Journal of Agricultural and Applied Economics, 39,3(December 2007):571–579

Determining Returns to Storage: Does Data Aggregation Matter?

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

Aggregate data are commonly used to determine returns to storage. However, recent studies have shown that aggregating data may lead to underestimated returns. This article compares aggregate and elevator data from Oklahoma to determine if aggregate data underestimate returns. We find no difference between the mean returns estimated with aggregate data and the mean returns estimated with transaction level data from grain elevators in Oklahoma.

Key Words: aggregate data, data collection, information loss, returns to storage

JEL Classifications: Q13

There has been a long debate in both finance and agricultural economics literature as to when it pays to store grain (i.e. Chang; Musser, Patrick, and Eckman; Zulauf and Irwin; Schroeder et al.). While researchers typically agree that there is little gain in trying to predict prices because markets are efficient, they disagree on how well producers are actually performing. For example, Anderson and Brorsen (2005) found that producers tend to perform above the market average. However, Hagedorn et al. found that farmers underperformed the market. An important difference between these two studies is the data used by the researchers. While Hagedorn et al. used aggregate data from the U.S. Department of Agriculture (USDA), Anderson and Brorsen used transaction-level data from grain elevators in Oklahoma. So could data aggregation explain the difference in the findings in these two studies?[1]

Brennan, Williams, and Wright argue that "no stocks are held at a monetary loss" and that "any apparent loss is an illusion from spatial aggregation" (p. 1009). Therefore, based on this view, what may appear to be storage at a loss in Hagedorn et al. could be the artifact of data aggregation (Benirschka and Binkley; Brennan, Williams, and Wright; Wright and Williams). An alternative explanation of why storage at a loss might occur is convenience yield (Kaldor; Working 1948, 1949).[2] But the concept of convenience yield is being challenged since an inverse carrying

Joni M. Klumpp is a credit analyst with US Bank and a former graduate student, B. Wade Brorsen is regents professor and Jean & Patsy Neustadt Chair, and Kim B. Anderson is Charles A. Breedlove Professor and extension economist, Department of Agricultural Economics, Oklahoma State University.

Helpful comments from Emílio Tostão are gratefully acknowledged.

[1] USDA wheat price data are collected through recall surveys on elevators around the 15th day of each month. The price data are weighted by state-level survey quantity data to create state-level prices. So the averages reported by elevator managers during recall surveys are further aggregated by the USDA to yield state-level price and quantity. Aggregation of spatial data reduces both producer heterogeneity and degrees of freedom.

[2] Convenience yields are the implicit benefits that accrue to the owners of physical stocks (processors) but not to the owners of contracts for future delivery (producers) (Yoon and Brorsen).

Table 1. Example of Aggregation Bias in Geographically Dispersed Market

Location	Period 1 Price	Period 2 Price	Interest at 10%	Storage Costs	Net Price for Period 2	Net Returns to Storage
Location *A*	3.20	3.71	0.32	0.20	3.19	−0.01
Location *B*	3.00	3.51	0.30	0.20	3.01	0.01
Aggregate[a]	3.20	3.51	0.32	0.20	2.99	−0.21

[a] Data are aggregated assuming that all of location *A* is sold in period 1 and all of location *B* is sold in period 2.

charge can be explained by risk aversion (Chavas) and by transactions costs (Benirschka and Binkley; Brennan, Williams, and Wright; Chavas, Despins, and Fortenbery; Wright and Williams).

The objective of this article is to determine differences in returns to storage estimated with transaction-level data and returns to storage estimated with market-level data. Market-level data are from the Oklahoma Department of Agriculture data, and microlevel data are from three grain elevators in Oklahoma. We also look at frequency of sales over time to determine if producers continue to store after prices have peaked. Results indicate that storage at a loss does occur in Oklahoma and that it cannot be explained by data aggregation.

