POULIN • HUGIN
PATTERNS & PREDICTIONS

Predicting The Daily Closing Price of a Futures Contract

In this example (which is one of the model templates included free with Poulin-Hugin Professional software) you'll see how our Bayesian technology may be used for predicting the daily frozen concentrated orange juice (FCOJ) contract close level based on average daily temperatures in more than 150 different US cities. The application is designed to identify patterns in the correlation between average daily temperatures that impact the level of the futures contract's closing price (e.g., the FCOJ indicator).

The knowledge base in this model template can be used to make inferences about FCOJ pricing and perform various types of analysis such as value of information analysis and scenario based sensitivity analysis. Importantly, the system does not need information about all data to be useful; but with more data, the more accurate the predictions will be.

The main component of this model is a knowledge base that depicts the correlation between the FCOJ indicator and the average daily temperature in each city. The knowledge base is constructed automatically based on the data from the Average Daily Temperature and the FCOJ contract close databases. The factors of the model are joined in a Bayesian Network that describes the correlation between the FCOJ indicator and the average daily temperature in each city. Not only does the knowledge base in Poulin-Hugin's model describe possible interdependence relations between factors, it also quantifies the strengths of the relationships. The dependence relations are quantified by conditional probabilities.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')}$$

The Temperature Data

The data used originates from the "Average Daily Temperature Archive" at University of Dayton and the New York Board of Trade, and covers the period from June 1, 1999 - February 28, 2005. In this example, the FCOJ indicator used is the price quotation (in cents and hundredths of a cent per pound solid), and temperatures are measured in Fahrenheit.

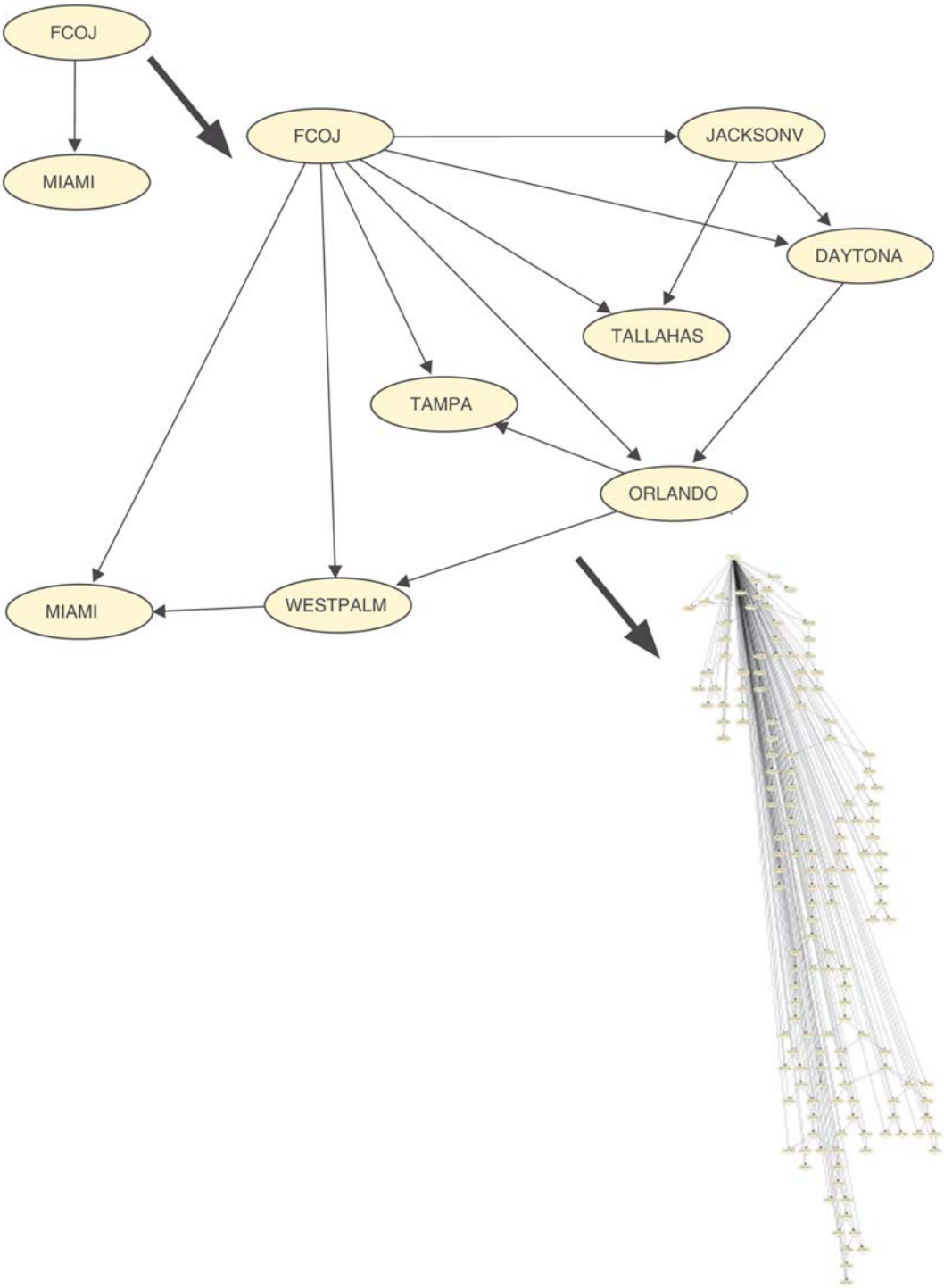
The data has the format shown in the table below.

