United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Development Division
Washington DC 20250

RDD Research Report
RDD-09-08

October 2009

Using FSA Data to Improve the Prices Received for Grain Program

William Cecere

EXECUTIVE SUMMARY

This paper looks at the feasibility of using Farm Service Agency (FSA) data to reduce response burden associated with the National Agricultural Statistics Service (NASS) prices received for grain questionnaire. Currently, NASS surveys grain elevators on the prices paid to farmers for commodities for the previous month and the middle of the current month. The FSA daily price estimates on these commodities were examined to determine whether they performed as well or better at predicting the NASS whole-month price than the NASS summarized data from the mid-month price portion of the questionnaire. If so, the NASS mid-month price section could be removed from the survey, thus reducing respondent burden.

The NASS mid-month and whole-month price data used for this research consisted of a price for every state involved in a given commodity's prices received program for each month over a three year period. The FSA data contained prices for every day of the year (except weekends and holidays) for every county in each of the states for a given commodity. To get an accurate state-wide price to correspond to the NASS data, the county prices were weighted by production and then averaged. To get an FSA mid-month price, prices for the three days out of each month closest to the 15th were selected and averaged. Also, another FSA price was evaluated that averaged all FSA daily prices up to the 17th of the month.

Two prediction models were examined to possibly replace the NASS mid-month prices in forecasting the whole-month price. These models used the FSA mid-month price and the FSA 17-day average price. The models were evaluated on their percent of variation explained (R-squared) and their mean squared error (MSE). Model validation was performed on the best models for each commodity using 2005 NASS and FSA data.

The best performing models were chosen for each commodity and the overall performance was determined to be good. Models tended to be more accurate for commodities such as corn and soybeans with more states in their prices received programs and less accurate for ones such as sunflower for oil with fewer. All models met their normality assumptions.

Following validation, models were assessed using mean squared prediction error (MSPE) to compare to the models' original MSE and examine whether there was a large variance inflation resulting in potential bias. Also, predicted prices were compared to actual prices in order to check for consistency. With the exception of soybeans, the NASS mid-month price outperformed the FSA price indicators for all commodities, indicating that the mid-month portion of the prices received questionnaire should not be removed.

RECOMMENDATIONS

1. This project's findings support that while the NASS mid-month price is sometimes inconsistent at forecasting the whole-month price for a few commodities, it is still the best known way to predict the NASS whole-month price. Therefore NASS should keep the mid-month portion of the prices received questionnaire.

2. Before considering the implementation of any future models that replace the mid-month portion of the prices received questionnaire, determine whether this will significantly reduce respondent burden.

Using FSA Data to Improve the Prices Received Questionnaire

William Cecere[1]

Abstract

The United States Department of Agriculture's National Agricultural Statistics Service (NASS) conducts a prices received for grain survey each month. This survey is used to assess monthly prices received by farms for a given commodity at grain elevators. Prediction modeling was performed that compared NASS mid-month prices to Farm Service Agency (FSA) prices in order to determine which did the best job of forecasting the NASS whole-month price, potentially reducing respondent burden if the NASS mid-month price was not surveyed. This was performed on nine commodities, comparing three models on each.

The prediction modeling and validation testing resulted in models for several commodities that accurately predict the mid-month price over time with low model variance and little or no bias. However, it was determined that the FSA prices did a relatively poor job at predicting the NASS whole-month price and should not be used.

Key Words: prediction modeling; validation; model variance.

[1] William Cecere is a mathematical statistician in the USDA's National Agricultural Statistics Service (NASS) – Research and Development Division (RDD), located in Room 305, 3251 Old Lee Highway, Fairfax, VA 22030.

1. INTRODUCTION

One of the primary goals of the National Agricultural Statistics Service (NASS) is to reduce respondent burden associated with its survey questionnaires. Another goal of the agency is to use available administrative data effectively to enhance or supplement NASS data. This paper addresses both of these goals in the context of the prices received survey.

Before the 15th of each month, NASS sends out a prices received questionnaire to sampled grain elevators in the states involved in the prices received program for a targeted group of commodities. The commodities for which the price information is requested varies by state. The survey collects the average price farmers received per unit (usually bushel) on certain commodities for the middle three days of the current month, as well as the quantity and value of commodities for the entire previous month. A goal of this paper is to evaluate the feasibility of reducing respondent burden by eliminating the mid-month portion of the survey instrument.

Currently, the NASS mid-month numbers are used to forecast the whole month's prices. Therefore, if the mid-month section of the questionnaire were removed, an alternate way of obtaining mid-month estimates to forecast whole-month prices would need to be in place. Using similar mid-month prices from other available data sources is one possibility. To assess the potential of such a replacement model, price data from the Farm Service Agency (FSA) were used as a source for testing. FSA price data contain a price, primarily gathered from auctions, for commodities each day of the month except weekends and holidays. This paper examines ways to use these data to compare with the NASS mid-month price and determines if they could potentially be used to better predict the NASS whole-month price.

The commodities used for this research are corn, soybeans, barley, oats, sorghum, durum wheat, flaxseed, sunflower for oil, and sunflower not for oil.

2. METHODS

To investigate this problem, a series of statistical models were examined to determine how to best predict the NASS whole-month price. These models are regression models where the response is the whole-month price and the regressors vary depending on the model. Price data from the NASS prices received questionnaire and FSA price data were used for the years 2001-2004.

First Model

The first model that was investigated used the NASS mid-month price as a predictor for the whole-month price. In this model (shown below), mprice is the mid-month price and wprice is the whole-month price. This may be used as a benchmark for any model involving FSA data, i.e., a model with FSA variables would not be considered unless it performs at least as well as a standard model using only NASS data.

