Forecasting Organic Food Prices: Testing and Evaluating Conditional Predictive Ability

Timothy Park[a]
Tatiana Gubanova
Luanne Lohr
and
Cesar Escalante

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24–27, 2005

[a]Timothy Park (tpark@agecon.uga.edu) and Luanne Lohr (llohr@agecon.uga.edu) are Associate Professors, Tatiana Gubanova (tgubanova@agecon.uga.edu) is a Research Associate, and Cesar Escalante (cescalante@agecon.uga.edu) is Assistant professor in the Department of Agricultural and Applied Economics, the University of Georgia. Tatiana Gubanova gratefully acknowledges financial support from the Muskie Graduate Fellowship Program, U.S. Department of State. Support for this project was provided through a partnership of the University of Georgia and U.S.D.A. Risk Management Agency.

Forecasting Organic Food Prices: Testing and Evaluating Conditional Predictive Ability

Timothy Park, Tatiana Gubanova,

Luanne Lohr, Cesar Escalante

May 15, 2005

Abstract

Organic farmers, wholesalers, and retailers need reliable price forecasts to improve their decision-making practices. This paper presents a methodology and protocol to select the best-performing method from several time and frequency domain candidates. Weekly farmgate prices for organic fresh produce are used. Forecasting methods are evaluated on the basis of an aggregate accuracy measure and several out-of-sample predictive ability tests. Combining forecasts to improve on individual forecasts is investigated.

Keywords: Organic produce, Price forecasting, ARMA, Exponential smoothing, Spectral decomposition, Forecast Evaluation, Forecast Combination

1 Background and Objectives

The efficiency of production decisions by farmers, wholesalers and retailers that either specialize in organic products or introduce organic items to their product lines, depends critically on their expectation of future prices. Based on a national survey by the Organic Farming Research Foundation (OFRF), organic farmers have identified constraints on finding the best price as a most significant marketing barrier to adopting organic agriculture (Walz 2004). Nearly a quarter of OFRF survey respondents cited prices, market quotes, or determining a fair price as their top marketing information need.

In an established market, futures prices serve as a good predictor. The availability of hedging facilities themselves reduces the risk in the whole chain, thus reducing the prices. For the emerging organic market, however, such mechanisms are not yet in place, nor are there any accurate forecasts broadly available. This leads to a high degree of uncertainty about the future revenues and, accordingly, to sub-optimal output and pricing decisions by all marketing participants involved.

The organic market represents a serious challenge for price forecasting. Price signals are subject to complex periodicity overlaid with non-periodic components, which are characteristic for a new market. Price forecasting for agricultural commodities, both conventional and organic, is usually performed with the use of equilibrium-based forecasting systems (Tomek and Myers 1993; Park and Lohr 1996). The development and operation of those is costly, time-consuming, and requires large arrays of information that industry executives may not have access to. Meanwhile, simple, self-contained price forecasting systems can provide relatively accurate and easily computable forecasts to meet the industry's day-to-day needs.

The primary objective of the study is to analyze empirically groups of methods that may be used by the organic industry's decision-makers. These are the family of exponential smoothing methods, autoregressive moving average (ARMA) methods, and spectral analysis. The methods fit a variety of models where there are cyclical patterns present. The methods make it possible to include or extract an aperiodic trend. In addition, the methods are computationally inexpensive and allow the analyst to generate forecasts for many time series in one step. Organic fresh produce prices are examined. They are of particular interest because organic fresh produce is the most popular group with consumers. It accounted for nearly a half of all organic food sales in 2003 (OTA).

When several forecasts of an uncertain event are available, a forecast user attempts to identify the most reliable forecast. This is typically done on the basis of an aggregate measure of forecasting method accuracy. The best forecast is then used for decision-making, while the others are ignored. Several scholars (Diebold and Mariano 1995; West 1996; McCracken 2000; Giacomini and White 2003) have argued that rather than seeking the best overall accuracy, the distribution of the sampling prediction error should be examined. This technique is known as the out-of-sample predictive ability testing.

An out-of-sample forecasting experiment allows to determine whether the entire forecasting method in question is potentially useful for forecasting the variable of interest under a chosen loss function. There are two main approaches to the out-of-sample predictive ability testing. The studies on forecast evaluation by Diebold and Mariano, West, and McCracken focus solely on the forecast model. The forecast model is the only entity which is considered to affect the method's performance. A situation when a good model produces bad forecasts because its parameters have been badly estimated or change over time has been neglected in the literature. In the study by Giacomini and White, the object of the evaluation is the forecasting method. It includes the forecast model, the estimation procedure, and possibly the choice of an estimation window.

An approach proposed by Giacomini and White, which they termed the test of conditional predictive ability, has several advantages over the conventional out-of-sample predictive ability testing. First, the use of a rolling estimation window instead of an expanding one avoids the problem of arbitrary sample division between estimation and evaluation parts of the data set. The rolling window cuts off all dated information, which may keep contaminating results when the data-generating process has already changed. Second, the predictive ability test is conditional on the values of parameter estimates in the model, not their probability limits. This matters when the researcher is unsure about the model itself. Finally, the approach can be easily used to conduct different predictive ability tests. We provide an adaptation of the well-known Henrikkson-Merton test in addition to conducting the original test by Giacomini and White.

We perform our empirical analysis of price forecasts for organic fresh produce performs in four steps. The data-generating process for the prices is discussed and a structural forecasting model of price determination is obtained. Second, the best-performing method is chosen out of several competing methods on the basis of their conditional predictive ability. Third, we examine the performance of methods across commodities is examined with the use of non-parametric ANOVA.

