



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

JOURNAL OF THE

Northeastern Agricultural Economics Council

WAITE MEMORIAL BOOK COLLECTION
DEPARTMENT OF AGRICULTURAL AND APPLIED ECONOMICS
232 CLASSROOM OFFICE BLDG.
1994 BUFORD AVENUE UNIVERSITY OF MINNESOTA
ST. PAUL, MINNESOTA 55108



PROCEEDINGS ISSUE
VOLUME IX, NUMBER 2
OCTOBER, 1980

AN APPLICATION OF THE BOX-JENKINS APPROACH TO FORECASTING THE COLD STORAGE OF FRENCH FRIES

Richard P. Beilock and James W. Dunn

Potatoes suitable for frozen french fries are harvested in the fall and then apportioned for the rest of the year by processors. This involves storage in either fresh or frozen form. Participants in both the fresh and french fry market consider the level of frozen french fry stocks to be an important determinant of activity in these markets and therefore use stock information to form price expectations [Phillips *et al.*, 16-18]. The ability to forecast the future level of stocks with some accuracy would undoubtedly be an additional aid in developing market strategies. This paper provides one means of forecasting frozen french fry stocks.

The forecasting of stocks or of any variable may be approached by using three general methods: Delphic, causal, or pattern recognition. The first method, which probably receives more use than is generally admitted, consists of nothing more than collecting the educated guesses of experts. The major source of uneasiness with this method, particularly for economists, is that the methods by which the forecasts are calculated are generally subjective, or at least not specified.

The causal or structural approach is seen in much more favorable light precisely because it attempts to define the relationships which affect the variable in question. Several models, such as Gustafson, or Brennan, have been used to explain why a stock in time t is at level S_t using such variables as current and expected commodity prices and storage costs. The value of such models for understanding the system and how it works cannot be denied. However, there are two serious problems with using such models for forecasting one component of the system. First, there often are practical problems of data limitations both with respect to defining the variables needed and to the existence of the data. Suppose for example, that the price paid for frozen french fries is a determinant of stocks held. What price or prices are appropriate? The retail price in New York or Dallas? Supermarket wholesale prices? The prices paid by fast food restaurants? Are the same prices relevant to companies that own potato fields, processing plants, storage facilities, and retail outlets? Even if all the above questions could be answered and modeled, much of the above data is nonexistent or unavailable.

The second problem is that causal models are generally not predictive models at all. Suppose that a model is constructed with storage, S_t , as a function of price, P_t , which is here assumed to be exogenous. What will be the level of storage two periods ahead? Without knowing P_{t+2} the model provides no clues. What the model does is to substitute for the problem for forecasting S_{t+2} the problem of forecasting P_{t+2} . Clearly this would only be desirable if there was greater confidence in the ability to forecast prices. Admittedly, if S_t were only a function of variables lagged k periods, then S_t could be predicted for k periods into the future. However, it is doubtful that k would be large in most models.

Pattern recognition operates on the assumption that a series of observations on a variable over time, say $(S_{t-n}, S_{t-n+1}, \dots, S_t)$, is a

specific outcome of random variables which are jointly distributed across time. That is, the observations are related to each other over time, and therefore, inferences can be made about probable future values from knowledge of current and past values by detecting and projecting past patterns. At first this may seem as conceptually deficient as the Delphic approach. It is subject to error due to possible structural changes in the future. But what predictive technique is not subject to such error? Pattern recognition does not necessarily provide answers about the decisionmaking mechanism. But it does not purport to explain phenomena; it only purports to predict them (just as more traditional econometric models explain but do not predict).

The use of pattern recognition techniques need not be devoid of a theoretical basis however. Indeed, theoretical considerations may be used to determine whether it would be reasonable to assume that observations on a variable might exhibit patterns over time and to guide the analyst as to the form of those patterns to be expected, thereby aiding in model specification and/or validation. For example, a seasonal pattern might be expected due to the nature of the demand, as with turkeys, or the nature of the supply, as with fresh strawberries. Autoregressive relationships might be anticipated if it is known to be impractical or infeasible to radically change the level of some variable in a single time period.

In the sections which follow, end-of-month stocks of frozen french fries are predicted using one class of univariate pattern recognition techniques—those developed and popularized by Box and Jenkins. The next section presents a brief, nonmathematical overview of the Box-Jenkins (B-J) approach. This is followed by a statement of *a priori* theoretical expectations regarding the form of an appropriate model. Next, these expectations, along with the more mechanistic techniques of the B-J approach, are used to identify (specify) and estimate a model for prediction. Finally, the predictions of the model are compared with actual values.