Model 1: $$wprice = \beta_0 + mprice\beta_1$$

Second Model

The second model evaluated uses an FSA mid-month equivalent price to predict the NASS whole-month price. Every county in every state had an FSA daily price for each crop whether or not the crop was actually produced in the county. Also, each county had a price for every day of the month except weekends and holidays. To obtain an overall state price for each month from the FSA data sets, the prices were weighted by county production using NASS county production estimates from 2001 through 2004. The weighted FSA county prices were then averaged to obtain an overall state price for each date.

To get a number that was compatible with the NASS mid-month price, the three days closest to the 15th of each month were averaged to obtain an FSA mid-month price. For example, if the weekend fell on the 15th and 16th of a certain month then the prices from the 13th, 14th, and 17th were averaged and called the FSA mid-month price. This is somewhat similar to way the NASS mid-month price is determined from the questionnaire, which collects the prices received for each commodity purchased on the three days in the middle of the month.

Once the FSA mid-month prices were averaged, the data set was merged with the NASS data set. In this second model (indicated below), the FSA mid-month price is labeled as FSA_mprice.

Model 2: $$wprice = \beta_0 + FSA\_mprice\beta_1$$

Third Model

An alternative way that the FSA price data were used was to obtain an average price over all the days of each month up to the 17th instead of only using the middle three days. This was then used as the predictor of the NASS whole-month price. The FSA 17 day average price is denoted as FSAavg_price.

Model 3: $$wprice = \beta_0 + FSAavg\_price\beta_1$$

To check the consistency of the best selected models for each commodity, a validation process was performed using data from 2005. Depending on the commodity, some months were missing. In order to assess the results, charts were used to display differences between actual and predicted prices for US level prices. One commonly used statistical measure of model validation is the mean squared prediction error (MSPE). This is calculated as the mean of the squared predicted errors where the predicted value, $\hat{Y}_i$, is obtained using the original prices received model. Then, this is compared to the mean squared error (MSE) of the model. The MSE is used as an unbiased estimate for the variance of the original fitted regression model and the MSPE is used as a variance estimate for the validation. If the MSPE is much higher than the MSE, then the mean squared predicted error will be used. However, if they are relatively close, then the model is stable and not noticeably biased.

3. RESULTS

The modeling results are summarized in Table 1.

Table 1: Modeling results for the best model by commodity

Commodity	Best Model	Model #	R-squared	MSE
Corn	wPrice = 0.1469 + 0.9382 mPrice	1	0.942	0.0067
Soybeans	wPrice = 0.5038 + 0.9166 FSAavg_price	3	0.960	0.0296
Barley	wPrice = 0.4153 + 0.8355 mPrice	1	0.662	0.0748
Oats	wPrice = 0.2202 + 0.8929 mPrice	1	0.748	0.0299
Sorghum	wPrice = 0.2007 + 0.9527 mPrice	1	0.867	0.047
Durum Wheat	wPrice = 0.1952 + 0.9611 mPrice	1	0.840	0.056
Flaxseed	wPrice = 0.1738 + 0.9806 mPrice	1	0.966	0.0714
Sunflower Oil	wPrice = 1.3534 + 0.892 mPrice	1	0.857	0.7281
Sunflower Non-oil	wPrice = 2.727 + 0.7943 mPrice	1	0.731	1.4525

The models tested using corn data produced relatively strong results in terms of fit. This may be in part to the relatively large number of states (18) involved with the prices received estimation for corn. The models using FSA variables produced a significantly poorer fit in the R-squared values as well as increased model variance.

Soybean model results showed much promise as well. Like corn, a larger number of states were used in the pricing estimates when compared to the rest of the crops. Here the third model with the FSA 17 day average price was chosen over the NASS mid-month. The model fit is just as good as with Model 1, while the model variance is decreased significantly. The soybean models were the only ones where FSA variables were comparable or better than the NASS mid-month price at predicting the NASS whole-month price.

All remaining commodities (barley, oats, sorghum, durum wheat, flaxseed, sunflower oil, and sunflower non-oil) showed that the FSA data did nothing in terms of improvement from the first model. Therefore, Model 1 was selected for those commodities. Appendix A shows the results of all models for each commodity. The normality assumptions are shown to be valid for the chosen models by probability plots or qq plots.

The results of variance estimation after model validation are shown in Table 2. Most MSPEs are relatively close, indicating low bias and relative stability for those models. However, it is evident from the validation process that the durum wheat, flaxseed, and sunflower oil models are unstable and have significant bias. The corn and soybean models look as though there may be problems for future predictions because of the higher than expected model variance.

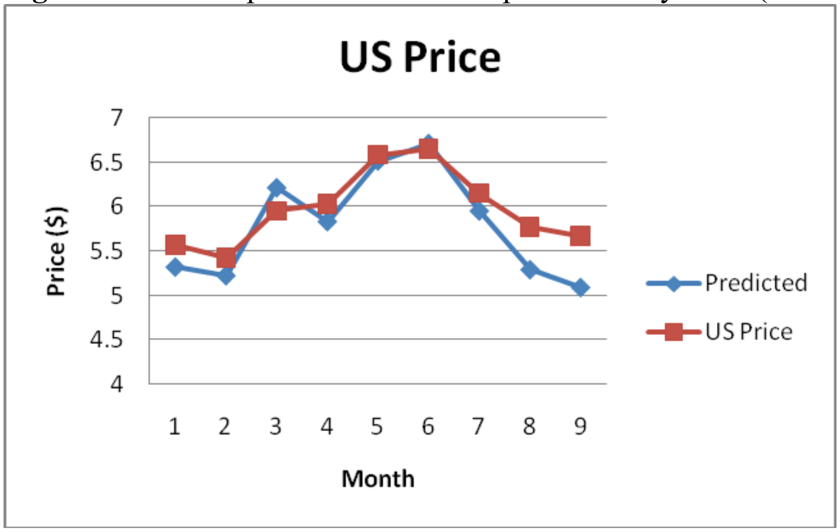
Table 2: Results of model validation

Commodity	MSPE	MSE
Corn	0.0129	0.0067
Soybeans	0.1087	0.0296
Barley	0.0829	0.0748
Oats	0.0363	0.0299
Sorghum	0.0494	0.047
Durum Wheat	1.2961	0.056
Flaxseed	1.085	0.0714
Sunflower Oil	6.4209	0.7281

