Forecasting Organic Food Prices:
Emerging Methods for Testing and Evaluating Conditional Predictive Ability

Tatiana Gubanova,
Luanne Lohr,
and
Timothy Park*

Paper Presented at the NCR-134 Conference on Applied Commodity Price Analysis,
Forecasting, and Market Risk Management
St. Louis, Missouri, April 18–19, 2005

*Tatiana Gubanova (tgubanova@agecon.uga.edu) is a Research Associate in the Department of Agricultural and Applied Economics, the University of Georgia. Luanne Lohr (llohr@agecon.uga.edu) and Timothy Park (tpark@agecon.uga.edu) are Associate Professors. Tatiana Gubanova gratefully acknowledges financial support from the Muskie Graduate Fellowship Program, directed by the Bureau of Educational and Cultural Affairs, U.S. Department of State. Support for this project was provided by a partnership with the USDA Risk Management Agency.

Forecasting Organic Food Prices:
Emerging Methods for Testing and Evaluating Conditional Predictive Ability

Organic farmers, wholesalers, and retailers need 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. A seasonal autoregressive method is recommended for all planning horizons. The role of better price forecasts for the agents who deal in less common organic produce is highlighted. A confirmation for the claim that the organic produce industry needs better farmgate price forecasts to grow is provided.

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

1 Background and Objectives

As consumers become interested in healthy and natural foods, organic farming in the U.S. continues to expand at a rapid pace. Almost 950,000 hectares are currently managed organically. The U.S. has the world's fourth largest acreage of certified organic farmland. The number of certified organic growers in the country has increased almost by a half during the last decade. An increasing number of agricultural producers turn to the organic market, to exploit high price premiums and new target customer strata.

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. 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 parties involved.

The organic market represents a serious challenge for price forecasting. Price signals are subject to complex periodicity overlaid with non-periodic components akin to shot noise, 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 forecasters provide fast, relatively accurate and easily interpretable forecasts, which can 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.

Applying a forecasting method to raw price data may call for a number of procedures to perform before the forecasting begins and after it is completed. These may include re-grouping the data to make them evenly spaced, dealing with missing observations, testing for stationarity and outliers, etc. This leads to the secondary objective of the study: to develop, describe, and implement a protocol to be used with method applications.

Organic fresh produce is the most popular group with consumers; it accounted for nearly a half of all organic food sales in 2003 (OTA). Accordingly, it is organic fresh produce that was selected out of all organic products to become the focus of this study.

