

**Examining the Relationships Between  
Cooperatives And Soybean Production  
for Biodiesel in TN**

**Zhang Yu  
University of Tennessee  
Statistics 571  
July 31, 2003**

## Table of Contents

Introduction.....	2
Methodology.....	5
<i>Graphical Analysis</i> .....	5
<i>Examining Correlation</i> .....	7
<i>Model Identification</i> .....	8
<i>Interpretation</i> .....	22
Conclusion.....	23
Appendix.....	25



to manage and operate a biodiesel facility would have a great effect on the cost competitiveness of biodiesel production.

Given this growing interest in biodiesel as an alternative fuel and the importance of soybean production to West Tennessee agriculture, In February of 2003, a mail survey was sent to producers in Tennessee. (Appendix) The listing of soybean producers was provided by Tennessee Agricultural Statistics Service (TASS).

The objective of this project is to examine the relationship of the number of bushels the soybean farmers in TN would be willing to sell to a cooperatively owned plant (**bushels to coop**) with variables that I think may affect it:

- **Coop or not**---whether they are currently a member of an agricultural cooperative
- **Acres harvested**---acres of soybeans harvested in 2001 (open ended)
- **Storage capacity**---the number of soybean bushels they can store on-farm (open ended)
- **Farm years**---years experience farming (open ended)
- **Net farming income**---net income from farming in 2001 (ten fixed categories from negative to greater than or equal to \$150,000)
- **Debt per \$100**---dollars financed with debt for every \$100 (nine fixed categories from \$0 to greater than \$70)
- **Income not from farming**---percent of their household's income came from off farm sources in 2001 (open ended)
- **Educational level**---highest education level (five fixed categories)
- **Biodiesel market**---the attitude the producers take toward whether biodiesel production will provide an important national market for soybeans in the next 10 years (five fixed categories)
- **Types of plants**---types of plants they would like to process soybeans (three fixed categories)
- **Contract**---percent they typically sell through contracts (open ended)

The variables that will be used in the following analysis are summarized in the Table 1 below.

**Table1: Summary of variables used in the project**

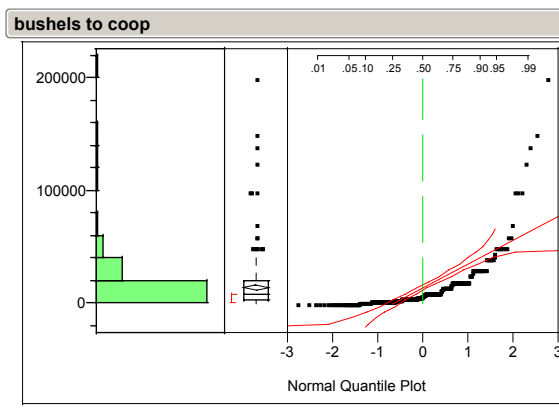
variable	type	unit
Coop or not	dummy variable	N.A.
biodiesel market	categorical variable	N.A.
acres harvested	numerical	acre
storage capacity	numerical	bushels
farm years	numerical	year
net farming income	categorical variable	N.A.
debt per \$100	categorical variable	N.A.
income not from farming	numerical	percentage
educational level	categorical variable	N.A.
types of plants	categorical variable	N.A.
bushels to coop	numerical	bushels

The data used in this project are from the survey. Totally, we got 577 effective responses. Since I have sufficient data, I cleared the observations with missing data in any of the above variables of interest. After clearing up the data set, I end up with 347 observations.

**Methodology:**

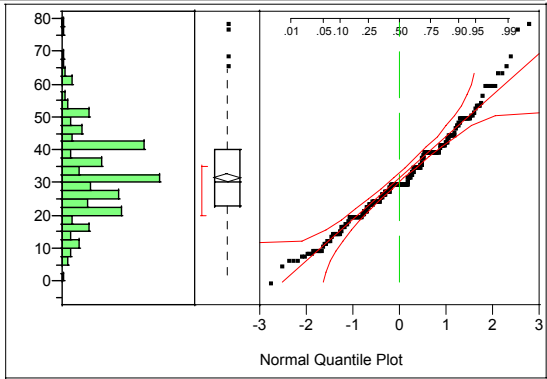
**I. Graphical Analysis:**

► *Familiarize the data by looking at the distribution and outlier boxplot first*

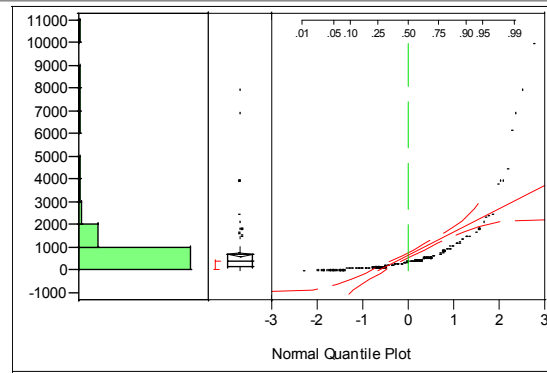


From the graph, the dependent variable (bushels to coop) is not normally distributed and has many outliers. And the dependent variable may need transformation.

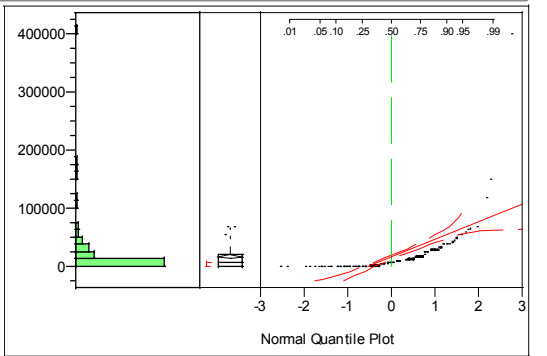
**farm years**

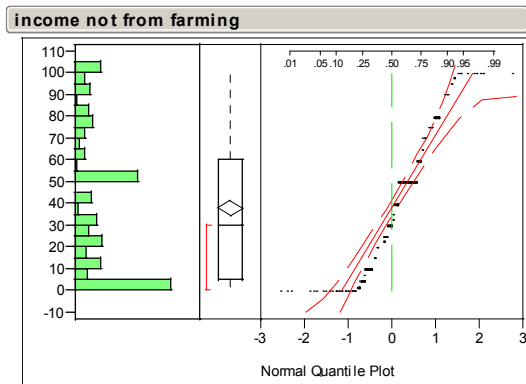


**acres harvested**



**storage capacity**





Also, from the plots of those numerical independent variables---acres harvested, storage capacity, farm years and income not from farming, I also found many outliers. Therefore, outliers would be a major problem for this data set. And except farm years, all the other three independent variables are not normal and may need transformation.

