Determining Returns to Storage: USDA Data versus Micro Level Data

Joni M. Klumpp, B. Wade Brorsen, and Kim B. Anderson

J.M. Klumpp is a graduate student (joni.klumpp@okstate.edu), B.W. Brorsen is a Regents professor and Jean & Patsy Neustadt Chair (wade.brorsen@okstate.edu), and K.B. Anderson is a professor and extension economist (anderso@okstate.edu), Department of Agricultural Economics, Oklahoma State University.

Selected paper prepared for presentation at the American Agricultural Economics Association annual meetings, Providence, Rhode Island, July 24 - 27, 2005.

Abstract

USDA data are commonly used to determine producers' returns to storage. Aggregating data may result in a loss of information, leading to underestimated returns. This study compares USDA and elevator data from Oklahoma to determine how much USDA data underestimates returns. Results indicate USDA data only slightly underestimate returns to storage.
Keywords: aggregate data, returns to storage, information loss, data collection

Agricultural economists typically use aggregate data from the United States Department of Agriculture (USDA) instead of micro level data when conducting research on returns to storage (Hagedorn and Irwin; Brorsen and Irwin). This is mainly attributed to the fact that micro level data can be difficult to obtain and can be costly in both time and money. However, concerns about using USDA data in research regarding returns to storage do exist. One concern is the potential for information loss during the aggregation process that may ultimately result in the underestimating of the returns to storage. Another concern regarding the use of USDA data relates to the method used to collect the data. The USDA relies mainly on surveys of elevator managers for information regarding prices received, bushels produced and sold, and sale dates. It is possible that elevator managers do not supply accurate information on the surveys. For example, they may report an average or rounded price instead of an exact price or they may give a rough estimate on the number of bushels sold or produced. Thus, using USDA data as opposed to actual elevator data could result in inaccurate research conclusions, such as underestimating the relative profitability of farmer marketing strategies and a need for research that compares USDA calculated returns with returns based on micro level data exists.

The prices received by producers decline with distance to the market due to the decreased transportation costs at closer locations. Therefore, the producers with the highest prices, storing

close to the market will sell first, while the producers further away from the market are more apt to store their crops. Given that the opportunity cost of storing also declines with distance, Benirschka and Binkley found that locations farther from the market have a slight advantage in commodity storage. In other words, locations farther from the market should enjoy higher returns to storage than locations closer to the market. Wright and Williams suggest that this effect of location on returns to storage may be less when aggregate data are used to determine returns. If this is the case, then research that uses USDA data may underestimate the returns to storage received by producers.

Farmer marketing strategies are an important part of the farm management process and have been researched extensively throughout the years (i.e. Musser, Patrick, and Eckman; Zulauf and Irwin; Schroeder et al.). Researchers typically agree with the efficient market hypothesis that suggests that little profit can be made from trying to beat the market. Instead, farmers will receive an average price over the crop year. However, a recent view on farmer marketing decisions is that farmers actually do worse than average. The research on producer performance is limited to a few studies with different results. Hagedorn and Irwin found that farmers do tend to underperform the market; while a study by Brorsen and Anderson found that farmers perform above the market average. An important difference in these two studies is the data used by the researchers. Hagedorn and Irwin used USDA data and Brorsen and Anderson used micro level farm data. Further, the lower farmer returns found by Hagedorn and Irwin are due primarily to farmers storing too long. If USDA data is indeed limited by the aforementioned concerns, then the study by Hagedorn and Irwin may have underestimated the returns farmers received and underestimated their marketing abilities.

So the question remains, "How much does using USDA data underestimate returns to storage"? Thus, the objective of this study is to determine how much lower returns to storage based on USDA data are compared to returns based on micro level data. This will be accomplished by comparing Oklahoma Department of Agriculture data with rare micro level data obtained from three Oklahoma elevators. The accuracy of the aggregation method will be tested along with comparing the returns to storage computed for each dataset. The Oklahoma wheat market provides a strong test of aggregate data because of the significant price differences within the state. Seasonality of wheat sales will also be addressed in order to determine if producers are making inefficient marketing decisions by continuing to store after prices have peaked.

