Cointegration between Prices of Pecans and Other Edible Nuts: Forecasting and Implications

Wojciech Florkowski and Yue Lai*
Department of Agricultural and Applied Economics
University of Georgia
Georgia Station
Griffin, GA 30223
(770)228-7231

* Contact Author

Cointegration between Prices of Pecans and Other Edible Nuts:
Forecasting and Implications

Abstract

The use of cointegration relationship among prices of pecans, almonds, and walnuts is found to forecast pecan prices more accurately than some of the best time series models. This indicates that cointegration also exist among substitutes. The findings can be used by the pecan industry in decision making.

Introduction

The state of Georgia produces the most pecans in the world. Pecan buyers and sellers, however, often face price uncertainty because of price fluctuations. Pecans can be substituted by other nuts in selective uses. In response to changing market conditions, food manufacturers and consumers not only select between grades of the same nut, but also substitute one type of nut for another. Economic theory indicates that prices of substitutable commodities are related because of substitution effects. Therefore, we may expect pecan prices are related to prices of other edible nuts.

Since the development of the concepts of cointegration, considerable interest has been generated in testing economic relationships among variables. Long-run relationships result from the tendency of economic variables to move together. The finding of cointegration between economic variables indicates the existence of a long-run relationship between the variables. Moreover, such cointegration relationships can be used in making forecasts. Previous studies (e.g., LeSage, 1990; Kim and Mo, 1995; Shoesmith, 1995) have found that error correction models based on cointegration relations often outperformed other time series models. Past research (e.g., Ardeni, 1989; Bessler et al., 1991; Chowdhury, 1991), however, has focused mostly on fundamentally related variables. Little empirical work in examining long-run price relationships among substitutes has been seen in the literature.

The primary objective of this paper is to examine the cointegration relationship between pecan prices and prices of almonds and walnuts and to develop an error correction model (ECM) based on the cointegration relationship to forecast pecan prices.

The forecast performance of the error correction model is then compared with those of a Bayesian vector autoregressive model (BVAR) and a restrictive vector autoregressive model (RVAR). BVAR and RVAR have been frequently used in time series forecasting and have been found to be among the best multi-variate time series models (e.g., Bessler et al. 1986; Kaylen, 1988; Zapata et al., 1990). Successful forecasts will provide valuable insights into the general performance of the pecan market for growers, shellers, and end users. Shellers could anticipate possible price changes in the near future, while growers could adjust price negotiations for their in-shell pecans. This paper is presented as follows: The next section describes the data. The modeling procedures are then discussed. Section three and four present and analyze the forecasts. Implications and discussion are given in the final section. **Data**

The price data are those of two grades of shelled pecans ("fancy-halves" grade and "fancy" grade), two grades of almonds ("supreme" grade and "selective-sheller-run" grade), and two grades of walnuts ("combination-half-and-pieces" grade and "combination-light-half-and-pieces" grade). Prices of two grades of almonds are averaged and so are the prices of two grades of walnuts. Therefore, four price series -- prices of higher-grade pecans (denoted as grade-1 pecan prices), prices of lower-grade pecans (denoted as grade-2 pecan prices), average (means of two grades) prices of almonds (denoted as almond prices), and average prices of walnuts (denoted as walnut prices) -- are used in model building. The data are of weekly price quotes from "The Food Institute Report" covering the period of February 7, 1994 through June 17, 1996 consisting of 119 observations. All models are estimated using the first 70 observations

through June 26, 1995. Out-of-sample forecasts and evaluation are made for the period of July 3, 1995 through June 17, 1996 (with a total of 49 observations). **The models**

The Phillips-Perron (1988) unit root tests indicate that all the price series are nonstationary. The price series, however, became stationary after first differencing, indicating integration of order 1 or I(1). Therefore, first-differenced data are used for model estimation for all the models.

1. Cointegration and error correction models (ECM)

The notion of a cointegration relationship is that while some related economic variables follow a random walk process, they may move together in the long run, forming an equilibrium relationship or cointegration relationship. If a variable moves away from the equilibrium, it will return to the equilibrium. This process is called error correction. Thus if we find the cointegration relationship among variables, we can use it to forecast the movements of these variables.