## II. Examining the correlation between the dependent variable and each of the independent variable.

**Table 2: Correlation between dependent variable and each independent variable:**

bushels to coop	biodiesel market	type of plants	acres harvested	coop or not
Correlation	-0.13	0.129	0.58	-0.038
bushels to coop	storage capacity	net farming income	contract	farm years
Correlation	0.4841	0.3264	0.038	-0.1553
bushels to coop	debt per \$100	educational level	income not from farming	
Correlation	0.1932	0.0768	-0.1024	

From the correlation table, I find the independent variables of coop or not, contract, and almost have nothing to do with dependent variable, which are beyond my expectation, especially the variable of whether they are member of a cooperative at present. Because as I assumed before, if they are affiliated to a cooperative, they are more likely to sell most of their soybeans to the cooperative first, or at least it would be the first choice.

### III. Model Identification:

#### ► Stepwise method on bushels to coop vs. all independent variables:

##### Stepwise Fit

Response: bushels to coop

##### Stepwise Regression Control

Prob to Enter 0.250

Prob to Leave 0.100

Direction Forward

##### Current Estimates

	SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
	8.0659e10	340	237232794	0.4854	0.4763	5.7714795	6698.668
Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Intercept	-18471.36	1	0	0.000	1.0000
<input type="checkbox"/>	<input type="checkbox"/>	biodiesel market	0	1	1.4444e8	0.608	0.4360
<input type="checkbox"/>	<input checked="" type="checkbox"/>	type of plants	3128.32703	1	1.2798e9	5.395	0.0208
<input type="checkbox"/>	<input type="checkbox"/>	coop or not	0	1	3.1144e8	1.314	0.2525
<input type="checkbox"/>	<input checked="" type="checkbox"/>	acres harvested	9.27899273	1	2.518e10	106.133	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	storage capacity	0.2157301	1	1.326e10	55.879	0.0000
<input type="checkbox"/>	<input type="checkbox"/>	contract	0	1	62743358	0.264	0.6078
<input type="checkbox"/>	<input type="checkbox"/>	farm years	0	1	4.2929e8	1.814	0.1789
<input type="checkbox"/>	<input checked="" type="checkbox"/>	net farming income	767.888982	1	9.8383e8	4.147	0.0425
<input type="checkbox"/>	<input checked="" type="checkbox"/>	debt per \$100	1085.32641	1	2.962e+9	12.486	0.0005
<input type="checkbox"/>	<input type="checkbox"/>	income not from farming	0	1	2966212	0.012	0.9112
<input type="checkbox"/>	<input checked="" type="checkbox"/>	educational level	2258.51207	1	1.9621e9	8.271	0.0043

##### Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	acres harvested	Entered	0.0000	5.273e10	0.3364	93.868	2
2	storage capacity	Entered	0.0000	1.569e10	0.4365	29.973	3
3	debt per \$100	Entered	0.0003	3.2882e9	0.4575	18.163	4
4	educational level	Entered	0.0050	1.9435e9	0.4699	12	5
5	type of plants	Entered	0.0143	1.4516e9	0.4791	7.9036	6
6	net farming income	Entered	0.0425	9.8383e8	0.4854	5.7715	7

After including the variables that have Prob>F less than 0.05, I ran my first model, with an Rsquare of 0.485401 which is low and a residual plot which has many outliers and is not normal. However, all the variables pass the T test and are significant.

### Summary of Fit

RSquare	0.485401
RSquare Adj	0.47632
Root Mean Square Error	15402.36
Mean of Response	14024.17
Observations (or Sum Wgts)	347

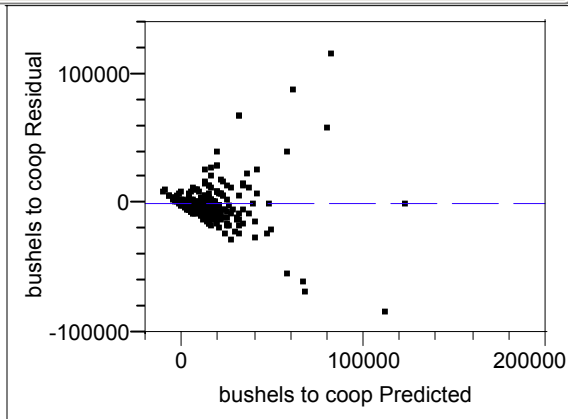
### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	7.60828e10	1.268e+10	53.4515
Error	340	8.06591e10	237232794	Prob > F
C. Total	346	1.56742e11		<.0001

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-18471.36	4529.616	-4.08	<.0001
type of plants	3128.327	1346.875	2.32	0.0208
acres harvested	9.2789927	0.900691	10.30	<.0001
storage capacity	0.2157301	0.028859	7.48	<.0001
net farming income	767.88898	377.0732	2.04	0.0425
debt per \$100	1085.3264	307.1537	3.53	0.0005
educational level	2258.5121	785.3247	2.88	0.0043

### Residual by Predicted Plot



Since there is a suspicious cone shape in the residual by predicted plot, I would like to transform the response variable. However, when I ran the Box-Cox, it said the program cannot do Box-Cox because of negative or zero Y value 0. So I created a new variable of bushels to coop+1.

► *Stepwise method on sqrt (bushels to coop+1) vs. sqrt (acres harvested), sqrt (storage capacity) and all other independent variables:*

After I ran the previous model with bushels to coop+1, I got a transformation suggested by the Box-Cox:  $\frac{(bushelstocoop + 1)^{0.4} - 1}{0.00264859999187482}$ . I simplified by taking square root of the dependent variable: bushels to coop+1. And because among the numerical variables in the first model, namely, acres harvested, storage capacity, and farm years, the former two have obvious curves as shown in the previous normal quantile plots, I decided to take transformation on both of them. I tried square root on both of them.