Figure 1 shows the actual US price (red) and the best model's predicted US price (blue) for soybean prices received for several months in the calendar year of 2005. It is evident that the predicted prices are somewhat consistent with the actual prices and appear to capture changes in price fairly well with the exception of the last two months. The rest of the price validation plots are in Appendix B. There was no validation done on the sunflower non-oil model due to lack of data.

Some commodities show little change between predicted and actual US prices, indicating a strong prediction capability. However, the models for barley, flaxseed, durum wheat, and sunflower oil showed a significant bias and little consistency in predicting the actual US level prices. A possible reason for the relatively poor performance of the flaxseed model may be that there is only one state, North Dakota, involved in the modeling.

Figure 1: Plot of predicted vs. actual prices for soybeans (dollars per bushel)



4. CONCLUSIONS

Prices received results on all commodities are important not only to NASS but also to grain elevators and consumers that use NASS's published price data. Therefore, it is essential that published estimates from all prices received surveys be as accurate as possible.

In order to replace the mid-month section of the prices received questionnaire with auxiliary FSA data to predict the NASS whole-month price, the model must be consistently accurate over time and be able to reflect upcoming economic changes for the entire month. The models using the FSA mid-month and the 17 day average price did not perform as well as the NASS mid-month price at predicting the NASS whole-month price for any commodity except soybeans. The best performing model for each crop is indicated in Table 3. Many of the best fitting models are unstable when predicting the whole-month price.

A major factor that limits the net benefit of these results is the extent to which the respondent burden would be reduced if the mid-month section of the questionnaire were to be replaced by an FSA indicator. It does not appear that taking a small section of a short questionnaire will significantly reduce respondent burden. As seen in Appendix C, the questionnaire is already short (two pages) and the respondents would still have to be contacted for the remaining data. As this was one of the two agency goals listed in this paper, it is important to have information as to the nature of the respondent burden. If there would be no significant respondent burden reduction due to elimination of the mid-month section, then the accuracy of the estimates would be jeopardized while gaining very little.

Table 3: Best Performing Models

Crop	Model #	Stable after validation (Yes/No)
Corn	1	Yes
Soybean	3	No
Barley	1	Yes
Oats	1	Yes
Sorghum	1	Yes
Durum Wheat	1	No
Flaxseed	1	No
Sunflower Oil	1	No
Sunflower Non-Oil	1	N/A

5. RECOMMENDATIONS

From the results of this study, the following recommendations are made:

1. This project's findings support that while the NASS mid-month price is sometimes inconsistent at forecasting the whole-month price for a few commodities, it is still the best known way to predict the NASS whole-month price. Therefore NASS should keep the mid-month portion of the prices received questionnaire.

2. Before considering the implementation of any future models that replace the mid-month portion of the prices received questionnaire, determine whether this will significantly reduce respondent burden.

6. REFERENCES

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied Linear Statistical Models: Fourth Edition.* Boston, MA: McGraw-Hill

APPENDIX A: Results of model testing by commodity

*- indicates selected model

Table A1: Corn

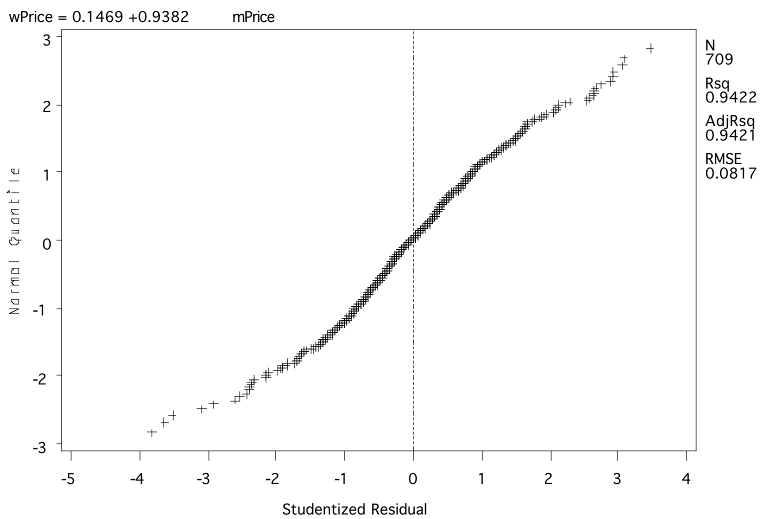
Model number	**R^2 (n=709)**	**Root MSE**	**QQ plot**
1 *	0.942	0.0817	
2	0.799	0.1383	
3	0.824	0.1293	

