



Munich Personal RePEc Archive

Forecasting seasonality in prices of potatoes and onions: challenge between geostatistical models, neuro fuzzy approach and Winter method

Amiri, Arshia; Bakhshoodeh, Mohamad and Najafi, Bahaeddin

GREQAM CNRS – ORS PACA – INSERM U912, 23 rue Stanislas Torrents, F-13006 Marseille, France, Departement of Agricultural Economics, College of Agriculture, Shiraz University, Shiraz, Iran

13. October 2011

Online at <http://mpra.ub.uni-muenchen.de/34093/>
MPRA Paper No. 34093, posted 13. October 2011 / 19:26

Forecasting seasonality in prices of potatoes and onions: challenge between geostatistical models, neuro fuzzy approach and Winter method

Arshia Amiri ^{a, b, *}, Mohamad Bakhshoodeh ^b and Bahaeddin Najafi ^b

^a *GREQAM CNRS – ORS PACA – INSERM U912, 23 rue Stanislas Torrents, F-13006 Marseille, France*

^b *Departement of Agricultural Economics, College of Agriculture, Shiraz University, Shiraz, Iran*

Abstract

This paper, we studied the ability of geostatistical models (ordinary kriging (OK) and Inverse distance weighting (IDW)), adaptive neuro-fuzzy inference system (ANFIS) and Winter method for prediction of seasonality in prices of potatoes and onions in Iran over the seasonal period 1986_2001. Results show that the best estimators in order are winter method, ANFIS and geostatistical methods. The results indicate that Winter and ANFIS had powerful results for prediction the prices while geostatistical models were not useful in this respect.

Keywords: Price; Geostatistical model; Kriging; Inverse distance weighting; Winter's method; Adaptive neuro fuzzy inference system; Potatoes; Onions; Iran

* **Corresponding author.** Postal address: block 14, street 24, Bolvare bahonare shomali, Shiraz, Iran
Postal code: 71768-83967, Tel.: +98 917 308 57 22; fax: +98 711 822 42 22.
E-mail addresses: arshiaamirishiraz@gmail.com (A. Amiri).

1. Introduction

Time series (extrapolatory) methods including moving average, exponential smoothing methods and trend estimation, causal (econometric) methods containing autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA), and Judgmental methods including Delphi method are the most popular methods of forecasting. As stated by Dieng (2008), these wide range of methods may be classified into parametric methods (such as exponential smoothing and ARIMA model) and non-parametric methods. The choice of method to be used highly depends on the structure of time series, and as stated by Umar (2007) to simplicity, accuracy and stability of the data. Choosing the best forecasting method for any particular situation is not a simple task (Armstrong and Green, 2005), and sometimes more than one method may be appropriate based on accuracy measures such as MSE and MAD, etc. However, food price forecasts can gain from a combination of these methods (Joutz, et al. 2000). Exponential smoothing is commonly applied to economic data and contains models with single, double and triple smoothing parameters and take trend and seasonality into account to perform Holt's Linear Exponential Smoothing and Holt-Winter's methods respectively. Both trend and seasonality can be additive and or multiplicative and the latter version of the model is the more widely used method.

Apart from above methods, there are some other approaches such as artificial neural networks and adaptive neural fuzzy interference system (ANFIS) for prediction of economic variables, which ANFIS was used for price prediction in here.

In this paper, a new application of using geostatistical methods for forecasting out of sample data in agricultural economics is suggested. Geostatistical methods are the ordinary methods for forecasting the locations and making map in water engineering, environment, environmental pollution, mining, ecology, geology and geography. There are some evidences (Van Kuylenberg et al. (1982), Voltz and Webster (1990), and, Bishop and McBratney (2001)) that results of geostatistical methods (kriging and IDW) will be exact and supportive than ordinary linear methods of forecast, such as, Geostatistical models improve the probably effects of choosing linear and nonlinear structures or electing functional form for prediction, because they have ability to choose best functional form that can show relations between exogenous variables and endogenous variables. In other words, geostatistical models have ability to mix different functional forms in different structures (Linear, Linear to sill, Spherical, Exponential and Gaussian – see David (1977), Krige (1981), Cressie (1985, 1991), Isaaks and Srivastava (1989) and Hill et al. (1994)).

