Frost Occurrence Risk Management for Pistachio Industry in Rafsanjan

Alireza Hosseini[1], Mohammad-Saber Fallah Nezhad[1],
Yahya Zare Mehrjardi[1], Reza Hosseini[2]

1 Department of industrial engineering, Yazd University, Iran

2 Division of Biostatistics, University of Southern California, USA

Abstract

This work develops a statistical model to assess the frost risk in Rafsanjan, one of the largest Pistachioproduction regions in the world. These models can be used to estimate the probability that: a frosthappens in a given time-period in the year; a frost happens after 10 warm days in the growing season etc.These probability estimates then can be used for: (1) assessing the agroclimate risk of investing in thisindustry; (2) pricing of weather derivatives. Autoregressive models with different seasonal componentsand lags are compared using AIC, BIC, AICc and cross validation criterions. The optimal model is AR(1)with 12 terms from Fourier series. The long-term trends are also accounted for and estimated from data.The optimal models are then used to simulate future weather from which the probabilities of appropriatehazard events for pistachio yield are estimated.

Keywords: Pistachio; Frost; Weather derivative; Minimum temperature; Time-varying autoregressivecoefficients.

1. Introduction

The greater Rafsanjan area in north of Kerman Province in Iran is a region with the largest pistachioproduction in the world and most of the region's economy relies on pistachio production. In therecent years the most important risk factor for pistachio producers and industry (e.g. farmers,distributers) has been frosts that have destroyed a large proportion of the yield. Therefore methodsthat can estimate the probability of such events is useful. In particular such methods can: (1) assess the agroclimate risk of investing in this industry; (2) pricing of weather derivatives. In factweather derivatives, which may be created as part of a risk management program, can be writtenin terms of the attainment or non-attainment of specific target-values stipulated in the contract.Temperature-related trades account for 80% of the transactions among all weather derivatives [1]. Most of the work in this area has focused on HDD/CDD (heating degreedays/cooling degree days) (e.g. [2,3]). In this paperwe focus on the occurrence of frosts an issue recently considered in [4], foragricultural crops in Canada.

The models developed in this paper can be applied to estimate: the probability that a given period is frost-free; the probability that a given day is the start of a long frost-free period;the distribution of the length of the frost-free period and so on. The same model can be used tocompute the probability that a given day of the year is the beginning of the growing season (the first day that the mean temperature is higher than 5 degrees for 5 consequent days) as well as thelength of the growing season which are important for agricultural applications. For example in thisstudy we estimate the probability of a useful event: *"the minimum temperature goes below zero atleast one day in the period March 27th-April 20th"*. This is an important event because it coincideswith the general flowering time of pistachio trees. Throughout this paper, temperature is measuredin degrees Celsius. Let us denote the minimum temperature series by {Y (t)}, t = 0,1,2,…, wheret denotes time. We let F to be the investor's defined frost which we take it to be zero in this work.Then we can define the binary frost process:

$$Y_F(t) = \begin{cases} 1 & Y(t) \leq F \text{ (deg C)} \\ F & Y(t) > F \text{ (deg C)}. \end{cases}$$

In order to study frosts we can use these approaches among others: (a) Fit the continuous-valued Markov model to the Y (t) chain; (b) Fit a binary Markov model to the YF (t) chain. Hosseinisuggests using binary Markov models to avoid assumptions regarding the distribution of temperature and gain robustness for modeling frosts in Alberta, Canada[4]. They show time-varyinghigh-order Markov models with complex seasonal structure are needed and therefore their computations become challenging. Here we investigate Method (a) in fitting such chains and calculatingthe probabilities of frost events. The advantages of Method (a) are: (1) Thefitting can be donewith standard packages such as R with less computational problems; (2) only this method canestimate the probability of complicated events. One such complicated event is: *"the temperaturein March-April is above 5 (deg C) for at least 3 consecutive days and is below zero after"*. A comparison of the two methods in terms of estimation when they are both applicable is left to futureresearch.

2. Data and statistical models

The data in this study are daily minimum temperature values collected at Rafsanjan weather station from 1992 to 2010. At the moment we do not have access to more data from other stationsin the area but we hope to acquire those data for future studies to offer more local predictions. Inorder to model frost occurrences, we introduce statistical models for minimum daily temperaturein Rafsanjan. Several features of the temperature process should be considered in modeling: (1) seasonal trends over time; (2) long-term trends (possibly a result of global warming or volcanicevents etc) (3) dependence over time. Let {Y (t)}, t = 0,1,2,…, T denote the daily minimumtemperature process in centigrade, where t denotes the day starting from March 1st 1992 toDecember 28th 2010. Here we consider autoregressive models with a seasonal component andvarious lags:

$$Y(t) \quad = \quad \mu(t) + \epsilon(t), \quad \mu(t) = a_0(t) + \sum_{i=1}^{r} a_i Y(t-i)$$