Following previous work (e.g., Johansen, 1988), let y_t be a vector of m time series. Each series is integrated of one I(1) and thus the first-differenced series are stationary, i.e. I(0). If the series are cointegrated, a representation of error correction model can be expressed as

$$A(B)(1-B)y_t = -bz_{t-1} + u_t \qquad\qquad (1)$$

where A and B are defined as in (1), z_{t-1} is a r by 1 vector of error correction terms based on r cointegration relationships $z_t = a'y_t$ (r is called the rank of the cointegrating vectors a'), and u_t is the disturbance term. Hence, the error correction model in (1) is essentially a VAR in differences with r lagged error correction terms in each of the equations

(Shoesmith, 1995).   The Johansen test for cointegration using maximum likelihood is preferred to the Engle-Granger two-step procedure for more than two variables (Shoesmith, 1995). The Johansen procedure can not only find multi-cointegrating vectors and test for their statistical significance, but also fully capture the underlying time series properties of the data in the form of VAR (Kim and Mo, 1995).

The Johansen procedure is applied to the price data series to test for a cointegration relationship. An optimal lag of 7 is used for the autoregressive lag structure in (1) as selected on the basis of the Tiao-Box procedure applied previously.   The approach suggested by Johansen (1991) is employed to examine the appropriateness of the inclusion of intercepts in the cointegrating vectors. The approach involves estimating both the restricted model (without intercepts) and the unrestricted model (with intercepts), computing the eigenvalues of both models, and using a χ^2 statistic to test the hypothesis that the inclusion of intercepts has inflated the eigenvalues (and therefore the number of cointegrating vectors) in a statistically significant way.   The test results rejected the hypothesis.   Therefore, intercepts are included in the cointegrating vectors in the cointegration tests. Table 1 presents the results of the Johansen cointegration test.   For the null hypotheses of r=0, r=<1, and r=<2, the test statistics are all greater than the 90% critical value, indicating the existence of cointegration relationships with a rank greater than 2. For the hypothesis of r =<3, however, the statistic is less than the critical value, indicating the acceptance of the null hypothesis. Therefore, a cointegration rank of 3 is concluded for the price series. Table 2 shows the three cointegrating vectors.

To examine the forecast performance of the error correction model, vectors 1 through 3 of the cointegration relations (table 2) are used in equation (1) to generate forecasts. High mean squared errors from the model are observed. Suspecting that overparameterization of the ECM may have caused the large forecast errors of the model, lag length is then reduced one by one and the corresponding mean squared forecast errors are computed and compared. The number of cointegrating vectors is also reduced to see if the forecast performance of the model can be improved. The results show that the ECM with only one cointegrating vector and no autoregressive independent variables produces the lowest mean squared forecast errors. Among the three cointegrating vectors, the first one produces the lowest RMSE, while the third one generates the highest RMSE indicating the dominance of individual cointegrating vectors in the order of 1 to 3. Therefore, the ECM with only one error correction term based on the first cointegrating vector is used in forecasting.

2. Bayesian vector autoregression (BVAR)

The Bayesian methodology in VAR modeling was introduced by Litterman (1979, 1986) to alleviate the problem of overparameterization. Litterman observed that most economic variables could be reasonably approximated as following a random walk process. By specifying some priors, the closeness to the random walk, the impact of increasing lags, and the interaction between variables can be dictated. The modeling procedure used in this study follows that by Bessler and Kling (1986). Let a VAR in differenced form take the following form

$$(1-B)y_t = A(B)(1-B)y_t + u_t \qquad\qquad (2)$$

where A are the coefficient vector with $A_{ij}(k)$ (for equation I and variable j at lag k) distributed independently and normally, B is the lag operator, and u_t is the disturbance term. The standard deviation of A_{ij} is then specified as

$$\lambda/k^Y \qquad\qquad \text{if } I=j$$

$$\delta=$$

$$\lambda w s_i /k^Y s_j \qquad \text{if } I \neq j$$

where s_i and s_j are standard deviations of residuals from univariate autoregression of series I and j. λ is the prior standard deviation of the coefficient of the first lag of the own variable, Y is the prior controlling the decay speed of standard deviations of coefficients of own lagged variables after the first lag, and w is the prior dictating the degree of interaction between variables.