A huge challenge exists between geostatistical methods, ANFIS and Winter method (which can show the variety of seasonality, trend, cyclical effects and irregularity effects) for forecasting out of sample data – see Black (1997) and Levin et. al (1989).

Based on Central Bank of Iran data, the price of onion and potato in Iran had a high fluctuation over the period of 1986 to 2001, therefore, we decided to choose this period for comparing four methods of forecast.

The rest of this paper is organised as follows: section 2 describes methodology including geostatistical, ANFIS and Winter methods following by the data specification for predicting the out of sample data. The results are discussed in section 3 and the paper ends with policy implications and suggestions.

2. Methodology

In this study, we applied geostatistical analysis and ANFIS together with Winther method to seasonal prices of onions and potatoes from 1986 to 2001 collected from Central Bank of Iran.

2.1. Geostatistical analysis

In here, each variable such as independent and dependent, and its lags, are defined with a dimension in spatial structure. For example, if we want to determinate an unrestricted structure of VEC with one lag we face a 4D space for investigation with geostatistics approaches. In other word, in geostatistics the characteristics of location are the same as variables (exogenous and endogenous) in econometrics.

Geostatistics can be used to determine an unknown value, estimate endogenous variables, produce a map of parameters and confirm sampling process and make a more accurate sample. The first step is to analyze the spatial structure in which semivariogram is the essential tools. Describing and modeling are two parts of analysis structure for predicting semivariogram. The semivariogram is a mathematical description of the relationship between the variance of pairs of observations and the distance separating them (h or dependent variable), i.e. for a 3D space (one endogenous and two exogenous variables), it explains the relationships between population variance within a distance class (y-axis) according to the geographical distance between pairs of populations (x-axis). The semivariance is an autocorrelation statistic defined as:

$$g(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (1)$$

where: $g(h)$ is the semivariance for interval distance class, $N(h)$ is the whole number of sample pairs of observations separated by a distance h , $Z(x_i)$ is the measured sample value at point i , $Z(x_i + h)$ is the measured sample value at point $i+h$. Semivariance is evaluated by calculating $g(h)$ for all possible pairs of points in the data set and assigning each pair to a lag or distance interval class h .

It can provide better resolved variograms when there are sufficient pairs of points at shorter separation distances. In Figure 6, there exists a shape of semivariance calculated in a 3D space where sill is $(C + C_0)$, the nugget variance (or constant amount) is (C_0) and the scale (or differences between nugget and observations separated by distance) is (C) .

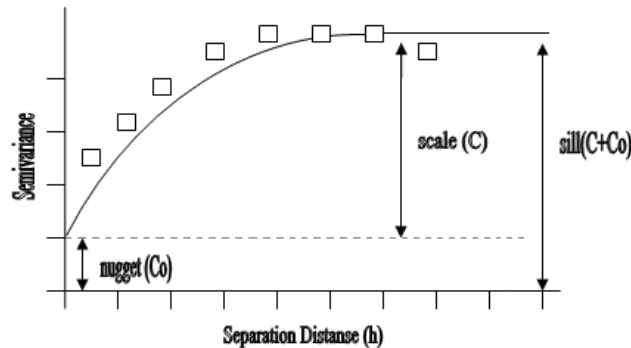


Figure 1. semivariance parameters in on surface.

In spatial structures we can calculate uncounted Semivariance in every degree. Collection of four semivariances in space is called variogram¹. The next step is to analyse the variogram and find the type of variogram for our observation.

To create a 'trustworthy' variogram, different steps must be respected. Different lag distances have to be tested until a sufficient number of pairs to represent the model are found. Four representative groups of pairs are sufficient to represent a relevant variogram with a significant R^2 and a good 'nugget-to-sill' ratio. The effective lag distance cannot be more than half of the maximum distance between data (see Isaaks and Srivastava, 1989).

Burgos et al. (2006) explain that direct dependence has to be tested in the spatial autocorrelation. The isotropic (no directional dependence) or anisotropic (directional dependence) characteristic of the variogram has to be determined. If no anisotropy is found, it means that the value of the variable varies similarly in all directions and the semivariance depends only on the distance between sampling points.