With forecasts from different methods in hand, market participants may want to use them all, attaching a certain degree of credibility to each method. Each forecast is a lottery, and it was established that combined forecasts usually outperform individual forecasts. The "diversification" problem is to determine what weights should be assigned to each individual forecast. Different approaches were suggested. Bates and Granger (1969) and Markidakis and Winkler (1983) suggested using the average of the forecasts; that is, using equal weights. A way of choosing unequal weights was proposed by Deutsch et al. (1994) and Granger and Ramanathan (1984). A Bayesian approach to combining economic forecasts was proposed by Holden and Peel (1988). As the fourth step in our analysis, we explore several approaches to obtain the weights for averaging forecasts. All these approaches are model-free, subjective rules that utilize the resampling nature of the rolling window experiment.

As a result of this study, organic market participants are provided with the methodology of application, evaluation, comparison, and combination of forecasting methods. The results yield direct recommendations for industry marketing professionals to guide the specification of price forecasting models, the identification of the best-performing models, and ways to interpret forecasting results. The output provides producers with information on predicted price level and variability. This can assist farmers in developing multi-year farm plans required for certification by the National Organic Standards. Agribusinesses using components of the price forecasting model will be able to set appropriate offer and sell prices for organic commodities and plan for future price variability. Conventional farmers considering entry into organic markets will be better able to evaluate the revenue risk associated with different crop strategies and allocation to transitional acreage.

2 Structural Data-Generating Process

Agricultural product markets are commonly assumed to be competitive and in equilibrium (Tomek and Myers 1993). In an equilibrium, the quantity and farm-level price of a commodity are determined simultaneously. Price and quantity uncertainties are thus closely interrelated. Both conventional and organic farming depend crucially on many natural processes, which are periodic either seasonally or on a multi-annual basis. Organic production primarily relies upon natural processes and, potentially, social cycles, such as seasonal workforce. One can see that cyclical natural and social phenomena bring about an output uncertainty that may effect price uncertainty. This leads to a conjecture that seasonality and cyclicity should be paid special attention in this forecasting application.

To obtain a structural forecasting model of organic price determination, we adhere to the decomposition approach. We first re-write the simultaneous equations supply/demand system used by Park and Lohr (1996) in a geometric lag form for the price. The dynamic supply and demand equations from the partial adjustment model (ibid.) can be written, with a change of notation, as

$$q_t = [1 \ y_t \ q_{t-1} \ \sin(\omega t) \ \cos(\omega t) \ y_t^* \ \boldsymbol{W}_t]\boldsymbol{\beta}_s + \epsilon_s \tag{1a}$$

and

$$y_t = [1 \ q_t \ y_{t-1} \ \sin(\omega t) \ \cos(\omega t) \ \boldsymbol{D}_t]\boldsymbol{\beta}_d + \epsilon_d \tag{1b}$$

where y_t and q_t are the equilibrium price and quantity for the organic item at time t, respectively, $\sin(\omega t)$ and $\cos(\omega t)$ are harmonic terms of a preset angular frequency ω to account for seasonal effects, $\boldsymbol{W}_t$ is a vector of weather variables, y_t^* is the supply-shifting price of the conventional counterpart to the organic item, $\boldsymbol{D}_t$ are demand-shifting factors that include a price premium for the organic item on the wholesale market, wholesaler's transportation and labor costs, etc., and $\boldsymbol{\beta}_s$ and $\boldsymbol{\beta}_d$ are the supply and demand coefficients to estimate, respectively.

Substituting repeatedly the supply equation (1a) into the demand equation (1b) and regrouping terms, y_t is expressed as

$$y_t = \mu + \theta \sum_{i=1}^{\infty} \lambda^i y_{t-i} + s(t, \boldsymbol{\theta_s}) + g(t) + \epsilon_t \tag{2}$$

or

$$B(L)y_t = \mu + s(t, \boldsymbol{\theta}_s) + g(t) + \epsilon_t$$

where μ is the mean of the series; λ and $\boldsymbol{\theta} = [\theta, \boldsymbol{\theta}_s]$ are parameters that depend on $\boldsymbol{\beta}_s$, $\boldsymbol{\beta}_d$, and also parameters of the y_t^* process; $s(t, \boldsymbol{\theta}_s)$ is a cyclical signal that incorporates $\sin(\omega t)$, $\cos(\omega t)$ as well as extracted periodic components from future levels of other variables in Equation (1); $g(t)$ is an unknown aperiodic stochastic process that reflects the cumulative effect of all explanatory variables that cannot be forecast; and $B(L)$ is a polynomial in lags of y_t.

The structural model of Equation (2) comes from the simultaneous equations model and does not require any additional information other than past values of the price series. This allows the use of a self-contained forecasting method. The exact data-generating process for organic fresh produce prices remains unknown. No exact specification of the data-generating process for the prices is assumed in the study. In order to choose price forecasting methods, some major components of the price-generating process need to be detected.

Assuming that the changing nature of the organic sector breaks the infinite memory geometric lag process and, accordingly, that the order of the polynomial $B(L)$ becomes finite, the structural model in Equation (2) clearly shows that the three major components of the price-determination process are: an ARMA component in y_t and ϵ_t; a seasonal component $s(t)$; and an aperiodic stochastic process $g(t)$ of unknown form. Price forecasting techniques which will be used in the study must be able to work with ARMA and seasonal components.