### Stepwise Fit

Response: sqrt(bushels to coop+1)

### Stepwise Regression Control

Prob to Enter 0.250

Prob to Leave 0.100

Direction: Forward

### Current Estimates

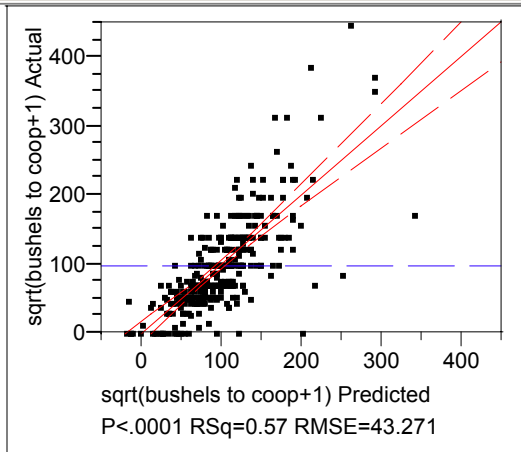
	SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC		
	630999.95	340	1855.8822	0.5759	0.5685	6.3534013	2618.49		
Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Intercept	-58.421315	1	0	0.000	1.0000		
<input type="checkbox"/>	<input type="checkbox"/>	biodiesel market	0	1	1575.413	0.848	0.3576		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	type of plants	25.9733137	1	88516.19	47.695	0.0000		
<input type="checkbox"/>	<input type="checkbox"/>	coop or not	0	1	311.7546	0.168	0.6825		
<input type="checkbox"/>	<input type="checkbox"/>	contract	0	1	45.8654	0.025	0.8754		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	farm years	-0.3871734	1	7487.062	4.034	0.0454		
<input type="checkbox"/>	<input type="checkbox"/>	net farming income	0	1	5768.236	3.128	0.0779		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	debt per \$100	2.90713446	1	19883.23	10.714	0.0012		
<input type="checkbox"/>	<input type="checkbox"/>	income not from farming	0	1	132.4261	0.071	0.7898		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	educational level	6.14446749	1	14125.77	7.611	0.0061		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	sqrt(acres harvested)	2.73130177	1	384039.6	206.931	0.0000		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	sqrt(storage capacity)	0.15098943	1	51071.19	27.519	0.0000		

### Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	sqrt(acres harvested)	Entered	0.0000	640414.7	0.4304	112.82	2
2	type of plants	Entered	0.0000	108935	0.5036	56.236	3
3	type of plants	Removed	0.0000	108935	0.4304	112.82	2
4	sqrt(storage capacity)	Entered	0.0000	58223.9	0.4695	83.508	3
5	type of plants	Entered	0.0000	102932.8	0.5387	30.151	4
6	sqrt(storage capacity)	Removed	0.0000	52221.63	0.5036	56.236	3
7	sqrt(storage capacity)	Entered	0.0000	52221.63	0.5387	30.151	4
8	debt per \$100	Entered	0.0001	29663.81	0.5586	16.197	5
9	educational level	Entered	0.0019	18254.87	0.5709	8.38	6
10	farm years	Entered	0.0454	7487.062	0.5759	6.3534	7

Since farm years only contribute 0.005 Rsquare. I decided to simply the model by excluding farm years also. So this time, I got the independent variables: type of plants, sqrt (acres harvested), educational level, sqrt (storage capacity), and debt per \$100. This is in accordance with the correlation table before. The variables with very low correlation rates with dependent variable are excluded from the model afterwards, such as coop or not and contract.

### Actual by Predicted Plot



### Summary of Fit

RSquare	0.570903
RSquare Adj	0.564611
Root Mean Square Error	43.27119
Mean of Response	98.67651
Observations (or Sum Wgts)	347

### Analysis of Variance

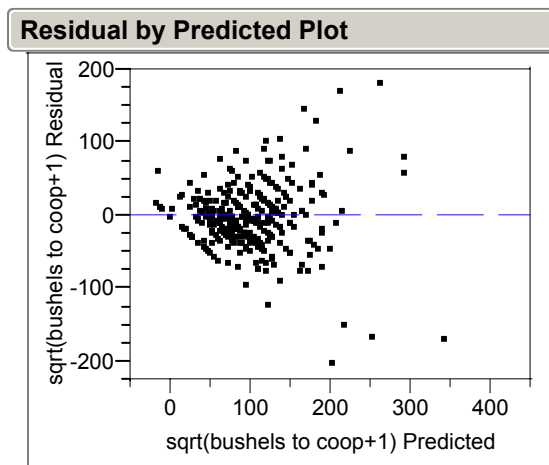
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	849490.0	169898	90.7383
Error	341	638487.0	1872	Prob > F
C. Total	346	1487977.0		<.0001

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t	VIF
Intercept	-76.12806	12.30453	-6.19	<.0001	.
sqrt(acres harvested)	2.7650672	0.189964	14.56	<.0001	1.1819589
sqrt(storage capacity)	0.1527379	0.028897	5.29	<.0001	1.1605634
type of plants	26.280035	3.774475	6.96	<.0001	1.0100609
debt per \$100	3.3180231	0.868128	3.82	0.0002	1.0353808
educational level	6.8878855	2.205948	3.12	0.0019	1.0080216

### Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
sqrt(acres harvested)	1	1	396703.40	211.8694	<.0001
sqrt(storage capacity)	1	1	52308.68	27.9368	<.0001
type of plants	1	1	90768.77	48.4773	<.0001
debt per \$100	1	1	27352.00	14.6080	0.0002
educational level	1	1	18254.87	9.7495	0.0019



**Sequential (Type 1) Tests**

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
sqtr(acres harvested)	1	1	640414.70	342.0295	<.0001
sqtr(storage capacity)	1	1	58223.90	31.0959	<.0001
type of plants	1	1	102932.75	54.9738	<.0001
debt per \$100	1	1	29663.81	15.8427	<.0001
educational level	1	1	18254.87	9.7495	0.0019

