2002

Contract design in soybean production

Nandini Ramesh
Iowa State University

Recommended Citation

Ramesh, Nandini, "Contract design in soybean production" (2002). *Retrospective Theses and Dissertations*. 17192.
https://lib.dr.iastate.edu/rtd/17192

Contract design in soybean production

by

Nandini Ramesh

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Economics

Program of Study Committee:
Roger Ginder, Major Professor
Brent Hueth
Robert Jolly
Charles Hurburgh
Kenneth Koehler

Iowa State University

Ames, Iowa

2002

Graduate College

Iowa State University

This is to certify that the master's thesis of

Nandini Ramesh

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Soybeans are currently produced and marketed through the commodity system. In such a system, soybeans from numerous producers are commingled at the local elevator and sold as commodity grain. Profits from production depend on the total output of soybeans. The incentive system therefore encourages growers to select high yielding varieties of soybeans. However, the value of soybeans to the processor depends on the quantities of soybean meal and oil that can be produced from a bushel of soybeans. The outputs of soybean meal and oil depend on the components, protein and oil contained in the soybeans. In an efficient market the price of soybeans reflects the value of protein and oil contained in the soybeans. Such a system in which the price of a commodity is set in accordance with the sum of the value of its components is called component pricing (Perrin, 1980). Such a price would give incentives to growers to select varieties with higher protein and oil contents in them. Although technology to measure protein and oil content using near infrared technology (NIR) is easily available, component pricing based on protein and oil is not commonly used in the soybean industry [†]. One possible reason is that the gains from such a system are not certain. The potential inverse relationships between soybean yield and protein (Hanson, 1991) and between protein and oil (Leffel, 1998) have an impact on the total value from the soybean crop. Moreover, the production of components varies considerably across regions and years (Hurburgh, 1998). This brings in additional risk in production that will have to be shared between the processor and the grower. This study attempts to address these issues through contract theory. A set of component prices is developed in a scenario where high value soybeans are grown under a contract between the processor and the grower. A processor would be willing to pay these prices to growers as incentives for growing soybeans with increased protein and oil. Since processors would want to maximize profits per bushel

---

[†] One processor Ag Processing Inc. (AGP) constituted a premium based on oil content in 1999. This was extended to include premiums for protein in 2001.

of soybeans processed, the incentive system would encourage growers to plant varieties that have higher components in them. The objective of this study is to design a contract to produce soybeans based on protein and oil content. This would include a set of component prices as incentives to grow soybeans with higher protein and oil content and a rule to share production risk between the processor and the grower. An estimate of the increase in expected social surplus from contracting is made and a method to select varieties that processors would want to implement by contracting is proposed.

CHAPTER 2. LITERATURE REVIEW

One of the earliest attempts to price soybeans based on its protein and oil content was made by Updaw(1976). He used regression equations to calculate the quantities of meal and oil from the protein and oil percentages of the soybeans. The average oil yield was subtracted from the sample oil yield to give an appropriate discount or premium. The protein percentages of the meal were then calculated from the estimated meal quantities by the use of a regression equation and historic USDA data. An average value for the soybean meal yield and protein percentage was calculated and subtracted from the sample value to obtain a discount or premium. The oil and protein premiums and discounts were then summed for each sample to assign a total premium or discount. This method does not take into account the meal and oil prices and therefore does not measure accurately the marginal value of soybeans in processing.

The development of the Estimated Processed Value (EPV) model (Brumm 1990) improved accuracy in estimating the quantities of meal and oil from soybeans with known protein and oil content. The model predicts the yield of crude soybean oil and soybean meal from the processing of soybeans in a solvent extraction plant under given soybean composition and processing conditions. The yields and prices of the components, meal, oil and hulls were used to calculate the estimated processed value of soybeans in dollars.

$$EPV = \frac{P_m W_m}{2000} + P_o W_o + \frac{P_h W_h}{2000} \qquad (1)$$

where EPV is the estimated processed value in dollars per bushel

P_m is the meal price after discounts from trading rules in dollars per ton

P_o is the oil price in dollars per pound

P_h is the hull price in dollars per ton

W_m is the weight of soybean meal in pounds per bushel

and W_h is the weight of hulls in pounds per bushel

The weight of hulls can be either positive or negative for removal or addition. The key feature of the model is that it includes National Oilseed Processors Association (NOPA) soybean meal trading rules. The rules limit the amount of hulls that can be included in the meal, because exceeding the fiber limit is so costly that a processor will chose to give away protein in preference to receiving a 10:1 fiber discount. Higher soybean protein contents allow the processor to include more hulls in the meal while still meeting the protein specifications. This results in greater meal yield. When hulls are added to the meal W_h is positive, otherwise W_h is negative. As protein content is increased, a point is reached where there is no additional value gained from increased protein content. This is because of the fiber limitation on soybean meal. The processor cannot add more hulls, resulting in a greater meal yield without exceeding the fiber specification. Additional value comes from increase in oil content alone. This method accurately measures the marginal benefit of increasing protein and oil in soybeans and with the prices of meal and oil included in the model, provides an improved estimate of the marginal value product of soybeans. The EPV also reflects the current market conditions through changes in the price of soybean meal and oil.

At the implementation level, only one processor, AGP Inc., has instituted a pricing scale based on protein and oil content. The structure of the premiums for the year 2001 is shown in Table 1 below. Premiums are awarded for oil above 19.5 % and higher and protein percentages 37 % and higher. The minimum oil required to receive the protein premium is 19.5%.Protein and oil are expressed on an as is moisture basis and components can be increased with drying. This premium system was found to be inadequate in providing effective incentives for growers because the protein levels required to qualify for a protein premium are so high that they are unlikely to be achieved. Also, there is a minimum oil level to get a protein premium. Protein percentages above 37% qualified for a premium of 3 cents per bushel. Therefore the incentives for protein are not likely to compete with that of oil when there is an inverse relationship between the two components. To overcome and correct these limitations, a pricing system for soybeans is proposed that would reflect the marginal value of protein and oil in soybeans. It will also include base premium payments to compensate for the yield drag and account for variability in the production of components across regions and years.

Table 1: Component Premium Schedule 2001

Percent Oil*	Oil Premium	Percent protein	Protein premium (37% and above)
19.4 or less	None	36.9% and less	None
19.5 to 19.8	2.0 cents	37% and higher	3.0 cents
19.9 to 20.1	3.0 cents		
20.2 to 20.4	4.0 cents		
20.5 to 20.7	5.0 cents		
20.8 to 21	6.0 cents		
21.1 and higher	7.0 cents		

* As is moisture

Such an incentive system could encourage growers to include protein and oil in their variety selection decisions. The key to component pricing is to provide adequate incentives to producers that would enable them to take actions to increase components from the soybean crop, assuming that the technology to increase components is available.

Interest in component pricing stems from the important question of how much value could be added to the soybean crop through component pricing and whether this added value is economically significant. Prominent among the studies that have estimated the social costs and benefits from pricing components in soybeans are those by Perrin (1980) and Updaw (1980). Perrin estimated moderate gains while Updaw estimated losses for the society as a whole as a result of the introduction of component pricing. Updaw developed an economic model of the production of soybeans and market equilibrium in the soybean oil and meal markets.

He estimated the production possibility frontier of meal and oil with the varieties of soybean that were used in production between 1971-75. The elasticity of transformation between protein and oil was estimated as infinity and the economically relevant frontier, a single point. He concluded that the effective value of the elasticity of transformation is zero. Component pricing would therefore not be able to induce the desired effect of increasing the protein and oil content of soybeans with the existing varieties of soybeans grown between 1971-75. He argued that the production of protein and oil would not change as a result of component pricing. With costs of pricing components at $0.81 cents a bushel, component pricing would result in net losses to the society. Assuming that component measurement occurs 2.5 times as soybeans progress through domestic marketing channels, for the average year during 1971-75 the annual social costs were estimated to be $12,585,375.

Perrin assumed elasticities of transformation of –0.5 and 1.5 and estimated the net gains at most about $125 million which is a modest value of about 1.4 per cent of the value of the annual soybean crop. He estimated the consumer's surplus and producer's surplus of the change in price that would result because of component pricing. His results show that oil consumers gain $ 628 million and protein consumers lose $495 million from the introduction of component pricing. The impact on producers was estimated to be negligible. Additional costs of component pricing were not taken into account in the study.

These models took the general equilibrium two good economy approach to modeling the soybean market. The change in quantities supplied for a change in price was estimated using the production possibilities frontier. The changes in prices and quantities were used to measure the consumer's and producer's surplus. These changes would be economy wide and under the assumptions of the model, component pricing would have to be implemented on an economy wide basis. The results however might be different in a single processor-grower coordination effort where the actions of one processor or one grower would not affect equilibrium prices and quantities in the market.

Hurburgh and Huck (1998) used data from the Iowa Soybean yield trials from the year 1997 to estimate the increase in value due to components in soybeans. The EPV was used as a measure of the value of protein and oil contained in the soybeans. A method for selecting varieties with higher yields and components was proposed. Among the highest yielding varieties those with the highest sum of protein and oil percentages were selected. EPV was regressed over the sum and it was found that the sum could predict EPV with a margin of error of ±0.3 units of standard deviation. The varieties were then ranked by yield and the top half was selected. Those varieties that were 0.5 and 1 standard deviation higher than the average sum were selected and their EPVs were predicted. The difference in EPVs from the average in the top half of the yields ranged between 12-24 cents a bushel. Hurburgh and Huck therefore claimed that an increase in value of 12-24 cents a bushel could be achieved without a reduction in yield.