Table A2: Soybean

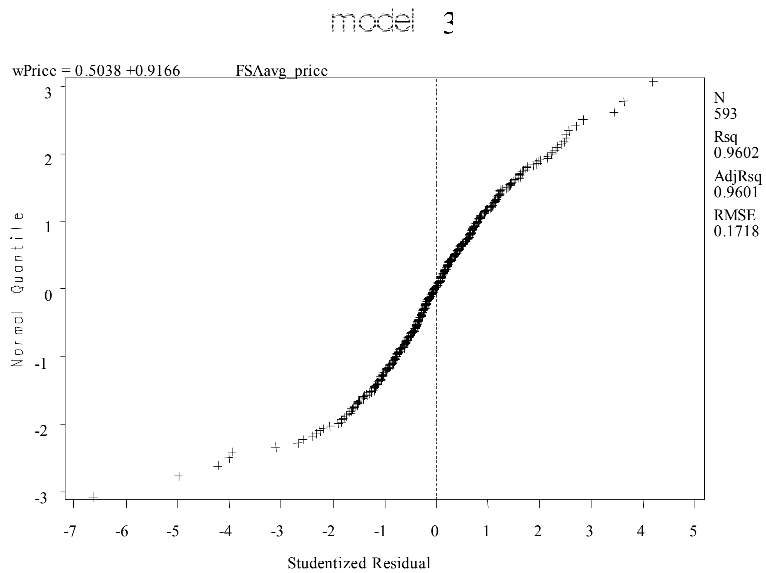
Model number	R^2 (n=593)	Root MSE	QQ plot
1	0.968	0.262	
2	0.959	0.173	
3 *	0.960	0.172	

Table A3: Barley

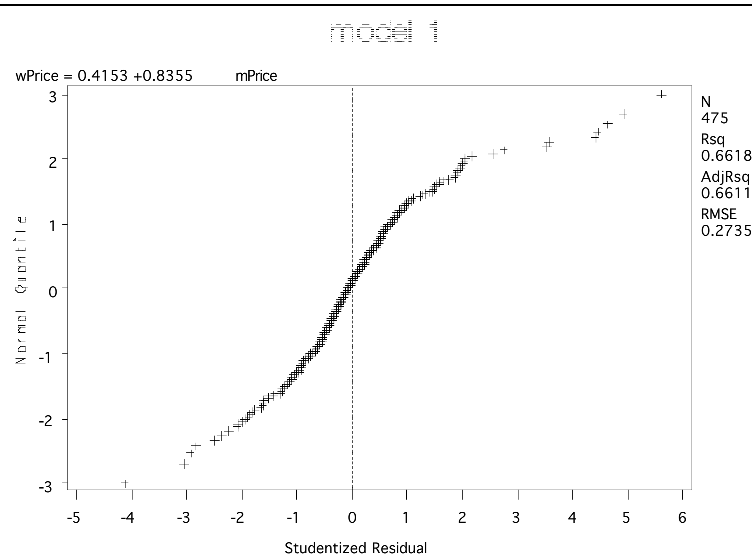
Model number	R^2 (n=475)	Root MSE	QQ plot
1 *	0.662	0.2735	
2	0.25	0.4081	
3	0.251	0.4076	

Table A4: Oats

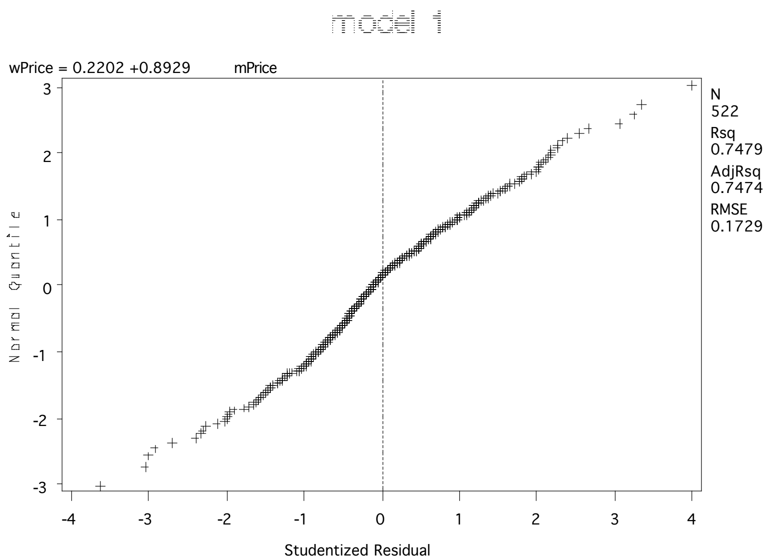
<table>
<tr><th>Model number</th><th>R^2 (n=522)</th><th>Root MSE</th><th>QQ plot</th></tr>
<tr><td>1 *</td><td>0.748</td><td>0.1729</td><td>
</td></tr>
<tr><td>2</td><td>0.328</td><td>0.278</td><td></td></tr>
<tr><td>3</td><td>0.306</td><td>0.2879</td><td></td></tr>
</table>

Table A5: Sorghum

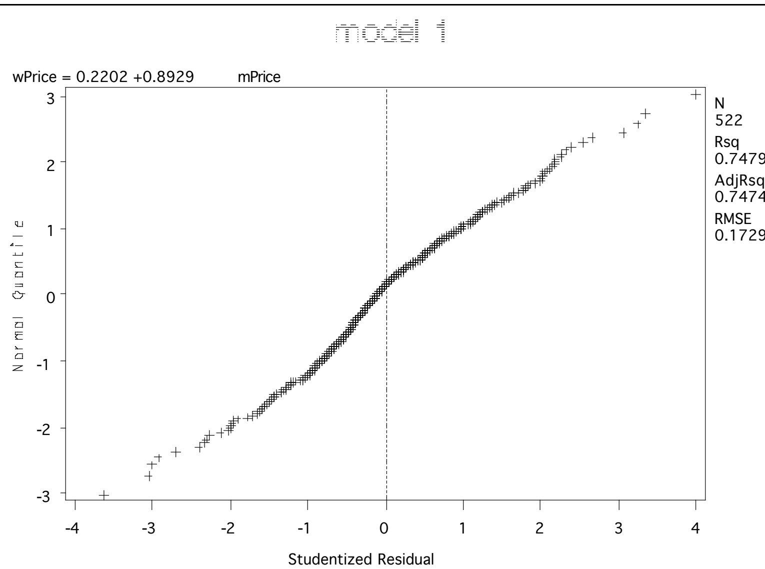
<table>
<tr><th>Model number</th><th>R^2 (n=331)</th><th>Root MSE</th><th>QQ plot</th></tr>
<tr><td>1 *</td><td>0.867</td><td>0.2169</td><td>
</td></tr>
<tr><td>2</td><td>0.754</td><td>0.2865</td><td></td></tr>
<tr><td>3</td><td>0.768</td><td>0.2782</td><td></td></tr>
</table>