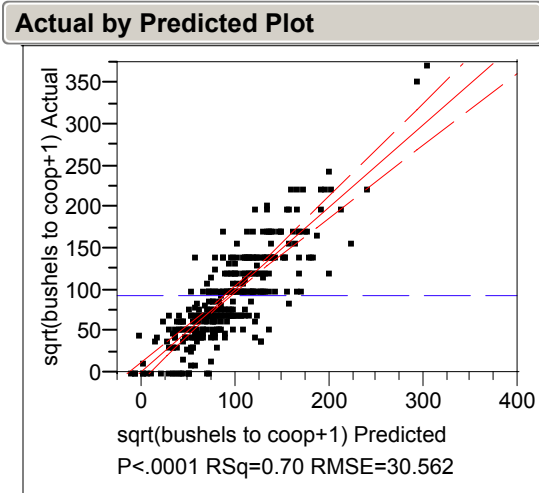
**Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation
1.9706941	347	0.0137

The second model:  $\sqrt{\text{bushels to coop}+1} = \text{type of plants, } \sqrt{\text{acres harvested}}, \sqrt{\text{storage capacity}}, \text{educational level, and debt per } \$100$  has a much better  $R^2$  of 0.575934, and the residual by predicted plot shows no obvious cone shape now. All the variables have significant coefficients. The Durbin-Watson is 1.9706941 and the autocorrelation is 0.0137, which shows autocorrelation is not a problem. However, this data are not related to time, so I don't need to worry about autocorrelation anyway. From the type I error in Sequential tests and the type III error in Effect tests are all significant. VIFs are all less than 10, so there is no multicollinearity. The residual by predicted plot and the leverage plots show evident proof that there are many outliers in the data which affect the fitness of the regression.

By looking at the Studentized residuals and Cook's D influence, I hid and excluded the observations whose absolute value of studentized residuals are greater than 2 and have an exceptionally large cook's D influence value as well.

► *Run the model  $\text{sqrt}(\text{bushels to coop}+1) = \text{type of plants, sqrt}(\text{acres harvested}), \text{sqrt}(\text{storage capacity}), \text{educational level, and debt per } \$100 \text{ after hiding and excluding the outliers}$*



#### Summary of Fit

RSquare	0.701232
RSquare Adj	0.696622
Root Mean Square Error	30.56242
Mean of Response	93.80462
Observations (or Sum Wgts)	330

#### Analysis of Variance

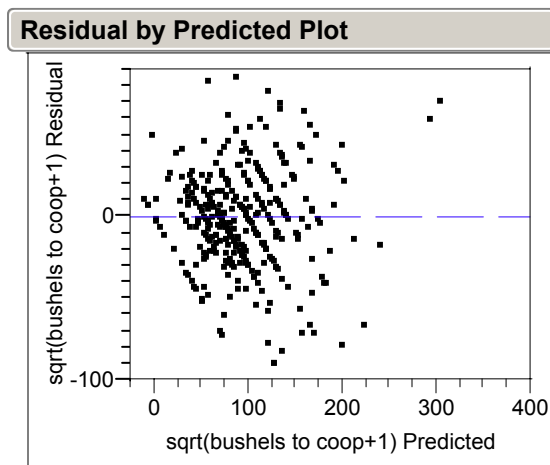
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	710310.9	142062	152.0908
Error	324	302635.9	934	Prob > F
C. Total	329	1012946.7		<.0001

#### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t	VIF
Intercept	-60.11201	8.964993	-6.71	<.0001	.
sqrt(acres harvested)	3.415567	0.179582	19.02	<.0001	1.2864642
sqrt(storage capacity)	0.1273686	0.02256	5.65	<.0001	1.2148829
type of plants	22.785274	2.788774	8.17	<.0001	1.0190615
debt per \$100	1.3738216	0.639259	2.15	0.0324	1.0762906
educational level	2.8424945	1.588497	1.79	0.0745	1.0019605

#### Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
sqrt(acres harvested)	1	1	337888.20	361.7409	<.0001
sqrt(storage capacity)	1	1	29772.66	31.8744	<.0001
type of plants	1	1	62352.99	66.7547	<.0001
debt per \$100	1	1	4314.02	4.6186	0.0324
educational level	1	1	2990.90	3.2020	0.0745



Press  
316984.58438

#### Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
sqrt(acres harvested)	1	1	603028.48	645.5984	<.0001
sqrt(storage capacity)	1	1	33650.08	36.0256	<.0001
type of plants	1	1	66175.22	70.8468	<.0001
debt per \$100	1	1	4466.17	4.7815	0.0295
educational level	1	1	2990.90	3.2020	0.0745

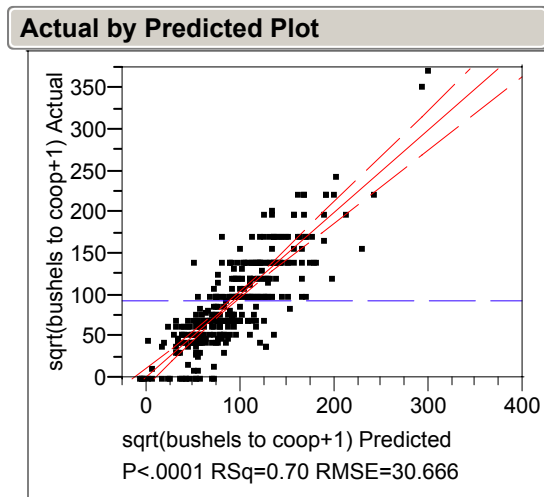
#### Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation
1.981441	330	-0.0465

After hiding and excluding some potential outliers, the variable: educational level becomes insignificant in both type I and type III error tests. And since it has only 0.07 cross-correlation with the dependent variable, I would like to try a simpler model without the variable of educational level. If the Rsquare doesn't change much, I would use the simpler model.

However, this model further improves the Rsquare to 0.701232, and the residual by predicted plot shows constant variance without any particular shape. The Durbin-Watson is 1.9882347 and the autocorrelation is -0.0465, which shows autocorrelation is not a problem. VIFs are all less than 10, so there is no multicollinearity problem.