OVERVIEW OF THE BOX-JENKINS APPROACH

The B-J approach assumes that there are basically two types of intertemporal relationships, autoregressive and moving average. A relationship is autoregressive if the observation in time t depends upon observations in past periods. For example a variable, S_t , is autoregressive of order two (AR2) if it is a function of the previous two observations, S_{t-1} and S_{t-2} . Such a situation might arise if it takes two periods to adjust to S_t . A relationship is of the moving average type if a variable is dependent upon some past disturbances, $(U_{t-n}, U_{t-n-1}, \dots)$, assumed to have a constant variance and mean zero. The sense of this, loosely speaking, is that some aberration(s) from the "norm" or expected value in the past may affect present performance. For example, if a grocery store experienced unusually heavy demand for turkeys last November (seasonally adjusted) the manager may well react to this past disturbance by stocking unusually large quantities of turkeys this November. If we were using monthly data to examine the stocks of turkeys from this grocery store, the appropriate model would be a moving average model of order twelve (MA(12)). The order is twelve even though the coefficients on the first eleven lagged disturbance terms may be zeroes.

Richard P. Beilock and James W. Dunn are Research Assistant and Assistant Professor, respectively, at The Pennsylvania State University. The authors wish to thank George Heitmann and our anonymous reviewers for their valuable criticism and comments.

It should, perhaps, be noted that the same information can usually be represented as either a moving average or an autoregressive model. As the two are equivalent, the choice between them depends largely upon consideration of parsimony.

Stationarity

The use of B-J techniques requires that the time series be stationary. Stationarity means that the joint probability distribution from which the series was assumed to have been generated did not change over time. In other words, there is no determinate drift or trend to the series, and the variance and covariance between equally spaced observations are constant.

In symbolic terms:

$$E(S_T) = E(S_{T+m}) \quad \text{all } m, T$$

$$\text{VAR}(S_T) = \text{VAR}(S_{T+m}) \quad \text{all } T \text{ for fixed } m$$

$$\text{COV}(S_T, S_{T+k}) = \text{COV}(S_{T+m}, S_{T+m+k}) \quad \text{all } T, m \text{ for fixed } k$$

A nonstationary series must be transformed so as to approximate a stationary series prior to estimation. This transformation is often accomplished via differencing. For example, some nonstationary series become stationary series by first differencing, i.e. subtracting an observation in one period from the observation in the previous period. Likewise, differencing "across a cycle" or seasonal differencing has the effect of transforming seasonal and other repetitive patterns into stationary series. To get a quick sense of how this would be accomplished, imagine that a series approximates a sine wave. Subtract each value from the preceding observation that was at a similar point in the cycle; a peak observation from the preceding peak observation, a trough from the preceding trough, etc. What would remain is a stationary series. In some cases there is a need for both regular differencing (usually of 1st or 2nd order) and seasonal differencing.

In situations where differencing is not needed, but the series has a nonzero mean, an explicit term for the mean must be included. In cases where the differenced series has a nonzero mean, an explicit term, called a trend term must be included. Suppose, for example, that first differencing results in a mean value of two for a series, this means that an observation is, on average, two units larger than the preceding observation. The "two," therefore, reflects a positive trend of two units per period.

The Model

Since January of 1956, the Crop Reporting Board of the Statistical Reporting Service of the U.S.D.A. has collected and reported the end-of-month stocks of frozen french fries, FFF, held in the United States. Virtually all of these stocks are held by potato processors.

The period used for estimation was from January 1956 to March 1978, 267 observations. The stocks over this period are shown in Figure 1. The March 1978 cutoff was made in order to leave one year of the data (April 1978 to March 1979) with which to compare to the forecasts derived. While this strategy may be costly, as it excludes the latest observations from use in the construction of the forecasting tool, it was felt to be appropriate as a validation measure.

EXPECTATIONS OF THE FORM OF THE MODEL

Frozen potato processing plants are generally located in production rather than consuming centers; in particular, in those states where the potato harvest comes in the late Summer or Fall. Once stocks of potatoes from local sources become exhausted, plants typically close down rather than import supplies from other areas. Since potatoes can only be stored for nine to eleven months, plants plan their production in order to deplete supplies before the

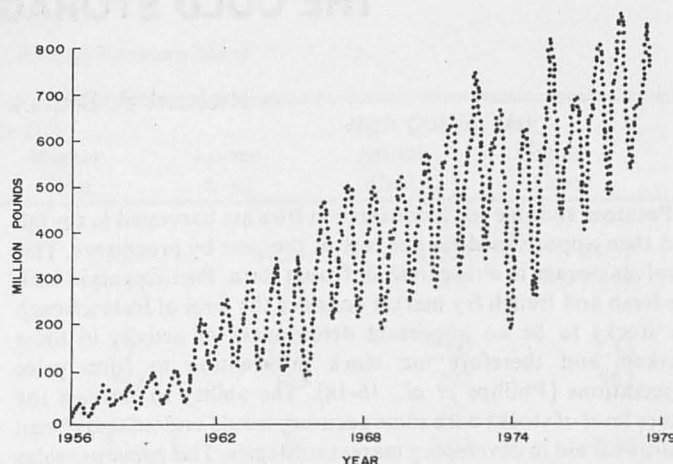


FIGURE 1.
End-of-Month Stocks of Frozen French Fries

next harvest. Large cold storage facilities are located at the plant sites to hold inventories to meet off-season demands. These facts, coupled with the knowledge that the U.S. does not import FFF, suggest that the data will exhibit marked annual cycles necessitating seasonal differencing.