Among methods operating in the time domain, the well-studied exponential smoothing and ARMA methods satisfy these criteria. Spectral decomposition, a frequency domain representative, also fits the needs of industrial application. Exponential smoothing, ARMA, and spectral decomposition are among the simplest forecasting methods available today. They can be implemented with mainstream statistical or all-purpose software, do not contain any proprietary algorithms, and do not require intense computing power. This makes the methods broadly qualify as the industry-oriented tools.

Accuracy-wise, the three methods are quite different in their handling of the data-generating process components in Equation (2). As its name suggests, ARMA is to work with the ARMA component. While it is moderately robust to noisy data, its capabilities in terms of modelling seasonality and cyclical patterns are limited to seasonal coefficients. Spectral decomposition, on the contrary, can extract a periodic signal of complex form from the data. However, it cannot handle autoregressive processes and therefore may produce very poor results for short-run forecasts. Exponential smoothing encompasses some kinds of ARMA process, can account for a simple seasonal pattern and trends, and is capable of producing satisfactory predictions in the presence of a considerable amount of noise.

3 Data and Forecast Competition

Nine produce items were chosen for implementing price forecasting methods, based on their large consumption and acreage shares; see Table 1.

Farm-level price series of organic commodities are currently available from three sources — Hotline Printing and Publishing, Inc. (Hotline), the Rodale Institute, and the Organic Farmers Agency for Relationship Marketing (OFARM). Hotline is the only for-profit firm providing weekly organic price information at the national level. According to the mail survey of existing and prospective subscribers to the Hotline Organic Commodity FAX Service (Lohr 2005), more than half of the respondents are satisfied with the overall quality of the information provided. Most of those who are able to make a comparison with other price reporting services rate the Hotline service as either "good" or "excellent."

The price data for the study were collected by Hotline through weekly telephone interviews of brokers and farmers throughout the United States. The list of sources is confidential and cannot be disclosed. Weekly prices were averaged for all locations from which data has been collected. The methods used to assess representativeness were based on statistical testing and qualitative comparison of the states in the source list with geographic distribution of production acreage and brokers (Lohr 2005).

To make the exposition simpler, method evaluation is dowtailed with a protocol developed for its use with forecasting methods. Its detailed, step-by-step description follows.

Data Rearrangement

In order to allow for a seasonality adjustment, series of weekly price observations were regrouped into ten-days periods. As a result, 36 observations per year were made available for estimation and forecasting. Regrouping weekly data into ten-days periods allowed to avoid the unevenly-spaced data problem that plagued the initial series. Missing ten-days values were linearly interpolated, using the available boundary points. Missing observations at the beginning and end of a series were cut off.

Table 1: Commodity Description

Item	Description	Series Length	Missing Observations	Data Availability
APPLES	Red Delicious, $/lb	479	28	August 1994 –June 2004
AVOCADOS	Hass, $/lb	454	90	April 1995 –June 2004
CABBAGE	Green, $/lb	418	2	January 1996 –June 2004
LEMONS	$/lb	451	22	April 1995 –June 2004
LETTUCE	Romaine, $ each	479	1	August 1994 –June 2004
ONIONS	Yellow, $/lb	521	31	January 1991 –June 2004
POTATOES	Red, A grade, $/lb	479	8	August 1994 –June 2004
STRAWBERRIES	$/pint	455	146	August 1990 –June 2004
TOMATOES	Roma, $/lb	408	37	March 1996 –June 2004

The entries in the last three columns are related to the initial data.

Pre-Testing

Each series was tested for the white noise with Bartlett's version of the Kolmogorov-Smirnov test, and for stationarity with the Augmented Dickey-Fuller test at the maximum lag order of 18. The p-values of the Kolmogorov-Smirnov statistic were all reported to be less than 10^{-4}, which lead to the rejection of the white noise null hypothesis for all commodities. The p-values of the Dickey-Fuller statistic fell in the range of 10^{-3} to 10^{-2}, which lead to the rejection of the non-stationarity null hypothesis for all commodities. As a result, the series were considered non-white noise and stationary. A preliminary analysis of price series revealed autoregressive and seasonal components in the data. No significant moving average processes were detected.

Method Specification

The three forecasting methods selected all allow to deal with data featuring seasonal variation. A seasonal autoregressive (AR) model was chosen out of the ARMA class. The additive version of the Holt-Winters (HW) exponential smoothing was chosen out of the exponential smoothing family. Each of the selected forecasting methods — AR, spectral decomposition (SD), and HW — were implemented to operate in a fully automatic way in order to provide the level ground for their competition.

Window Specification

The width of the rolling estimation window was set at two years ($m = 72$ observations), so that every observation in the year cycle would have its year-long lag included in the estimation data set. The forecast horizons were chosen according to the type of the price forecast user. Farmers might be interested in 6–9 months price forecasts at the time of planting, and 10–30 days forecasts at harvest time. In turn, a periodical 1–2 months price forecast is the likely need of wholesalers and retailers, depending on their product replenishment strategy. Therefore, four forecast horizons — next decade, next month, two months ahead, and six months ahead — were selected as being reasonable for the purpose of comparing method performance in short-, mid-, and relatively long-term perspectives. The squared prediction error was used as the loss function with all methods and lags.