FCOJ	ROCKFORD	MIAMI	TALLAHAS
93.20	67.4°F	N/A	76.6°F
90.70	63.1°F	78.9°F	78.2°F

Each line in the table represents the values of each factor on a particular date. The data has been preprocessed whereby the average daily temperatures for a particular date have been associated with the value of the FCOJ indicator the following business day. This implies that the data has missing values for all non-business days where the FCOJ indicator value is not available. The first line of data specifies the average daily temperatures for June 1, 1999 and FCOJ indicator level for June 2, 1999 while the last line specifies the average daily temperatures on February 27, 2005 and FCOJ indicator level on February 28, 2005.

Looking For Correlations

An extrapolated example, constructed only for the purpose of illustrating the underlying idea, is shown in the below figure. Illustrating progressive complexity, each oval represents a variable or factor. The factor with label "FCOJ" represents the level of the FCOJ indicator–while each of the other factors represents the average daily temperature in a certain city.



Discovery

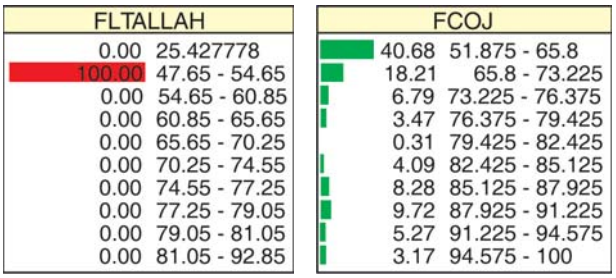
So, which city is the most informative city when trying to predict the FCOJ indicator? A value of information analysis can give the answer to this question. The table below shows the value of information score (VOI score) for each city in the model (the VOI score is, in this example, a measure of the correlation between a particular city and the FCOJ indicator).

City	VOI score
Rockford, IL	0.161853
Oklahoma City, OK	0.160159
Tallahassee, FL	0.131722
Miami Beach, FL	0.125184
San Francisco, CA	0.100514

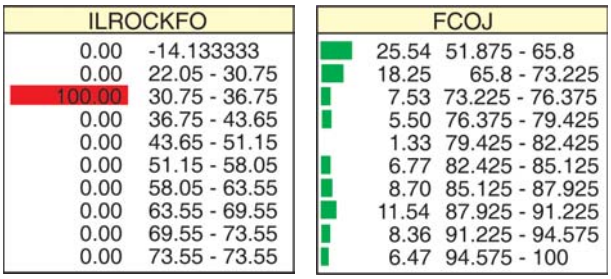
The table lists the VOI scores for five different cities. It shows that the most informative FCOJ indicator city for temperature measurement is Rockford, IL and the least informative city is San Francisco. Thus, much to our surprise, the most informative temperature reading when predicting the FCOJ indicator (when no other cities have been observed) turns out to be the temperature in Rockford, IL. The VOI score is dynamic and will change with the information available to add to the model.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')}$$

What happens to the distribution as we make observations on the average daily temperatures in different cities? The answer is that the probability of the distribution of the indicator will become more concentrated. The model is highly predictive with this city and other high VOI cities. The figure below shows how an observation on the most informative city, Rockford, IL concentrates the probability mass on the lower temperature ranges of the FCOJ indicator.