Theory

The theory of optimal storage over time and space discussed here is largely derived from Benirschka and Binkley. Because of transportation cost, storage costs are minimized when storage occurs at the point of production. So when producers are spatially dispersed, those who are closer to the market tend to sell first because their higher grain price results in higher opportunity cost of forgone interest relative to producers farther from the market who have lower prices. To illustrate, consider two grain producing regions A and B. Let c_A and c_B be the transportation costs to ship grain from A and B to the market. Assuming that the market is competitive and producers face similar storage costs, the market price at any arbitrary time t is defined as

(1) $$p_m^t = \min[p_A^0\exp(rt) + c_A, p_B^0\exp(rt) + c_B],$$

where p_m^t is the market price at time t, p_A^0 and p_B^0 are grain prices in regions A and B at the beginning of the marketing season, c_A and c_B represent transport costs to ship grain from region A and region B, and r is the rate of interest at time t.

Equation (1) suggests that as long as $c_A > 0$ and $c_B > 0$, then the "cheapest" suppliers will sell first. Since transportation costs and returns to storage increase with distance, the "cheapest" suppliers are producers closer to the market. Thus, producers closer to the market sell early in the marketing season, and those farthest from the market sell later in the marketing year.

To illustrate how data aggregation underestimates returns to storage, we consider a two-period grain market supplied by two regions A and B. Region A is assumed to be closer to the market than location B, so the price at the closer location A is higher than that at the farther location B in both time periods (see Table 1). Assuming an interest cost of 10% and storage cost of \$0.20 at both locations, net returns to storage at location A (−\$0.01) are less than net returns at location B (\$0.01). This is consistent with the notion that returns to storage increase with distance from the market. However, if all of location A sold in period 1 and all of location B sold in period 2 and the data are aggregated, then results are much different. The aggregate price is \$3.20 in period 1 and \$3.51 in period 2, and the net return to storage calculated with aggregate data is −\$0.21. The example demonstrates how data aggregation may lead to underestimating returns to storage. But the interest costs assumed are larger than historical and the difference in prices across locations is larger than actually occurs within a state, yet the difference in storage costs is only two

Table 2. Descriptive Statistics for Elevator Data and USDA Data

Descriptive Statistics	South	Central	North	USDA	USDA-Like
Average price received ($/bu.)	3.42	3.33	3.44	3.34	3.40
June price ($/bu.)	3.36	3.24	3.36	3.28	3.35
Percent June sales	59%	23%	13%	24%	34%
Average bushels per transaction	819	1,491	2,148		
Number of observations	12,253	6,135	6,385		

cents. Thus, the example also suggests that the effect of aggregation on data within a state may be small.

Data and Procedures

The microlevel 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 and record transactions of individual producer wheat sales at each elevator.[3] 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 beginning harvest date was identified as the date when the number of bushels sold increased noticeably and stayed relatively high for an extended period of time. The southern elevator has an earlier harvest that typically begins around the end of May.[4] Harvest at the central and northern elevators is slightly later, beginning around the first of June and the middle of June, respectively. Commercial storage is widely available near producing areas, and over 90% of the grain produced in Oklahoma is stored commercially (National Agricultural Statistic Service). Small amounts of grain are held on farm mainly for seed.

The data likely do not include all transactions for some sellers. The data also do not include derivatives such as futures contracts, options, and forward contracts, so the data represent an incomplete picture of producers' marketing decision. These data weaknesses, however, should have little effect on measuring returns to storage.

The returns to storage are calculated with elevator data and with USDA aggregate data obtained from the Oklahoma Department of Agriculture. These data 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. Plots and formal statistical tests are used to determine any differences in mean net returns estimated with aggregate data and mean return estimated with disaggregate data. Formal statistical tests are the parametric paired t-test and its corresponding nonparametric version—the Wilcoxon signed-rank test.

Table 2 contains descriptive statistics for each elevator as well as the USDA and USDA-like data. Average price received is the average nominal price that producers received over eight crop years. Since the USDA data contain only monthly averages, averages are presented for June rather than a shorter harvest period so that means are comparable. These average prices are weighted within each year by the number of bushels sold. Percent of June sales is the percent of sales that occurred in June compared to sales for the whole year. As can be seen from Table 2, producers at the southern elevator

[3] Because of missing transactions at the northern elevator, the 1998 crop year was deleted from all data sets.