A positive trend is also anticipated, reflecting the tremendous growth that FFF have experienced over the last 22 years (the average per capita consumption of frozen potatoes in 1956 was 2.9 pounds versus 37.9 pounds in 1977). In addition, a positive, regular AR pattern would not be surprising since an unusually high or low inventory in any one month probably must be adjusted for over several months.

In the last few years per capita FFF consumption has begun to level off. Should this trend continue, the model may not adequately account for it. Longer range forecasts may, therefore, tend to overshoot the mark. This should be a consideration when using the model as a forecasting tool.

Identification

In the B-J procedure, models are identified by examining the shape of the estimated autocorrelation function (ACF). Additional clues may be acquired by studying the partial autocorrelation function (PACF).¹ Certain shapes are known to correspond to certain model specifications.

The first objective is to determine if the series is stationary. A nonstationary series is revealed by an autocorrelation that remains large, even when lag lengths are great, e.g. 15 to 30 periods. Moreover, a plotting of such functions is often scalloped in appearance. Stationarity may generally be achieved if the series is differenced L periods, where L is the number of periods between the peaks of the scallop pattern. As expected, this was the case with the FFF series, with L equalling 12 periods, suggesting one year cycles.

Having achieved stationarity, the estimated ACF and PACF of the differenced series was examined. The ACF pattern resembled a dampening sine wave. This form indicated an AR model of order two or greater. The PACF corresponded to that of an AR model of

¹Both the autocorrelation and the partial autocorrelation functions are functions of the length of the lag. The ACF is the simple correlation of an observation of a variable with a lagged observation of the same variable. The partial autocorrelation has a particularly complex formulation which will not be presented here.

order two or, possibly, three or four. In addition, the PACF gave evidence that something was happening at lag 12. As the nature of this activity at lag 12 was unclear, it was ignored in the first fitting procedure.

The first model, then, was differenced 12 periods and had two regular AR terms. The short-hand notation for this is: SARIMA (2,0,0) \times (0,1,0). "SARIMA" stands for Seasonal Autoregressive Integrated Moving Average. The word "integrated" represents the differencing procedure. The numbers of each set of parentheses represent the number of AR terms, differences, and MA terms respectively. The second set of parentheses refer to seasonal aspects. In this case seasonal means 12 periods. The sense of this is that one may think the total system as being the joining of two subsystems. The regular subsystem describes the impact of recent shocks upon current behavior. The seasonal subsystem describes the impact of what was happening one full cycle ago on current behavior.

Estimation

Two items are of primary interest in the estimation of this model: (1) are the parameters all significant, and (2) is anything clearly happening at lag 12 of the residual ACF? The answer to both of these questions is yes. The ACF of the residuals that remains after applying the model reveals a single seasonal MA terms, MA^s1. Therefore a SARIMA (1,0,0) \times (0,1,1) is estimated.

This model appears to be quite satisfactory as it passed the white noise test at the 5 percent level of significance (i.e., could not reject the hypothesis that only white noise or random elements remained as residuals) and all of the parameters are significant as is shown in Table 1.

Because the evidence regarding the order of the AR portion of the model is unclear, overfitting was done for orders 3 and 4. The final parameter for the order 4 model was insignificant, and is not presented. The results of the order 3 model is shown in Table 2.

Final Choice

As the model performs slightly better, in terms of residual sum of squares and residual mean square, when the third AR term is added; the SARIMA (3,0,0) \times (0,1,1) model was chosen. The equation for this model is:

$$S_t = S_{t-12} + 1.41 (S_{t-1} - S_{t-13}) - .70 (S_{t-2} - S_{t-14}) \\ + .16 (S_{t-3} - S_{t-15}) - .56 U_{t-12} + 3.67.$$

It should be noted, however, that the SARIMA (2,0,0) \times (0,1,1) would have been chosen if more weight were placed on parsimony, since the difference in performance is fairly small with respect to the proportion of the total sum of squares explained.