The lag length of the VAR in equation (1) is determined using the Tiao-Box (1981) likelihood-ratio test

$$M(k_2,k_1)=[n-0.5-(k_2-k_1)m][\ln S(k_1)-\ln S(k_2)]$$

where n is the effective sample size, k_1 and k_2 is the shorter and longer lag length, respectively, m is the number of series, and $S(k_1)$ and $S(k_2)$ are the determinants of the residuals of the VAR for k_1 and k_2, respectively. $M(k_1, k_2)$ has an asymptotical χ^2 distribution with $m^2(k_2-k_1)$ degrees of freedom. Based on a sample size of 58, the use of this test leads to the selection of 7 as the optimal lag length.

A mean of zero is set for the coefficient of the first lag of own variable in equation

(2) since the data have been first-differenced (Kaylen, 1988). A three dimensional search for the optimal values of (λ, Y, w) is conducted over the values of (0.00, 0.01, 0.25, 0.50, 0.75, 1). For λ, it is found that the optimal value is over the interval of 0 and 0.01. Therefore a search with an increment of 0.001 from 0.00 to 0.01 is conducted for the value of λ. Based on the criterion of minimum log determinant of the error covariance matrix of out-of-sample one-step-ahead forecasts for 12 period, the set of (λ=0.001, Y=1, w=0.25) are selected as the symmetric Bayesian priors. Previous studies (e.g., Bessler and Kling, 1986) have shown that the use of asymmetric Bayesian priors based on prior knowledge of economic relationships led to better forecasts. Therefore the asymmetric Bayesian priors are specified by keeping the values of λ and Y as 0.001 and 1, respectively, and setting the values of w as in table 3. The selection of w in table 3 is based on prior knowledge of the relationships between these nuts.

3. Restricted VAR (RVAR)

Another approach to deal with the problem of overparameterization of VAR is not to use the same lag length for all the endogenous variables in a VAR. That is to restrict the lag length of individual series based on some model selection criteria. Various studies (e.g., Kaylen, 1988) have reported good forecast performances from such models. The selection of individual lag length in this study follows Hsiao's procedure (Hsiao, 1979) based on Akaike's criterion of final prediction error (FPE). The results of model selection are given in table 4.

Forecast formulation

The error correction model, the Bayesian vector autoregression and the restricted autoregression are constructed using the first 70 observations. Out-of-sample forecasts are made 1 through 5 steps ahead based on these models. The models are updated continuously by including newly available observations using the Kalman filter and forecasts are generated from the updated models. This forecasting process continues through the end of the data period. For the error correction model, the cointegrating vectors are updated once every five periods. This forecast procedure produces out-of-sample forecasts of 49 for one-step ahead, 48 for two-step-ahead, 47 for three-step-ahead, 46 for four-step-ahead, and 45 for five-step ahead, respectively.

Forecast analysis

The root mean squared errors (RMSE) of forecasts from individual models are shown in table 5. The error correction model produced the lowest RMSE for both grades of pecans for all the forecast period. While the Bayesian vector autoregression (BVAR) was more accurate than the restricted vector autoregression (RVAR) in forecasting one-step and two-step ahead, the latter did better in making longer-step-ahead forecasts. The relative performances of the various models are consistent for both grades of pecans. That is, if a model outperformed another model in forecasting one grade of pecans, it also did better for another grade of pecans.

While the RMSE criterion indicates the overall performance of a forecast model, this descriptive statistic may be sometimes misleading because MSE is the average of squared forecast errors. Ashley, Granger, and Schmalensee (1980) developed an approach to test for significance of differences of mean squared errors from individual

forecast models. Defining $\_t = e_{1t} - e_{2t}$ and $\Sigma_t = e_{1t} + e_{2t}$, the AGS test involves the

$$\Delta_t = \alpha_1 + \alpha_2 [\Sigma_t - m(\Sigma_t)] + \mu_t$$

following least square regression

where e_{1t} and e_{2t} are the forecast errors from models with higher RMSE (model 1) and

lower RMSE (model 2), respectively; $m(\Sigma_t)$ is the sample mean of Σ_t.; μ_t is the error term

assumed independent of Σ_t; α_1 is the difference in mean squared errors from model 1 and 2

while α_2 is proportional to the difference in forecast error variance from the two models

(Bradshaw and Orden, 1990). The test is then on the null hypothesis $\alpha_1 = \alpha_2 = 0$ against the

alternative that both coefficients are nonnegative and at least one is positive. If either

coefficient is significantly negative, model 2 can not be judged as statistically better than

model 1 in terms of RMSE reduction; if one coefficient is insignificantly negative, a one-

tailed t test can be used on the other coefficient estimate; if both coefficient estimates are

positive, a four-tailed F test can be employed (Ashley, Granger, and Schmalensee, 1980).