Theory

In a geographically dispersed market commodity prices decrease as distance to the market increases because of the increase in transportation costs. As mentioned in the introduction, the opportunity cost of storing also decreases as distance increases, which results in producers further from the market receiving higher returns to storage. Due to this observation, Benirschka and Binkley suggest that commodities stored at two different locations be treated as two different commodities. Aggregation of the commodities may result in a loss of information, creating a biased dataset that underestimates the returns to storage.

In order to further explain how the aggregation of data could create bias imagine a geographically dispersed market consisting of two time periods where location A is closer to the market than location B. As can be seen from table 1, the price at the closer location (A) is higher than that at the further location (B) for both time periods. Assuming an interest cost of 5% and storage cost of $0.10 at both locations, the nominal ($0.20) and net returns (- $0.06) to storage at

location A are less than the returns to storage at location B ($0.30, $0.05). This is consistent with the belief that returns increase as distance from the market increases. However, if all of location A sold in period one and all of location B sold in period two and the data is aggregated the results are much different. The aggregate price will be $3.20 in period one and $3.30 in period two and the nominal and net returns to storage are $0.10 and - $0.16, respectively. Thus, aggregating the data resulted in lower returns to storage than the disaggregated data and reported negative net returns even though the net returns at location B are positive. The example demonstrates how using aggregate data may lead researchers to underestimate the returns to storage.

Data

The micro level data for this study come from three elevators located in the southern, central, and northern regions of western Oklahoma. The data span nine crop years, from the spring of 1992 through the spring of 2001[a], and contain transactions of individual producer wheat sales at each elevator. Each transaction includes the number of bushels sold, the nominal price received per bushel, and the date of the sale. Harvest is a three-week period with beginning and ending dates that vary by elevator as well as by year. The harvest start date was determined by reviewing the daily transactions that occurred around the end of May or beginning of June. The date when the number of bushels sold increased noticeably and stayed relatively high for an extended period of time was used as the beginning harvest date. The southern elevator typically has an earlier harvest that begins around the end of May. Harvest at the central and northern elevators is slightly later, beginning around the first of June and the middle of June, respectively.

The returns to storage will be calculated with elevator data and with USDA aggregate data obtained from the Oklahoma Department of Agriculture. The aggregate data span from the

[a] Due to missing transactions at the northern elevator, the 1998 crop year was deleted from all datasets.

harvest of 1992 through the harvest of 2000 and contain statewide monthly average wheat production statistics. These statistics include the price received, total number of bushels produced, and the percent of wheat sold each month. Average number of bushels sold each month was calculated by multiplying the number of bushels produced by the percent sold each month. Since the USDA data contain only monthly averages, harvest is assumed to be the month of June.

Table 2 contains descriptive statistics for each elevator, as well as the USDA data. Average price received is the average nominal price producers received over the nine crop years. Harvest price is the average price received during the three week harvest period. These average prices are weighted within each year by the number of bushels sold. Percent of harvest sales is the percent of sales that occurred during the three week harvest, compared to sales for the whole year. As can be seen from table 2, producers at the southern elevator sell slightly more than half of their wheat at harvest. This is likely due to the earlier harvest date at the southern elevator. Producers may be trying to sell before the Kansas and Nebraska harvests begin and prices hit harvest lows. It is also interesting to note that harvest prices are higher than the average prices received. This agrees with Benirschka and Binkley that southern locations close to the market, such as Oklahoma, may experience smaller returns to storage than northern locations further away from the market.

Procedures

In order to compare returns to storage calculated with micro level data with returns calculated with aggregate data, the elevators' daily prices must be converted to monthly prices. This was done using a weighted average to calculate monthly prices across years for each elevator, where price was weighted within each year by the number of bushels sold. Average

harvest prices were then computed for each elevator, as well as the USDA data, based on the aforementioned harvest dates. Monthly returns to storage from harvest for each elevator and the USDA data are calculated using the following equation:

$$rtrns_i = price_i - hrvst \quad (1)$$

where $rtrns_i$ is the returns to storage from harvest for month i, $price_i$ is the nominal weighted-average price received per bushel for month i, and $hrvst$ is the weighted-average harvest price for each dataset. For example, the returns to storage from harvest for the month of August at the northern elevator would equal the average August price minus the average harvest price ($3.35). As previously discussed, the harvest price differs for each elevator, as well as for the USDA data.