Forecasting

Forecasts are performed by plugging the appropriate values into the above equation. Actual values for stock levels and errors are used if they are in the past or present, and therefore known. Forecasted values are used if the terms refer to the future. For example, if t represents the present period, S_{t+1} may be estimated using known values. S_{t+2} may then be estimated by using known values and the forecast for S_{t+1} . Zero is always the forecast value for future error terms.

Forecasts were made for one year forward and compared with the actual data. The results are depicted in Figure 2. The forecasts appear to be reasonably accurate. Fully 61 percent of the squared deviations from the mean are explained. The last four months are

seriously over-estimated, perhaps verifying the previous assertion that the estimated trend would be too large. Unfortunately, however, there is no clearcut way to revise the forecasts to incorporate the belief that the positive trend is moderating. In addition, the sharp turnaround in stock levels in August is predicted to have occurred in September. This error was understandable, however, as an examination of the data revealed that in eight out of nine of the preceding years, the turnaround had been in September.

Table 1.

Estimated Model for SARIMA (2,0,0) \times (0,1,1)

| Parameter | Estimate | 95% Confidence Interval | |
|-------------------|----------|-------------------------|-------------|
| | | Upper Limit | Lower Limit |
| AR1 | 1.33 | 1.44 | 1.22 |
| AR2 | -.47 | -.36 | -.58 |
| MA ^s 1 | .51 | .63 | .40 |
| Trend | 4.27 | 6.37 | 2.17 |

Residual Mean Square—537.23

Residual Sum of Squares—133,771

Chi-Square for the White Noise Test—110.24 (88 d.f.) 5% Critical Value—113

Proportion of Total Sum of Squares Explained (Roughly the R²)—801

Table 2.

Estimated Model for SARIMA (3,0,0) \times (0,1,1)

| Parameter | Estimate | 95% Confidence Interval | |
|-------------------|----------|-------------------------|-------------|
| | | Upper Limit | Lower Limit |
| AR1 | 1.41 | 1.29 | 1.54 |
| AR2 | -.70 | -.90 | -.50 |
| AR3 | .16 | .04 | .29 |
| MA ^s 1 | .56 | .45 | .67 |
| Trend | 3.67 | 5.73 | 1.61 |

Residual Mean Square—528.267

Residual Sum of Squares—130,482

Chi-Square for the White Noise Test 104.55 (87 d.f.) 5% Critical Value—110

Proportion of the Total Sum of Square Explained (Roughly the R²)—806

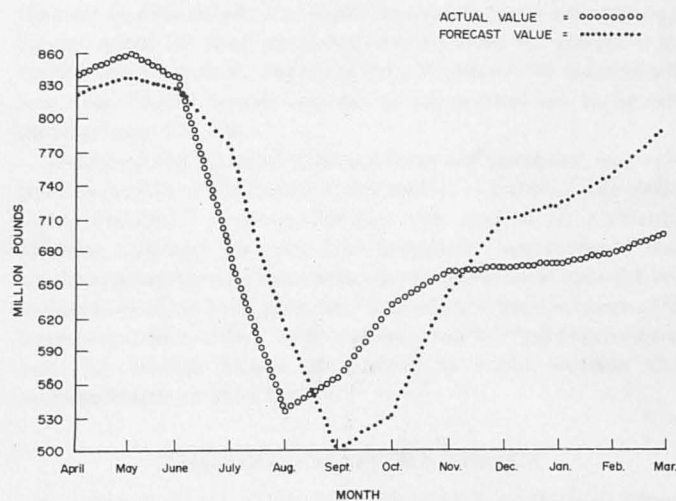


FIGURE 2.

Actual and Estimated Levels of End-of-Month Stocks of Frozen French Fries from April 1978 to March 1979

SUMMARY AND CONCLUSIONS

In this paper a pattern recognition model has been estimated for use in predicting the end-of-month stocks of FFF. It has been demonstrated that theoretical considerations can be important in specifying and validating such models. The form of the estimated model conformed with theoretical expectations and predicted with reasonable accuracy.

REFERENCES

Box, G. and G. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day Inc., 1976.

Brennan, M., "The Supply of Storage," *American Economic Review*, 47(1958):52-67.

Gustafson, Robert L., *Carryover Levels for Grains: A Method for Determining Amounts that are Optimal Under Specified Conditions*, U.S.D.A. Tech. Bul. 1178, 1958.

Nelson, C., *Applied Time Series Analysis for Managerial Forecasting*, San Francisco: Holden-Day Inc., 1973.

Phillips, M., T. Sporleder, J. Baarda, and G. Biggs, *Processed Potato Growers' Associations: Information and Organization Needs*, U.S.D.A., F.C.S., Research Report 35, January 1976.

U.S.D.A., *Summary of Regional Cold Storage Holdings*, C.R.B., S.R.S., Monthly January 1956-March 1979.