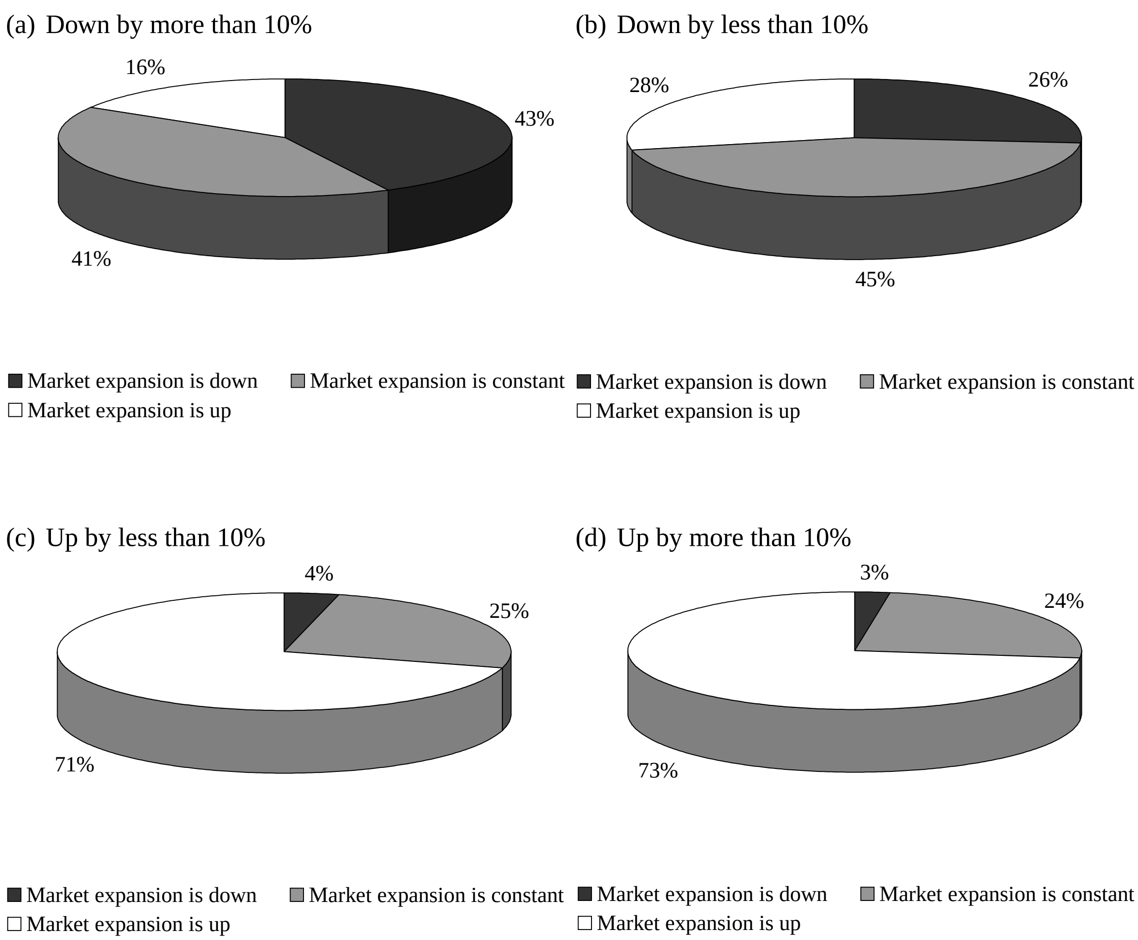
In the rest of the paper, Section 2 provides an overview of organic fresh produce markets; Section 3 discusses the structure of the data-generating process for prices series; Section 4 provides a general description of the forecasting methods to be applied; Section 5 describes the data and fleshes out the forecast evaluation methodology and protocol; Section 6 provides a discussion of obtained results; and Section 7 concludes with a summary of findings and suggestions for future research.

2 Organic Produce Markets

Packaged and precut organic vegetables and fruit are occupying more shelf space in the produce department as they continue to gain acceptance by consumers. By 2003, fresh produce had become the most popular category among consumers, accounting for about 42 percent of organic food sales (OTA 2004). As consumers become interested in organically-grown foods, organic farming in the U.S. continues to expand at a rapid pace. Almost 950,000 hectares are managed organically, which amounts to a 0.23 percent share of the total agricultural area. The U.S. is the fourth country in the world with respect to certified organic farmland acreage, following Australia, Argentina, and Italy. The number of certified organic growers has increased by 38 percent since 1997, and, in 2001, there were 6,949 organic farms in the country (Yussefi 2004).