This method overestimates the benefits from component pricing in soybeans for several reasons.

- The yield of soybeans that is important in determining the total amount of meal and oil produced per acre of soybeans. The tradeoffs between yield and oil and protein and oil are important in determining the total value from the soybean crop. On closer examination of data it was found that there is also a significant variation in the yield and components even in the top half of the yields in the sample; therefore it cannot be assumed that the increased components can be achieved without a reduction in yield. The measure of surplus from component pricing should include the yield per acre because that would determine the total meal and oil that can be processed from an acre of soybeans produced.
- It also does not take into account the costs of processing soybeans. Costs of processing are measured per bushel of soybeans and an improvement in the quality of soybeans is akin to increasing the processing capacity.

The costs of measurement and segregation were estimated as 3 cents a bushel in 1991. Since then, the cost of the tester has been halved and accuracy improved so the costs were projected at 2 cents a bushel (Hurburgh 1997). With costs having fallen from $0.81/bu to $0.02/bu since 1971-75 (Updaw 1980) , the benefits of component pricing could exceed the costs of measurement and implementation.

Grower response to component pricing would be to select varieties that would maximize their gross profits. These profits would depend on the specific component pricing regime. Component pricing of soybeans complicates the problem of variety selection for producers. This occurs because varieties differ not only in their yields but also in their protein and oil content. Another factor that producers need to take into consideration is the nature of the physical relationships among the components. Significantly high negative correlations have been found between protein and oil content and also between the seed yield and protein content. Producers will need to estimate their gross return per acre before deciding to plant a particular variety. This value will be influenced not only by differences among varieties in the yields, percentages of protein and oil but also the relative prices of meal and oil. To aid in the decision making process of producers several measures have been developed to rank varieties based on the economic value of their components.

Significant among the measures developed to rank varieties is the Constituent Yield Index (CONY) (Helms and Watt, 1991). The CONY is defined as $CONY=(P+O)\ Y$ where P,O are the percentages of protein and oil and Y is the grain yield. The CONY index is independent of the relationship between protein and oil prices and can be used as a simple and approximate measure of the value of soybeans. It was found to give the same ranking of soybean varieties as the GVA or gross value per acre that is defined as $GVA= PY$ where P is the price of the soybean when its components are priced and Y is the yield of the soybeans. Such indices were developed to aid the variety selection decision for producers.

Leffel(1998) compared high protein varieties and traditional varieties for their gross returns per acre. He utilized the formulae of Updaw and Nichols and determined the total discounts and premiums per bushel for each variety in each environment. He found that protein premiums were largely eliminated by oil discounts for the highest protein varieties. High protein soybeans were not more profitable than conventional varieties when their gross returns per acre were compared. The value of soybean varieties was consistent over a wide range of protein and oil prices. The result of the study can be attributed to the lower yield of the high protein lower oil soybean and further reiterates the importance of the tradeoffs between protein content and oil content and protein content and yield in the production of soybeans.

Leffel(1990) compared alternative economic models used by Updaw and Nichols, the EPV model and the Approximate Process Value (APV) model (Hanson, 1989). The APV is defined in terms of soybean components and their values as

$$APV = P_o X + \left(\frac{P_{44}}{0.44}\right) Y \tag{2}$$

where P_0 is the oil price in \$/lb, P_{44} is the 44% protein meal price in \$/lb , Y and X are the protein and oil percentages at 13 percent moisture. Leffel used these models in making recommendations for soybean variety selection. Leffel showed that all three methods provided approximately the same valuation and recommended that soybean producers produce the highest yielding, high-oil soybeans assuming that the price of oil to the price of meal bears a 2:1 ratio.

Several methods have been proposed to aid variety selection for growers, and each of these methods would depend on the component prices that are being used in the market. In this study, a method of selecting varieties is developed. Scarce data on variety and environmental effects prevent a more sophisticated econometric analysis. Data on yields of selected soybean varieties grown in different environments over several years will help to find the variety that will maximize surplus and reduce risk when soybeans are priced based on their components. The objectives of the study are:

- Develop a set of component prices or a grid that provides incentives for variety selection in soybeans that would maximize surplus per acre of soybeans produced.
- Estimate the potential surplus from variety selection and compare this surplus across several regions in Iowa and Illinois to see where the maximum benefits from component pricing will be captured.

- Find a decision rule that would help growers select the varieties that would maximize surplus for the growers and the processors.

The incentive payments for protein and oil are developed in a contract theory framework. The motivation to move from the traditional commodity market system to that of a contract based system comes from coordination benefits and risk sharing (Milgrom and Roberts, 1992). In a typical principal agent setting the actions of the agent are not observable to the principal. The contract is then designed in such a way to ensure that the agents would take actions that are desired by the principal. Such models are called hidden action or moral hazard models. In such models, the actions that are contracted upon affect the profits of the principal. The profits of the principal are assumed to be stochastic. Though the actions of the agent are not observable, they can be inferred through a vector of signals. In the contract design problem of interest here the principal is assumed to be the processor, the grower is the agent who caries out actions on his behalf. The stochastic nature of production of components and the associated risk in production, hidden actions of the agents, the fact that the incentive payments can be conditioned on the quality measures that are observable by both parties and measurable by NIR; all these features make the framework very convenient to develop a set of incentive payments conditioned on observed and measurable yields, protein and oil outcomes in soybean production. This is how the risk incentive framework of the Principal Agent theory (Grossman and Hart, 1983) is used to design an optimal contract that can be used by the processors to provide incentives to growers to select varieties that can maximize EPV per bushel. It is assumed that growers will be able to take actions that can influence the quality of output. Under the commodity system, growers maximize yield per acre but under the contract they would maximize EPV per bushel. A variety selection rule that would maximize EPV per bushel from the soybean crop is developed. This surplus is estimated across different regions to find where the maximum benefits from component pricing can be captured.

CHAPTER 3. DATA

All the data used in the study are the outcome of trials conducted and reported in the Varietal Information Program for Soybeans. The tests are conducted on as uniform a soil as is available in the testing area. Small plots are used to reduce the chance of soil and climatic variations occurring between one variety plot and another. Performance results for yield and 31 other important attributes including protein (% per dry weight), oil (% by dry weight), isoflavones, fatty acids and amino acids are conducted. The protein and oil content of the entries are determined with an Infratec near-infrared transmittance analyzer.

Test Program in Iowa:

Iowa Crop Performance tests for soybeans are conducted each year to provide information growers need to select the best varieties for their production conditions. Seed companies, Iowa growers and the Iowa Crop Improvement Association include entries in the Iowa Crop Performance Test. Entries in the test are evaluated at three regions in each district. The northern district includes the northern three tiers of counties, the central district includes the central three tiers and the southern district includes the southern three tiers of counties in Iowa. The individual locations and the counties are shown in the table below:

1. Northern District: Sioux Rapids (Buena Vista county), Luverne(Humboldt county), and Greene(Butler county)
2. Central District: Arcadia(Carroll county), Boxholm (Boone county) , and Keystone(Benton county)
3. Southern District: Griswold(Cass county), Winterset(Madison county) , Crawfordsville(Washington county)

Entries are grown in four-row plots with a row spacing of 27 inches. Three replications of the entries are used at each location. A preplant herbicide is applied at all locations. Two post emergence herbicide tests are offered for each district. Both conventional and Roundup varieties are grown. The trials are conducted identically each year and data from Iowa for the years 1998, 1999 and 2000 are used in this study. Data from all the locations were not available for all the years. Griswold was lost to a severe winter storm in 2000. In all, 8580 observations were used in the study. Since data were available for individual locations it was possible to look at the variation of components within and across regions and make an estimate of the effect component pricing would have on farm revenues across regions. The only information on the variety was the commercial names and had no information on the genotype. This made the data on variety confusing and unmanageable. On this account, most of the analysis ignores varietal effect. The crucial limitation of the data were the lack of adequate information to group similar varieties together. Without this information it is very difficult to find a rule to select varieties that maximize surplus and thereby measure accurately the gains from component pricing. It was also found that the life cycle of varieties is 2-3 years making it difficult to look at the performance of a single variety across several years to assess its performance.

Table 2: Number of observations by region and year

Region	1998	1999	2000
Northern District	1089	1176	1118
Central District	1065	1175	1130
Southern District	681	768	478

CHAPTER 4. DATA ANALYSIS

Summary Statistics:

In this section the data described earlier are used to compare yields and components across regions in Iowa and estimate the increase in expected social surplus from improved variety selection. The data are divided into Northern, Central and Southern regions using the convention adopted in the Iowa Soybean Yield Trials.

The mean and coefficient of variation for yield in bu/acre, protein and oil in percentages are reported in table 3 below. There is considerable variability in the characteristics within regions as shown by the coefficient of variation. The average yield of soybeans ranged from 50.10 bu/acre to 63.16 bu/acre, protein content from 35.24 % to 37.12%, and oil content from 17.79 % to 19.79% between regions.