► ***Further try the previous model without educational level after hiding and excluding the outliers:  $\text{sqrt}(\text{bushels to coop}+1) = \text{type of plants, sqrt}(\text{acres harvested}), \text{sqrt}(\text{storage capacity}), \text{and debt per } \$100$***



After I excluded the variable of educational level, the Rsquare only drops from 0.701232 to 0.69828, which is very little. So, educational level is not necessarily needed. And I simplified by model delete the unnecessary variable.

**The best model I get becomes:  $\text{sqrt}(\text{bushels to coop}+1) = 22.878245 \cdot \text{type of plants} + 3.4099344 \cdot \text{sqrt}(\text{acres harvested}) + 0.1264139 \cdot \text{sqrt}(\text{storage capacity}) + 1.3975387 \cdot \text{debt per } \$100$**

#### Summary of Fit

RSquare	0.69828
RSquare Adj	0.694566
Root Mean Square Error	30.66578
Mean of Response	93.80462
Observations (or Sum Wgts)	330

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	707320.0	176830	188.0390
Error	325	305626.8	940	Prob > F
C. Total	329	1012946.7		<.0001

The model has a good Rsquare of 0.69828, which means 69.828% of the variability in sqrt (bushels to coop+1) is accounted for by type of plants, sqrt(acres harvested), sqrt(storage capacity) and debt per \$100.

It passes the F test:

Ho:  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$  vs. H1: at least one of them is not equal to 0.

$$F_{cpt}=188.3090 > F_{4,325,0.05} = 2.37$$

So, we reject Ho, and there is at least one parameter estimate is not equal to zero. And the model has a good overall fit.

#### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t	VIF
Intercept	-51.98349	7.755079	-6.70	<.0001	.
sqrt(acres harvested)	3.4099344	0.180162	18.93	<.0001	1.286069
sqrt(storage capacity)	0.1264139	0.02263	5.59	<.0001	1.2142035
type of plants	22.878245	2.79772	8.18	<.0001	1.0187078
debt per \$100	1.3975387	0.641284	2.18	0.0300	1.0758279

For individual parameter estimate, all of them pass the T test:

Ho:  $\beta_i = 0$  vs. H1:  $\beta_i \neq 0$

Since all the P value is less than 0.05, we reject the Ho, and the coefficients of all the regressors and the intercept are significantly different from 0.

#### Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
sqrt(acres harvested)	1	1	336878.21	358.2324	<.0001
sqrt(storage capacity)	1	1	29344.42	31.2045	<.0001
type of plants	1	1	62884.69	66.8709	<.0001
debt per \$100	1	1	4466.17	4.7493	0.0300

#### Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
sqrt(acres harvested)	1	1	603028.48	641.2536	<.0001
sqrt(storage capacity)	1	1	33650.08	35.7831	<.0001
type of plants	1	1	66175.22	70.3700	<.0001
debt per \$100	1	1	4466.17	4.7493	0.0300

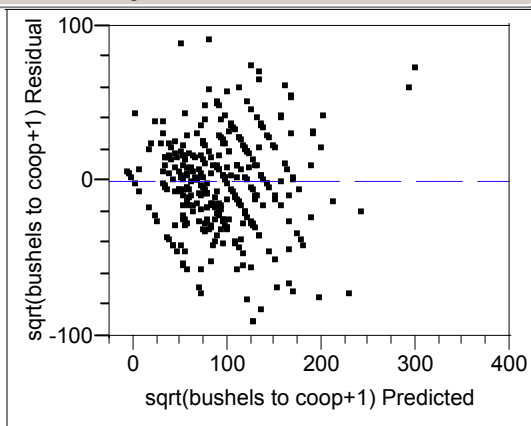
The type I error and type III error tests demonstrate that each independent variable are significant at  $\alpha=0.05$  level. Therefore, each independent variable provides additional information about the dependent variable sqrt (bushels to coop+1).

### Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation
1.9882347	330	-0.0488

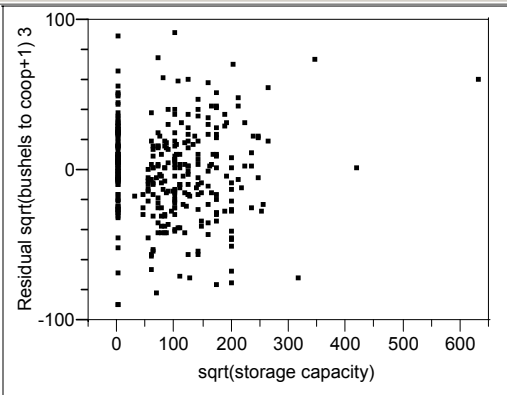
The Durbin-Watson is 1.9882347 and the autocorrelation is -0.0488. So we could not reject the  $H_0$  that there is no autocorrelation present. However, this data are not related to time, so I don't need to worry about autocorrelation anyway. All the VIF are much less than 10, so there is also no multicollinearity problem.

### Residual by Predicted Plot

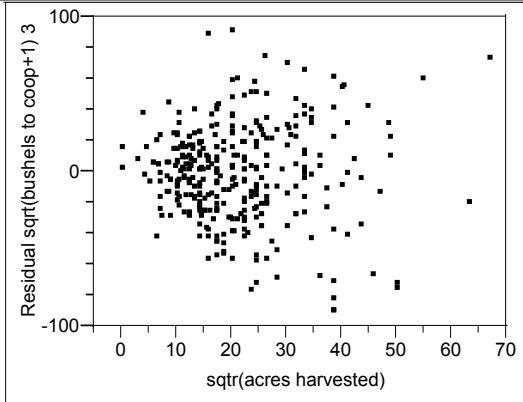


The residual by predicted plot shows no obvious cone shape and the dispersion of the residuals are approximately constant with respect to the predicted values. Therefore, the constant variance assumption is satisfied.