Estimation

Estimation of the AR model was performed in two stages. In the first stage, monthly constants were estimated by regressing the price on a set of 12 month indicators. Residuals from the first stage regression were used to estimate the autoregressive part of the model. The latter was estimated by least squares. The appropriate autoregressive order, up to 3 lags, was chosen in each case by using the minimum Akaike Information Criterion (AIC) method. Once the optimum lag order was identified, the forecast was produced by recombining the first-stage monthly constant estimate and the predicted value from the autoregressive part of the model. SAS IML/TIMSAC modules were used to program the method (SAS Institute Inc. 1999b).

Estimation of the SD model was performed by using the Finite Fourier Transform (FFT) of the series and obtaining smoothed spectral density estimates. $m/2$ Fourier cosine and sine coefficients were used to obtain the respective values of the amplitude periodogram according to the following equation:

$$J_k = \frac{T}{2}(a_k^2 + b_k^2) \tag{3}$$

where J_k is the amplitude periodogram, a_k and b_k are the Fourier coefficients, and T is the number of observations in the series.

Since the periodogram J_k is a volatile and inconsistent estimator of the spectrum, spectral density estimates were produced by smoothing the periodogram. A triangular symmetric kernel with three points on each side was used for smoothing. A simple form of model identification in the frequency domain was chosen, based on the identification of peaks in spectral density. A spectral density estimate $\widehat{s}_k,\ k = 1 \ldots m/2$ was considered to be a peak if its value was greater than its neighbors; that is, if $\widehat{s}_k > \widehat{s}_{k-1}$ and $\widehat{s}_k > \widehat{s}_{k+1}$. Correspondingly, amplitude coefficients for all non-peak harmonics were set to zero. Thus modified coefficients were used to obtain the forecast value. In case the spectral density were found monotone, only the series mean would have been used as the forecast for all periods. SAS ETS/SPECTRA procedure was employed to program the method (SAS Institute Inc. 1999a).

Monthly seasonal factors were used for the HW method, one for each month in the year. The starting values for the seasonal factors were computed from seasonal averages over the first complete seasonal cycle of 36 observations. The weights for updating the seasonal factors were set at $\omega_3 = \omega_2 = 0.25$ and $\omega_1 = 0.2$. SAS ETS/FORECAST procedure was employed to program the method (SAS Institute Inc. 1999a).

After a forecast was generated at any point of the rolling window, squared residual for the last observation in the estimation window, and squared forecast error at the specified lead were stored. This information was used at the next stage to conduct the Giacomini-White and Henriksson-Merton tests.

Evaluation

Forecast quality in aggregate was evaluated using the root mean squared error (RMSE). The conditional predictive ability tests conducted included the Giacomini-White and Henriksson-Merton tests.

In order to assess the economic value of forecasts, the direction-of-change test proposed by Merton (1981) and Henriksson and Merton (1981) was conducted. The null hypothesis of the Henriksson-Merton test is that the probability limit of the Henriksson-Merton criterion is one; that is

$$\mathrm{H}_0 : \mathrm{plim}_{n\to\infty}\left(\frac{n_{ii}}{n_i} + \frac{n_{jj}}{n_j}\right) = 1 \tag{4}$$

against the alternative of the left-hand side being greater than one. i denotes the "up" state (an increase from the last observed value) and j indicates the "down" state (a decrease) into which forecasts and realizations fall. n_i and n_j are the numbers of actual price "ups" and "downs," respectively, recorded by moving the data window n times. n_{ii} and n_{jj} are the numbers of correctly forecast price realizations. Under H_0, n_{jj} follows a Hypergeometric distribution with parameters $(n_j, n, n_{\cdot j})$, where $n_{\cdot j}$ is the number of forecast "downs." Henriksson and Merton (1981) assert that a forecast has an economic value if their criterion is greater than one.

In order to compare pairs of methods on the basis of their conditional predictive ability, the already mentioned Giacomini-White test was conducted. For a horizon τ and a fixed estimation window of length m that has been moved n times, the test statistic is a Wald-type statistic of the following form:

$$T^h_{n,m,\tau} = n\overline{Z}'_{m,n}\widehat{\Omega}_n^{-1}\overline{Z}_{m,n} \tag{5}$$

where $\overline{Z}_{m,n} = n^{-1} \sum_{t=m}^{T-\tau} h_t \triangle L_{t+\tau}$, $\triangle L_{t+\tau}$ is the difference of loss functions at $t+\tau$, h_t is a vector of test functions, and $\widehat{\Omega}_n$ is the estimated covariance matrix of $\overline{Z}_{m,n}$. In practice, the test function is chosen by the researcher to embed elements of the information set that are believed to have potential explanatory power for the future difference in predictive ability. In the present research, the test function is $h_t = (1, \triangle L_t)$, corresponding to the difference of squared residuals in the last period in the window. A level α rejects the null hypothesis of equal conditional predictive ability whenever $T^h_{n,m,\tau} > \chi^2_{q,1-\alpha}$, where $q = 2$ is the size of h_t and $\chi^2_{q,1-\alpha}$ is $(1-\alpha)$-quantile from the χ^2_q distribution.

4 Discussion of Results

The obtained RMSE values are reported in Table 2; Henriksson-Merton results are presented in Table 3. The precision of AR forecasts is notably better in both magnitude and direction-of-change sense. RMSE of AR forecasts are smaller, for all commodities and all horizons, than those for SD and WH forecasts, sometimes by two or three times. Values of the Henriksson-Merton criterion are significantly greater than unity for most commodities, with both AR and SD model, while those for WH model were often found insignificant. The reason for mostly poor fits with WH model appears to be an autoregressive rather than moving-average nature of the data-generating process and problems with the automatic choice of smoothing weights. Although RMSE does point at the best-performing method for the considered data series, this aggregate measure does not allow formal testing. Therefore, a statistical technique, such as the Giacomini-White test, must be employed to verify if the method yielding the minimum RMSE can indeed boast a better predictive ability.