Table 3: Summary Statistics for Yield, Protein and Oil

		Yield (bu/acre)			Protein (percent)			Oil (percent)	
Region	1998	1999	2000	1998	1999	2000	1998	1999	2000
Northern	57.85	53.52	57.72	36.48	35.45	36.08	19.10	18.49	18.89
	(15.89)	(7.19)	(6.91)	(2.97)	(3.99)	(3.16)	(3.05)	(3.82)	(3.33)
Central	63.16	54.56	50.10	35.57	35.24	36.68	19.79	18.19	18.42
	(8.81)	(7.83)	(8.32)	(3.49)	(2.94)	(3.27)	(3.99)	(4.22)	(4.23)
Southern	55.98	53.67	50.53	35.99	36.55	37.12	19.37	17.79	18.65
	(8.54)	(9.51)	(8.39)	(3.38)	(2.89)	(2.79)	(4.15)	(4.42)	(3.20)

Mean

(Coefficient of Variation in percentages)

Estimation of the increase in expected social surplus from contracting:

It is assumed that under the commodity pricing system growers select varieties with relatively high yield per acre. Under a contract, processors provide incentives for growers to select varieties that have relatively high EPV per bushel. Let the action that growers take be variety selection. Let a_L be the action of selecting varieties that are the highest yielding and let a_H be the action that growers take to select varieties with the highest EPV per bushel. Let the EPV generated by taking action a_L be denoted by $V(r,q| a_L)$ where r =protein, q=oil. Similarly $V(r,q| a_H)$ represents the EPV generated when action a_H is taken by the grower. The processor wants to implement action a_H by contracting. The difference between the expected values of the EPV from taking action a_H over taking action a_L is an estimate of the increase in expected social surplus from contracting. Table 4 and 5 show the summary statistics of the data under the two proposed actions.

Table 4: Summary Statistics for Yield, Protein and Oil under action a_L (Top quartile in yield/acre)

		Yield (bu/acre)			Protein (percent)			Oil (percent)	
Region	1998	1999	2000	1998	1999	2000	1998	1999	2000
Northern	68.01	58.32	62.55	36.76	35.15	62.55	18.84	18.57	18.78
	(3.20)	(2.41)	(2.95)	(2.17)	(2.99)	(2.95)	(2.25)	(3.09)	(3.52)
Central	70.28	60.00	55.00	35.68	34.95	55.00	19.49	18.37	18.66
	(3.98)	(2.88)	(3.12)	(2.43)	(2.54)	(3.12)	(3.02)	(3.73)	(4.27)
Southern	61.66	60.14	55.87	35.82	36.70	55.87	19.55	18.31	18.64
	(2.54)	(3.42)	(3.38)	(3.78)	(2.86)	(3.38)	(4.09)	(3.38)	(3.37)

Table 5: Summary Statistics for Yield, Protein and Oil under action a_H (Top quartile in EPV/bushel)

		Yield (bu/acre)			Protein (percent)			Oil (percent)	
Region	1998	1999	2000	1998	1999	2000	1998	1999	2000
Northern	58.93	51.78	55.35	37.49	37.18	37.31	18.95	18.01	18.82
	(15.13)	(7.43)	(7.55)	(2.33)	(2.19)	(2.05)	(3.52)	(3.53)	(2.93)
Central	60.00	53.44	48.69	36.82	36.11	37.77	19.50	18.38	18.29
	(8.46)	(8.45)	(10.03)	(1.99)	(2.22)	(2.39)	(3.26)	(2.92)	(3.54)
Southern	57.11	56.98	49.79	37.16	37.44	38.05	19.28	18.31	18.66
	(7.62)	(8.88)	(8.47)	(2.21)	(2.26)	(2.19)	(3.65)	(3.10)	(3.48)

$$S = E((V(r,q \mid a_H)) - E(V(r,q \mid a_L)) \tag{3}$$

The estimation of the social surplus from contracting is outlined below:

1. Regression Analysis: Two samples from the entire data set of protein and oil outcomes are selected corresponding to varieties selected under action a_L and under action a_H. To represent the sample selected under action a_L, varieties in the top quartile of the yields in each region and year are selected. The varieties in the top quartile in EPV per bushel for each year and region are selected to represent the sample that growers would select under

action a_H. Two regressions one with protein and the other with oil as the dependent variable are run on indicator variables representing region, year and the type of soybean (Conventional/Roundup) conditional on the action taken by the grower.

2. Joint density estimation: Residuals from these regressions are used to estimate the joint distribution of (q , r) conditional on each action. The indicator variables are added back to get the distributions corresponding to each region, year and soybean type. The marginal densities show the improvement in the distribution of protein and oil percentages that can be brought about through a contract.
3. Calculation of the EPV: The EPV was calculated using SPROC V 2.42 (Brumm and Hurburgh, 1990) for the protein oil combinations corresponding to each region and year combination. Each protein and oil combination gives a unique value of the EPV.
4. Estimation of the social surplus from contracting: The estimated joint distribution of protein and oil is used to calculate the expected value of EPV from a bushel of soybeans conditional on each action. The difference between the expected value of EPV conditional on action a_H over that conditioned on action a_L gives an estimate of the expected increase in social surplus from contracting a bushel of soybeans. This is multiplied by the yield per acre to get the expected social surplus per acre of soybeans grown.

Regression analysis:

Under action a_L, growers select the top yielding varieties of soybeans. The sample corresponding to action a_L is the set of varieties that are in the top quartile in yield for each year and region. Growers taking action a_H select varieties that have high EPV per bushel. The sample corresponding to the action a_H is the top quartile of observations for each region and year in their EPV per bushel . Since there is considerable variability in yield and components across regions and years, it becomes necessary to estimate the distribution of protein and oil across the three regions and years under the assumed actions taken by the growers. Geographic differences in composition are primarily caused by environmental factors (Hurburgh, 2000). The most important factor in explaining differences in composition in soybeans has been temperature. Soybeans from northern and western soybean growing states (North Dakota, South Dakota, Minnesota, Iowa , Wisconsin) contained about 1.5 –2% less protein and 0.2 –0.5% more oil than soybeans from Southern States (Texas, Arkansas, Louisiana, Mississippi, Tennessee, Kentucky, Alabama, Georgia, South Carolina, North Carolina) (Hurburgh et.al. 1990) .

Two linear regressions one with protein and the other with oil as the dependent variable are run on each of these samples with year, region and type as indicator variables. Two indicator variables are used to identify the three regions. Similarly two indicators are used to identify the three years. The regression equation was used to analyze the data for the significance of the region and year variables. The regression equation took the form

$$Y_{ij} \mid a_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_{ij} \qquad (4)$$

where $Y_{ij} \mid a_i$ is the dependent variable, the percentage of protein or oil in the soybean sample conditional on the given action.

X_{1i} is the indicator variable for region that take the values

$X_{1i} = 1$ if Central district otherwise $X_{1i} = 0$

$X_{2i} = 1$ if Southern district otherwise $X_{2i} = 0$

X_{3i} and X_{4i} are the two indicator variables used to identify the years 1999 and 2000.

$X_{3i} = 1$ if year=1999 and $X_{3i} = 0$ otherwise

$X_{4i} = 0$ if year=2000 and $X_{4i} = 0$ otherwise

X_{5i} is an indicator or a dummy variable for the type of soybeans, Conventional or Roundup and takes the following values, $X_{5i} = 1$ if Conventional otherwise $X_{5i} = 0$

The term ε_{ij} is the random error and captures the effect on protein and oil that is unexplained by the indicators. No assumptions are made on the distribution of errors. If all the indicator values are set to zero the value of protein and oil correspond to the year 1998, Conventional and located in the Northern district. The relevant indicator values can be set to one to get the mean values for each region, year and type combination. The residuals from the regression are then used to estimate the bivariate distribution of protein and oil. By using the residuals from the regression to estimate the bivariate distribution of protein and oil it is being assumed that the distributions differ only in their mean across regions and years. The significant coefficients are added back to get the distribution corresponding to each location and year. The regression of protein and oil using the sample conditional on action a_L is reported in Table 6 and 7. The coefficient for the central district (β_2) was not found to be significant over the base region (Northern district). The coefficient for the Southern district was significantly different from zero. The indicator variables for the year 1999 and 2000 were significant. The indicator β_5 was significant indicating that there is significant difference between the Conventional and Roundup varieties. In the regression of oil on the indicator variables the region variable β_1 was not significant but the variables β_2, β_3 and β_4 are significant at the 5% level. All the year indicators and the Conventional/Roundup indicators are significant. It can be concluded that there is significant difference between years and across the Conventional and Roundup varieties. The Southern district was significant in its effect on protein and oil values. A similar regression was run when action a_H is taken by the grower. The sample of observations used for this regression is the set of varieties with the highest EPV per bushel in each of the regions and years.

The results from the regression of protein and oil on the indicator variables showed that the location variables are not significant but the years are significant. The Conventional/Round up indicator turned out to be significant. All the indicators turned out to be significant in the regression of oil on the indicators. It can be concluded that among the varieties with the highest EPV, region indicators are not significant in explaining variations in protein, but they explained variation in oil values. The indicators for the years are significant in protein and oil suggesting that year to year variation in protein and oil values is significant. The results of the regression of protein and oil under action a_H are reported in Table 8 and 9.