Where $\mu(t)=E\{Y(t)|Y(t-1),Y(t-2),\ldots\}$ is the conditional mean of minimum temperatureat time t; $\varepsilon(t)$ are independent identically distributed normal errors $\varepsilon(t) \sim N(0,\delta^2)$; $a_0(t)$ is thefixed trend coefficient; $a_1, a_2,\ldots,a_r$are autoregressive coefficients. We allow $a_0(t)$ to include bothseasonal and long-term effects by using a Fourier series with period, $\omega = \frac{2\pi}{366}$, and a quadratictrend:

$$a_0(t) = \{\alpha_0 + \sum_{j=1}^{k} \alpha_j \cos(j\omega t) + \beta_j \sin(j\omega t)\} \quad + \quad \{\gamma_1 t + \gamma_2 t^2\},$$

The estimation of the parameters is done by maximizing the (partial) likelihood of the data asdiscussed in [4]. The nice property of the gaussian error assumption is themaximization can be done in closed form exactly in the same way as minimizing the mean-squareerror of classical regression problem originally solved by Gauss and therefore the estimation is fastin statistical packages such as "R"(a free widely used software by statisticians and practitioners).

3. Statistical model selection

In the above we introduced several autoregressive models of: (1) various lags; (2) various seasonal complexity (number of Fourier terms); (3) various long-term trends. Therefore we need to usesome criteria to select an optimal model. The problem of model selection is an important onein statistical theory and application. Various criteria are suggested in the literature for example:AIC in [5]; BIC in [6] and AICc in [7]. Denote thelikelihood of the data by L (in this paper the "partial likelihood"), the number of covariates by pand the sample size by n. Then we have

$$AIC = 2p - 2\ln(L), \quad AICc = AIC + \frac{2k(k+1)}{n-k-1}, \quad BIC = p\log(n) - 2\ln(L).$$

Since n in our data is large compared to k, AIC and AICc are very close. When we compared themodels using these criteria, AIC and AICc give rise to the same optimal model while BIC pickeda simpler model. In Table 1 we have compared these optimal models using cross-validation errorand cross-validated correlation. The cross-validation proceeds by: (1) taking an existing data pointout; (2) fitting the model; (3) predicting the value of the point we took out (validation). Then thecross-validation error (CVE) is the mean square error of the predictions and the cross-validationcorrelation (CVR) is the correlation between the predictions and the observed. Table 1 shows thatwhile the CVE and CVR are very close for the two models, the model picked by AIC/AICc slightlyoutperforms the one picked by BIC and therefore we use that model for estimation.

Criterion	optimal Model: Z_{t-1}	CVE	CVR
AIC and AICc	$sin(\omega t), cos(\omega t), \cdots, sin(6\omega t), cos(6\omega t), Y_{t-1}, t, t^2$	**2.691**	**0.9458**
BIC	$sin(\omega t), cos(\omega t), Y_{t-1}$	2.696	0.9456

Tab 1. We compare the optimal model picked by AIC and AICc (first row) with the optimal model picked by BIC, (second row) using cross validation error and cross-validated correlation.

4. Applications in frost risk assessment

Previous section found an optimal fit to the data from which estimating the probability of anydesired (possibly complex) event is possible by performing multiple simulations. In order to find out the probability of frost in any given day during 2011-2012 we have done 10000 simulationsfrom the model for 2011-2011 and then for each day we have calculated the proportion of frostdays (number of frost days divided by 10000). The results are plotted in Figure 1.As we pointed out in the introduction because the flowering time of different varieties ofpistachios in Rafsanjan is generally between March 27th to April 20th, it is important to investigate the frost-occurrence during this period which we call the hazard period. Figure 2 shows thedistribution of the "number of frost days" during the hazard period of 2012, where the frequencyout of 10000 of any "number of frost days" is plotted. We observe that while it is most likely thatno frost occurs in that period, there is a considerable probability that there are at least one frosts.This probability turns out to be about 9 percent which is a plausible number with our experience of pistachio damages caused by frosts in the past 20 years.