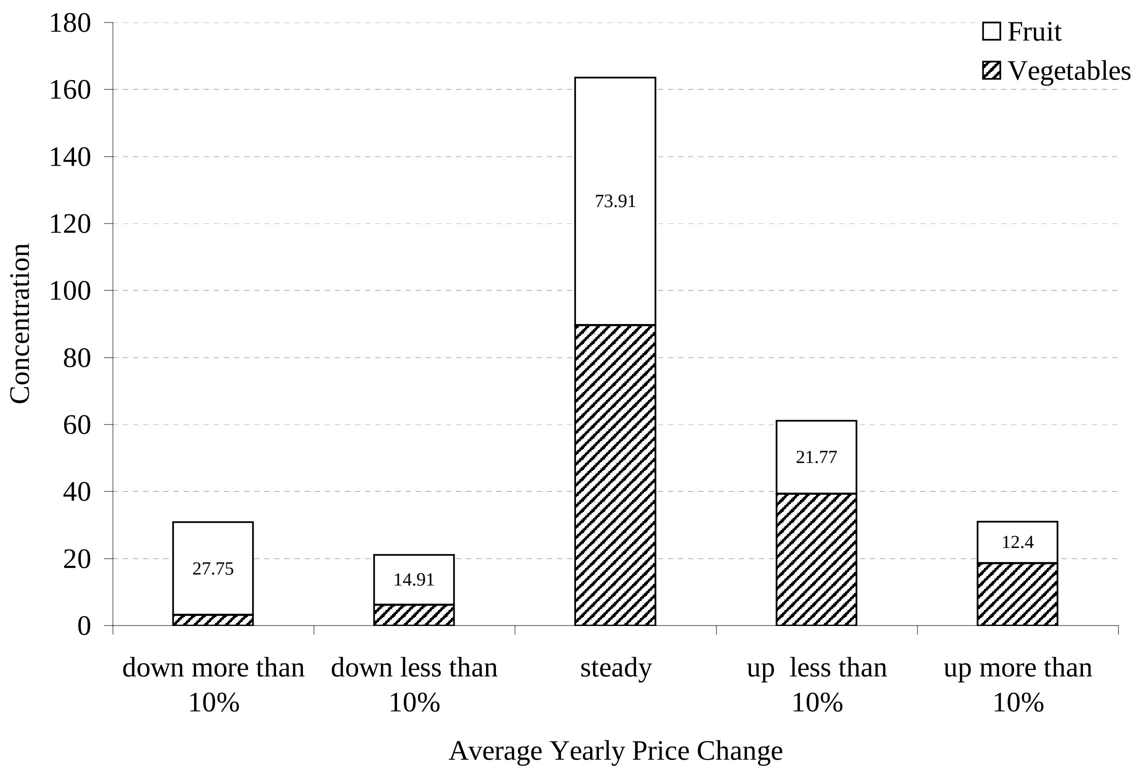
Farmers nation-wide allocated 2.3 million acres of cropland and pasture to organic production systems in 2001, which is by 74 percent more than in 1997. Over 1.3 million acres were used for growing crops versus over 1 million acres that were used for pasture and rangeland. For a comparison, the percentage of cropland acres versus pasture in 1997 was 63 percent to 37 percent, respectively. Colorado, California, North Dakota, Montana, Minnesota, Wisconsin, and Iowa have the largest share of organic cropland. Colorado, Texas, and Montana have the largest expanses of organic pasture and rangeland.

Figure 1: Price Effect on Market Expansion, 2001



Source: Fourth National Organic Farmers Survey, OFRF

Figure 2: Production Concentration Versus Price Change, 2001



Source: Fourth National Organic Farmers Survey, OFRF

Organic grain crop acreage accounted for the largest 19.4 percent of total organic acreage in 2001. Organic fruit and vegetable acreage constituted 2.4 and 3.1 percent of the total organic acreage, respectively.

According to the Fourth National Organic Farmer's Survey conducted by the Organic Farming Research Foundation (OFRF) in 2001, organic production grows constantly in the U.S. and little by little substitutes conventional food production. The market of 51 percent organic producers expanded in 2001; the market of 39 percent of respondents held steady, and it contracted for only 9 percent. Two thirds of those who scaled down operations (6 percent versus 3 percent) suffered a price drop of more than 10 percent during the year. Two thirds (39 percent to 12 percent) of the farmers who expanded enjoyed a price increase of 10 percent and more.

Fifty six percent of the OFRF survey respondents indicated that prices held steady in 2001, 28 percent indicated that prices went up, and 16 percent said prices went down. The price decreases reported were both small, less than 10 percent for 7 percent of the survey respondents, and large — more than 10 percent for 9 percent of farmers. Price increases were distributed less equally: 18 percent of farmers faced an increase of less than 10 percent while prices went up by more than 10 percent for only 10 percent of producers. The distribution of average price change for farmers of different income implies more favorable growth conditions for smaller farms: high-income farmers faced more price decreases than lower-income farmers.