[4] Errors in the southern elevator data were found and removed; thus, if any erroneous data were missed, the data quality for this elevator may not be as good as for the other two elevators.

sell slightly more than half their wheat in June. This is likely due to the earlier harvest date at the southern elevator.[5] 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 at the southern elevator are slightly higher than those at the other elevators, partly because of an earlier harvest but also because the elevator is closer to the market. The differences in prices across the three dispersed markets is small, which alone casts some doubt on the likely importance of data aggregation.

Storage and interest costs were calculated for all elevators as well as the USDA and USDA-like data sets. The storage cost, set by the elevators, averages \$0.00085 per day and \$0.0255 per month.[6] Interest cost is calculated using the prime rate of the given year plus 2%. The prime rate is based on the prime rate charged by banks in June of the given year and is quoted from the Federal Reserve Bank of Kansas City. Daily interest costs at each elevator are calculated by multiplying the interest rate by the elevator's average harvest price and dividing the product by 365 days. The monthly interest cost for the USDA data set is determined using the same method, except the product is divided by 12 months. The cost of carry (storage cost plus interest cost) is then figured per day for the elevators and per month for the USDA data sets. Storage and interest charges begin accumulating immediately after the three-week harvest period ends at each elevator. Thus, the southern producers start accumulating storage and interest costs on June 15, the central producers start on June 25, and the northern producers start on July 1.

Gardner and Lopez argue that federally subsidized loan rates stimulate an increase in stocks held, especially in years when prices are higher than subsidized loan rates. Federally subsidized marketing loans may also distort the optimal marketing pattern by providing incentives to store longer. However, the quantity of Oklahoma wheat under federal loan is small. Less than 5% of the total wheat production in Oklahoma was under federal loan between 1995 and 2001 crop years, except in 1998, which is not included here, when the quantity of wheat under loan was about 15% of total production (Oklahoma State Farm Service Agency). The selling prices net of interest and storage at each elevator are

$$(2) \quad netprice_{tdk} = P_{dk} - days_k\left(\frac{hp_{tk}(z_t + .02)}{365} + S_{dk}\right),$$

where t is the year, d is the day, k is the elevator, $netprice_{tdk}$ is the net price, P_{dk} is the nominal price received on day d at elevator k, $days_k$ is the number of days after harvest at elevator k, hp_{tk} is the average harvest price at elevator k, z_t is the prime interest rate for year t, and S_{dk} is the storage cost per day. The net prices for the USDA data set are calculated using the following equation:

$$(3) \quad netprice_{ti} = P_i - mon\left(\frac{hp_t(z_t + .02)}{12} + S_i\right),$$

where t is year, i is month, $netprice_{ti}$ is the net price, P_i is the monthly price received, mon is the number of months after harvest, hp_t is the harvest price, z_t is the prime interest rate, and S_i is the monthly storage cost.

In order to compare returns to storage calculated with microlevel 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 on the basis of the aforementioned harvest dates. Monthly nominal returns to storage from harvest for each elevator and the USDA data

[5] Farms in south Oklahoma are about three times smaller than farms in north Oklahoma. Wheat is a minor crop, and crop share is common in southern Oklahoma. Many wheat sellers are landowners selling their share.

[6] The interest rates are for a one-year period so that the same interest rate can be used throughout the year. There is no minimum storage cost for grain in Oklahoma.

are calculated using the following equation:

$$(4) \quad rtrns_i = price_i - hrvst,$$

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 data set. 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).

The monthly net returns to storage from harvest for each elevator and the USDA data set are determined such that

$$(5) \quad netrtrns_i = netprice_i - nethrvst,$$

where $netrtrns_i$ is the net returns to storage from harvest for month i, $netprice_i$ is the net price for month i, and *nethrvst* is the average harvest net price.

The microlevel data were aggregated using the same aggregation method as the USDA. The individual producer data were 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 data set similar to the USDA data set. Monthly nominal returns to storage from harvest were calculated for the USDA-like data set using Equation (4) and assuming the harvest price to be equal to the average June price. Monthly net returns to storage from harvest were calculated using Equations (3) and (5).

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 notably 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.

We also looked at frequency of sales over time to determine if grain producers in Oklahoma continue to store grain after prices have peaked just like corn and soybean producers in Illinois.[7] The frequency of sales in each month was calculated for each elevator as well as for the USDA data using the following equation:

$$(6) \quad freq_i = \frac{sales_i}{\sum_i sales_i},$$

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.