The micro level data was aggregated using the same aggregation method as the USDA data. The individual producer data was aggregated by month and year and weighted monthly averages were computed using the same method as that mentioned above. Then, the bushel weighted monthly averages were aggregated by year in order to get an aggregate dataset similar to the USDA data set. Monthly returns to storage from harvest were calculated for the USDA-like data set using equation (1) and assuming the harvest price to be equal to the average June price.

The monthly returns to storage from harvest at each elevator were compared to the returns to storage from harvest calculated using the USDA data. If the returns computed using the USDA data are significantly less than the returns computed using the elevator data, then using aggregated data to determine returns to storage may result in smaller returns than are actually the case. It is also likely that using aggregate USDA data in storage research may result in a significant loss of information.

Due to the fact that Oklahoma wheat producers typically sell the majority of their crop close to or at harvest, seasonality of wheat sales is also an important factor. The frequency of sales in each month was calculated for each elevator, as well as for the USDA data using the following equation:

$$freq_i = \frac{sales_i}{\sum_i sales_i} \quad (2)$$

where $freq_i$ is equal to the percentage of total wheat sales that occurred in month i and $sales_i$ is equal to the total number of sales that occurred in month i. Comparing the seasonality of wheat sales with returns to storage will allow us to observe whether producers are continuing to store their crop after price has reached its peak.

Results

Figure 1 graphs the monthly nominal returns to storage from harvest for each elevator, as well as the USDA dataset and the USDA-like dataset. Clearly, the returns calculated using the USDA data are not much different than the returns calculated using the micro level data. This indicates that the USDA data only slightly underestimates the returns to storage. The USDA-like data closely resembles the actual USDA data, showing that the method used to aggregate the elevator data was consistent with the USDA method. The similarity of the USDA-like dataset with the that of the USDA dataset also indicates that the data collection process used by the USDA produces data that is consistent with actual elevator data.

Returns to storage are low close to harvest and start increasing around September, reaching their peak during November and December. The low returns during July and August are likely due to the beginning of the Kansas and Nebraska harvests. The northern market no longer has a demand for Oklahoma wheat due to the availability of local wheat. Thus, the

increase in wheat supply results in lower prices. One possible explanation for prices falling off in late December/early January is the occurrence of two world harvests. It is possible that due to the beginning of harvest in the southern-hemisphere the export demand for U.S. wheat decreases. The domestic demand for U.S. remains the same, but the available supply increases, driving down price. While the two world harvests theory is a possibility, it has not been found to be reflected in export shipment data.

Figure 2 graphs the frequency of wheat sales by month at each elevator and for the USDA dataset. As expected, the southern producers do most of their marketing at or very close to harvest. The central and northern elevators also exhibit a high percentage of producer wheat sales during the harvest months of June and July. Prices peak around late November and early January (figure 1), so storing past these months would be uneconomical for producers. The results in figure 2 show that very few wheat sales take place from February through the May.[b] The sales that do occur after January are relatively small in terms of bushels sold. Thus, there appears to be only a small number of sales that take place during the uneconomical time period.

Conclusions

This study is based on the belief that the aggregation of data can result in a loss of location information which causes returns to storage to be under estimated by USDA data. The objective was to determine how much USDA data underestimates the returns to storage compared to returns based on micro level data. The results indicate that the use of USDA data only slightly lower returns to storage and that the USDA data accurately reflects actual elevator transactions. Therefore, USDA data appear to be accurate and almost as reliable as micro level data.

[b] Due to its earlier harvest, the southern elevator does show increased sales in May.

The seasonality of wheat sales was also addressed in the research. The belief that producers store too long was not supported by the results. Oklahoma wheat producers tend to sell very close to harvest, likely due to an earlier harvest date. Only a small amount of wheat sales took place after prices hit their peak in November/January. Thus, there is some indication of producers storing grain longer than is economical, but it is likely a small number of producers.

References

Benirschka, M., and J.K. Binkley. "Optimal Storage and Marketing over Space and Time." *American Journal of Agricultural Economics* 77(1995): 512-524.

Brorsen, B.W., and K. Anderson. "Actual Farmer Market Timing." Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, Missouri, April 22-23, 2002. Available at http://agecon.lib.umn.edu.

Hagedorn, L.A., S.H. Irwin, and D.L. Good. "The Marketing Performance of Illinois Corn and Soybean Producers." Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, Missouri, April 19-20, 2004. Available at http://agecon.lib.umn.edu.