The results of the Giacomini-White test of equal conditional predictive ability for three forecasting methods appear in Table 4. AR is broadly the best forecasting method as compared to both SD and HW methods, for all produce items and all horizons. Results of pairwise comparisons of SD and HW based on the Giacomini-White test are presented in Table 5. HW appears to be the best forecasting method for the ten-days-ahead forecast horizon. SD outperforms HW for medium and relatively long-term forecasts.

Table 2: Root Mean Squared Error for Price Forecast

	Forecast Horizon											
	ten days			one month			two months			six months		
	AR	SD	HW	AR	SD	HW	AR	SD	HW	AR	SD	HW
apples	0.094	0.152	0.112	0.120	0.138	0.145	0.139	0.154	0.166	0.163	0.190	0.273
avocados	0.254	0.387	0.272	0.339	0.348	0.344	0.369	0.378	0.386	0.405	0.412	0.469
cabbage	0.081	0.120	0.116	0.118	0.121	0.153	0.134	0.140	0.176	0.153	0.172	0.290
lemons	0.150	0.341	0.172	0.204	0.344	0.218	0.216	0.378	0.243	0.214	0.326	0.346
lettuce	0.178	0.232	0.218	0.226	0.240	0.279	0.245	0.267	0.321	0.252	0.277	0.456
onions	0.080	0.143	0.111	0.115	0.139	0.145	0.135	0.157	0.167	0.152	0.157	0.231
potatoes	0.085	0.124	0.105	0.102	0.122	0.133	0.114	0.141	0.153	0.118	0.148	0.223
strawberries	0.400	0.737	0.559	0.592	0.720	0.760	0.709	0.834	0.895	0.892	0.919	1.435
tomatoes	0.148	0.201	0.188	0.205	0.218	0.238	0.214	0.242	0.263	0.214	0.260	0.346

The entries are root mean squared errors for seasonal autoregression (AR), spectral decomposition (SD), and the additive Holt-Winters (HW) method.

Table 3: Henriksson-Merton Criterion for Organic Produce

	Forecast Horizon											
	ten days			one month			two months			six months		
	AR	SD	HW	AR	SD	HW	AR	SD	HW	AR	SD	HW
APPLES	1.04	1.06	1.04	1.17**	1.12**	1.08*	1.24**	1.20**	1.13**	1.53**	1.31**	1.09*
AVOCADOS	1.04	1.03	1.03	1.24**	1.17**	1.13**	1.42**	1.35**	1.33**	1.35**	1.36**	1.27**
CABBAGE	1.24**	1.10**	1.07	1.33**	1.28**	1.06	1.41**	1.35**	1.16**	1.45**	1.30**	0.87
LEMONS	1.07	1.01	1.06	1.36**	1.11**	1.24**	1.68**	1.19**	1.58**	1.82**	1.79**	1.75**
LETTUCE	1.15**	1.19**	1.11**	1.27**	1.22**	1.11**	1.18**	1.08*	0.95	1.36**	1.21**	0.93
ONIONS	1.04	1.08**	0.99	1.28**	1.28**	1.12**	1.42**	1.44**	1.31**	1.63**	1.54**	1.31**
POTATOES	1.22**	1.21**	1.20**	1.28**	1.26**	1.28**	1.46**	1.35**	1.39**	1.58**	1.44**	1.30**
STRAWBERRIES	1.14**	1.19**	0.90	1.34**	1.24**	1.00	1.42**	1.20**	1.16**	1.58**	1.41**	1.29**
TOMATOES	1.17**	1.23**	1.14**	1.37**	1.35**	1.24**	1.56**	1.42**	1.35**	1.45**	1.37**	1.08

** — significant at 5% level; * — significant at 10% level.

The entries are the Henriksson-Merton criterion for seasonal autoregression (AR), spectral decomposition (SD), and the additive Holt-Winters (HW) method.

Table 4: Giacomini-White Test Comparison for Three Methods

	Forecast Horizon			
	ten days	one month	two months	six months
apples	AR	AR	AR	AR
avocados	AR/HW	AR/HW	AR/HW/SD	AR/SD
cabbage	AR	AR/SD	AR/SD	AR
lemons	AR	AR/HW	AR	AR
lettuce	AR	AR/SD	AR/SD	AR
onions	AR	AR	AR	AR/SD
potatoes	AR	AR	AR	AR
strawberries	AR	AR	AR	AR
tomatoes	AR	AR/SD	AR	AR

The entries indicate the best-performing method among seasonal autoregression (AR), spectral decomposition (SD), and the additive Holt-Winters (HW) methods according to the Giacomini-White test. "AR/HW/SD" indicates the equivalence of the corresponding methods. Tests were conducted at a 5% significance level.