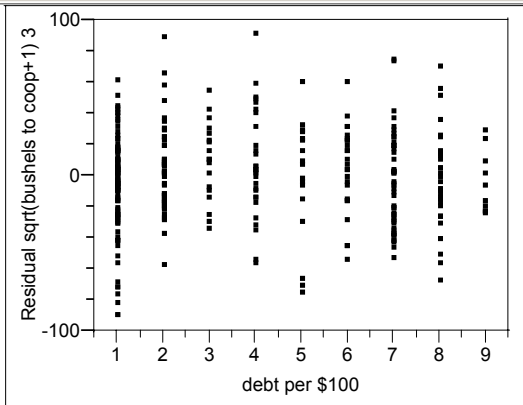
### Bivariate Fit of Residual sqrt(bushels to coop+1) 3 By sqrt(storage cap



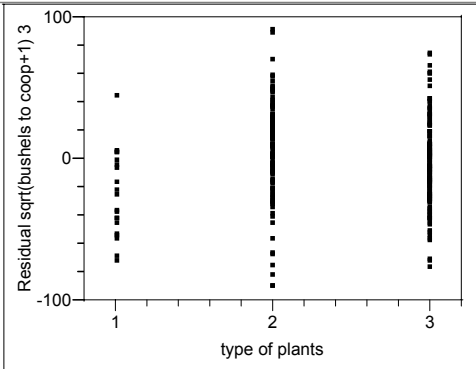
**Bivariate Fit of Residual sqrt(bushels to coop+1) 3 By sqtr(acres harve**



**Bivariate Fit of Residual sqrt(bushels to coop+1) 3 By debt per \$100**

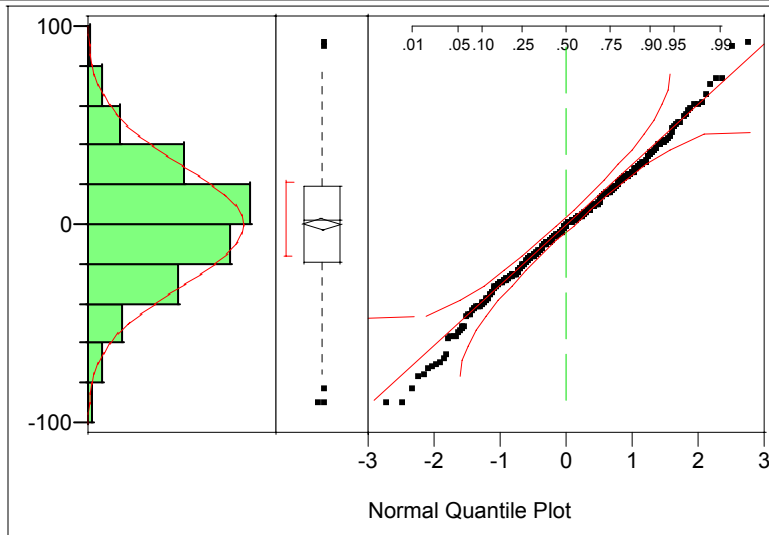


**Bivariate Fit of Residual sqrt(bushels to coop+1) 3 By type of plants**



From the residuals vs.  $X_i$  plot, all of them exhibit random scatter around zero, and there is no systematic pattern. So the linearity assumption is satisfied.

**Residual sqrt(bushels to coop+1) 3**



— Normal(-9e-14, 30.4788)

**Quantiles**

100.0%	maximum	93.00
99.5%		91.65
97.5%		61.78
90.0%		37.81
75.0%	quartile	19.28
50.0%	median	2.08
25.0%	quartile	-19.15
10.0%		-38.07
2.5%		-69.05
0.5%		-88.49
0.0%	minimum	-88.49

**Moments**

Mean	-9.31e-14
Std Dev	30.478792
Std Err Mean	1.6778023
upper 95% Mean	3.3005738
lower 95% Mean	-3.300574
N	330

**Fitted Normal**

**Parameter Estimates**

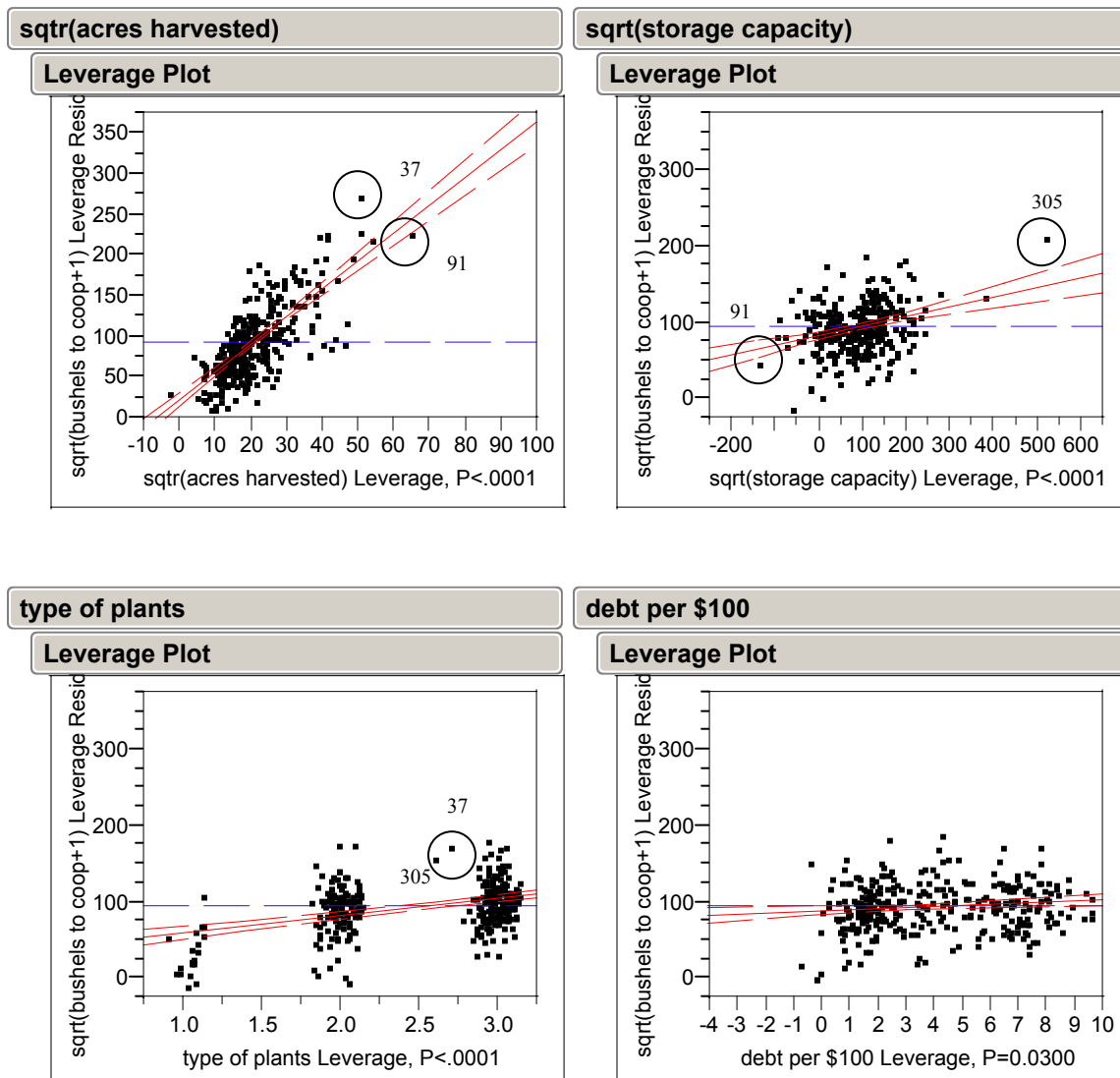
Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	-0.0000	-3.3006	3.30057
Dispersion	Sigma	30.4788	28.3173	33.00028