Table A6: Durum Wheat

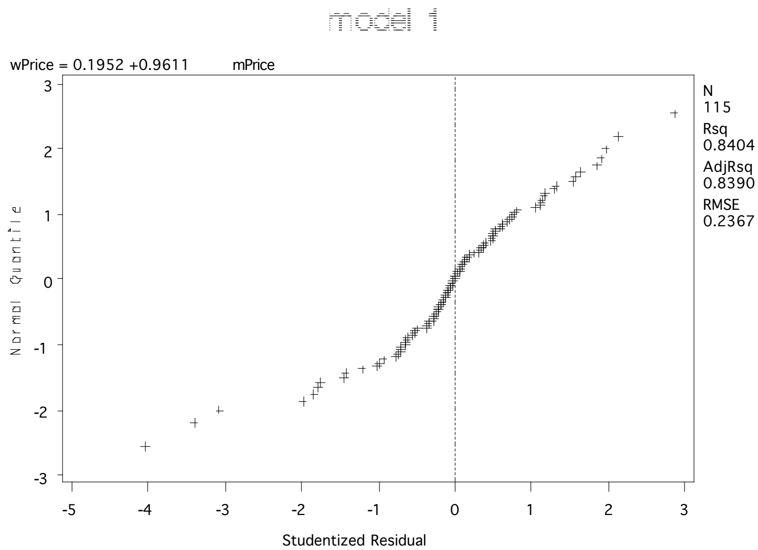
Model number	R^2 (n=115)	Root MSE	QQ plot
1 *	0.840	0.2367	
2	0.524	0.3848	
3	0.524	0.3641	

Table A7: Flaxseed

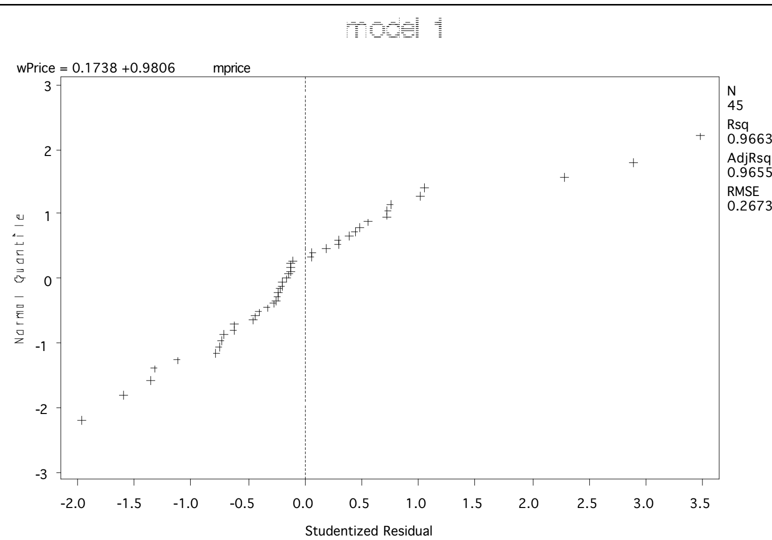
Model number	R^2 (n=45)	Root MSE	QQ plot
1 *	0.966	0.2673	
2	0.868	0.2756	
3	0.91	0.3606	

Table A8: Sunflower Oil

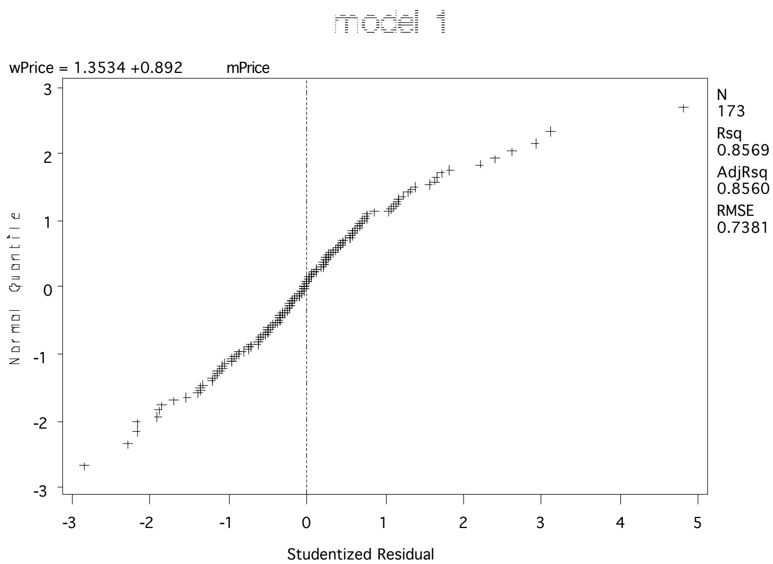
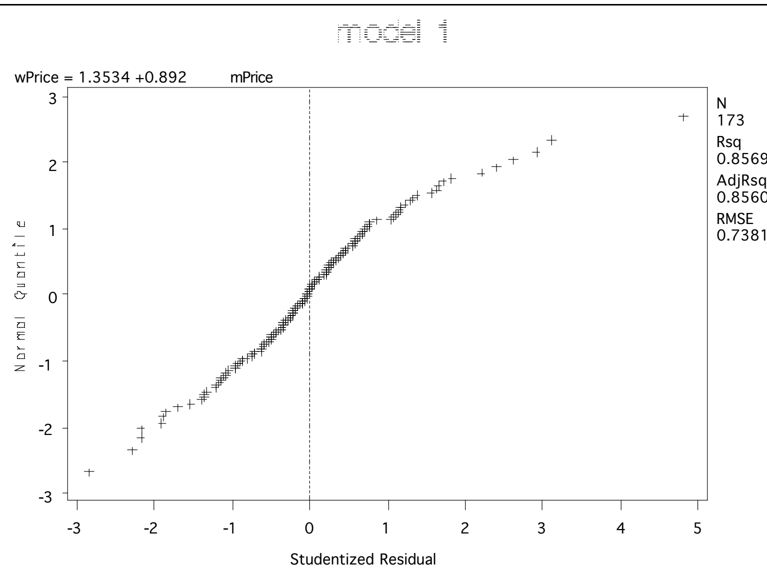
Model number	R^2 (n=173)	Root MSE	QQ plot
1 *	0.857	0.7381	
2	0.790	0.9459	
3	0.804	0.9131	