At last the best variogram model (exponential, linear, etc.) and its parameters (nugget, sill, scale, range, etc.) have to be determined in order to validate the modeling of the spatial autocorrelation through the variogram's parameter optimization. The last step is to challenge between ordinary geostatistical methods (kriging and IDW) for predicting dependent variable.

2.1.1. Ordinary Kriging

Kriging provides a means of interpolating values for points not physically sampled using knowledge about the underlying spatial relationships in a data set to do so. Variograms provide this knowledge. Kriging is based on regionalized variable theory and is superior to other means of interpolation because it provides an optimal interpolation estimate for a given coordinate location, as well as a variance estimate for the interpolation value (Gamma Design Software, 2004). In kriging, before determining the models, it is necessary to evaluate variogram to realize whether it is isotropic or anisotropic. The best way to evaluate anisotropy is to view the anisotropic semivariance surface (Semivariance Map), if anisotropic semivariance surface was symmetrical variogram would be isotropic, and if it was asymmetrical variogram would be anisotropic. The differences between variogram types, isotropic and anisotropics, lead to calculate same or various weights in space for kriging model. After the variogram estimation, the interpolation between the measurement points was carried out. To do this, ordinary kriging method was used to interpolate a great number of local scour maps of exogenous and endogenous variables². Geostatistical and spatial correlation analyses of basic infiltration rate redistribution were performed with version 5.1 of GS^+ software (Gamma Design Software, 2004).

2.1.2. Inverse distance weighting

Inverse Distance Weighting (IDW) is interpolation techniques in which interpolated estimates are made based on values at nearby spatial locations of our observation weighted only by distance from the interpolation location. IDW does not make assumptions about spatial relationships except the basic assumption that nearby points ought to be more closely related than distant points to the value at the interpolate location. Similar to kriging, inverse distance

¹ In geostatistics it is ordinary to calculate four semivariances in 0, 45, 90 and 135 degrees.

² For more explanation of Kriging method see Isaaks and Srivastava (1989).

weighting (IDW), exactly implements the hypothesis that a value of an attribute at an unsampled location (variable) is a weighted average of known data points within other local neighborhoods surrounding the unsampled location (Robinson and Metternicht, 2006). In other word an improvement on simplicity giving equal weight to all samples is to give more weight to closet samples and less to those that are farthest away. One obvious way to do this is to make the weight for each estimated as follows:

$$\hat{Z}(x_0) = \frac{\sum_{i=1}^n Z(x_i) d_{ij}^{-r}}{\sum_{i=1}^n d_{ij}^{-r}} \quad (2)$$

Where x_0 is the estimation point and x_i are the data points within a chosen neighborhood. The weights (r) are related to distance by d_{ij} , which is the distance between the estimation point and the data points. The IDW formula has the effect of giving data points close to the interpolation point relatively large weights whilst those far away exert little influence.

2.2. Neuro-fuzzy model structure

Neuro-fuzzy modeling is a powerful problem-solving methodology with many applications in embedded control and information processing. It provides a simple way to draw definite conclusions from vague, ambiguous or imprecise information. In a sense, neuro-fuzzy modeling resembles human decision making with its ability to start from approximate data and find precise solutions; also the method is similar in nature to that of neural networks. From the ‘cross-pollination’ between fuzzy logic and neural networks a new modeling approach has emerged: the Adaptive Neuro-fuzzy Inference System (ANFIS). ANFIS takes a given input/output data set and constructs a fuzzy inference system whose membership function parameters are tuned, or adjusted, using either a back propagation algorithm alone or a combination of the latter with a least squares type of method. The basic structure of ANFIS consists of three conceptual components: a ‘rule base’ which contains a selection of fuzzy rules, a ‘data base’ which defines the membership function (MFs) used in the fuzzy rules, and ‘reasoning membership’ which performs the inference procedure upon the rules to derive an output as shown in Figure. 2.