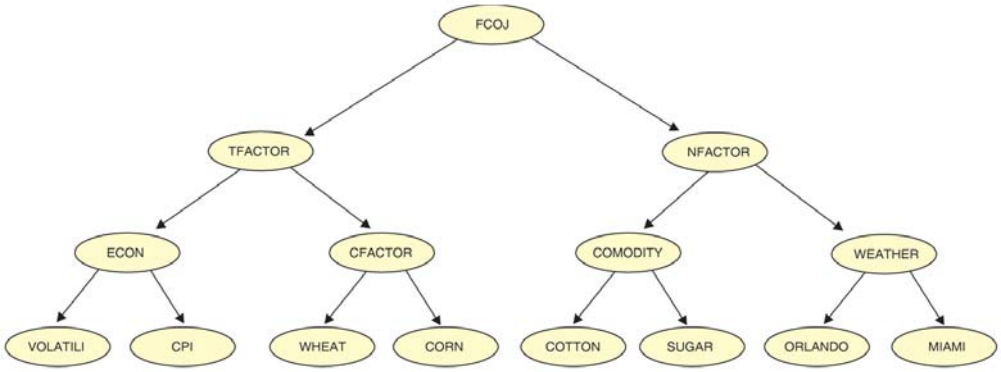
By adding to the observation on the average daily temperature of the average daily temperature in Tallahassee, FL to that of Rockford, IL we will most likely see a further concentration of the probability mass.



As we observed, with the average daily temperature in Tallahassee, FL being in the low 50s, the probability mass is, as expected, further concentrated on the range 51.875-73.225 degrees Fahrenheit. That is, the Poulin-Hugin model, with the two observations, now predicts that the level on the FCOJ index price for the next day will be in the range of 51.875-73.225 degrees Fahrenheit with almost 60% probability certainty.

If the user is not satisfied with the strengths of the correlations between factors, they may manually adjust the values of the conditional probability distributions estimated from the data.

Discovery of New Factors



The discovery of new factors is one of the potential benefits from using a hierarchically structured model. In each case the new factors represent a relationship between the factors below it. Therefore, we 'discover' the factors COMODTY, NFACTOR, and TFACTOR. These relationships may or may not be accounted for by publicly available information or even by generally accepted economic theory.

- COMODTY = A 'Commodity' factor that accounts for a variety of non-FCOJ farm raised commodity prices. Any number of commodities can be entered to test the relevance of a good to this custom indicator.
- NFACTOR = A 'New' undefined factor that "discovered" the relationship between Weather and Commodity data. This factor accounts for the unknown similarities in TTfarm raised commodity prices, including FCOJ.
- TFACTOR = A trading factor that accounts for BOTH market volatility and econometric factors. We would then assume that this factor was in fact representative of other active trading also responding to these subfactors. It's called TFACTOR because the non-linear relationship is probably due to behaviour of other automated trading programs that we can't quantify.

Note: The real system should be imagined as a somewhat more complex model. Based on the available information at a given time, the Poulin-Hugin system template will calculate the probabilities over the FCOJ indicator.

Model Summary

The FCOJ/Weather model predicts the FCOJ historical price with a high degree of accuracy (almost 60% probability). By combining the use of Value of Information analysis and What-If Sensitivity analysis, we can focus in greater detail on the parts of the model that are more predictive, thereby improving accuracy. This technique can be used to isolate parts of a data feed (such as Floridian towns and, oddly, Rockford, IL) that are worthy of more attention, and thereby further increasing the accuracy of our models. Finally, by using hierarchical analysis–a key feature of Poulin-Hugin's Bayesian-based modelling capability–we can quantify unknown factors influencing events.