Table 5: Giacomini-White Test Comparison for Two Methods

	Forecast Horizon			
	ten days	one month	two months	six months
apples	HW	SD	HW/SD	SD
avocados	HW	HW/SD	HW/SD	SD
cabbage	HW	SD	SD	SD
lemons	HW	HW	HW	HW/SD
lettuce	HW/SD	SD	SD	SD
onions	HW	SD	HW/SD	SD
potatoes	HW	SD	SD	SD
strawberries	HW	SD	SD	SD
tomatoes	HW	HW/SD	HW/SD	SD

The entries indicate the best-performing method among spectral decomposition (SD) and the additive Holt-Winters (HW) methods, as found with the Giacomini-White test. "HW/SD" indicates the equivalence of the two forecasting methods. Tests were conducted at a 5% significance level.

Given the available data and the quadratic loss function, the results indicate that a forecast user would be better off using the seasonal autoregressive model as a forecasting technique for all forecast horizons. For the purpose of short-term forecasting, such as ten days ahead, the additive Holt-Winters method can be reasonably employed along with a seasonal autoregressive model, whereas spectral decomposition would likely have resulted in decreased forecast accuracy. For mid-term and long-term forecasts, however, spectral decomposition along with a seasonal autoregressive model would promise better forecasts than the additive Holt-Winters method.

5 Forecast Accuracy Across Commodities

In order to see the complete picture, one should also look at method performance across commodities. The question we pose is: are there any commodities for which the methods perform better and, if so, what might be the reason?

To answer the first part in a statistically valid way, the Friedman test (non-parametric ANOVA) was performed. This test (Conover 1999) is similar to the usual parametric method of testing the null hypothesis of no treatment difference (two-way ANOVA). Friedman's method makes use of only ranks of observations within each block, not their actual values. This makes their distribution immaterial. For the purpose of the test, commodities were considered treatments and methods played the role of blocks. Commodity RMSE were averaged across all forecast horizons and normalized by average commodity prices. The Friedman test allows for correlation between treatment effects, which is useful when dealing with complement or substitute goods. The $\chi^2[8]$ distributed test statistic was 21.51, which leads to the rejection of the null hypothesis of no forecast quality difference among the nine commodities. The null hypothesis is rejected at any reasonable confidence level, since it has the p-value of 0.006. Therefore, we can conclude that prices for some commodities can be better predicted with any method than others.

To see how the performance differs across commodities, Dunn's post-test pairwise comparisons (Conover 1999) were conducted. This particular implementation of the post-test makes use of the asymptotic t-distribution of the absolute difference of ranks across blocks. At the borderline tolerance of 0.006 above, three groups can be identified. Apples and potatoes feature the highest forecast quality. These are followed by cabbage, lemons, onions, and tomatoes. The prices for avocados, strawberries, and lettuce turn out to be the least predictable.

This grouping from the pairwise comparison does not lend itself to any evident explanation. It does not align with the OFRF survey results, where it may appear that fruit producers who experienced more market shrinkage than vegetable producers should be facing more unpredictable prices. A significant relationship comes to light when analyzing the correlation between normalized commodity RMSE and commodity-specific factors. It was found that the correlation between the transformed RMSE and the consumption share of the commodity in total consumption is -0.6. The correlation of the transformed RMSE and the standard deviation of price series is 0.5. Both values indicate the presence of relatively high correlation.

Larger organic produce markets appear to have less price volatility and behave in a more predictable way. This result is broadly in line with economic theory which states that larger markets with many agents more resemble the perfect competition environment (Ferris 1997). Information is more freely available in larger markets. Farmers that supply to large markets are less subject to the oligopsonic market control by retailers (McLaughlin 2004). The predictability of price is positively related to the commodity's market size. This emphasizes the role of better price forecasts for the agents — farmers and traders — who deal in less common organic produce. Economic theory tells us that better price information improves profits of the producer. This confirms the claim made at the very beginning of the study: organic producers do need better farmgate price forecasts to grow.

Another factor is the varying perishability of produce. Supply and demand for more perishable commodities experience more sporadic shocks, which boosts price volatility in these markets and makes the prices less predictable. Buyers and sellers of fresh produce generally tend to be averse to opportunistic transactions and do engage in contractual agreements. This stabilizes prices. One can therefore recommend producers of the commodities in the least predictable group to consider forward contracting to damp down shocks and improve their profits.

The influence of commodity consumption share also reveals an important role of demand factors in the farmgate price formation. This study deals with the prediction of farmgate prices only; the downstream effects of pricing behavior of wholesalers and retailers are not considered. Forecasting prices at a wholesale and retail level coupled with the farmgate price may thus improve the general accuracy of forecasts. More than half of farmers that participated in the OFRF survey stated that they had not experienced much price volatility. One can expect, because of the positive correlation between the forecast accuracy and price volatility, that at least 50 percent of these farmers would receive price forecasts of relatively high quality. Since the distribution of farmers income is roughly symmetric in volatility categories, price forecasts cannot be expected to influence a particular income category of organic farmers.

6 Combination of Forecasts

The resampling framework of the rolling window experiment allows testing different subjective rules of forecast combination. We partitioned the entire set of forecasts from the three methods into two subsets. The first $n1$ forecast triplets were used as the training dataset; that is, the dataset to obtain the knowledge for implementing rules. The last 20 triplets were used to test the rules, to see which rule would provide the smallest RMSE on this evaluation dataset. Thus, $n1$ was calculated as $n1 = n - 20$, where n is the number of rolling window instances.

We chose four different methods to establish the rules. Using the best-performing method (BM), i.e. the method giving the minimum RMSE on the training dataset puts all the credibility, the unity weight, on the best method forecast; other forecasts get zero weights. Averaging forecasts (AV), just the opposite, assigns equal weights of 1/3 to all forecasts. The weighted averaging (WA) that we considered weights forecasts by the reciprocals of the respective RMSE from the training dataset.