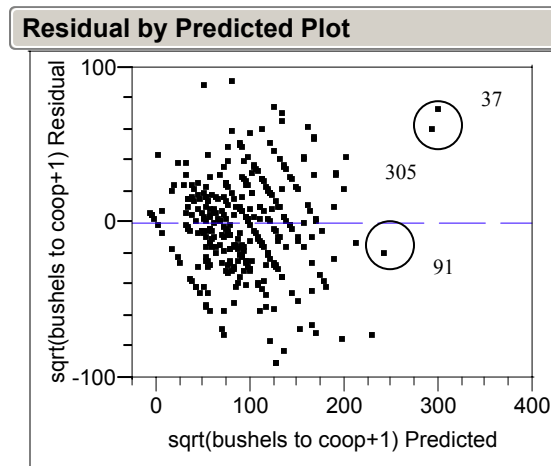
**Goodness-of-Fit Test**

Shapiro-Wilk W Test		
W	Prob<W	
0.984138	0.5126	

The Normal Quantile Plot has clear linear shape, which indicates the residuals are normally distributed. In the Goodness-of-Fit Test, the Shapiro-Wilk W Test also proves that.  $H_0$ : the residuals are normal.  $H_1$ : the residuals are not normal. The attained significance level is 0.5126 which is not small. Consequently, we should not reject the  $H_0$  and the residuals normality assumption is also met.

There is rule of thumb to test influential observations. Any  $h_{ii} > 2 \cdot (k+1)/n$  has high leverage. So if  $h_{ii}$  is larger than 0.06, it is influential observation. Actually there are 3 influential observations: #37, #91, #305. However, they are not outliers. These dots can be seen in the leverage plots and the residuals plot.





**IV. Interpretation of the Best Model:  $\text{sqrt}(\text{bushels to coop}+1) = 22.878245 \cdot \text{type of plants} + 3.4099344 \cdot \text{sqrt}(\text{acres harvested}) + 0.1264139 \cdot \text{sqrt}(\text{storage capacity}) + 1.3975387 \cdot \text{debt per } \$100$**

The coefficients of all these independent variables are positive, which mean the more acres of soybeans harvested and the larger the storage capacity on farm and the more dollars are financed with debt per \$100 and people prefer to sell their soybeans to a cooperatively owned processing plant or have no preference, the more bushels of soybeans people are willing to sell to a cooperative.

It makes a lot sense. If people have large acres of soybeans harvested, they are certain to have more soybeans sell to anywhere. And people who have a large storage capacity on farm usually have more land growing soybeans and harvest more as well. For people who finance more with debt tend to turn to a cooperative which usually provides service and whose benefits are derived and distributed equitably on the basis of use. So to join a cooperative could reduce some risk to those starters who are not familiar with the business and usually have more debt in finance. Type of plants is a categorical variable. 1: prefer to sell soybeans to a privately owned processing plant; 2: prefer to sell soybeans to a cooperatively owned plant and 3: no preference. So if people prefer cooperatives

than privately owned plant, they are more likely to consider cooperatives first when selling their products.

I still would like to analyze some variables which look much related to the dependent variable, but do not enter the regression equation. Take the coop or not for example, it seems natural to me at the first glance that people who belong to a cooperative tend to sell more soybeans to cooperatives. However, it is not true. The reason is in Tennessee, the cooperatives haven't developed fully as in some other states in the North and the West, like Minnesota. Here the cooperatives are not very efficient and are mostly traditional cooperatives. So people may be member of a cooperative and buy something from the cooperative but don't have the responsibility to sell their products to the cooperatives as an exchange. Thus, people usually choose the processing plant which is near from their farm and would incur less transportation fee. Or they would sell their soybeans to some other places which give a much better price. However, in the North and the West, the new generation cooperatives are much more popular, where the members have the liability of selling required amount of products to the cooperatives at a presettled price no matter whether they harvested enough soybeans or not. Therefore, after thinking a little longer, I would find it reasonable that whether farmers in the survey belong to a cooperative or not would not actually affect their decisions on where to sell their products.

And biodiesel market---the attitude the producers take toward whether biodiesel production will provide an important national market for soybeans in the next 10 years, also has little to do with people's choice of how much to sell to the cooperatives. I think it might affect whether they would increase the number of acres to grow soybeans and whether to expand their soybean business scale.