Table A9: Sunflower Non-oil

Model number	R^2 (n=159)	Root MSE	QQ plot
1 *	0.7310	1.2052	
2	0.426	1.6246	
3	0.433	1.6146	

APPENDIX B: Validation plots at US level by commodity, calendar year 2005

Figure B1: Plot of predicted vs. actual corn prices (dollars per bushel)

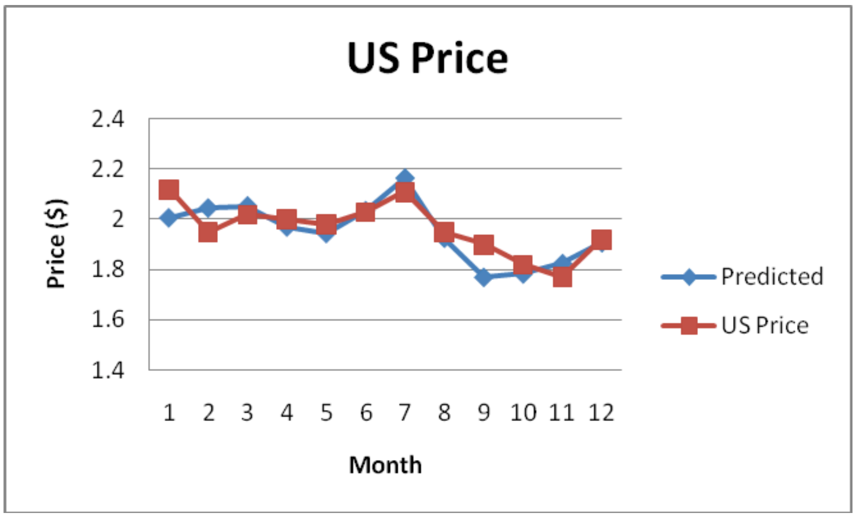


Figure B2: Plot of predicted vs. actual prices for soybeans (dollars per bushel)

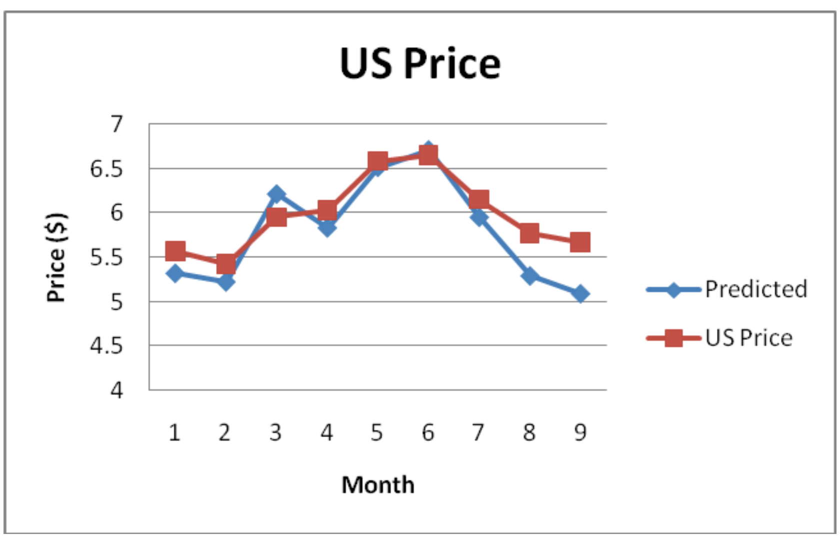


Figure B3: Plot of predicted vs. actual prices for barley (dollars per bushel)

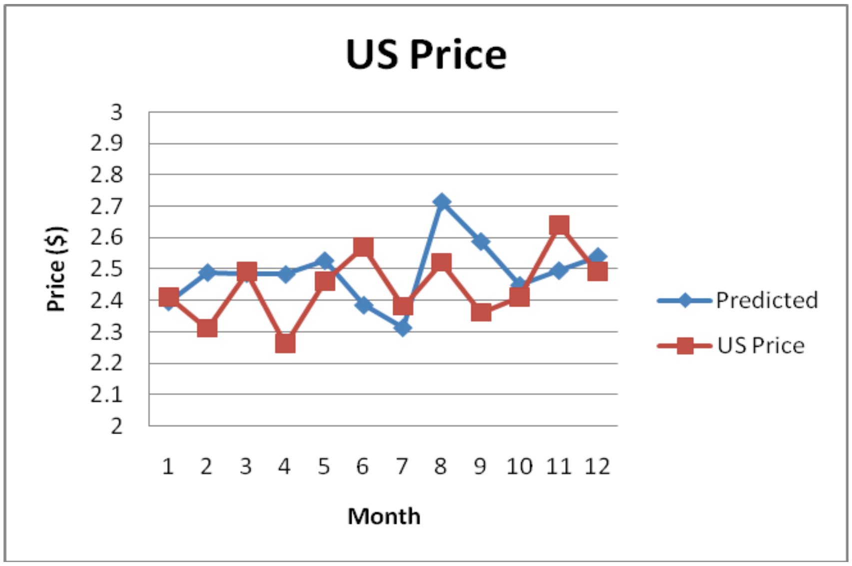


Figure B4: Plot of predicted vs. actual prices for durum wheat (dollars per bushel)