Musser, W.N., G.F. Patrick, and D.T. Eckman. "Risk and Grain Marketing Behavior of Large-Scale Farmers." *Review of Agricultural Economics* 18(1996):65-77.

Oklahoma Department of Agriculture. *Oklahoma Agricultural Statistics 1996.* Oklahoma City, Oklahoma, 1996.

Oklahoma Department of Agriculture. *Oklahoma Agricultural Statistics 2001.* Oklahoma City, Oklahoma, 2001.

Schroeder, T.C., J. Parcell, T. Kastens, and K. Dhuyvetter. "Perceptions of Marketing Strategies: Producers versus Extension Economists." *Journal of Agricultural and Resource Economics* 23(1998):279-293.

Wright, B.D., and J.C. Williams. "A Theory of Negative Prices for Storage." *The Journal of Futures Markets* 20(2000):59-71.

Zulauf, C.R., and S.H. Irwin. "Market Efficiency and Marketing to Enhance Income of Crop Producers." *Review of Agricultural Economics* 20(1998):308-331.

Table 1. Example of Aggregation Bias in Geographically Dispersed Market

	Period One Price	Period Two Price	Interest @ 5%	Storage Costs	Net Price for Period Two	Net Returns to Storage
Location A	3.20	3.40	0.16	0.10	3.14	- 0.06
Location B	3.00	3.30	0.15	0.10	3.05	0.05
Aggregate	3.20	3.30	0.16	0.10	3.04	- 0.16

Table 2. Descriptive Statistics for Elevator Data and United States Department of Agriculture Data

Descriptive Statistics	South	Central	North	OKDA
Average price received ($/bu.)	3.41	3.33	3.43	3.30
Harvest price ($/bu.)	3.38	3.25	3.36	3.21
Percent harvest sales	53.21 %	17.31 %	13.05 %	24 %
Average bushels sold at harvest	961	1728	1770	18,825
Number of observations	14470	7089	6389	108
Average beginning harvest date [a]	May 25	June 3	June 11	June 1

[a] Harvest is 3 weeks long and beginning and ending dates vary by year.