### **Conclusion:**

This project has examined the relationship of number of bushels the soybean farmers in TN would be willing to sell to a cooperatively owned plant with variables that I think

may affect it, such as whether they are currently a member of an agricultural cooperative; acres of soybeans harvested in 2001; the number of soybean bushels they can store on-farm; years experience farming; net income from farming in 2001; dollars financed with debt for every \$100; percent of their household's income came from off farm sources in 2001; highest education level; the attitude the producers take toward whether biodiesel production will provide an important national market for soybeans in the next 10 years; types of plants they would like to process soybeans; percent they typically sell through contracts.

The model established here is an inferential type model and sort of data summary type model rather than a predictive model. After finding out the model, the strength of relationships between the response variable and individual predictor variable becomes much clearer and the large set of data are summarized by a single equation. The characteristic of the data also becomes much obvious.

It seems that the cooperatives in TN do not play an active role in promoting soybean production for biodiesel at present. So, if possible, some changes in law or a new statute should be made based on the suitability of cooperatives organized under the statute to develop and finance processing projects requiring substantial capitalization.

## Appendix



### growers' views on biodiesel production & Marketing

The purpose of this study is to measure soybean producers' attitudes about biodiesel markets and formation of a cooperative to produce biodiesel from soybean oil in the West Tennessee area. Your response is important for obtaining an accurate measure of producers' views. Your participation is *completely voluntary* but will help us to serve you and other soybean growers. Only summaries of responses from the survey will be reported. Only researchers conducting the study will have direct access to the data. This questionnaire will take about 15 minutes to complete. You will be provided with an opportunity to request a copy of the study summary at the end of this questionnaire. We appreciate your participation.

#### Section I. Biodiesel Markets

Biodiesel is an alternative fuel that can be made from soybean oil. Blends of up to 20% biodiesel mixed with petroleum diesel fuels (B20) can be used in nearly all diesel equipment and are compatible with most storage and distribution equipment.

1. Please circle the rating that most closely matches your opinions on the following statements.

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
a. The U.S. markets for biodiesel will grow rapidly in the next 10 years	1	2	3	4	5
b. Biodiesel production will provide an important national market for soybeans in the next 10 years	1	2	3	4	5
c. If priced competitively with conventional diesel, I would be interested in using biodiesel from soybeans in a 20 percent blend on my farming operation.	1	2	3	4	5

2. Do you believe that biodiesel from soybeans could be profitably produced in West Tennessee? (Circle the answer.)

a. YES (Continue on to question 3.)

b. NO Please indicate reasons why \_\_\_\_\_ (Go to question 11)

3. Would you be willing to sell some or all of your soybeans directly to a biodiesel processing plant?

a. YES (Continue on to question 4)

b. NO (Go to question 11)

4. I would prefer to sell my soybeans to a processing plant that is (Please circle the answer):

a. Privately owned (Continue on to question 5)

b. Cooperatively owned (Go to question 6)

c. No preference (Go to question 6)

5. Please indicate the number of bushels you (on an average year) would be willing to sell to a plant, then go to question 11.

a. Through marketing contracts \_\_\_\_\_ bushels

b. On a spot basis (no contract) \_\_\_\_\_ bushels

6. Please indicate the number of bushels you (on an average year) would be willing to sell to a plant (Continue on to question 7).

a. Through marketing contracts \_\_\_\_\_ bushels

b. On a spot basis (no contract) \_\_\_\_\_ bushels

## Section II. Cooperative Production of Biodiesel

7. Would you be interested in participating in a new generation cooperative to produce biodiesel?

a. YES (Continue on to question 8)

b. NO Please indicate reasons why \_\_\_\_\_ (Go to question 11)

8. What is the minimum percent per year you would find acceptable on any investment you made in a biodiesel production facility? \_\_\_\_\_% (for example, an

### New Generation Cooperatives (NGC)

- Major focus is on "value-added" products, rather than commodities.
- Members sell raw products grown to their cooperatively-owned processing plant. Cooperative profits are then distributed back to members in proportion to raw product delivered.
- Shares allocate delivery rights so each share entitles & obligates a member to deliver soybeans to the cooperative
- Restricted or closed membership limits soybean deliveries to the amount the plant can absorb

8% return would mean for every \$100 invested, you would receive \$8 per year). (Continue on to question 9)

9. At the rate of return indicated above, and assuming a share price of \$2.25 per bushel with minimum purchase of 2,500 shares or rights to deliver 2,500 bushels (\$5,625 investment), would you be willing/able to purchase the minimum number of shares?
- a. YES (If Yes, continue on to question 10)
  - b. NO (If No, go to question 11)
10. At the rate of return you stated in question 9, please indicate the number of additional shares you would be willing and able to purchase above the minimum of 2,500 (if none, indicate by answering '0'). \_\_\_\_\_ additional shares (Continue on to question 11)

**Section III. Farming Operation and Producer Characteristics**

11. Are you currently a member of an agricultural cooperative?
- a. YES
  - b. NO

12. Acres of soybeans harvested in 2001 \_\_\_\_\_.

13. Please indicate the number of soybean bushels you can store on-farm.  
\_\_\_\_\_ bushels.

14. What percent of your soybeans do you typically sell through contracts?  
\_\_\_\_\_ %.

15. Other acreage harvested or used on the farm(s) you operate in 2001
- |   |       |                            |       |
|---|-------|----------------------------|-------|
| a. Cotton                                     | _____ | e. Pasture                 | _____ |
| b. Corn                                       | _____ | f. Woodland                | _____ |
| c. Wheat                                      | _____ | g. Other (Please describe) | _____ |
| d. Other Crops Harvested<br>(Please describe) | _____ |                            | _____ |

16. Number of livestock on your farming operation
- |              |       |                            |       |
|--------------|-------|----------------------------|-------|
| a. Beef Cows | _____ | e. Broilers                | _____ |
| b. Milk Cows | _____ | f. Sows                    | _____ |
| c. Yearlings | _____ | g. Horses                  | _____ |
| d. Heifers   | _____ | h. Other (Please describe) | _____ |

17. Your age in years \_\_\_\_\_.  
Years experience farming \_\_\_\_\_.

18. For the farm(s) I operate, I am (Please circle the best answer)
- a. A full owner (sole proprietorship)

- b. A part owner in a partnership, family held corporation, or other corporation
- c. A renter
- d. Other (please describe): \_\_\_\_\_.