        The AGS test is applied to the forecast errors of individual models and the results

of AGS test for equality of RMSE are presented in table 6.   The error correction model

generated significantly lower RMSE than the Bayesian vector autoregression in one

through three step ahead forecasts for both-grade prices of pecans and in four-step ahead

forecasts for prices of grade-two pecans. The relative performances between the error

correction model and the restricted vector autoregression are similar to those between

ECM and BVAR except for the three-step-ahead forecasts for prices of grade-1 pecans in which the difference in RMSE is insignificant. While the Bayesian vector autoregression significantly outperforms the restricted vector autoregression in only one-step-ahead forecasts for prices of grade-2 pecans, the latter produces significantly lower RMSE than the former in three and four-steps-ahead forecasts for both grades of pecans and five-step-ahead forecasts for grade-one pecans.

Concluding Remarks

This study found that a cointegration relationship existed between prices of pecans, almonds, and walnuts during the data period. The superior forecast performance of the error correction model based on the cointegration as compared to other forecast models has further verified the appropriateness of modeling such a relationship between those nuts. These findings have important implications. Theoretically and intuitively, relationships among substitutes are quite different from those among fundamentally related variables such as homogeneous commodities in different regions. Unlike fundamentally related variables, the lack of a long run relationship among substitutes does not indicate the existence of arbitrage opportunities which have been frequently related to market inefficiency. On the other hand, it is changes in demand and supply conditions among substitutes that lead to substitution which drives prices. Therefore it is interesting to note that market forces can also bring prices of substitutes to a long run equilibrium relationship. The findings of the study indicates that markets for pecans, almond, and walnuts are related to such an extent that prices of these three nuts move towards an equilibrium in the long run. Theses findings also have important implications for the pecan

industry. Pecan prices have been influenced by the supplies of other edible nuts. The finding of a cointegration relationships between pecans and other edible nuts will help the pecan industry in making decisions on production, storage, marketing, and forecasting. For example, pecan buyers and sellers can make more accurate forecasts of pecan prices based on available price information of almonds and walnuts using the cointegration relation among these nuts. The use of error correction models in forecasting has minimal requirement on data because past price information on edible nuts can be easily obtained. Such forecast information can be used to guide decision making on pecan production, storage, and marketing, thus increasing economic efficiency.

Table 1. Statistics of the Cointegration Test

Eigenvalue	Ho: r_	L-max	L-max 90	trace	trace 90
0.4799	0	40.53	18.03	81.54	49.92
0.3056	1	22.61	14.09	41.01	31.88
0.1909	2	13.13	10.29	18.40	17.79
0.0815	3	5.27	7.50	5.27	7.50

Notes: r is the number of cointegrating vectors. L-max and trace are the test statistics based on the formulas given in Johansen (1988). L-max 90 and trace 90 are the corresponding critical values.

Table 2. Eigenvectors of the Cointegration Relationship

Eigenvectors	Pecan 1	Pecan 2	Almond	Walnut	Constant
1	-33.622	38.842	-4.318	37.999	-74.738
2	-9.411	9.853	-1.225	-4.654	12.014
3	-25.641	22.896	6.892	-14.475	27.123

Note: Pecan 1 and pecan 2 refer to higher and lower grades of pecans, respectively.

Table 3. Tightness Priors (w) for the Asymmetric BVAR

Equation	Lagged Variables			
	Pecan 1	Pecan 2	Almond	Walnut
Pecan 1	1.0	0.3	0.1	0.7
Pecan 2	0.3	1.0	0.1	0.7
Almond	0.1	0.1	1.0	0.1
Walnut	0.7	0.7	0.7	1.0

Note: The overall tightnessλ and the decay priorΥ are 0.001 and 1, respectively.

Table 4. Results of Lag Selections for the Restricted VAR

Equation	Lagged Variables			
	Pecan 1	Pecan 2	Almond	Walnut
Pecan 1				6
Pecan 2				6
Almond				
Walnut			4	

Note: The numbers refer to the number of lags.