Results

Figure 1 shows the monthly nominal returns to storage from harvest for each elevator as well as the USDA data set and the USDA-like data set. Figure 2 shows the monthly net returns to storage from harvest for each data set. Both figures show that the returns calculated using the USDA data are not much different than the returns calculated using the microlevel data. The paired *t*-tests (Table 3) show no difference between the mean net return estimated with the aggregate data and the mean return estimated with disaggregate data. The Wilcoxon signed-rank test led to the same conclusion and is not reported here. Thus, while Brennan, Williams, and Wright and Benirschka and Binkley argue that data aggregation is the likely explanation of the storage-at-a-loss puzzle, we find no differences in mean storage returns between aggregate and disaggregate data. Therefore, the difference between the findings in Hagedorn et al. and Anderson and Brorsen (2005) may not be explained by data aggregation.

As can be seen from Figure 2, negative net returns to storage are common. One explanation for the negative returns is the presence of processor convenience yields. Since processors receive a convenience yield from holding

[7] It is possible that stocks held reflect producer's cash needs rather than speculation or psychological biasness such as myopic loss aversion. In any case, if aggregation of quantity data does not matter, then the frequency of sales with aggregated data should not differ from the frequency of sales with disaggregate data.

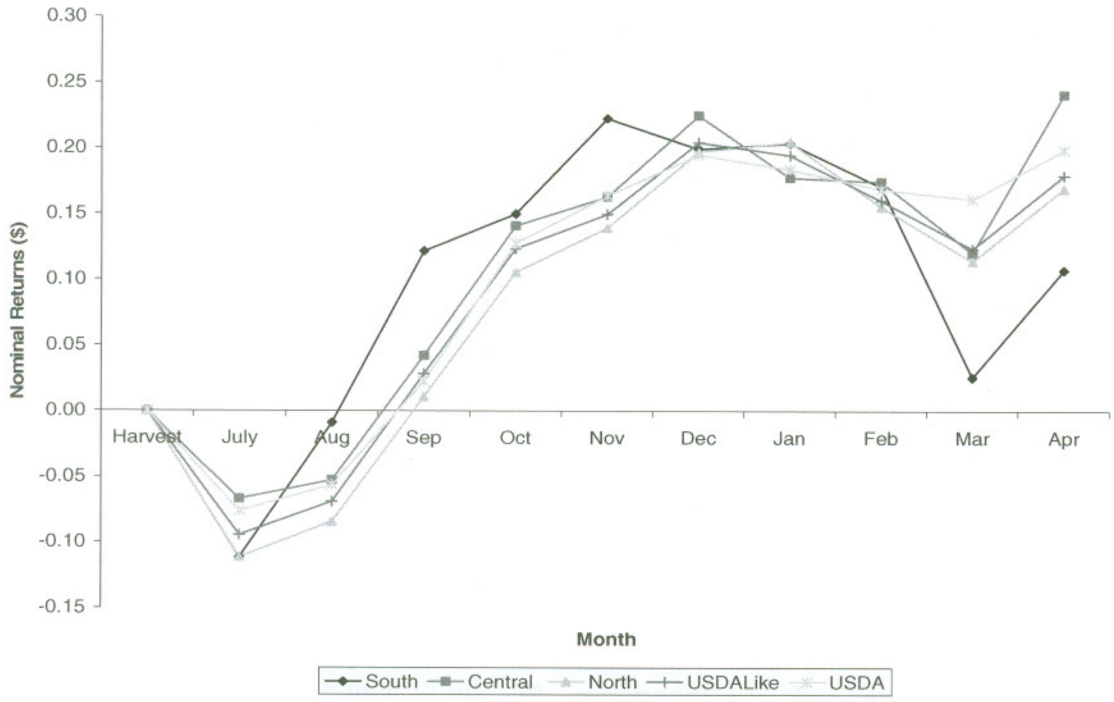

Figure 1. Nominal Returns to Storage from Harvest

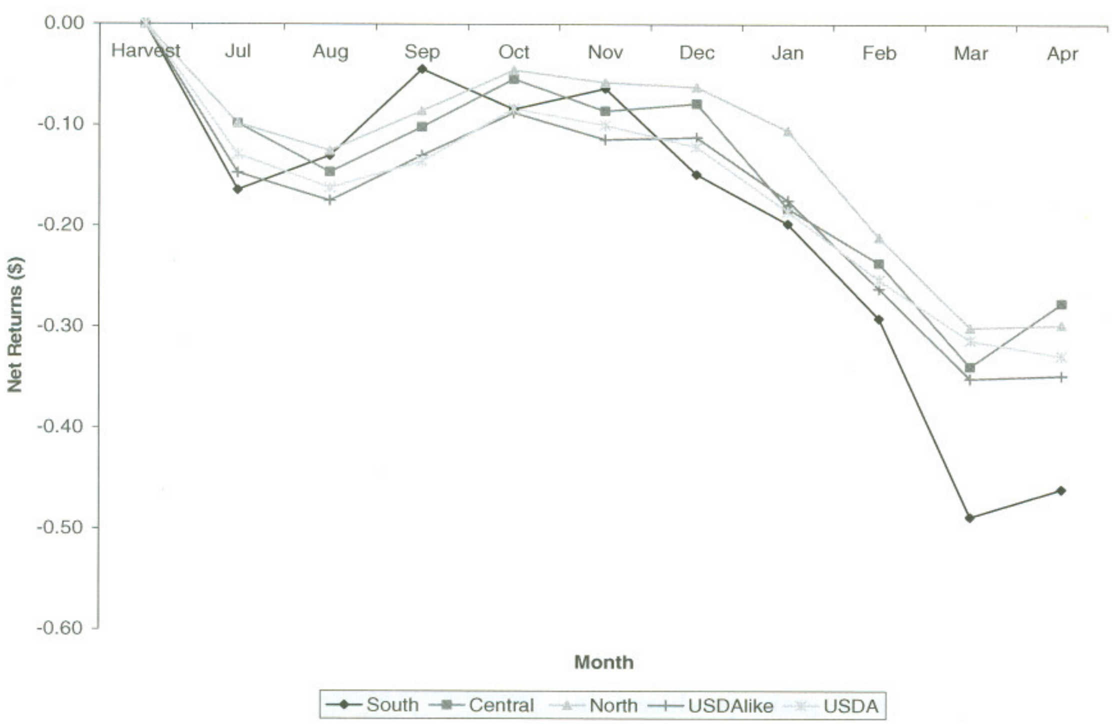

Figure 2. Net Returns to Storage from Harvest

Table 3. Difference between Average Monthly Net Return Estimated with USDA Data and Mean Return Estimated with Elevator Data