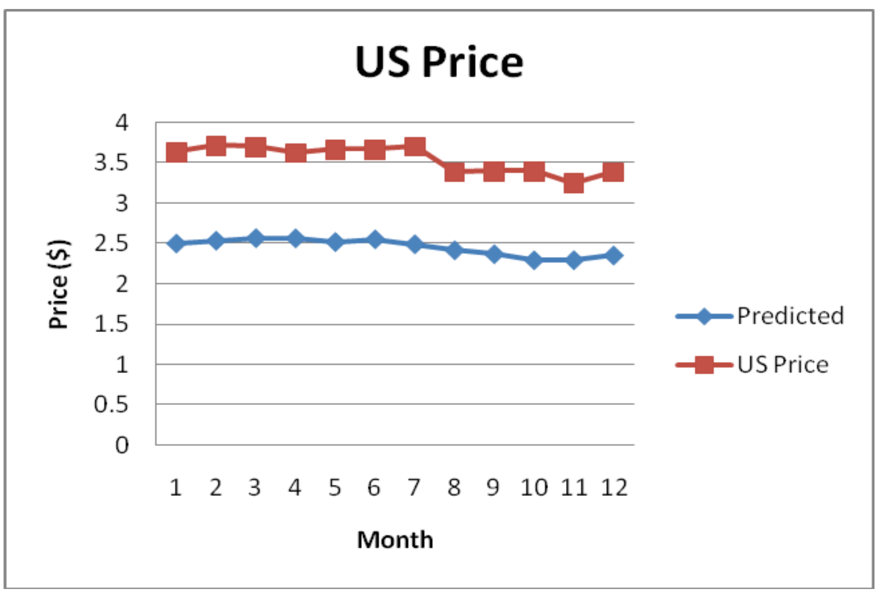


Figure B5: Plot of predicted vs. actual prices for flaxseed (dollars per bushel)

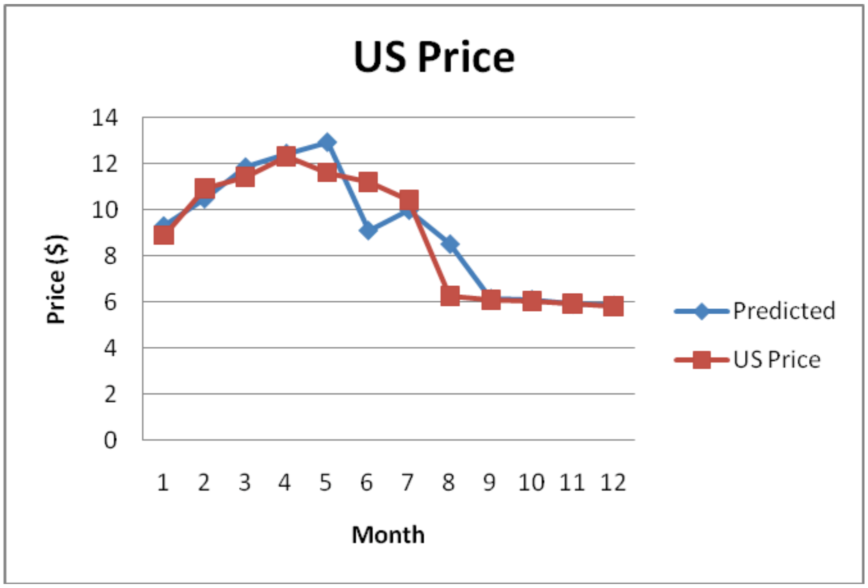


Figure B6: Plot of predicted vs. actual prices for oats (dollars per bushel)

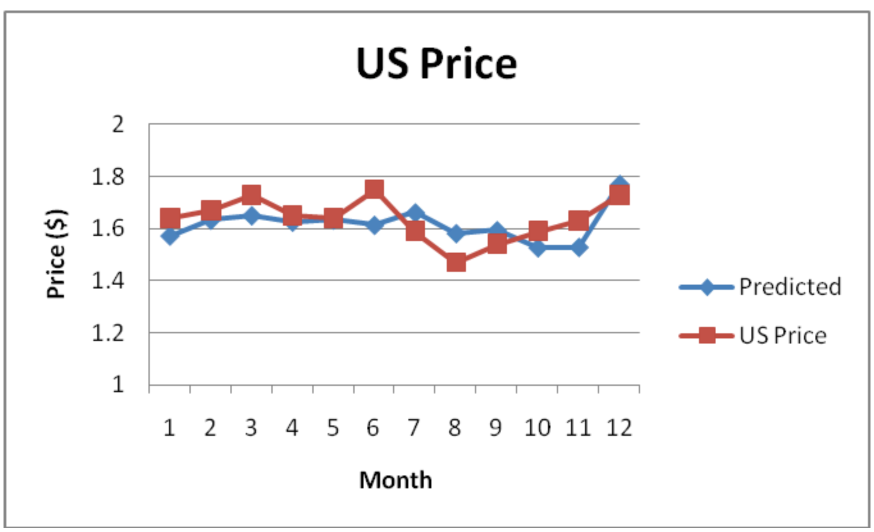
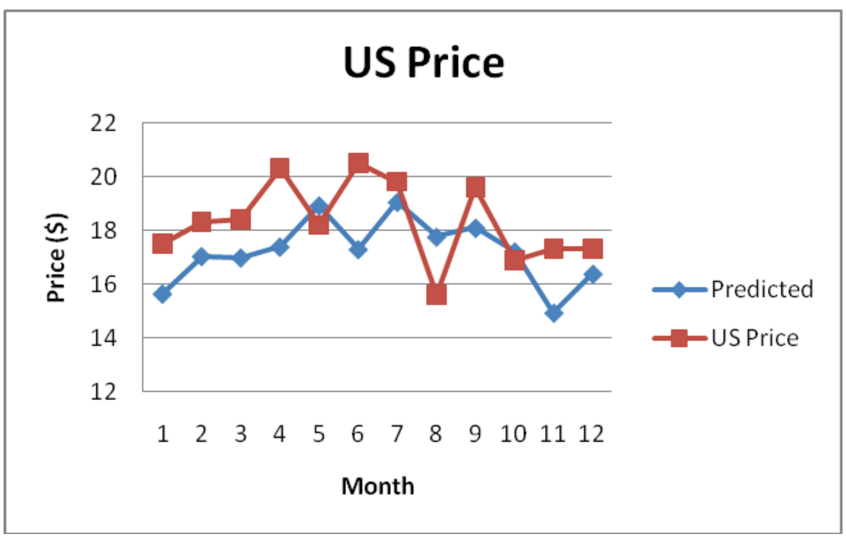