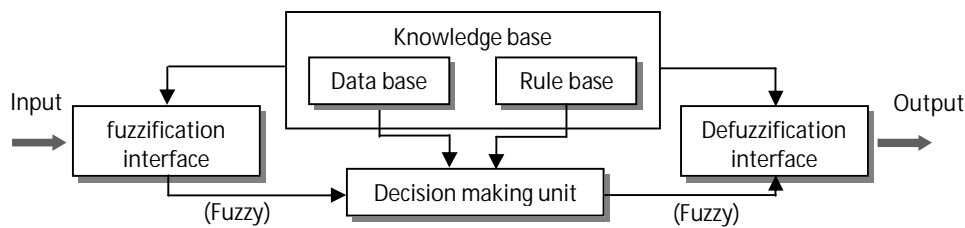


Figure 2. Fuzzy Inference System (Source: Jang, 1993)

ANFIS has been introduced by Takagi and Sugeno (1983). Readers are referred to Jang (1993) for a thorough description of the system and to Mathworks (2001) for guidelines concerning its practical implementation.

2.3. Winter method

The equations for multiplicative seasonality version of Holt-Winter's Method are

$$\begin{aligned}
 L_t &= a \frac{Y_t}{S_{t-s}} + (1-a)(L_{t-1} + b_{t-1}) \\
 b_t &= b(L_t - L_{t-1}) + (1-b)b_{t-1} \\
 S_t &= g \frac{Y_t}{L_t} + (1-g)S_{t-s} \\
 F_{t+m} &= (L_t + b_t m)S_{t-s+m}
 \end{aligned} \tag{3}$$

where L_t and b_t are respectively (exponentially smoothed) estimates of the level and linear trend of the series at time t , s is the number of periods in one cycle of seasons e.g. number of months or quarters in a year. F_{t+m} are the linear forecast from t onwards. The parameters a, b, g should lie in the interval $(0, 1)$, and can be selected by minimising forecast accuracy measures such as MAD, MSE or MAPE, etc.

In this study, the performances of all Winter method, ANFIS and geostatistical approaches configurations were assessed based on calculating the mean absolute error (MAE), and ordinary least squares (RSQ). The coefficient of determination, R^2 , of linear regression line between the predicted values from the Winter method, ANFIS and geostatistical approaches and the desired output was also used as a measure of performance. The two statistical parameters used to compare the performance of the various geostatistical approaches configurations, ANFIS and Winter method are:

$$\begin{aligned}
 MAE &= \frac{1}{N} \sum_{i=1}^N |O_i - t_i|, \\
 R^2 &= 1 - \frac{\sum_{i=1}^N (O_i - t_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2}
 \end{aligned} \tag{4}$$

Where O_i and t_i are observed and predicted for the i^{th} output (observed output that comes from catching-up effect and predicted output that Winter method and geostatistical approaches forecasts it), and \bar{O}_i is the average of predicted, and N is the total number of events considered. The models that minimized the two error measures described in the previous section (and maximize RSQ) were selected as the optimum. The whole analysis was repeated several times. In here, two exogenous variables were selected for the inputs of ANFIS or X and Y coordinates in OK and IDW. Which one is a lag of dependent variables; another is a dummy variable for definition a shock which was happened on spring in all years.

3. Results

In this section we focus on the results of Winter method, ANFIS and geostatistical approaches for forecasting out of sample data (about 20% of observations for 3 last years of the data) with two ordinary tests of prediction.

The results of two methods of geostatistical analysis show as follows: the best type of variogram for both onions and potatoes were isotropic, and best model of variogram was investigated with spherical for onions and linear for potatoes.

In this study, two types of ANFIS models were used: the grid partition ANFIS method and the subtractive clustering ANFIS method. Clustering and grid partition can be very effective techniques to identify natural groupings in data from a large data set, thereby allowing concise representation of relationships embedded in the data (see Chiu, 1994). A small cluster radius usually yields many small clusters in the data – resulting in many rules – and vice versa. In general, as the number of rules increases, the difference between the estimated values and experimental values decreases, and more complex relations can be modeled with a larger number of rules. In grid partition, an investigation of all available membership functions (MF) was conducted in an attempt to determine the optimum type of function for use with the present model. In addition to comparing the different types of MF's, the number of MF's used in the fuzzy model was also studied. In general, the model error is reduced as the number of MF's increases. However, there are limitations to increasing the number of MF's. The more MF's are selected, the longer the model will take to compute the fitting parameters of each MF. It was therefore decided to minimize the number of MF's, since the marginal improvement from using higher numbers was judged not to be significant.