Elevator	Average Difference in Net Returns between USDA Data and Elevator Data	Parametric t-Statistic[a]	Two-Tailed p-Value
North	−0.02	−0.27	0.79
Central	−0.002	−0.02	0.98
South	0.03	0.44	0.66

[a] The Wilcoxon signed rank is not reported since it also led to failure to reject the null hypothesis of no difference between mean returns estimated with aggregate and disaggregate data at the 5% level.

stores of commodities (i.e., wheat) used in the production of other commodities (i.e., flour), they will look to purchase contracts for future delivery. If the price for the deferred delivery is below the harvest price, then negative returns to storage may arise. Brennan, Williams, and Wright, and Wright and Williams propose data aggregation as an explanation for negative returns to storage. However, our results are inconsistent with the data aggregation explanation since we find no difference between returns to storage estimated with aggregate data and returns to storage estimated with disaggregate data. Figure 2 shows that we still find storage at a loss even when microlevel data are used to calculate returns.

Cumulative returns to storage are low close to harvest and start increasing around September, reaching their peak during November and December. The negative returns during July and August are likely due to the beginning of the Kansas and Nebraska harvests. One possible explanation for prices falling off in late December/early January is the occurrence of two world harvests. It is possible that because of the beginning of harvest in the Southern Hemisphere, the export demand for U.S. wheat decreases. The domestic demand for U.S. wheat remains the same, but the available supply increases, driving down price. While two world harvests is a plausible explanation, we were unable to find any seasonality in export shipment data.