Finally, we implemented a learning expert system to combine forecasts. This system obtains the weights by first finding the minimum-loss forecast in all training triplets and giving the winner each time a non-zero credibility score. The value of the score is bound between zero and one. It is calculated as a similarity measure, the product of Gaussian "memberships" of the training set forecasts in fuzzy sets centered on the evaluation set forecasts:

$$w_t^j = \begin{cases} \prod_{j \in \mathcal{F}} \phi\big(\hat{y}_t^j, \hat{y}_\tau^j, \sigma^j\big) & \text{if method } j \text{ has minimum loss at time } t, \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $\mathcal{F} = \{AR, SD, HW\}$, $t = 1 \dots n1$, $\tau = 1 \dots 20$, $\phi\big(\hat{y}_t^j, \hat{y}_\tau^j, \sigma^j\big) = \exp\big(-0.5(\sigma^j)^{-2}[\hat{y}_t^j - \hat{y}_\tau^j]^2\big)$, $\hat{y}_t^j$ is a forecast by method j in t-th position of the training dataset, $\hat{y}_\tau^j$ is a forecast by method j

Table 6: Root Mean Squared Error for Price Forecast Combination

	Forecast Horizon							
	ten days				one month			
	BM	AV	WA	FL	BM	AV	WA	FL
apples	0.115	0.101	0.102	0.101	0.156	0.121	0.123	0.125
avocados	0.168	0.145	0.146	0.147	0.272	0.186	0.187	0.184
cabbage	0.059	0.079	0.076	0.070	0.078	0.102	0.098	0.091
lemons	0.180	0.180	0.183	0.181	0.218	0.208	0.215	0.217
lettuce	0.149	0.167	0.165	0.164	0.202	0.201	0.198	0.201
onions	0.078	0.074	0.075	0.074	0.121	0.095	0.096	0.096
potatoes	0.045	0.052	0.049	0.049	0.043	0.055	0.054	0.056
strawberries	0.209	0.228	0.219	0.213	0.234	0.281	0.276	0.275
tomatoes	0.068	0.069	0.068	0.067	0.064	0.073	0.072	0.071
	two months				six months			
apples	0.185	0.144	0.147	0.151	0.210	0.206	0.203	0.205
avocados	0.295	0.229	0.229	0.227	0.297	0.230	0.232	0.233
cabbage	0.085	0.117	0.112	0.105	0.078	0.133	0.114	0.112
lemons	0.262	0.239	0.249	0.257	0.264	0.288	0.283	0.278
lettuce	0.209	0.227	0.222	0.217	0.207	0.207	0.197	0.196
onions	0.136	0.103	0.103	0.104	0.122	0.099	0.092	0.089
potatoes	0.043	0.061	0.060	0.063	0.043	0.059	0.061	0.056
strawberries	0.223	0.374	0.368	0.338	0.232	0.269	0.298	0.306
tomatoes	0.063	0.077	0.075	0.077	0.059	0.116	0.102	0.095

The entries are root mean squared errors for the using of the best forecasting method (BM), equal weight-averaging of the forecasting methods (AV), unequal weight-averaging of the methods (WA), and fuzzy-logic system (FL).

Table 7: Forecast Combination Results

	Forecast Horizon			
	ten days	one month	two months	six months
apples	FL	AV	AV	WA
avocados	AV	FL	FL	AV
cabbage	BM	BM	BM	BM
lemons	BM	AV	AV	BM
lettuce	BM	WA	BM	FL
onions	FL	AV	AV	FL
potatoes	BM	BM	BM	BM
strawberries	BM	BM	BM	BM
tomatoes	FL	BM	BM	BM

The entries indicate the best forecast combination method among the using of the best forecasting method (BM), equal weight-averaging of the forecasting methods (AV), unequal weight-averaging of the methods (WA), and deploying a fuzzy-logic system (FL), based on the lowest RMSE value.

in τ-th position of the evaluation dataset, and σ^j is the estimated standard deviation of forecasts by method j on the training dataset.

After all scores w_t^j are calculated with Equation (6), the weighted combination is obtained by using the cumulative credibility scores:

$$C_\tau = \left(\sum_{j \in \mathcal{F}} \sum_{t=1}^{n1} w_t^j \right)^{-1} \sum_{j \in \mathcal{F}} \hat{y}_\tau^j \sum_{t=1}^{n1} w_t^j \qquad (7)$$

This expert system qualifies as a fuzzy logic (FL) type-1 system (Mendel 2001). Its purpose is to look for any pattern in the way the forecasting methods err and, if any, to utilize this pattern for an intelligent forecast combination.

The results of testing the four forecast combination methods are presented in Table 6 and Table 7. Table 6 presents forecast combination results in terms of RMSE obtained on the evaluation subset. Table 7 presents forecast combination results in terms of the rules yielding the minimum RMSE. Using the best-performing method (BM) is optimal in roughly a half of instances. AV and FL exhibit equal performance and work best in a smaller number of cases. WA fires the worst.

The apparent reason for such an outcome appears to be high correlation of forecasts by different methods in more than 50 percent of cases. Coupled with relatively large variances of forecasts, especially that of the Holt-Winters method, this makes any averaging inferior to simply using the best-performing forecasting method. When there is no high correlation, a simple average seems to be the way to combine forecasts, that yields an improvement in accuracy and requires no computational expense at all.

