred[h])
cov3.pred[h] <- 1/(t13*t13*tau1.pred[h] + t23*t23*tau2.pred[h] + tau3.pred[h])
cov12.pred[h] <- t12*tau1.pred[h]
cov13.pred[h] <- t13*tau1.pred[h]
cov23.pred[h] <- t12*t13*tau1.pred[h] + t23*tau2.pred[h]

p.pred[h] <- 1/(1 + exp(delta[1] + delta[genp[h]] + delta[locp[h]] +
    delta[7]*inwtlogp[h] + delta[seasonp[h]]))

mu4.pred[h] <- beta4[1] + beta4[genp[h]] + beta4[locp[h]] + beta4[7]*inwtlogp[h]
 + beta4[seasonp[h]]
tau4.pred[h] <- exp(gam4[1] + gam4[genp[h]] + gam4[locp[h]] + gam4[7]*inwtlogp[h]
 + gam4[seasonp[h]])

mean4.pred[h] <- mu4.pred[h] + cov14.pred[h]*cov1.pred[h]*(ldmfcp[h]
    - mu1.pred[h]) + cov24.pred[h]*cov2.pred[h]*(lvcphp[h]
    - mu2.pred[h]) + cov34.pred[h]*cov3.pred[h]*(adgp[h] - mu3.pred[h])

cov14.pred[h] <- t14*tau1.pred[h]
```

```
cov24.pred[h] <- t12*t14*tau1.pred[h] + t24*tau2.pred[h]
cov34.pred[h] <- t13*t14*tau1.pred[h] + t23*t24*tau2.pred[h] + t34*tau3.pred[h]
cov4.pred[h] <- 1/(t14*t14*tau1.pred[h] + t24*t24*tau2.pred[h]
 + t34*t34*tau3.pred[h] + tau4.pred[h])

}

#Priors:
beta1[1:3] ~ dmnorm(beta11[], Tau11[,]), beta1[5] ~ dnorm(beta12, Tau12),
beta1[7:10] ~dmnorm(beta13[], Tau13[,]), gam1[1:3] ~ dmnorm(gam11[], STau11[,])
gam1[5] ~ dnorm(gam12, STau12), gam1[7:10] ~dmnorm(gam13[], STau13[,])
beta2[1:3] ~ dmnorm(beta21[], Tau21[,]), beta2[5] ~ dnorm(beta22, Tau22)
beta2[7:10] ~dmnorm(beta23[], Tau23[,]), gam2[1:3] ~ dmnorm(gam21[], STau21[,])
gam2[5] ~ dnorm(gam22, STau22), gam2[7:10] ~dmnorm(gam23[], STau23[,])
beta3[1:3] ~ dmnorm(beta31[], Tau31[,]), beta3[5] ~ dnorm(beta32, Tau32)
beta3[7:10] ~dmnorm(beta33[], Tau33[,]), gam3[1:3] ~ dmnorm(gam31[], STau31[,])
gam3[5] ~ dnorm(gam32, STau32), gam3[7:10] ~dmnorm(gam33[], STau33[,])
beta4[1:3] ~ dmnorm(beta41[], Tau41[,]), beta4[5] ~ dnorm(beta42, Tau42)
beta4[7:10] ~dmnorm(beta43[], Tau43[,]), gam4[1:3] ~ dmnorm(gam41[], STau41[,])
gam4[5] ~ dnorm(gam42, STau42), gam4[7:10] ~dmnorm(gam43[], STau43[,])
delta[1:3] ~ dmnorm(delta1[], DTau1[,]), delta[5] ~ dnorm(delta2, DTau2)
delta[7:10] ~dmnorm(delta3[], DTau3[,])
t12 ~dnorm(4,.1)
t13 ~dnorm(-4.4, .1)
t23 ~dnorm(1.4, .1)
t14 ~dnorm(.4,.1)
t24 ~dnorm(4.3, .1)
t34 ~dnorm(.8, .1)


beta1[4] <- 0, beta1[6] <- 0, beta1[11] <- 0, gam1[4] <- 0, gam1[6] <- 0
gam1[11] <- 0, beta2[4] <- 0, beta2[6] <- 0, beta2[11] <- 0, gam2[4] <- 0
gam2[6] <- 0, gam2[11] <- 0, beta3[4] <- 0, beta3[6] <- 0, beta3[11] <- 0
gam3[4] <- 0, gam3[6] <- 0, gam3[11] <- 0, beta4[4] <- 0, beta4[6] <- 0
beta4[11] <- 0, gam4[4] <- 0, gam4[6] <- 0, gam4[11] <- 0, delta[4] <- 0
delta[6] <- 0,delta[11] <- 0}

Data:
list(n=7598, n1=4135, m=3799, beta11 = c(0,0,0), beta21 = c(0,0,0), beta31 =
c(0,0,0), beta13 = c(0,0,0,0), beta23 = c(0,0,0,0), beta33 = c(0,0,0,0),
beta12 = 0, beta22 = 0, beta32 = 0, beta41 = c(0,0,0), beta43 = c(0,0,0,0),
beta42 = 0, gam11 = c(0,0,0), gam21 = c(0,0,0), gam31 = c(0,0,0), gam13 =
c(0,0,0,0), gam23 = c(0,0,0,0), gam33 = c(0,0,0,0), gam12 = 0, gam22 = 0,
gam32 = 0, gam41 = c(0,0,0), gam43 = c(0,0,0,0), gam42 = 0,
delta1 = c(0,0,0), delta3 = c(0,0,0,0), delta2 = 0,
Tau11=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
Tau21=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
Tau31=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
Tau41=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
Tau12=0.01, Tau22=0.01, Tau32=0.01, Tau42=0.01,
```

```
Tau13=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)),
Tau23=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)),
Tau33=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)),
Tau43=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)),
STau11=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
STau21=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
STau31=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
STau41=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)),
STau12=0.01, STau22=0.01, STau32=0.01, STau42=0.01,
STau13=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)),
STau23=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)),
STau33=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)),
STau43=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)) ,
DTau1=structure(.Data = c(0.01,0,0,0,0.01,0,0,0,0.01), .Dim = c(3,3)), DTau2=0.01,
DTau3=structure(.Data = c(0.01,0,0,0,0,0.01,0,0,0,0,0.01,0,0,0,0,0.01),.Dim=c(4,4)) )
```