Figure 3 shows the frequency of wheat sales by month at each elevator and for the USDA data set. Results with both disaggregate data and aggregate data indicate that grain producers in Oklahoma hold stocks

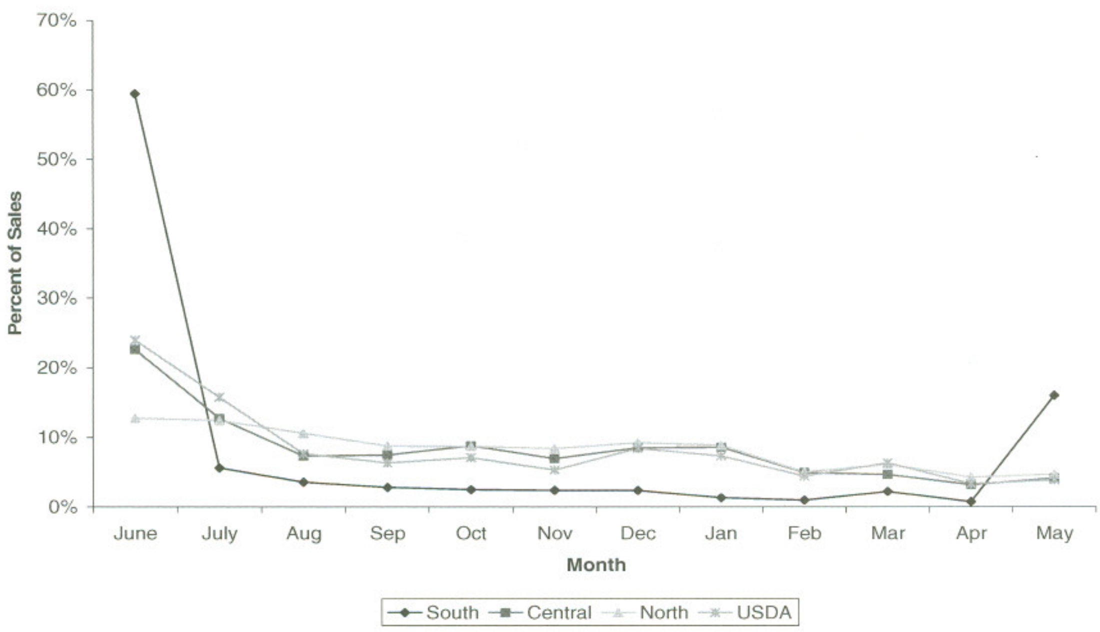

Figure 3. Frequency of Wheat Sales by Month

throughout the marketing year. This is inconsistent with Oklahoma State University Extension recommendations that advise producers to use mechanical marketing strategies, such as selling at harvest in the south and selling in lots of one-third on June 20, September 15, and November 15 (Anderson and Brorsen 2004). The results show net prices are highest at harvest (Figure 2), indicating that producers should sell at harvest regardless of location. However, net returns for September and November are close to harvest returns at the central and northern elevators, showing some support for the one-third/one-third/one-third marketing strategy.[8] Prices peak around late November and early January (Figures 1 and 2), so storing past these months would be uneconomical for producers. Figure 3 shows that most wheat sales occur before prices start declining in early February but that some wheat sales do take place during the more uneconomical time period of February to May.[9] One possible reason for holding stocks longer than it is economical is myopic loss aversion (Anderson and Brorsen 2005). Producers may store grain longer than is economical because they do not want to accept loss. But producers may also store grain for fear that prices will increase. In either case, psychological biasness would result in producers holding losing positions too long (Anderson and Brorsen 2005; Locke and Mann). Government programs are another possible explanation for producers' holding stocks after prices have peaked. However, little wheat was under federally subsidized loans in Oklahoma during the study period.

Conclusions

The storage-at-a-loss anomaly has long puzzled researchers. While most previous studies have offered the convenience yield explanation for "stocks held at a loss," some recent studies are questioning the convenience yield explanation and argue that the storage-at-a-loss puzzle is an illusion caused by aggregation of spatial data. Storage theory posits the optimal storage time increases with distance, so producers closer to the market sell first because they have lower returns to storage. However, while location differences among producers may explain why some producers store longer than others, data aggregation could mask spatial differences leading to the conclusion of "storage at a loss."