7 Conclusions

The organic food market is one of the most promising emerging sectors of the U.S. economy. A substantial consumer demand for organic produce leads to an increasing interest in this sector by farmers, wholesalers, and retailers. This emphasizes the importance of farm-level price information in decision-making.

Three forecasting methods — seasonal autoregression, spectral decomposition, and the additive Holt-Winters exponential smoothing — were selected, implemented and extensively tested at four planning horizons with nine produce items. A problem was considered that decision-makers face: how to select the best forecasting method from a set of several competing ones. Forecast quality is evaluated by using the RMSE for the comparison at an aggregate level, and the Henriksson-Merton test for the direction-of-change comparison. For comparing several forecasting techniques, a test of conditional predictive ability, proposed by Giacomini and White (2003), along with the conventional stochastic dominance analysis were discussed and implemented.

The best-performing method was found among these three industry-oriented forecasting techniques. Based on both the quantile analysis and the Giacomini-White test, seasonal autoregression is the best forecasting method, compared to spectral decomposition and the Holt-Winters exponential smoothing for all produce and all horizons.

A significant positive correlation between the forecast precision and market size and negative correlation between the precision and commodity price volatility were found. This emphasizes the role of better price forecasts for agents who deal in less common organic produce. A confirmation for the claim that the organic produce industry needs better farmgate price forecasts to grow was provided. The relevance of joint forecasting of prices in the whole marketing channel of the product was underlined.

Four computationally inexpensive methods to combine forecasts were considered. If forecasts by the available methods are expected to correlate, then the user would be better off going with the best-performing forecasting method itself. Otherwise, the simple averaging of forecasts is recommended.

Much further research needs to be done on the forecasting of organic produce prices. The forecasting methods need to be adapted for cases when the data are unevenly-spaced. Missing data are a common problem not only for agricultural data but for economic data in general. More effective techniques need to be implemented instead of linear spline interpolation used in the present research. Instead of applying a forecasting method to one commodity at a time, prices for a group of products can be forecast jointly, in order to account for an effect of substitution amongst commodities. Spectral decomposition and multivariate ARMA allow to conduct such a kind of analysis.

A separate direction comes from the insufficiency of price forecasts for the farmgate level only. Three price spreads (differentials) matter in the decision-making by organic industry agents; these are: farm-wholesale, wholesale-retail, and farm-retail spreads. The analysis in this study shows the importance of demand-driven factors in the farmgate price formation. This means that the above spreads should better be forecast together with the farm-level price rather than considering the latter in isolation. Such a joint forecasting would necessitate the development of an extensive forecasting system that takes into account mathematically the interaction between the farm, wholesale, and retail stages.

References

Bates, J. and C. Granger (1969, December). The combination of forecasts. *Operational Research Quaterly 20*(4), 451–468.

Conover, W. (1999). *Practical Nonparametric Statistics* (3rd ed.). New York, NY: John Wiley & Sons.

Deutsch, M., C. Granger, and T. Terasvirta (1994, June). The combination of forecasts using changing weights. *International Journal of Forecasting 10*(1), 47–57.

Diebold, F. X. and R. S. Mariano (1995, July). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–263.

Ferris, J. (1997). *Agricultural Prices and Commodity Market Analysis.* New York, NY: WCB/McGraw-Hill.

Giacomini, R. and H. White (2003, April). Tests of conditional predictive ability. Web: `http://ideas.repec.org/p/wpa/wuwpem/0308001.html`.

Granger, C. and R. Ramanathan (1984, April–June). Improved methods of combining forecasts. *Journal of Forecasting 3*(2), 197–204.

Henriksson, R. and R. Merton (1981, October). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business 54*(4), 513–533.

Holden, K. and D. Peel (1988, November). Combining economic forecasts. *The Journal of the Operational Research Society 39*(11), 1005–1010.

Lohr, L. (2005, February). An assessment of demand for new and existing organic price information products. (unpublished paper).

Markidakis, S. and R. Winkler (1983, September). Averages of forecasts: some empirical results. *Management Science 29*(9), 987–996.

McCracken, M. W. (2000, December). Robust out-of-sample inference. *Journal of Econometrics 99*(2), 195–223.

McLaughlin, E. (2004). The dynamics of fresh fruit and vegetable pricing in the supermarket channel. *Preventive Medicine 39*, 81–87.

Mendel, J. (2001). *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall.

Merton, R. (1981, July). On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business 54*(3), 363–406.

OTA (2004). The OTA 2004 manufacturer survey overview. Organic Trade Association. Researched and produced for the Organic Trade Association under contract by Nutrition Business Journal. Web: `http://www.ota.com`.

Park, T. A. and L. Lohr (1996, August). Supply and demand factors for organic produce. *American Journal of Agricultural Economics 78*(3), 647–655.

SAS Institute Inc. (1999a). *SAS/ETS User's Guide*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1999b). *SAS/IML User's Guide*. Cary, NC: SAS Institute Inc.

Tomek, W. G. and R. J. Myers (1993). Empirical analysis of agricultural commodity prices: a viewpoint. *Review of Agricultural Economics 15*, 181–202.

Walz, E. (2004). Final results of the fourth national organic farmers' survey: Sustaining organic farms in a changing organic marketplace. Web: `http://www.organicaginfo.org`.

West, K. D. (1996, September). Asymptotic inference about predictive ability. *Econometrica 64*(5), 1067–1084.