Table 6: Regression of protein on region, type and year under action a_L

Variable	Degrees of freedom	Parameter Estimate	Standard Error	T value	Pr>|t|
Intercept	1	35.8474	0.0455	786.57	<0.0001*
Central region	1	-0.0837	0.0754	-1.11	0.2670
Southern region	1	0.5643	0.0914	6.17	<0.0001*
1999	1	-0.3795	0.0611	-6.21	<0.0001*
2000	1	0.5272	0.0633	8.32	<0.0001*
Type	1	0.1338	0.0672	1.99	0.0467*

R^2=10.85, * Significant at the 5% level for a two-tailed test

Bivariate density estimation of protein and oil:

The residuals from the distribution of protein and oil conditional on the actions a_L and a_H are used to estimate the bivariate density function of protein and oil conditional on the given action. Non-parametric kernel density methods are used to estimate these distributions. Non- parametric kernel density methods help to describe data when parametric distributions do not effectively summarize a data distribution. This typically happens when data are skewed or are multimodal. Non-parametric kernel density estimation is used to smooth a density estimate that accounts for such characteristics. Kernel density estimation is a technique that averages a kernel function across observations to create a smooth approximation. The density function $f(r,q \mid a_i) \quad i = L,H$ was estimated using the non-parametric kernel density estimator (Silverman, 1996). The kernel density estimate is defined by the equation:

$$f(r,q \mid a_i) = \frac{1}{nh_r h_q} \sum_{i=1}^{n} K\left(\frac{r-R_i}{h_r}, \frac{q-Q_i}{h_q} \right) \tag{5}$$

Table 7: Regression of oil on region, type and year under action a_L

Variable	Degrees of freedom	Parameter Estimate	Standard Error	T value	Pr>|t|
Intercept	1	19.3746	0.0279	694.09	<0.0001*
Central region	1	-0.0502	0.0462	-1.09	0.2670
Southern region	1	0.2619	0.0560	4.68	<0.0001*
1999	1	-1.0456	0.0374	-27.92	<0.0001*
2000	1	-0.7044	0.0388	-18.15	<0.0001*
Type	1	-0.0637	0.0411	-1.55	0.0467*

R^2=28.04, * Significant at the 5% level for a two-tailed test

Table 8: Regression of protein on region, type and year under action a_H

Variable	Degrees of freedom	Parameter Estimate	Standard Error	T value	Pr>|t|
Intercept	1	35.5077	0.03218	1103.53	<0.0001*
Central region	1	-0.0269	0.04839	-0.56	0.5781
Southern region	1	0.0928	0.05944	1.56	<0.1182
1999	1	0.4959	0.04033	12.30	<0.0001*
2000	1	0.6634	0.04107	16.14	<0.0001*
Type	1	0.1006	0.03339	3.01	0.0026*

R^2=12.09, * Significant at the 5% level for a two-tailed test

Table 9: Regression of oil on region, type and year under action a_H

Variable	Degrees of freedom	Parameter Estimate	Standard Error	T value	Pr>|t|
Intercept	1	19.9200	0.0215	926.72	<0.0001*
Central region	1	-0.0787	0.0323	-2.44	0.0149*
Southern region	1	0.4577	0.0397	11.53	<0.0001*
1999	1	-1.4796	0.0269	-54.92	<0.0001*
2000	1	-0.8003	0.0274	-29.17	<0.0001*
Type	1	0.0457	0.0223	2.05	0.0404*

R^2=57.69, *Significant at the 5% level for a two-tailed test where n is the sample size, K is the standard normal density function and h_X and h_y are constants depending on the size of the sample that controls the amount of smoothing in the estimate. $h_X > 0$ and $h_y > 0$ are called bandwidth parameters and they have substantial effect on the value of the density estimate. The function K(r,q) is the bivariate normal density function

$$K(r,q) = \frac{1}{2\pi}\exp(-\frac{r^2+q^2}{2}) \tag{6}$$

The estimated joint distributions under action a_L and under action a_H for the Northern district in 1998 are shown in Figures 1 and 2 respectively. The marginal density functions for r and q are estimated using the joint density function of (r,q). The marginal density function for r is defined as,

$$f(r \mid a_i) = \int f(r,q \mid a_i)dq \tag{7}$$

Similarly, the marginal density function for q is defined as,

$$f(q \mid a_i) = \int f(r,q \mid a_i)dr \tag{8}$$

The estimated marginal distribution of protein and oil under the two actions for the Northern district in 1998 is displayed in Figures 3 and 4 respectively. The vertical axes represent the probability density of protein and oil. The significant coefficients from the regression are added back to the estimated distribution to get the distribution for the relevant region and year combination. It is assumed that the distributions for each region and year differ in their mean values alone. It can be seen that quality incentives cause a shift in the marginal density of protein and oil. This shift represents the improvement in quality from contracting. It was found that the mean value for oil increased but the mean protein fell from contracting. But the variation in both protein and oil values reduced considerably from contracting.

Table 10: Characteristics of the bivariate distribution of protein and oil under the two actions

	Mean (a_L)	Variance (a_L)	Mean (a_H)	Variance(a_H)
Protein	35.84	1.39	35.50	0.63
Oil	19.37	0.52	19.92	0.28

Calculation of the EPV:

- The EPV was used as a measure of the value that can be generated from processing a bushel of soybeans. The Estimated Processed Value was calculated through the program SPROC V 2.42. The prices used are $166.26/ton for soybean meal, $0.18/lb for soybean oil and $0.01/ton for hulls. These prices were the averages for the years 1998-2000 at Decatur, Illinois. All the NOPA (National Oilseeds Processing Association) rules were applied. The limitations of fiber content of 3.5 % for 48% meal was used. The discount for exceeding the maximum fiber specifications being 1% of the invoice price per 0.1 % fiber in excess of specification was used. A tolerance of 0.3 percentage points in fiber is allowed.
- The protein discount of 2 times the unit price of protein per 1% protein below minimum specifications was used.
- The EPV was calculated on a 13 % moisture basis and no premiums were awarded for producing meal with protein content in excess of the specifications.

Figures 5 and 6 summarize the relationship between EPV and oil at fixed levels of protein. Protein levels at 30,35,40 and 45 % were arbitrarily selected and the relationship with increased oil was examined. As oil is increased, EPV increases linearly when premiums are awarded for meal above 48%. When protein is not awarded with a premium for exceeding specifications above 48% meal, protein levels of 40 and 45% give the same line. This means that there cannot be an increase in