Though organic food markets are growing, farmers indicate some barriers to expansion. These are lack of information on prices and unavailability of price forecasts. The availability of future prices is important to decision-making for it helps farmers to make production decisions. Figure 1 illustrates the point by combining the market expansion and price change information. A year-long decrease by more than 10% in the received price has caused almost a half of farms to contract; see Panel (a). Smaller decreases bring this percentage down to a quarter; see Panel (b). Meanwhile, three out of four farms have responded to price increases by scaling up their operations; see Panels (c)–(d). A seemingly small change in price expectations can thus have a profound effect on the farmer's market expansion.

The state of organic fresh produce markets varied for commodity groups. Fruit producers experienced worse growth conditions in 2001 when compared to vegetable producers. Sixty four percent of fruit producers contracted when facing a price decrease of more than 10 percent, while this share was only 32 percent for the farmers that specialize in vegetables (that is, those having more than 50 percent of their land allocated to vegetable production). The picture is similar for smaller price decreases, 58 percent versus 20 percent, and for price increases.

Figure 2 displays farmers' concentration plotted against the price change categories. The concentration index on the value axis represents acreage percentages dedicated to fruit and vegetable production, summed up across all survey participants that grow either fruit or vegetables, or both. One can see that fruit production makes about three fourths of those farms that faced a price decrease over the year, while vegetable production makes two thirds of those farms that enjoyed an increase in the received price.

As a summary, the brief exploratory analysis of organic produce markets reveals several important market trends:

- A small change in the received price can have a large effect on the farmer's expansion or contraction.
- There are favorable growth conditions for lower-income, smaller farmers, since they receive higher prices for their produce.
- Market growth differs significantly across different types of organic produce.

3 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. We 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, one re-writes 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

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

and

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

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

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

or

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

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 the lag polynomial.

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.

4 Forecasting Methods and How to Evaluate Them

An industry-oriented forecasting method should meet a number of requirements: it should be easy to implement with conventional software; it should give quick price forecasts without much of an analytical input; and it ought not to require more information than is contained in the series being forecast.

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 we seek.

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.

It can be seen therefore that, if it so happens that a component of the data-generating process dominates the others for a majority of series, then the method which is the most suitable to deal with that component will be, overall, the best performing one. Otherwise, a more broadly applicable method should be preferred.

When faced with several forecasts of the same uncertain event, a forecast user attempts to identify the best point forecast. The best forecast is then used in the decision-making process, 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 point forecast, the sampling prediction error distribution should be examined. This technique is known as 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. Whereas comparisons of forecasts themselves are essentially an attempt to infer on the parameters of distribution on the basis of a single draw from it.

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. On the contrary, 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 dubious probability limits. This matters when the researcher is unsure about the model itself. Finally, the conditional predictive ability test is easily computed using standard regression packages.

5 Data and Method Evaluation Protocol

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

The price data were collected by weekly telephone interviews of brokers and farmers throughout the United States, as deemed appropriate for the particular commodity. The list of sources is confidential and cannot be revealed. 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).

Table 1: Commodity Description

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

Table describes the initial, untransformed data.

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.

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 general was evaluated using the root mean squared error (RMSE) as an aggregate measure of forecast accuracy.

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 $(n_j, n, n_{.j})$, where $n_{.j}$ 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.

Table 2: Root Mean Squared Error 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	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.

6 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.

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.

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 allow 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.

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 found 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.

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 a negative one between the precision and commodity price volatility were found. This emphasized 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 thus provided. The relevance of joint forecasting of prices in the whole marketing channel of the product was underlined.

Directions for future research were identified as follows.

- Adaptation of forecasting methods 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.
- Combining several methods. Even though the seasonal ARMA was found to be the best performing method, ARMA forecasts can be combined with those from the Holt-Winters method and spectral decomposition to further improve the forecast quality.

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

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

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. `http://ideas.repec.org/p/wpa/wuwpem/0308001.html`.

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.

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

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.

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.

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.

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

Yussefi, M. (2004). Development and state of organic argiculture worldwide. In H. Willer and M. Yussefi (Eds.), *The World of organic agriculture: Statistics and Emerging Trends* (6 ed.). Bonn, Germany: IFOAM.