This article contributes to this growing debate by showing that data aggregation is not the only explanation for the storage-at-a-loss anomaly since we find no difference between returns estimated with disaggregate data and aggregated data from the USDA. Thus, results are inconsistent with the recent argument advanced by Brennan, Wright, and Williams and Williams and Wright that storage at a loss is due to data aggregation. Results with both aggregate and disaggregate data also indicated that grain producers in Oklahoma hold grain stocks throughout the whole marketing year. A possible reason for limited responsiveness to returns to storage is myopic loss aversion.

The answer to the question "Could data aggregation explain the difference in the findings in Anderson and Brorsen (2005) and Hagedorn et al.?" is clearly no. Data aggregation does not appear to be a problem. The problem could be in the choice of the benchmark. Oklahoma is closer to the Gulf, and therefore producers in Oklahoma have an incentive to sell early in the marketing season. Figure 2 shows that net returns decrease sharply after December, so a 12-month benchmark is a poor choice for Oklahoma. But a 12-month benchmark might work reasonably well for corn produced in Illinois since it is near the center of U.S. corn production.

[Received February 2006; Accepted November 2006.]

[8] This study uses interest rates for borrowers. Since many farmers have little or no debt, their opportunity cost of interest is lower. Thus, storage might provide a greater return than a bank checking or savings account.

[9] Because of its earlier harvest, the southern elevator does show increased sales in May.

References

Anderson, K.B., and B.W. Brorsen. "Mechanical Marketing Strategies." Oklahoma State University Cooperative Extension Service, F-592, April 2004.

———."Marketing Performance of Oklahoma Farmers." *American Journal of Agricultural Economics* 87(2005):1265–70.

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

Brennan, D., J. Williams, and B.D. Wright. "Convenience Yield without the Convenience: A Spatial-Temporal Interpretation of Storage under Backwardation." *The Economic Journal* 107(1997):1009–22.

Chang, E. "Returns to Speculators and the Theory of Normal Backwardation." *Journal of Finance* 40(1985):193–208.

Chavas, J.P. "On Competitive Speculation under Uncertainty: An Alternative View of the Inverse Carrying Charge." *Journal of Economics and Business* 40(1988):117–28.

Chavas, J.P., P.M. Despins, and T.R. Fortenbery. "Inventory Dynamics under Transaction Costs." *American Journal of Agricultural Economics* 82(2000):260–73.

Federal Reserve Bank of Kansas City. *Bank Prime Loan*. Internet site: www.federalreserve.gov/releases/h15/data/m/prime.txt (Accessed on June 9, 2005).

Gardner, B.L., and R. Lopez. "The Inefficiency of Interest-Rate Subsidies in Commodity Price Stabilization." *American Journal of Agricultural Economics* 78(1996):508–16.

Hagedorn, L.A., S.H. Irwin, D.L. Good, and E.V. Colino. "Does the Performance of Illinois Corn and Soybean Farmers Lag the Market?" *American Journal of Agricultural Economics* 87(2005):1271–79.

Kaldor, N. "Speculation and Economic Stability." *Review of Economic Studies* 7(1939):1–27.

Locke, P.R., and S.C. Mann. "Professional Trader Discipline and Trade Disposition." *Journal of Financial Economics* 16(2005):401–44.

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.

National Agricultural Statistic Service. *Grain Stocks*. Internet site: http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1079. Multiple issues (Accessed September 2006).

Oklahoma Department of Agriculture. *Oklahoma Agricultural Statistics*. Oklahoma City: Oklahoma Department of Agriculture, 1996–2001.

Oklahoma State Farm Service Agency. PK9 (State Level) Loan Summary Inquiry. Report generated from electronic database. September 20 2006.

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):279–93.

Working, H. "Theory of Inverse Carrying Charge in Futures Markets." *Journal of Farm Economics* 30(1948):1–28.

———."The Theory of Price of Storage." *American Economic Review* 39(1949):1254–62.

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

Yoon, B.S., and B.W. Brorsen. "Market Inversion in Commodity Futures Prices." *Journal of Agricultural and Applied Economics* 34(2002):459–76.

Zulauf, C.R., and S.H. Irwin. "Market Efficiency and Marketing to Enhance Income of Crop Producers." *Review of Agricultural Economics* 20(1998):308–31.