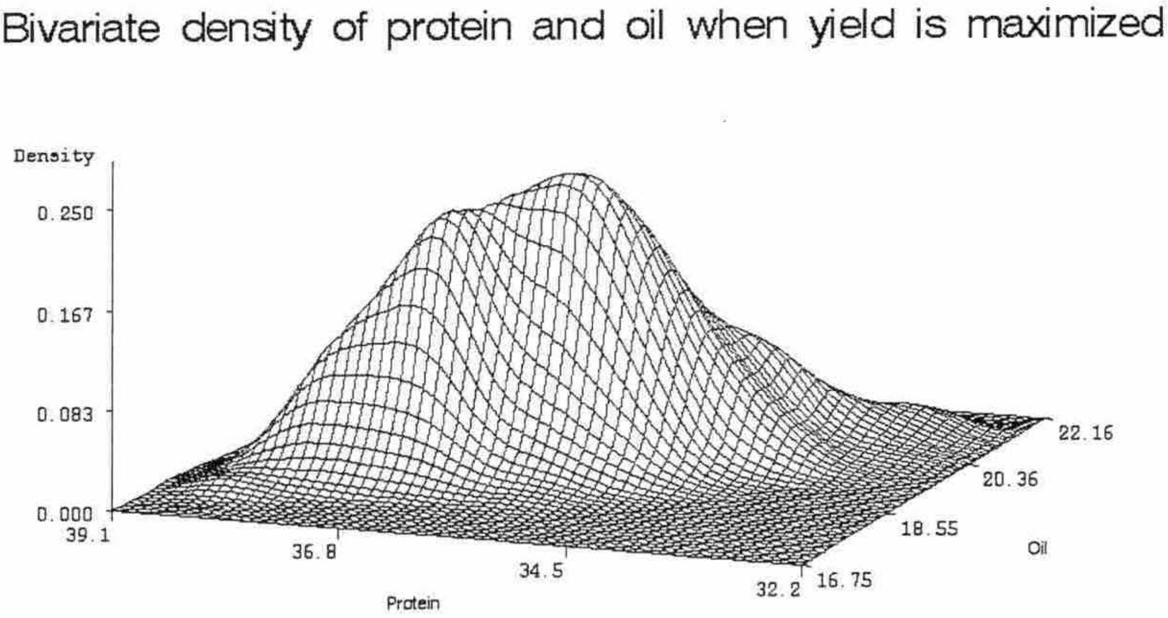

Figure 1: Bivariate density function of protein and oil under action a_L

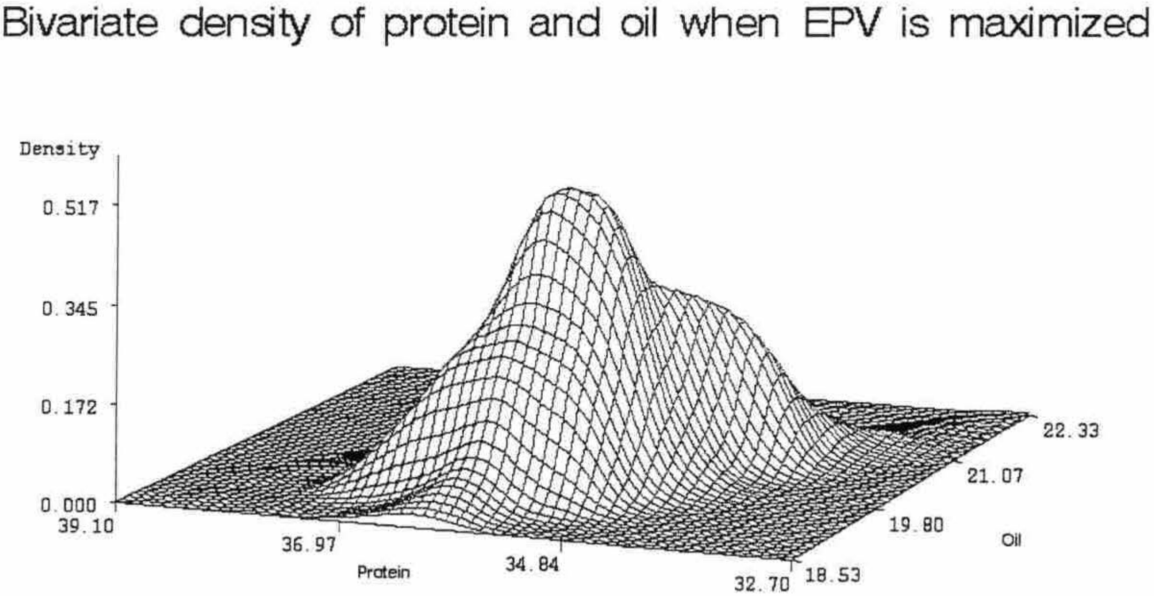

Figure 2: Bivariate density function of protein and oil under action a_H

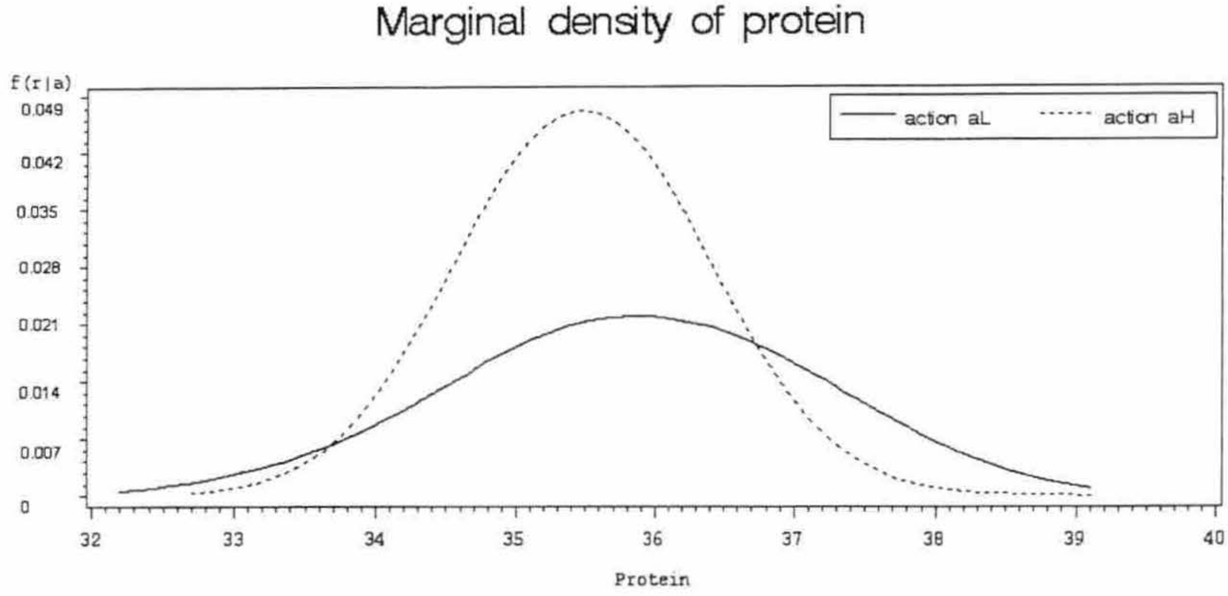

Figure 3: Estimated marginal distribution of protein

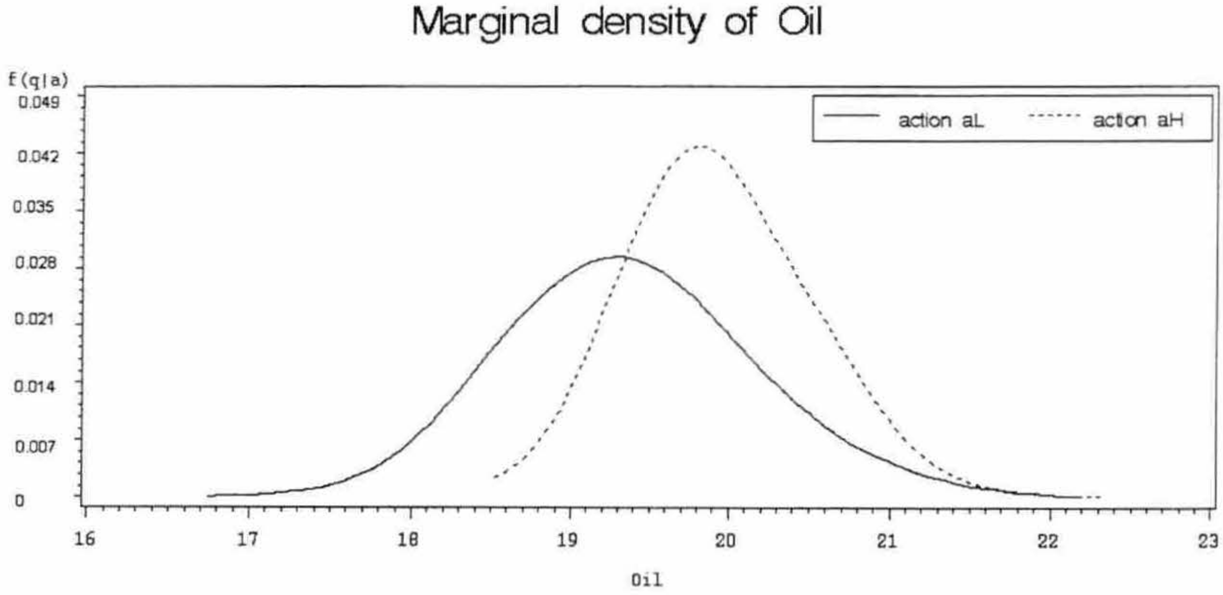

Figure 4: Estimated marginal distribution of oil

EPV for protein increases from 40 to 45%. This relationship can be understood more clearly from the EPV protein plots. To examine the relationship of EPV with protein, plots were made between EPV and protein with levels of oil arbitrarily fixed at 14 %, 18% and 22%. These are shown in Figures 7 and 8. The EPV increased linearly and then became constant at higher levels of protein when EPV was not awarded with a premium. Increase in protein above a certain point does not increase the EPV, although these levels of protein can be physically achieved in the soybeans. This is because adding hulls to increase meal yield will exceed the fiber requirement. The processor cannot add more hulls, resulting in a greater meal yield without exceeding the fiber specification. Additional value comes from increase in oil content alone. When protein is rewarded with a premium for exceeding specifications of meal above 48%, the rate of increase of EPV falls after the target meal specification of 48% is reached. The incremental benefit from increased oil is linear, but it is not so for protein.

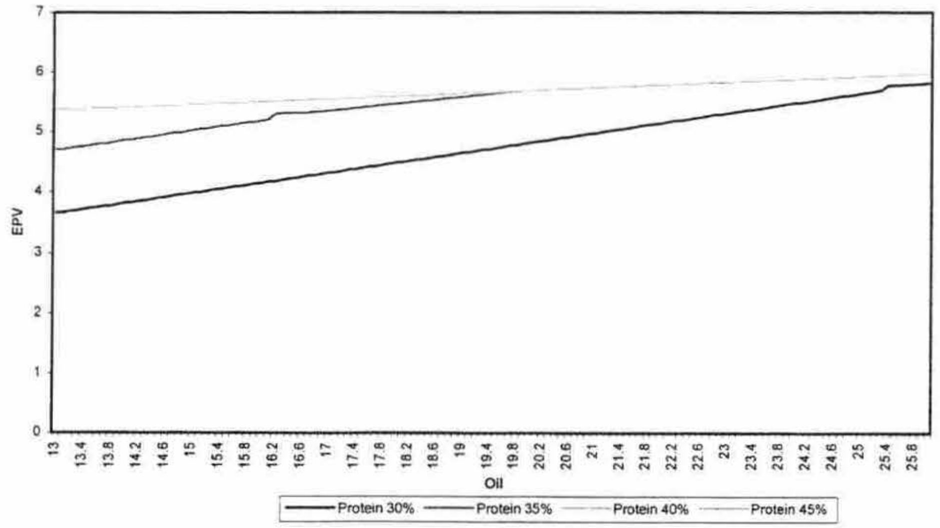

Figure 5 : EPV Vs Oil at fixed levels of protein

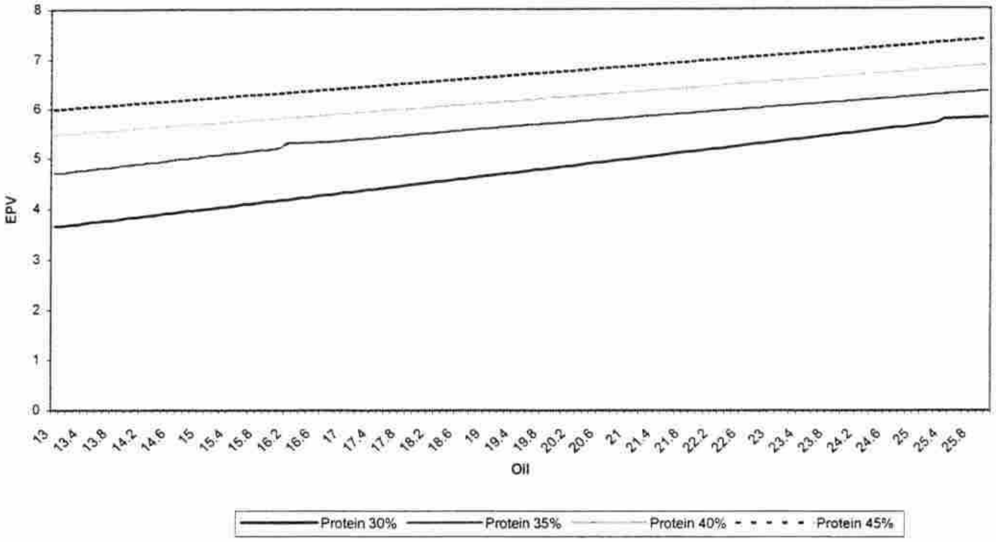

Figure 6: EPV Vs Oil at fixed levels of protein when premiums are awarded for exceeding