The Granger-Newbold test (see Granger and Newbold, 1976) was estimated for choosing best generating ANFIS between grid partition and subtractive clustering. In both, best structure of grid partition had better results instead of best result of subtractive clustering. For onions, best result in grid partition was obtained with 4 MF's and 16 rules, and for potatoes, with 3 MF's and 9 rules. Last, best results of winter were gained by 0.2 smoothing contact for level, trend and seasonal for both crops.

All results are available in Table 1, which shows that the best estimators in order are winter method, ANFIS, kriging, and IDW. Winter and ANFIS had useful results for prediction the prices but geostatistical models had weak results. ANFIS had a better investigation for onions instead of potatoes, but the results of winter are almost same for both crops, and more powerful than others methods. Results of training for winter is blanked for the reason of navigate of it. Performance of out of sample for winter is available in figure 3.

Table 1. The gist of results of methods

		Validation (out of sample)	
		MAE	RSQ
onions	Kriging	290.10	0.04
	IDW	200.06	0.00
	ANFIS	119.04	0.81
	Winter	44.83	0.96
potatoes	Kriging	119.39	0.04
	IDW	166.31	0.01
	ANFIS	95.20	0.44
	Winter	42.25	0.93

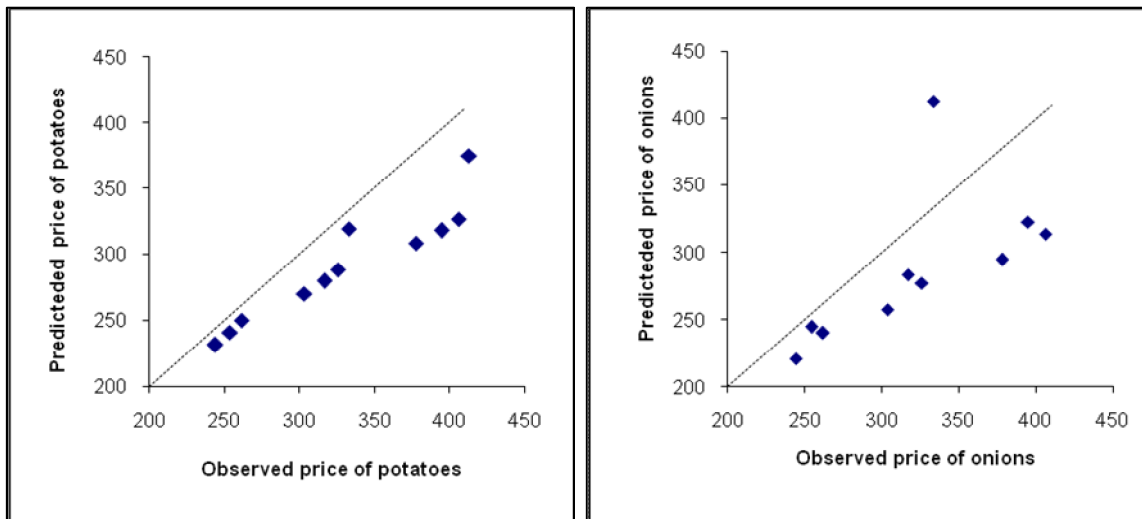


Figure 3. Performance of out of sample for winter method

4. Conclusion

In this paper a huge challenge exists between geostatistical methods, ANFIS and Winter method for predicting the price of potatoes and onions in Iran over the seasonal period of 1986_2001. The empirical results show that the best estimators in order are winter method, ANFIS, kriging, and IDW. Winter and ANFIS had useful results for forecasting the prices but geostatistical models were weak.