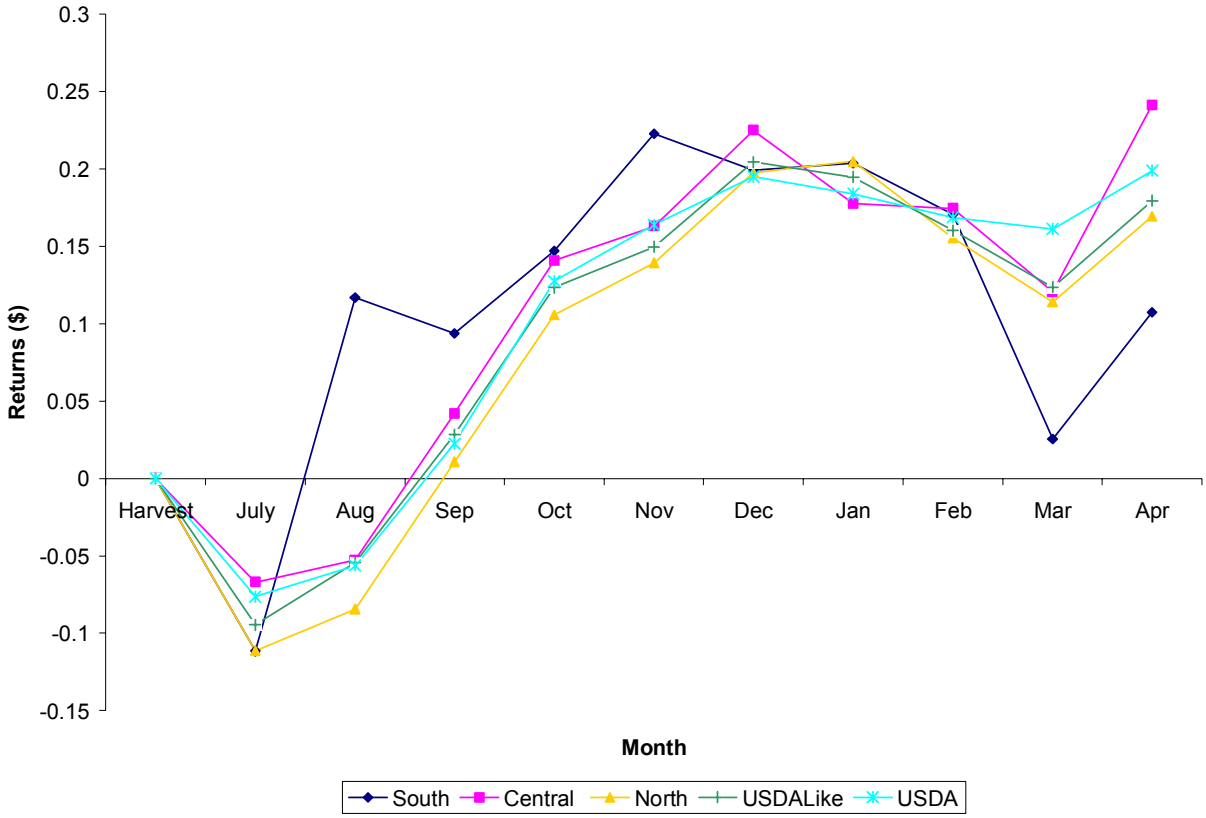

Figure 1. Nominal returns to storage from harvest

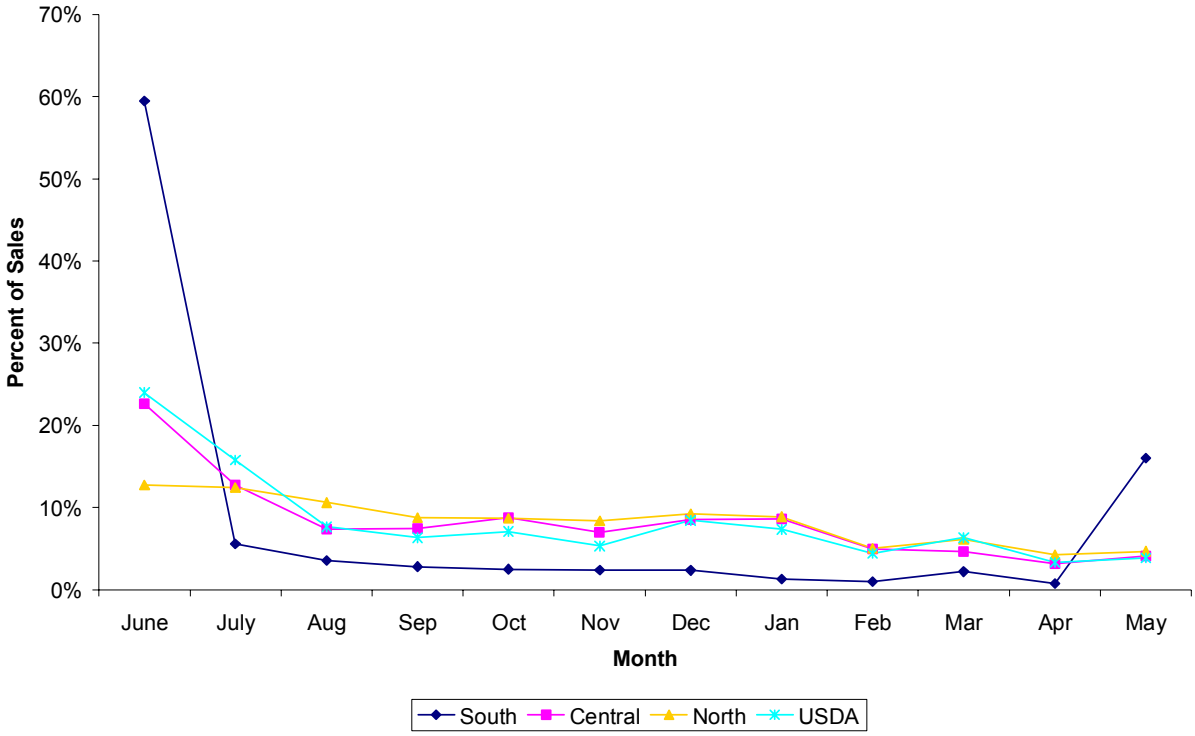

Figure 2. Frequency of wheat sales by month