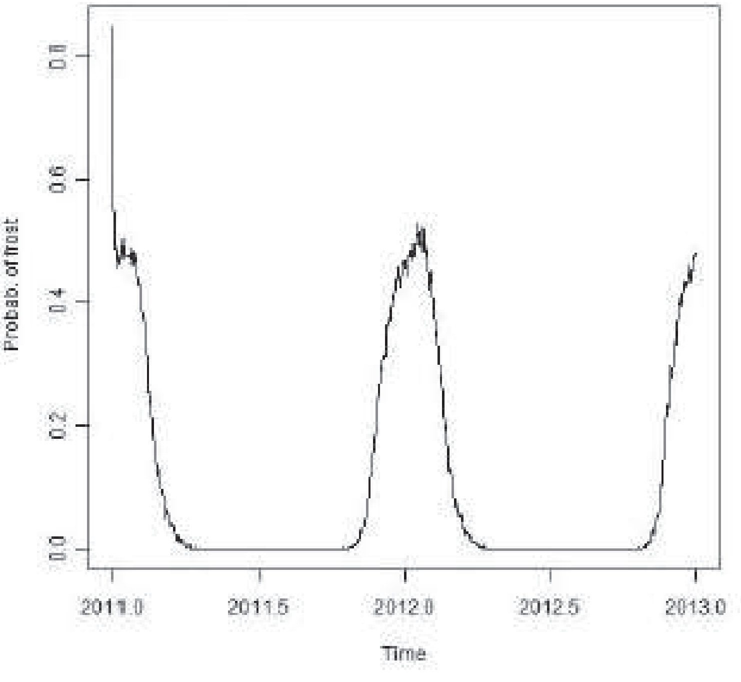

Fig 1. Estimated daily frost probability for 2011-2012 from the model, obtained using 1000 simulations of future weather.

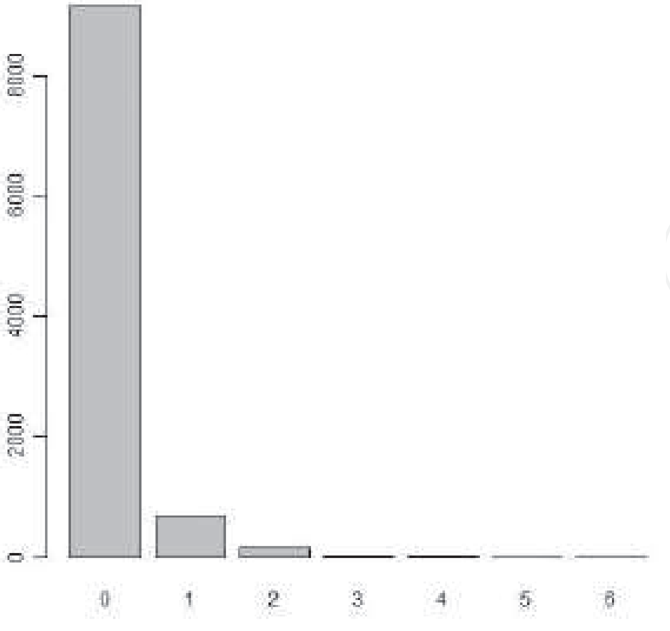

Fig 2. Distribution of frost days during the Hazard period (March 27th to April 20th 2012). This is based on 10000 simulations of the future chains. The probability of at least one frost based on this simulations is 0.0872 which is about 9 percent.

5. Summary and conclusion

This paper developed and compared several statistical models to estimate the probability of hazardfrost events for pistachio industry in Rafsanjan. Despite the importance of such risk factors, nosystematic studies and estimations of these risks are available in this region as far as we know;this paper is one of the first attempts in developing methods that can assess such risks. Assessingthe probabilities of the hazard events are useful in estimating the risk of investing in this industryfrom production to distribution and exporting. However here we have not investigated other riskfactors such as: extremely high temperature during summer; heavy short-time rain during floweringperiod; slow but long rain during the flowering time. For future studies we plan to acquire thedata for precipitation, maximum temperature and developing models that assess these other riskfactors.

Another important aspect of assessing the risk is relating the risk factors to the losses inyield or monetary values involved. For this study we relied on expert knowledge (by interviewingfarmers and agriculture engineers) to define our hazard period. However if for example data foryield per km2 becomes available for enough number of years and/or locations, one can develop astatistical model to relate the weather events to the losses in the yield in the same model.

6. Acknowledgements

We are indebted to Mr. islami for providing the data for this study.

7. References

[1] Cao, M. and Wei, J. Weather derivatives, valuation and market price of weather risk The Journal of Future Markets, 24(11):1065-1089, 2004.

[2] Richards, T. J., Manfredo, M. R. and Sanders, D. R. Pricing Weather Derivatives American Journal of Agricultural Economics, 86(4):1005-1017, 2004.

[3] Benth, F. E., and Benth, J. S. Volatility of temperature and pricing of weather derivatives Quantitative Finance, 7(5):553-561, 2007.

[4] Hosseini.R, Le.N, Zidek.J,2012. Time-Varying Markov Models for Binary Temperature Series in Agrorisk Management. Journal of Agricultural Biological and Ecological Statistics, Vol. 17,Num. 2,283-305.

[5] Kedem, B. and Fokianos, K. Regression Models for Time Series Analysis, Wiley Series in Probability and Statistics, 2002.

[6] [Akaike, H. A new look at the statistical model identi_cation. IEEE Transactions on AutomaticControl, AC- 19:716-723, 1974.

[7] Schwartz, G. Estimating the dimension of a model. Annals of Statistics, 6:461-464, 1978.

[8] Brockwell, P.J., and Davis, R.A. Time Series: Theory and Methods, 2nd edition Springer,2009.