Table 5. RMSE of Forecasts from Individual Models

Model	Grade	Forecast			Steps	
		1	2	3	4	5
ECM	Pecan 1	0.08708	0.11852	0.14105	0.16941	0.19382
	Pecan 2	0.08759	0.12375	0.15318	0.17409	0.19233
BVAR	Pecan 1	0.09188	0.13320	0.16805	0.20237	0.23569
	Pecan 2	0.08522	0.12411	0.16039	0.19593	0.23007
RVAR	Pecan 1	0.10478	0.14019	0.15778	0.17555	0.20198

Pecan 2	0.10125	0.13875	0.15635	0.17110	0.19391

Note: RW denotes the random walk model, ECM denotes the error correction model, BVAR denotes the Bayesian vector autoregression, and RVAR denotes the restricted vector autoregression.

Table 6. AGS Test for Significance of Differences of Forecast Errors from Alternative Models

		BVAR		RVAR	
		Pecan 1	Pecan 2	Pecan 1	Pecan 2
ECM	1	Y	Y	Y	Y
	2	Y	Y	Y	Y
	3	Y	Y	N	Y
	4	N	Y	N	Y
	5	N	N	N	N
BVAR	1			N	Y
	2			N	N
	3			Y	Y
	4			Y	Y
	5			Y	N

Notes: The numbers refer to the forecast steps. Y/ N denote significance/insignificance in the differences of RMSE.

References

Ashley, R., C. W. J. Granger, and R. Schmalensee. "Advertising and Aggregate

Consumption: An Analysis of Causality." *Econometrica* 48(1980):1149-1167.

Bessler, D. and J. Kling. "Forecasting Vector Autoregressions with Bayesian Priors."

American Journal of Agricultural Economics 68(1986): 144-151.

Bessler, D. And T. Covey. "Cointegration: Some Results on U.S. Cattle Prices." *The Journal of Futures Markets* 11(1991):461-474.

Doan, T. *RATS User's Manual*. Evanston, IL: Estima, 1992.

Engle, R. And C. W. J. Granger. "Cointegration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55(1987): 251-276.

Goodwin, B. "Forecasting Cattle Prices in the Presence of Structural Change." *Southern Journal of Agricultural Economics* 24(1992):11-22.

Hsiao, Cheng. "Autoregressive Modeling of Canadian Money and Income Data." *Journal of American Statistical Association* 74(1979):553-60.

Johansen, S. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12(1988): 231-254.

----------. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59(1991): 1551-1580.

Kaylen, M. S. "Vector Autoregression Forecasting Models: Recent Developments Applied to the U.S. Hog market" *American Journal of Agricultural Economics* 70(1988): 702-712.

Kim, B. and S. Mo. "Cointegration and the Long-run Forecast of Exchange Rates." *Economic Letters* 48(1995): 353-359.

LeSage, J. "A Comparison of the Forecasting Ability of ECM and VAR Models." *The Review of Economics and Statistics* 72(1990): 664-671.

Litterman, R. B. "Techniques of Forecasting Using Vector Autoregressions." Minneapolis MN: Federal Reserve Bank of Minneapolis Working Paper No. 115, 1979

---------- "Forecasting with Bayesian Vector Auto-regressions: Five Years of

Experience." *Journal of Business and Economic Statistics* 4(1986):25-38.

Park, T. "Forecast Evaluation for Multivariate Time-Series Models: The U.S. Cattle

Market." *Western Journal of Agricultural Economics* 15(1990): 133-143.

Phillips, P. and P. Perron. "Testing for a Unit Root in Time Series Regression." *Biometrica*

75(1988):335-46.

Shoesmith, G. L. "Multiple Cointegrating Vectors, Error Correction, and Forecasting

with   Litterman's Model." *International Journal of Forecasting* 11(1995): 557-567.

Tiao, G. C. and G. E. P. Box. "Modeling Multiple Time Series with Applications."

Journal   of the American Statistical Association 76(1981):802-816.

Zapata, H. O. and P. Garcia. "Price Forecasting with Time-Series Methods and

Nonstationary Data: An Application to Monthly U.S. Cattle Prices." *Western*

Journal   of Agricultural Economics 15(1990): 123-132.