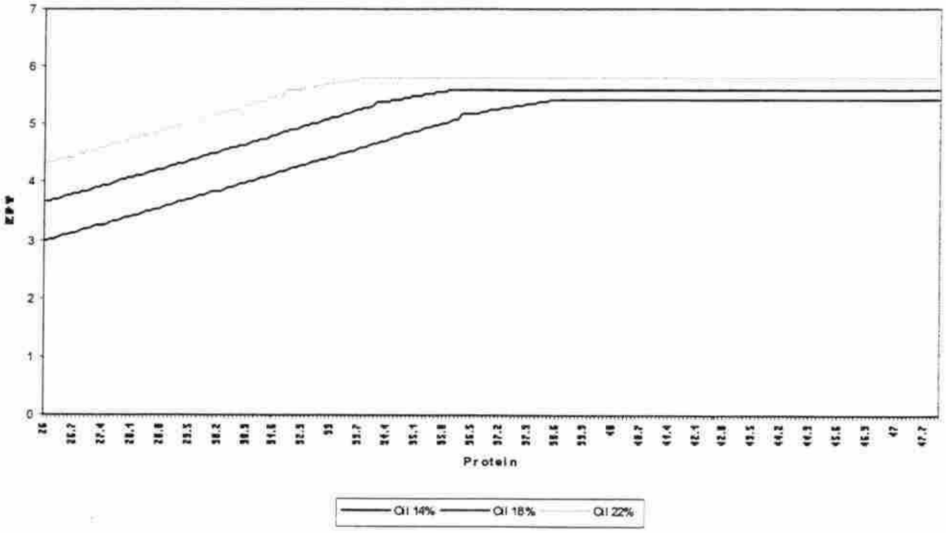

Figure 7: EPV Vs Protein at fixed levels of oil

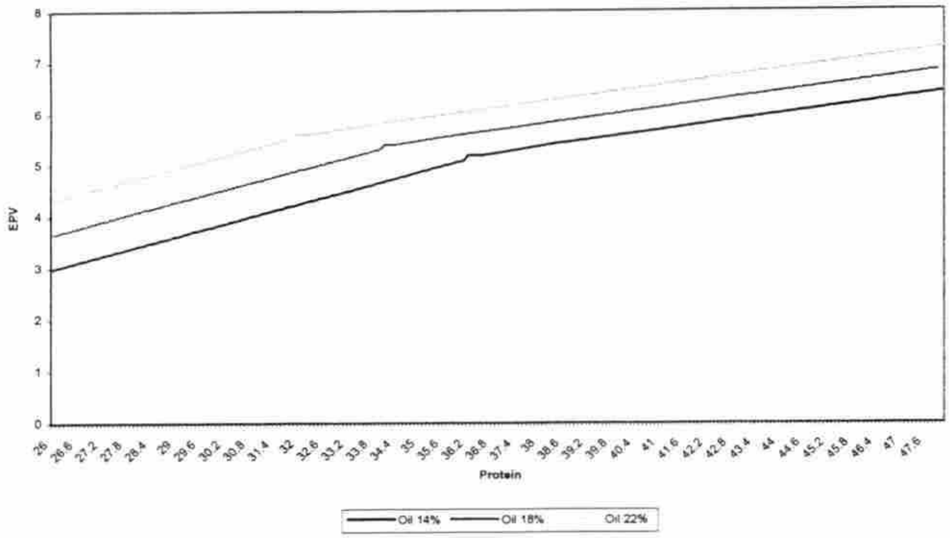

Figure 8: EPV Vs Protein at fixed levels of oil when protein is rewarded with a premium. 48% meal specifications

Estimation of the expected social surplus from contracting:

The estimation of the expected social surplus from contracting was done under the assumption that currently growers select varieties that have high yield per acre. Under a contract, growers select varieties that have high EPV per bushel. The surplus that can be obtained from contracting a bushel of soybeans is calculated as

$$S = \int V(r,q) f(r,q \mid a_H) drdq - \int V(r,q) f(r,q \mid a_L) drdq \tag{9}$$

where V(r ,q) is the EPV from processing a bushel of soybeans where r=percentage of protein in the soybeans , q=percentage of oil in the soybeans and $f(r,q|a_i)$ is the estimated distribution of protein and oil conditional on the action a_i.

The estimated distributions are used to calculate the expected values of the EPV conditional on action a_L and action a_H for Conventional varieties of soybeans for each region and year. It is assumed that the outcomes of protein and oil follow a discrete distribution. The EPV for each of these combinations of protein and oil are found. The estimated bivariate distribution of protein and oil is used to find the density at each of these points. These densities are multiplied by the EPV to get the expected value of the EPV for each region and year. These values are substituted into (9) to estimate the increase in expected social surplus from contracting.

Table 11: Expected increase in social surplus estimated by year and region in \$/bu

Region	1998	1999	2000
1	0.12	0.12	0.08
2	0.18	0.20	0.08
3	0.18	0.20	0.15
Average	0.16	0.18	0.10
Average across all regions			0.15

The increase in surplus obtained from growing an acre of soybeans can be defined as

$$S_a = S * y \qquad (10)$$

where S is the social surplus that be generated from contracting a bushel of soybeans and y is the yield per acre.

Assuming that there is no yield drag in the production of soybeans with higher components then y is the expected yield when action a_L is taken. The calculated value of y is the average yield in the top quartile of yields selected by region and year. The expected surplus estimated per acre is shown in table 12. When it is assumed that there is no yield drag in the production of soybeans with higher components, the surplus from contracting soybeans averaged to \$8.96 an acre. However, it is expected that there will be a fall in yield when soybeans with higher components are selected because of the inverse relationship between soybean yield and protein. The yield drag for each region and year combination was calculated as the percentage change in yield when action a_L is implemented vis-a-vis action a_H. The fall in yield averaged to -0.10 % when varieties are selected by the EPV per bushel rather than by yield. If y is the expected yield when action a_H is taken, the value of y is calculated as

Table 12: Surplus estimated by year and region in $/acre with no yield drag

Region	1998	1999	2000
1	8.16	6.99	5.00
2	12.65	12.00	4.40
3	11.09	12.02	8.38
Average	10.64	10.34	7.14
Average across all years			8.96

Table 13: Yield drag in percentages

Region	1998	1999	2000
1	-13	-11	-11
2	-14	-10	-11
3	-7	-5	-10
Average	-11	-8	-10
Average across all years			-10

Table 14: Surplus estimated by year and region in $/acre with a yield drag

Region	1998	1999	2000
1	7.07	6.21	4.42
2	10.80	10.68	3.89
3	10.27	11.39	7.46
Average	9.38	9.43	6.57
Average across all years			8.02

the average yield of the soybeans in the top quartile of EPV per bushel of soybeans selected. The social surplus that is generated per acre from contracting is then estimated in the table 14. The expected gains from contracting are estimated as $0.15 a bushel and $8.02 an acre of soybeans planted. This means that there are considerable gains from contracting and this value can be transferred from the processors to growers. A contract between processors and growers is designed that provides incentives for growers to include protein and oil in their variety selection decisions. The contract is designed based on the assumption that growers are risk-neutral. Therefore, the grower faces all the risk in the production of components.

CHAPTER 5. MODEL

A principal agent model is used to develop a contract between a single processor and grower. The contract determines the compensation for a single bushel of soybeans based on its protein and oil content. The outline of a basic contract design problem is as follows. If actions of the grower are observable to the processor, the contract would specify the actions to be taken by the grower and the compensation that the processor would provide in return. But actions taken by the grower are not observable to the processor (hidden actions), therefore the contract must be designed in such a way that indirectly gives the grower incentives to take the same actions that would be contracted for if his actions are observable. In this problem the grower's compensation scheme is designed to give him incentives to include protein and oil in his variety selection decisions. Since the actions of the grower are not observable to the processor, he will condition the grower's compensation based on some vector of observable signals through which the processor can infer actions taken by the grower. In the current problem, these observable signals are the protein and oil percentages in soybeans. Protein and oil content are measurable and observable to both the processor and the grower. Let $S \equiv \{(r,q), r \in R, q \in Q\}$ be the set of all possible realizations of s. Since s is fully observable by both parties, grower compensation can be specified for each level of s(r,q). Suppose that the processor offers a compensation schedule of the form $w(r,q) = V(r,q) - \alpha$

where α is some constant. This contract would give the grower compensation that equals V(r,q) except for a fixed payment α .