Figure B7: Plot of predicted vs. actual prices for sorghum (dollars per bushel)

Figure B8: Plot of predicted vs. actual prices for sunflower oil (dollars per bushel)



APPENDIX C: Prices received example questionnaire

Project Code 185 QID 050100

OMB No. 0535-0003: Approval Expires 12/31/2007

USDA

PRICES RECEIVED BY FARMERS FOR GRAIN

May 2006

NATIONAL AGRICULTURAL STATISTICS SERVICE

Montana Field Office
10 West 15th Street
Suite 3100
Helena, MT 59626
406-441-1240
Fax: 1-800-915-6277
E-mail: nass-mt@nass.usda.gov

«districtid» «countyid» «poid» strata: «targetstra» method: «method»
«edrsvycode»
«Opername»
«Wholename»
«Addrdelive»
«Placename», «STATEALPHA» «zip5»-«zip4»

Please make corrections to name, address and Zip Code, if necessary.

Dear Reporter:
The U.S. Department of Agriculture collects and publishes information on prices farmers receive for agricultural commodities. This information is used in estimation of farm income, the computation of counter-cyclical payments, State tax projections and rates, and for price studies and analysis of returns from various crops.

This survey is voluntary. Individual reports are combined to estimate State and U.S. prices. Information from individual reports is kept strictly confidential. Please fax your report or mail it in the enclosed postage paid envelope. Thank you.

Please fax 1-800-915-6277 or mail by May 16.

1. **Report grains and oilseeds purchased directly from U.S. Farmers in April 2006.**
   If no purchase of any commodity listed, check box ☐.

COMMODITY *(Report all varieties, grades, and qualities.)*	**TOTAL QUANTITY PURCHASED** Dry or "shrunk" basis Delivered any time prior to the end of April	**UNIT REPORTED** Circle one	**TOTAL VALUE** Gross dollars **plus** quality premiums **minus** quality discounts. 1/ **WHOLE DOLLARS**
Corn (*Yellow and White*)	011	012 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]	013 $
Feed Barley	051	052 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]	053 $
Malting Barley	071	072 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]	073 $
Oats	231	232 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]	233 $
Durum Wheat	111	112 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]	113 $
Winter Wheat	091	092 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]	093 $
Other Spring Wheat	131	132 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]	133 $

*1/ Add or deduct **quality factors** including-- test weight, protein content, foreign matter, damage, moisture content, and farmer delivery.*
***DO NOT** deduct check-off fees, drying, cleaning, handling, storage, grading or other marketing or service fees. For more information go to instructions.*

2. **MID-May PRICES: Report a mid-month price for commodities which were purchased May 12, 15, & 16, 2006.** If no purchase of any commodity listed, check box ☐.

COMMODITY *(Report all varieties, grades, and qualities.)*	**AVERAGE PRICE** *(Dollars and Cents)*	**UNIT REPORTED** Circle one
Feed Barley	054 $. __ __	055 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]
Malting Barley	074 $. __ __	075 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]
Oats	234 $. __ __	235 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]
Durum Wheat	114 $. __ __	115 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]
Winter Wheat	094 $. __ __	095 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]
Other Spring Wheat	134 $. __ __	135 Bu.[1] Lbs.[2] Tons[3] Cwt.[4]

Reported by: **Date** **Telephone** ( ) -

Title ____________ **Fax Number** ( ) -

OFFICE USE		ENTER ALL CODES THAT APPLY			
1 Re-contact/verified 2 Re-contact/ changed 3 Rounded 4 Circled Units	5 Changed Units 6 Office Error 7 Other- comment	801	802	803	804
		805	806	807	

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The time required to complete this information collection is estimated to average 10 minutes per response.

9910 MM DD YY

DATE: __ __ __ __ __ __

Response		Respondent		Mode		Enum.	Eval.	Office Use for POID
1-Comp 2-R 3-Inac 4-Office Hold 5-R – Est 6-Inac – Est 7-Off Hold – Est 8-Known Zero	9901	1-Op/Mgr 2-Sp 3-Acct/Bkpr 4-Partner 9-Oth	9902	1-Mail 2-Tel 3-Face-to-Face 4-CATI 5-Web 6-e-mail 7-Fax 8-CAPI 19-Other	9903	098	100	789 _ _ _ - _ _ _ - _ _ _ Optional Use 407 408

S/E Name