References

- Armstrong, J.S., Green, K.C., 2005. Demand Forecasting: Evidence-based Methods, Monash University, *Department of Econometrics and Business Statistics, Working Paper 24/05*, ISSN 1440-771X.
- Black, K., 1997. Business statistics, contemporary decision making, 2nd Ed., *West Publishing Company*.
- Bishop, T.F.A., McBratney A.B., 2001. A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma*, 103, 149-160.
- Burgos, P., Madejon, E., Perez-de-Mora, A. and Cabrera, F., 2006. Spatial variability of the chemical characteristics of a trace-element-contaminated soil before and after remediation. *Geoderma*, 130(1-2), 157-175.
- Chiu, S.L, 1994. Fuzzy model identification based on cluster estimation, *Journal of Intelligent and Fuzzy Systems*, 2, 267-278.

- Cressie, N., 1985. Fitting variogram models by weighted least squares. *Mathematical Geology*, 17, 563-586.
- Cressie, N.A.C., 1991. Statistics for Spatial Data. *John Wiley*, New York, USA.
- David, M., 1977. Geostatistical Ore Reserve Estimation, *Elsevier, Scientific Publishing Co.*, Amsterdam, The Netherlands.
- Dieng, A., 2008. Alternative Forecasting Techniques for Vegetable Prices in Senegal, *Revue sénégalais de recherches agricoles et agroalimentaires*, 1 (3), 5-10.
- Duffera, M., White, J.G. and Weiz, R., 2007. Spatial variability of Southeastern U.S. Coastal Plain soil physical properties: implications for site-specific management. *Geoderma*, 137, 327-339.
- Gama Design Software, 2004. GSp Version 5.1. Geostatistics for the Environmental Sciences, User's guide. *Gama Design Software, LLC* (160 pp.).
- Granger, C.W.J., Newbold, P., 1976. The use of R^2 to determine the appropriate transformation of regression variables. *Journal of Econometrics*, 4(3), 205-210.
- Hill, T., Marquez, L., O'Connor, M. and Remus, W., 1994, Artificial Neural Network Models for Forecasting and Decision Making. *International Journal of Forecasting*, 10, p.5-15.
- Isaaks, E.H., Srivastava, R.M., 1989. Applied Geostatistics. *New York Oxford University Press*, pp: 257-259.
- Jang, J.R., 1993. ANFIS: Adaptive-Neural-Network-Based Fuzzy Inference System. *Man and Cybernetics*. 23(3), 665-685.
- Joutz, F.L., Trost, R.P., Hallahan, C., Clauson, A., Denbaly, M., 2000. Retail food price forecasting at ERS: the process, methodology, and performance from 1984 to 1997. *Economic Research Service*, U.S. Department of Agriculture. Technical Bulletin No. 1885.
- Krige, D.G., 1981, Lognormal-de Wijsian geostatistics for ore evaluation. South African Institute of Mining and Metallurgy Monograph Series. *Geostatistics I. South Africa Institute of Mining and Metallurgy*, Johannesburg, South Africa.
- Mabit, L., and Bernard, C., 2007. Assessment of spatial distribution of fallout radionuclides through geostatistics concept. *Journal of Environmental Radioactivity*, 97, 206-219.
- Mathworks, 2001. "Fuzzy Logic Toolbox, for use with MatLab." *User's Guide*, Version 2.
- Robinson, T.P. and Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. *Computer and Electronics in Agriculture*, 50, 97-108.

Takagi, T., Sugeno, M., 1983. Derivation of fuzzy control rules from human operator's control actions. Proc. Of the IFAC Symp. on Fuzzy Information. *Knowledge Representation and Decision Analysis*, 55-60.

Taylor, J.W., 2003. Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19, 715-725.

Umar, M.A., 2007. Comparative study of Holt-Winter, double exponential and the linear trend regression models, with application to exchange rates of the naira to dollar. *Research Journal of Applied Sciences*, 2(5), 633-637.

Van Kuilemberg, J., De Gruitjer, J., Marsman, B., Bouma, J., 1982. Accuracy of spatial interpolation between point data on soil moisture supply capacity, compared with estimates from mapping units. *Geoderma*, 27, 311-325.

Voltz, M., Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. *Journal of Soil Science*. 41, 473-490.

Zou, H., Yang, Y., 2004. Combining time series model for forecasting. *International Journal of Forecasting*, 20, 69-84.