The processor is assumed to be risk neutral and his profits depend on the quality of inputs, and can be described by the function V(r,q), for simplicity can be written as V(s), an increasing and concave function. It is assumed that there is no price risk for both the processor and grower. Let a denote the action choice of the grower where $a = \{a_L, a_H\}$ where a_L denotes the action of selecting varieties that maximize yield per acre and a_H being the action of selecting varieties that maximize

EPV per bushel. Taking action a_H will involve certain costs for the grower over action a_L. These costs can be interpreted as the costs of searching and variety selection. Through the commodity market system growers take action a_L and it is assumed that the processor wishes to implement a_H through the contract. Action a_H leads to higher profits for the processor but entails greater difficulty for the grower. The processor's profits are affected by the actions of the grower but are not fully determined by it because of natural risk in the production of components. The processor's profits are assumed to be stochastic and are conditional on the actions of the growers as described by the following conditional probability density function $f(s| a_L)$ and $f(s| a_H)$ with $f(s \mid a\ ) > 0$ for all $a \in [a_L, a_H]$ and $f(s \mid a) > 0$ for all $s \in S$.

The assumption made here is that both the processor and the grower are risk-neutral. The utility function of the grower can be described as $u(w)=w-c(a)$ where w is the compensation paid to the grower and c(a) is the cost of action. Since the grower is neutral he does not need to be insured against risk. The optimal contract for the processor solves the following problem

$$\underset{a \in \{a_L, a_H\}, w(s)}{Max} \int (V(s) - w(s)) f(s \mid a) ds \tag{11}$$

subject to

$$\int w(s) f(s \mid a) ds - g(a) \geq \overline{u} \tag{12}$$

Equation (11) states that the processor maximizes his profits less wage payments subject to the constraint that the expected utility of the grower should be greater than his reservation utility. Reservation utility is the utility that the grower gets through the current pricing system, in this case being the price of soybeans in the commodity market system.

This problem is equivalent to minimizing the compensation costs of the processor

$$\underset{w(s)}{Min} \int w(s) f(s \mid a) ds \tag{13}$$

subject to

$$\int w(s)f(s \mid a)ds - g(a) \geq \bar{u} \tag{14}$$

The processor optimally specifies the action level $a = \{a_L, a_H\}$ that maximizes his expected profits after compensating the grower.

$$\underset{a\in\{a_L,a_H\}}{Max} \int V(s)f(s \mid a)ds - \bar{u} - g(a) \tag{15}$$

a* is the optimal action that solves for (15) . Equation (15) specifies the processor's profit and the grower receives an expected utility of at least $\bar{u}$. Using the specific form of the grower compensation

$w(s) = V(s) - \alpha$

The grower is willing to accept this contract as long as it gives him an expected utility of at least $\bar{u}$, as long as

$$\int V(s)f(s \mid a)ds - \alpha - g(a_H) \geq \bar{u} \tag{16}$$

Let α^* be the level of α at which the above equation holds with equality. The processor's return from the compensation w(V)=V-α^* is exactly α^* and we find that

$$\alpha^* = \int V(s)f(s \mid a\ )ds - g(a) - \bar{u} \tag{17}$$

and the optimal compensation scheme is w(s)=V(s)-α^*, which can be written as w(r,q)=V(r,q)-α^*. The processor finds the action level that maximizes his expected profits after compensating the grower. i.e. $\underset{a\in\{a_L,a_H\}}{Max} \int V(s)f(s \mid a)ds - \bar{u} - g(a)$ and the problem is solved in two stages.

Empirical Estimation of the Contract:

In the following section, the contract designed in the earlier section is estimated empirically in the case of soybean production. V(r,q) is the EPV of a bushel of soybeans for a given protein and oil content less processing costs. The costs of processing are set at 30 cents a bushel. To compute the compensation $w(r,q) = V(r,q) - \alpha *$, it is necessary to solve the equation (17) for α^*.

The expected value of EPV per bushel of soybeans is computed for each location and year using the empirical bivariate distributions of protein and oil. If the action a_H is optimal, then the expected return to the processor should be higher under this action. This means that

$$\int V(r,q)f(r,q \mid a_H)drdq - g(a_H) \geq \int V(r,q)f(r,q \mid a_L)drdq - g(a_L) \qquad (18)$$

Normalizing $g(a_L) = 0$ and assuming that the above equation holds with equality we have that

$$g(a_H) = \int V(r,q)f(r,q \mid a_H)drdq - \int V(r,q)f(r,q \mid a_L)drdq \qquad (19)$$

The expected value of V(r, q) under action a_H was found to be \$5.68/bu and the expected value of EPV per bushel is \$5.56 /bu under action a_L for the Northern district in year 1998. This gave an estimate of g(a_H) as 0.12 cents a bushel. g(a_H) is interpreted as the cost of implementing a_H over a_L. $\bar{u}$ = w_u is the reservation utility level or the wage payment from trading soybeans in the commodity market . This was assumed to be \$5.07 /bu , the average price of soybeans over the period 1998-2000 at Decatur, Illinois. Having identified values for the expression in (9), α^* was estimated as \$0.49 /bu. The payment grid for each level of r and q was developed as $w(r,q) = V(r,q) - \alpha^*$, by substituting the values for V(q,r) and α^*. Similarly, a grid of component prices can be developed for each region and year. The calculated values of α^* for each region and year are shown in Table 16. Under such a payment scheme, it was found that the returns per acre for the grower are less than they would get under the commodity market system, because the increased payment per bushel does not compensate for the fall in yield due to production of soybeans with higher components. Therefore, it becomes necessary to compensate growers for the yield drag. This is done through a base payment b that satisfies the condition that the expected total returns to the grower from contracting an acre of soybeans should equal the expected total returns to the grower from producing an acre of soybeans and selling it in the commodity market.

Table 15: Soybean prices based on protein and oil content in \$/bu

Oil Protein	18	18.5	19	19.5	20	20.5	21	21.5	22
30	4.00	4.08	4.17	4.24	4.33	4.42	4.49	4.57	4.66
31	4.20	4.29	4.36	4.45	4.54	4.62	4.71	4.79	4.87
32	4.42	4.49	4.58	4.66	4.75	4.83	4.91	4.99	5.11
33	4.62	4.70	4.79	4.87	5.01	5.04	5.10	5.16	5.22
34	4.82	4.93	4.98	5.04	5.10	5.16	5.22	5.27	5.30
35	4.99	5.04	5.10	5.16	5.20	5.23	5.25	5.27	5.30
36 and above	5.11	5.13	5.16	5.18	5.20	5.23	5.25	5.27	5.30

Table 16: Values of α^* for each location and year

Region	1998	1999	2000
1	0.49	0.49	0.56
2	0.34	0.32	0.42
3	0.47	0.45	0.53
Average	0.43	0.42	0.50
Average across all years			0.45

Table 17: Values of b for each location and year

Region	1998	1999	2000
1	0.18	0.04	0.02
2	0.35	0.11	0.16
3	-0.25	-0.37	-0.06
Average	0.09	-0.08	0.04
Average across all years			0.02

$$(E(w(r,q))+b)*E(y|a_H) = w_u*E(y|a_L) \qquad (20)$$

The value E(w(r,q)) is found through the payment scheme developed in table 13, and the sample means are used as estimators for expected values of the yield per acre conditional on the actions a_L and a_H. The base payments developed per bushel for each location and year are shown in table 17. The payments to the grower taking into account the values of b are shown in table 18 for the Northern district in 1998.

Table 18: Payments to the grower compensating for the yield drag in $/bu

Oil Protein	18	18.5	19	19.5	20	20.5	21	21.5	22
30	4.18	4.26	4.35	4.42	4.51	4.60	4.67	4.75	4.84
31	4.38	4.47	4.54	4.63	4.72	4.80	4.89	4.97	5.05
32	4.60	4.67	4.76	4.84	4.93	5.01	5.09	5.17	5.29
33	4.80	4.88	4.97	5.05	5.19	5.22	5.28	5.34	5.40
34	5.00	5.11	5.16	5.22	5.28	5.34	5.40	5.45	5.48
35	5.17	5.22	5.28	5.34	5.38	5.41	5.43	5.45	5.48
36 and above	5.29	5.31	5.34	5.36	5.38	5.41	5.43	5.45	5.48

Comparison between the AGP grid and the value based grid:

A market price was of $5.07/bu was used which was the average price of soybeans for the years 1998-2001 in Decatur, Illinois. All premiums are awarded over the market price. A premium of $0.03/bu was awarded for protein above 37%. A minimum of 19.5 % oil was needed to get the protein premium. For protein levels below 37% and oil levels below 19.4 % there are no premiums. There was no premium when protein crossed the 37% level but the oil was below 19.5%. Above the 19.5 % level oil, premium are granted irrespective of the protein levels, but to qualify for the protein premium, the level of oil had to be above 19.5%.

Comparison in returns per bushel between the AGP component prices and the value based grid at selected protein and oil values shown in Table 19 below. Since the value based grid has a return calculated for each realization of the signal s, which are a protein oil combination in our example, while the AGP grid does not, they are compared only at certain combinations of protein and oil. To compare returns when protein is below 37%, the returns from the value added grid at 33% was chosen arbitrarily. Similarly oil values are arbitrarily selected. Since there was no increase in premiums above 37% for either grid, they are compared at the 37% protein value.

The returns from the value-based grid are always higher than the AGP grid when the payments are calculated with a compensation for the yield drag. This compensation leaves the grower with the same returns per acre as with the commodity price. It gives the processor a return α^*.

Table 19: Returns from the AGP and the value based grid in $/bu

Oil Protein	19.4	19.6	20	20.1	20.3	20.6	20.9
Below 37%(AGP)	5.07	5.09	5.10	5.11	5.12	5.13	5.14
Value based 33%	5.21	5.23	5.28	5.29	5.32	5.35	5.39
37% and above(AGP)	5.1	5.12	5.13	5.14	5.15	5.16	5.17
Value based 37% and above	5.36	5.37	5.38	5.39	5.40	5.41	5.43

CHAPTER 6. CONCLUSIONS

This study examines component pricing in soybeans through a contract design perspective. It is shown through this approach, how a system of component prices can be designed and implemented to grow soybeans with higher protein and oil content. Preliminary examination of data and existing literature showed that there is considerable variability in the production of components both within and between regions and years. This makes it necessary to account for this variability in the contract design problem. This also brings in risk in the production of components into the problem. The optimal contract was designed using a Principal Agent model with a risk neutral processor and a risk neutral grower. In such a contract, the grower faces all the risk from the production of components. The contract returns full marginal value of protein and oil to the grower, less a constant. This constant is the processor's return from the contract. The EPV is used as a measure of marginal value of soybeans. Development of the contract under the assumption of risk aversion of the grower is left to future research.

The significant feature in the empirical development of the contract is the estimation of the bivariate distribution of protein and oil. All the data used are from the Iowa Soybean Yield trials for the period 1998-2000. Simple linear regressions are run one with protein and the other with oil as the dependent variable on indicator variables that identified the region and year. The values of the dependent variable are conditioned on assumed actions taken by the grower. Several of the region and year indicators turned out to be significant showing that year and region variables are significant in explaining variability in the production of components. Residuals from these regressions are used to estimate the bivariate distribution of protein and oil conditional on actions taken by the grower. These bivariate distributions for protein and oil are estimated using non-parametric methods. The significant coefficients for the region and year indicators are added back to the estimated distribution of protein and oil to yield the distribution corresponding to each region and year. The bivariate distribution of protein and oil are used to estimate the expected value from processing in a given region and year conditional on actions taken by the grower. This makes it possible to develop a different component pricing scheme for each region and year. This is how variability in the production of components is taken into account in the contract design problem. The marginal distributions of protein and oil show the improvement in the distribution of protein and oil that can be brought about through a contract.

The estimated distributions for each region and year are used to estimate the expected increase in social surplus from contracting. This value was estimated to be $0.15 per bushel which translated into a per acre value $8.02 when a yield drag due to production of soybeans with higher components is taken into account. The expected social surplus from contracting varied considerably across years and regions. It can be concluded that there are substantial gains to be made from contracting through improvements in the distribution of protein and oil. Currently there are no premiums for producing soy meal that exceeds the 48% meal requirement. This puts a limitation on the premiums for higher protein in soybeans. Therefore, the premiums for protein are constant above 37% protein. Unless processors can capture this value from the feed industry downstream, they will not be able to give higher premiums to the growers for their components.

All the data used in the analysis are from the Iowa Soybean Variety trials. Trial data are characterized by very small plots and superior crop management. It would be useful to compare it with the actual quality of yield and components that goes into the processors. Information on grower response with and without quality incentives will make it possible to estimate accurately the increase in social surplus or the improvement in quality from such incentives.

The biggest limitation is that the contract designed here only takes into account payments per bushel and payments made for contracting per acre could not be designed. Estimation of a multivariate distribution of yield and components will help estimate accurately the surplus from contracting and design a contract with payments per acre. Such a contract will help adjust for the fall in returns to the grower as a result of the reduced yield not being overcome by the incentives for components. One of the implementation issues in the design and implementation of a contract is the lack of adequate information on soybean variety. Isolating varieties with higher yield and components and studying their variability across regions and years would help estimating the distributions of their yields and components, thereby enabling a more accurate estimation of the gains from contracting. Identifying varieties useful for contracting would aid implementation of the contract.

REFERENCES

Brumm, T. J., and C. R. Hurburgh, Jr., "Estimating the Processed Value of Soybeans." *J. Amer. Oil Chem. Soc.* 67(1990): 302-307.

Brumm, T. J., Charles R. Hurburgh, Jr., James M. Guinn and Randy A. Hartwig, " Protein and Oil Patterns in U.S. and World Soybean Markets" *J. Amer. Oil Chem. Soc.* 67(1990): 966-973.

Component Premium Schedule, Ag. Processing Inc., Sheldon, IA, 2001.

Ginder, Roger G., "New Opportunities for Farmers in Value-Added Grains." *Iowa State University Extension Paper*, Ames, IA, 1998.

Grossman, S. and O. Hart, "An Analysis of the principal agent problem" *Econometrica* 51(1983): 7-45.

Helms, T. C., and David L. Watt, "Protein and Oil Discount/Premium Price Structure and Soybean Cultivar Selection Criteria." *J. Prod. Agr.* 4(1991): 120-124.

Hennessey, David A., and Brent Hueth "Contracts and Risk in Agriculture: Conceptual and Empirical Foundations", in *Comprehensive Assessment of the Role of Risk in Agriculture*",eds. Richard Just and Rulon Pope, Boston, Kluwer Academic Publishers, 2002.

Huck, Stephen L., Benjamin R. Winborn, and Charles R. Hurburgh, Jr., "Soybean Price Adjustments Based on Protein and Oil Variations." *Internal Discussion Paper, Iowa State University Grain Quality Laboratory*, Ames, Iowa, 1997.

Hurburgh Jr., Charles R., "Experiences with Value-Added Grains: Testing and Handling." Presented at GEAPS Exchange '97, Feb. 22-26, 1997, Minneapolis, MN.

Hurburgh Jr., Charles R., "Using Protein and Oil Content in Soybean Variety Selection." *Internal Discussion Paper, Iowa State Grain Quality Laboratory*, Ames, IA, 1997.

Hurburgh Jr., Charles R., "Localized Impacts of Component Pricing for Soybeans." *Internal Discussion Paper, Iowa State Grain Quality Laboratory*, Ames, IA, 1998.

Hurburgh Jr., Charles R., "Quality of the 2000 Soybean Crop from the United States" *Internal Discussion Paper, Iowa State Grain Quality Laboratory*, Ames, IA, 2000.

Leffel, R. C., "Economic Models and Breeding Strategies of Soybean Improvement." *J. Prod. Agr.* 3(1990): 582-586.

Leffel, R. C., "High Protein Lines and Chemical Constituent Pricing in Soybeans." *J. Prod. Agr.* 1(1988): 111-115.

Mas-Colell, Andreu, Michael D Whinston and Jerry R. Green *Microeconomic theory* ,New York, Oxford University Press Inc. 1995.

Milgrom, P. and J. Roberts (1992) *Economics, Organization and Management*. Englewood Cliffs, NJ, Prentice Hall, 1992.

Park, Hyesun S., and Charles R. Hurburgh, Jr., "Improving the U.S. Position in World Soybean Meal Trade." *Internal Discussion Paper, Iowa State Grain Quality Laboratory*, Ames, IA, 1998.

Plato, Gerald, " The Soybean Processing Decision : Exercising a Real Option on Processing Margins", U.S. Department of Agriculture, Econ. Res. Ser. *USDA Technical Bulletin Number 1897* , December 2001.

Perrin, R K. "The Impact of Component Pricing of Soybeans and Milk", *Amer. J. Agr. Econ.* 62(1980): 445-55.

Plummer, Robert, "Amid Low Margins Giant U.S. Soy Crushers Continue Slowdown." Chicago, *Bridge News*. December 17, 1999.

Silverman B.W. (1986) *Density Estimation for Statistics and data analysis*, London and New York, Chapman and Hall , 1986.

Updaw, N. J., J. B. Bullock, and T. E. Nichols, Jr., "Pricing Soybeans on the Basis of Oil and Protein Content." *S. J. Agr. Econ.* 8(1976): 129-32.

Updaw, N. J., "Social Costs and Benefits from Component Pricing of Soybeans in the United States." *Amer. J. Agr. Econ.* 62(1980): 647-655.

Westgate, M.E., E Piper, W.D. Batchelor, and C. Hurburgh, Jr " Effects of Cultural and Environmental Conditions During Soybean Growth on Nutritive Value of Soy Products", Soy in Animal Nutrition in James K Drackley ed., Savoy, IL, *Federation of Animal Science Societies*, 2000.

U.S. Department of Agriculture, Economic Research Service, Oil Crops Outlook OCS-0102 Jan 14, 2002.

ACKNOWLEDGEMENTS

I would like to thank my Major Professor Dr. Roger Ginder for his support, encouragement and patience throughout this project. The financial support that he provided helped me complete my graduate studies. I express my appreciation to all the members of my committee, Dr. Brent Hueth for the time he spent on advising me on the project and also for giving me the exposure to contract theory. His contribution to the project has been vital. I have enjoyed working with him and learnt a lot in the way of "thinking like an economist" from him. I would also like to thank Dr. Charles Hurburgh for his time and understanding of the soybean industry that he shared with me through his prompt answers for all the questions I had. I would also like to thank Dr. Robert Jolly and Dr. Kenneth Koehler for their time and co-operative efforts in the project.

I would like to thank all the members of my family for their contribution, my parents for their encouragement and confidence in me. My brother, Narayan for his willingness to think about any problem even though economics is very far from his area of expertise. He has been able to advise me through his experience. I would also like to make a special mention of my aunt Radha and uncle Nataraj who were always there for me. Finally, I would like to thank my fiancé Ganesh who supported me through the last stages of the thesis.

My stay in Ames was complicated by several problems that I faced, both within and out of my control. It is through the grace of the Almighty that I have been able to overcome all these difficulties. I would also like to thank my friends both inside and outside the department of economics for making my stay in Ames a rich experience. I would like to specially mention Deanna Ward who has been a very good friend (oops… given me a hard time!).